VOLUME XXVII APRIL, 1948 NO. 2

THE BELL SYSTEM TECHNICAL JOURNAL

DEVOTED TO THE SCIENTIFIC AND ENGINEERING ASPECTS OF ELECTRICAL COMMUNICATION

AMERICAN TELEPHONE AND TELEGRAPH COMPANY
NEW YORK

50¢ per copy $1.50 per Year

THE BELL SYSTEM TECHNICAL JOURNAL

Published quarterly by the
American Telephone and Telegraph Company
195 Broadway, New York, N. Y.

SUBSCRIPTIONS

Subscriptions are accepted at $1.50 per year. Single copies are 50 cents each. The foreign postage is 35 cents per year or 9 cents per copy.

PRINTED IN U. S. A.

The Bell System Technical Journal

Vol. XXVII *April, 1948* *No. 2*

Microwave Repeater Research

By H. T. FRIIS

Introduction

IT WAS some 80 years ago that Maxwell and Hertz demonstrated that free space is a good transmission medium for electromagnetic waves. Since this fundamental contribution, the radio art has advanced tremendously and a decade ago it had progressed to the point where it was possible to construct equipment suitable for quantitative propagation studies of microwaves. Such studies were made and they indicated that normal propagation over "line-of-sight" paths of signals of 10 to 20 centimeters wavelength was characterized by free space attenuation and freedom from atmospheric interference. These results, together with the facts that in this wavelength range wide bands of frequencies are available and it is possible to design small antennas having high directivity, encouraged us to start more comprehensive research work on microwave repeater circuits. This paper gives the present status of the work which was interrupted by our war efforts and resumed at the end of the war with the construction of an experimental New York-Boston system as an initial objective.

The first section will describe our propagation studies. It will be followed by sections on repeater circuit planning, antennas, radio frequency channel filters, the construction and testing of the repeater amplifier, and a concluding section on the whole repeater.

I. Propagation Studies*

That portion of the radio frequency spectrum represented by wavelengths shorter than about five meters has long been considered as the proper domain for point-to-point communication links, local broadcasting, and mobile radio communication. Since these ultra-short waves are not reflected by the ionosphere, their effective range is not much greater than the horizon distance and it therefore becomes possible for a number of stations, properly separated, to operate in the same frequency band; for the same reason, atmospheric interference is not an important factor in this wave-

* This section was prepared by A. B. Crawford who, with W. M. Sharpless, is at present engaged in microwave propagation studies.

length range. Also, as the wavelength decreases, it becomes possible to construct antennas large in comparison with the wavelength so that high antenna gains are obtained and the corresponding directivity further reduces the interference areas.

Since about 1930, with the exception of the war years, we have conducted fundamental studies in radio propagation, taking advantage of advances in the art to extend the wavelength range from about four meters (ultra-short wave region) in the beginning to 1.25 centimeters (microwave region) at the present time. A considerable portion of the effort of those engaged in propagation studies has, of necessity, been devoted to the development of measurement techniques and reliable measuring apparatus. The present discussion, however, will be concerned with the results of experiments rather than with a description of the apparatus and methods. Most of these results have been described in the literature; the following is a review intended to show the development of the background leading to the present field trial of a microwave repeater circuit.

The object in making propagation studies has been to evaluate and to understand the effects of the terrain and of the lower atmosphere upon the transmission of ultra-short-wave and microwave signals. The evaluation is usually obtained by amassing sufficient data on a particular transmission experiment so that a statistical analysis can be made. Efforts to understand the transmission phenomena usually take the form of experiments involving specially designed apparatus. These experiments are varied from time to time as information is obtained or as it becomes desirable to check the validity of such theories as may be devised. The hope is always present that an understanding of the phenomena may suggest a means for reducing the transmission difficulties.

The absence of ionospheric reflections at these frequencies suggested at the start that propagation studies would probably be concerned mainly with phenomena familiar in optics, namely: reflection, refraction and diffraction. Two of the early papers[1, 2] treated ultra-short-wave propagation from this viewpoint. It was soon observed that diffracted signals tended to be unstable in the shadow region; furthermore, as the wavelength is decreased the shadows cast by obstacles such as hills or the bulge of the earth itself become more sharply defined. For these reasons, a considerable part of our experimental work has been done on paths for which a line-of-sight exists between transmitter and receiver. The chief interest, therefore, has been in ground reflections and the effect of the atmosphere.

[1] J. C. Schelleng, C. R. Burrows and E. B. Ferrell, "Ultra-Short-Wave Propagation", *Proc. I. R. E.*, vol. 21, pp 427–463; March 1933.

[2] C. R. Englund, A. B. Crawford and W. W. Mumford, "Some Results of a Study of Ultra-Short Wave Transmission Phenomena", *Proc. I. R. E.*, vol. 21, pp 464–492; March 1933.

Ground Reflections

Some of our first experiments with ultra-short-waves showed that regular reflections were obtained locally from open, relatively flat fields. The reflection coefficients were in good agreement with theory. Later, measurements of propagation between a transmitter located on a hill top and a receiver carried in an airplane[2] showed that for near-grazing angles of incidence, the irregular and wooded terrain, typical of the New Jersey countryside, could give rise to regular reflections at wavelengths as short as four meters. The depth of the minima in received signal strength, caused by wave interference between the direct and ground reflected components, corresponded to a reflection coefficient of about 0.9. In 1939, unpublished results obtained over the 39-mile Beer's Hill-Lebanon optical path (See map of Fig. I-1) indicated that for a wavelength of 30 centimeters the reflection coefficient was still large, about 0.8.

More recently, microwave propagation studies have been made over the same type of terrain at wavelengths of 3.25 centimeters and 1.25 centimeters and the situation in regard to ground reflections seems to have changed somewhat. Experiments were conducted over the 12.6 mile Beer's Hill-Deal path in which the height of the transmitting terminal was varied and which also made use of narrow-beam scanning antennas to separate the direct wave from a possible ground reflected component. The results showed the apparent reflection coefficient to be of the order of 0.2 at 3.25 centimeters and to be even less at 1.25 centimeters. Figure I-2 shows typical curves of signal level versus transmitter heights for wavelengths of 3.25 and 1.25 centimeters. Actually, the shapes of the curves can be accounted for better by diffraction, for which the hill about two miles from Deal is considered to be a straight edge, than by reflection from an assumed average ground plane. The true picture is probably a combination of reflection and diffraction effects.

In an effort to minimize ground reflection, over-water paths were avoided in the layout of the New York-Boston microwave repeater circuit and as a final check a number of variable antenna-height tests* were made in the preliminary survey of all sites. A few curves obtained at a wavelength of 7 centimeters are reproduced in Fig. I-3. Similar results were observed during a survey of sites between Chicago and Milwaukee.

It is concluded, therefore, that although in the wavelength range down to 30 centimeters, at least, the effects of ground reflection must be taken into account in the choice of sites for an optical path radio circuit, in the lower microwave range, below say 10 centimeters, scattering and absorption of the reflected wave by rough terrain and vegetation usually results in substantially free-space propagation under normal conditions when the line of

* F. F. Merriam was in charge of this work.

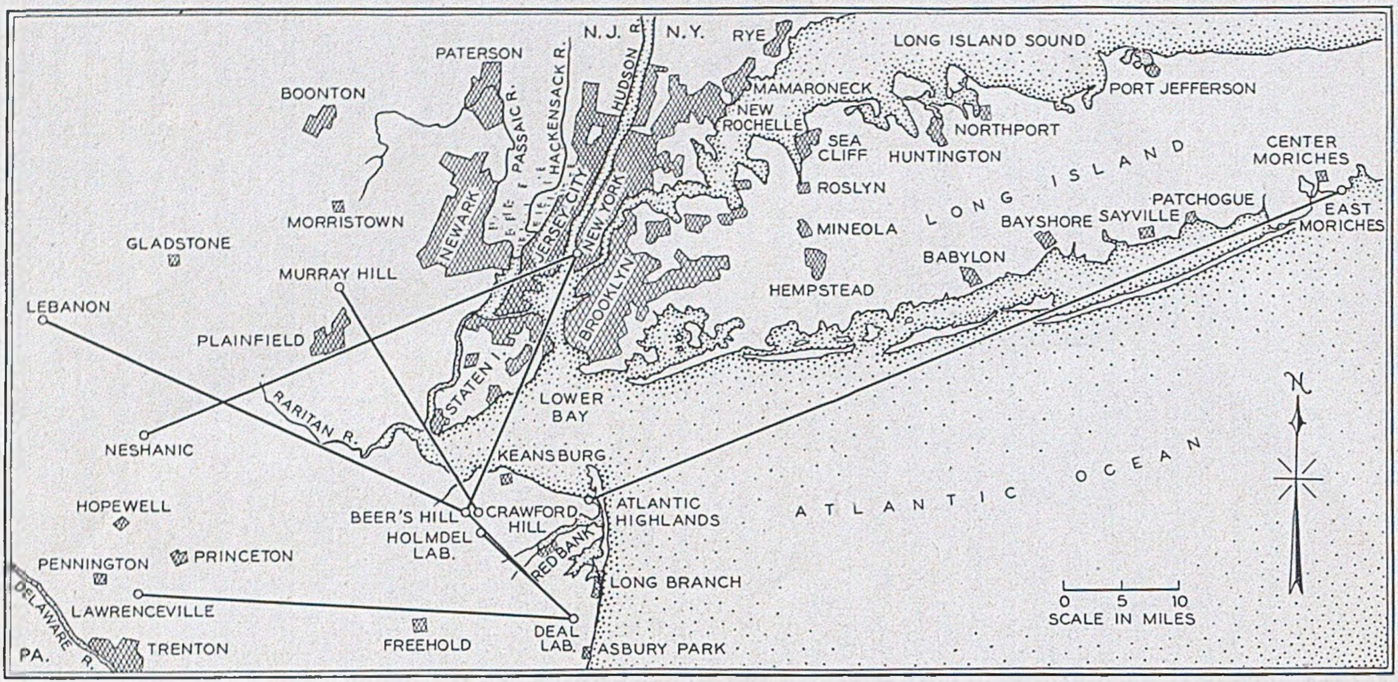


Fig. I-1.—Map showing principal Propagation Paths.

sight is well clear of intervening obstructions. In order to have a rule-of-thumb as to the amount of path clearance desirable, we have suggested that the first Fresnel region should be clear of all obstacles. The first Fresnel region for a given transmitter and receiver is bounded by points for which the length of the path, transmitter to point to receiver, is greater by one-half

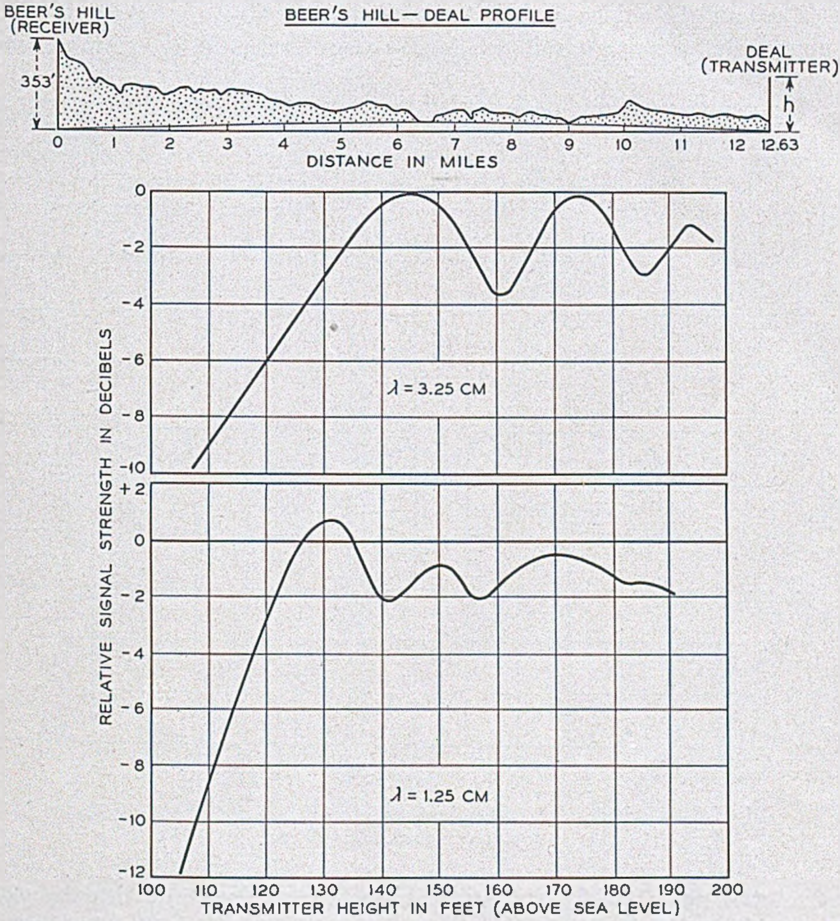


Fig. I-2.—Variable antenna-height tests on Beer's Hill-Deal Path.

wavelength than the direct path from transmitter to receiver; its cross-section by any plane perpendicular to the direct path is the first Fresnel zone in the sense used in optics. A wave can be transmitted with practically no loss through an opening whose area is of the order of the first Fresnel zone. Also, in the case of a smooth reflecting surface between transmitter and receiver, the first Fresnel zone clearance provides a maximum in received field strength since the half wavelength path difference plus the 180-degree phase change at reflection causes the direct wave and the reflected wave to arrive in phase at the receiver. In Fig. I-4, the first Fresnel region is sketched on the profile map of a typical microwave link for wavelengths of 3 meters and 3 centimeters.

It should be emphasized that the above remarks on ground reflections apply only for rough terrain and for the case of reflection at a distance from

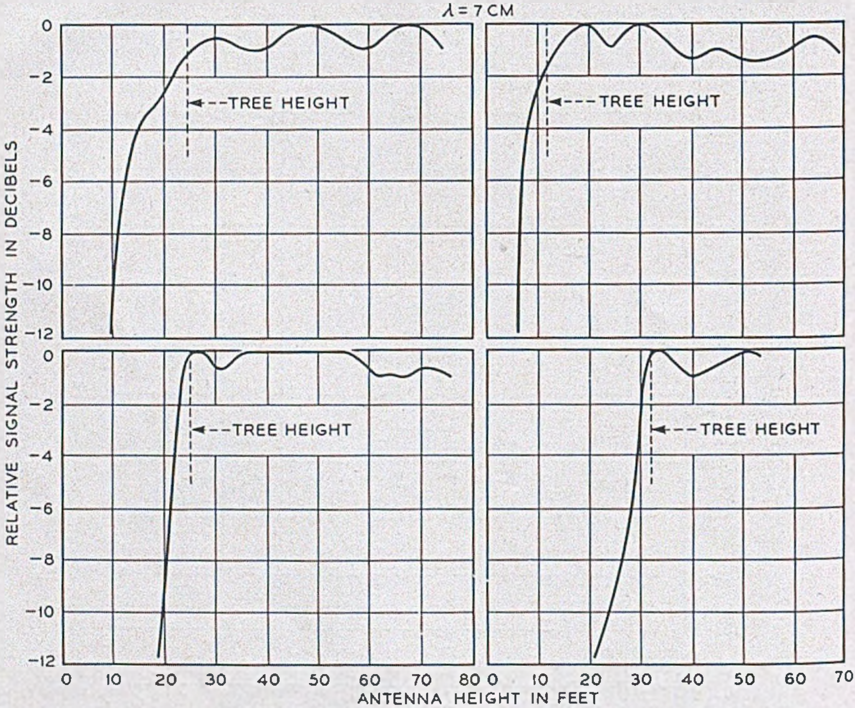


Fig. I-3.—Variable antenna-height tests on several of the New York-Boston repeater circuits. λ = 7 cm.

the terminals. Variable height experiments involving short distances over open flat fields reveal the presence of almost perfect ground reflections at wavelengths as short as 1.25 centimeters. For transmission paths over water, strongly reflected components are often observed. Reports[3] of experiments in the Arizona desert indicate a strong ground reflection at wavelength of 3 centimeters. In such locations, and most likely in the plains regions, the presence of substantial ground reflected components may prove to be troublesome.

[3] Report No. 6. Electrical Engineering Research Laboratory, The University of Texas, February 1, 1947.

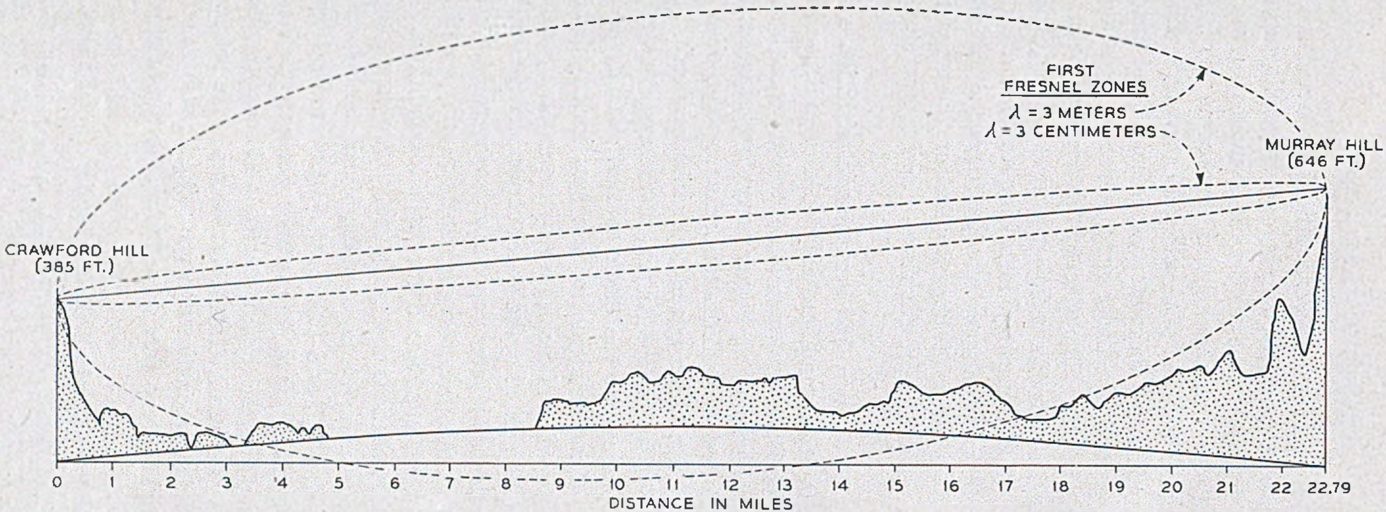


Fig. I-4.—Profile map of Murray Hill-Crawford Hill path showing first Fresnel regions for wavelengths of 3 meters and 3 cm.

Atmospheric Refraction

As in the case of ground reflection, the refractive effect of the atmosphere has been found to play a somewhat varying role, depending upon the wavelength employed. In some of the early work on ultra-short-wave propagation,[1, 2] the concept of average atmospheric refraction was found to bring about better agreement between observed and calculated results. Due to the variation of temperature and water vapor content of the atmosphere with height above ground, the dielectric constant of the atmosphere normally decreases with height. The effect of this negative dielectric constant gradient is to cause the path of a radio wave to be bent slightly downward toward the earth, thus effectively increasing the horizon distance. It has been suggested that a good approximation for average refraction was to assume the radius of curvature of the ray to be four times that of the earth.[1] This condition is used at the present time to describe a "standard atmosphere."

It was soon found, however, that atmospheric refraction could vary between rather wide limits depending chiefly on the gradient of water vapor with height.[4] Refraction effects were found to be greater in summer than in winter since the air contains a higher percentage of water vapor in the summertime. A diurnal variation in refraction was also observed on overland transmission paths. During the day, rising convection currents and surface winds, caused by surface heating of the earth, usually produce a well mixed atmosphere near the earth so that "standard" atmospheric conditions prevail. On clear nights, however, particularly if the wind velocity is low, radiation cooling of the earth may cause a temperature inversion in the lower atmosphere; if, also, the water vapor decreases with height, the combined temperature and water vapor effects may add to produce a steep negative gradient in the dielectric constant. Stormy weather and overcast skies usually result in standard atmospheric conditions.

Most of the signal variations observed during a two-year study of propagation of two and four-meter waves over the 39-mile over-land optical path between Beer's Hill, N. J. and Lebanon, N. J.[5] could be explained satisfactorily on the basis of wave interference between direct and ground-reflected radiations; the relative path lengths, and hence the phases, of these two components of the received field varied with the refractivity of the atmosphere. The fading on the two wavelengths was usually similar in major detail as might be expected from the geometry of the path. On the

[1, 2] Loc. cit.

[4] Englund, Crawford and Mumford, "Further Studies of Ultra-Short-Wave Transmission Phenomena", *B. S. T. J.*, vol. 14, pp 369–387; July 1935.

[5] Englund, Crawford and Mumford, "Ultra-Short-Wave Transmission over a 39 mile 'Optical' Path", *Proc, I. R. E.*, vol. 28, pp. 360–369; August 1940.

few occasions when the fading could not be accounted for in this simple fashion, it was assumed that signal components were arriving from above by virtue of reflections from small, relatively abrupt changes in the dielectric constant of the atmosphere. The existence of such reflections was demonstrated by a frequency-sweep method during propagation studies on the 70 mile over-water path between Highlands, N. J. and East Moriches, Long Island.[6]

With microwaves, where, as stated previously, ground reflections are usually small or absent, one might surmise that changes in atmospheric refraction would have a smaller effect on transmission than at ultra-short-wavelengths where strong ground reflections are present, and that fading should, therefore, be less. Actually the opposite is observed. Fading is found to be more frequent, faster, and deeper as the wavelength is decreased. This frequency effect may be explained in a qualitative fashion by a consideration of the relative sizes of Fresnel zones at ultra-short waves and at microwaves. It is known that the dielectric constant of the atmosphere usually does not vary with height in a smooth linear manner; on calm nights, particularly, very steep gradients in the dielectric constant may exist over small vertical ranges measuring only tens of feet. The effectiveness of these steep gradients would be expected to depend on their extent relative to the size of a Fresnel zone. Thus on a path such as that in Fig. I-4, a steep gradient extending over only a hundred feet would include practically the whole first Fresnel zone at 3 centimeters while it would cover only a small part of a zone at 3 meters wavelength; the effective gradient, therefore, would be considerably less at 3 meters than at 3 centimeters. Analyses based on wave theory show that atmospheric layers, in which the dielectric constant has a steep negative gradient, tend to confine or guide the radiation in much the same way as a waveguide, and that this "trapping" phenomenon, for a given layer thickness, becomes more pronounced as the wavelength is decreased.[7]

The mechanism of microwave propagation is certainly a complicated one, and a considerable amount of experimental work in the fields of radio and meteorology will be required to unravel it. However, it is very difficult to interpret the radio measurements in terms of meteorological data. The chief difficulty is that meteorological measurements often do not give an accurate picture of the atmosphere, particularly at those times when microwave fading indicates that rapid changes of some sort are occurring in the

[6] Englund, Crawford and Mumford, "Ultra-Short-Wave Transmission and Atmospheric Irregularities", *B. S. T. J.*, vol. 17, pp. 489–519; October 1938.

[7] H. G. Booker, in England, was the first to call attention to this phenomenon. For more recent work see: C. L. Pekeris, "Wave Theoretical Interpretation of Propagation in Low-Level Ocean Ducts, *Proc. I. R. E.*, vol. 35, pp. 453–462; May, 1947. This paper gives references to other work in this field.

atmosphere. The instruments used to measure temperature and humidity require a few seconds to reach equilibrium—a length of time comparable at times with the period of fading. To measure the variation of dielectric constant with height, the measuring instruments are usually carried aloft by means of captive balloons. A half hour may be required to measure to heights of six or seven hundred feet with the result that the final curve represents an unknown combination of variations of dielectric constant with height and with time. It is also extremely doubtful that the atmosphere is uniform in the horizontal plane—an assumption which is usually made in the theoretical treatment of microwave propagation. It seems likely that the lower atmosphere is far from being a homogeneous fluid but rather may contain small air masses or "boulders" with properties which differ considerably from those of the surrounding air. Reflections from these boulders may be the cause of radar echoes received from the lower atmosphere.[8] Scintillation fading of microwaves is another evidence of these inhomogeneties in the atmosphere. Scintillation fading, a rapid fluctuation in signal level about a more or less steady average value, increases as the wavelength becomes less and as the path length is increased.

In order to evaluate, on a statistical basis, the effect of atmospheric changes on a typical microwave circuit, extensive measurements of transmission were made over a 40-mile overland path between New York City and Neshanic, New Jersey. The tests covered a period of about two years. Most of the data were obtained at wavelengths of 10, 6.5, and 3.2 centimeters although some data were taken at wavelengths of 42 centimeters and 1.25 centimeters. The results are described in a recent paper.[9] In many respects, observations were in agreement with those made earlier on the 39-mile Beer's Hill-Lebanon path at wavelengths of 4 and 2 meters and on the 38-mile non-optical path between Deal, N. J. and Lawrenceville, N. J. at a wavelength of 2 meters.[10] The same seasonal and diurnal trends in fading were found; transmission was generally steady during the midday hours and during periods of windy or rainy weather; fading was the same on vertical and horizontal polarizations. However, the character of the fading was different; the fading at microwaves was much faster and deeper than that observed on the ultra-short-wave path. The average daily fading range for July on the New Yoik-Neshanic path was 20 db at 6.5 centimeters compared with a median daily fading range of 8.5 db for 2.0 meters observed in July on the Lebanon-Beer's Hill path.

[8] H. T. Friis, "Radar Reflections from the Lower Atmosphere", *Proc. I. R. E.*, vol. 35, pp. 494–495; May 1947 (Correspondence Section).

[9] A. L. Durkee, "Results of Microwave Propagation Tests on a 40-mile Overland Path", *Proc. I. R. E.*, vol. 36, No. 2, pp. 197–205, Feb. 1948.

[10] C. R. Burrows, A. Decino and L. E. Hunt, "Stability of Two-Meter Waves", *Proc. I. R. E.*, vol. 26, pp. 516–528; May 1938.

Other observations on the New York-Neshanic microwave path may be summarized as follows: While all the wavelengths were affected at times of anomalous propagation, the shorter wavelengths faded more severely and the character of the fading was different from that observed at the 42 centimeters wavelength; apparently the 3–10 centimeter range was more sensitive to the fine structure of the atmosphere, as pointed out previously. During non-fading periods, signal levels were very close to the free-space values with the exception of the 1.25 centimeter signal which was usually 15 db or more below the free space value because of atmospheric absorption effects. Some special tests showed that fading was considerably more severe when one of the terminals was lowered so that the transmission path was grazing slightly below line-of-sight. It was also found that fading was about twice as great, in decibels, on the whole path as on either half-section. A statistical analysis, on an hourly basis, of all the data on 6.5 centimeters showed that only one-half of one percent of the total hours had signal minima deeper than 20 db below the free space field. Also during August 1, the day of the most severe fading, the signal was more than 20 db below free space for about one-half of one percent of the time. It was also found that signals of the order of 10 db above free space were equally probable. From a consideration of these statistics, it was decided to engineer the New York-Boston repeater circuit with -20 to $+10$ db allowance for fading on each link.

Specialized Experiments

Much of our more recent work on microwave propagation has been of a specialized nature in which apparatus and experiments have been designed more for the purpose of studying the mechanism of anomalous propagation than for making a statistical analysis of the transmission. Perhaps the most informative experiments have been those in which narrow beam scanning antennas were used to explore the incident wave fronts.

The first of these antennas had an aperture of 20 feet and a beam width between half-power points of $\frac{1}{3}$ degree at the design wavelength of 3.25 centimeters. It was built for the purpose of establishing a practical limit to the size, and hence the directivity, of microwave repeater antennas from the standpoint of variations in the angle of arrival of the received wave. It had been realized, of course, that variations in the refractivity of the atmosphere would cause some deviations in the path of the wave. While these deviations should be negligible in comparison with the beam width of antennas normally used in the ultra-short-wave region, they might be comparable with the beam widths readily obtainable in the microwave region.

Using this antenna for measurements in the vertical plane and another identical antenna for measurements in the horizontal plane, angle-of-arrival data were obtained during the summer of 1944 over a twenty-four mile, partly over-water, path between Beer's Hill, N. J. and New York and over a thirteen mile over-land path between Beer's Hill, N. J. and Deal, N. J.[11] In the horizontal plane, deviations in the angle of arrival were rather uncommon and were not greater than $\pm$ 0.1 degree from the true bearing of the transmitter. In the vertical plane, angles of arrival above the true elevation of the transmitter were observed to be as much as 0.5 degree on the New York path and 0.3 degree on the Deal path during times of anomalous propagation. From these measurements it was concluded that microwave repeater antennas could be made highly directive in the horizontal plane but should have beam widths somewhat greater than $\frac{1}{2}$ degree in the vertical plane unless means for steering the beams are provided.

Although the $\frac{1}{3}$ degree beam width of these scanning antennas was sharp enough to permit the separation of the direct and the water-reflected components on the New York path, and to demonstrate the anomalous behavior of each, there was evidence that occasionally there were signal components so close together in angle that a sharper antenna would be required to resolve them. Consequently a scanning antenna of the metal-lens type was constructed for operation at 1.25 centimeters. The aperture of this antenna was 20 feet in the long dimension; the beam width was 0.12 degrees. Using this antenna and also the 3.25 centimeter scanning antennas, angle-of-arrival measurements were made in the summer of 1945 on the Deal-Beer's Hill path.[12] The most noteworthy result of these observations was the demonstration of multiple-path transmission. Two, three and, at times, four distinct signal components were observed simultaneously during one night when the transmission was extremely disturbed. These transmission paths generally were above the true direction of the transmitter; at one time, a weak signal was arriving at an angle of 0.75 degree relative to the line of sight. These components varied in angle of arrival and in signal amplitude. Wave interference among them caused severe fading on broad beam antennas that would accept all the wave paths.

Another significant result of these angle-of-arrival measurements was evidence that the transmission mechanism was very similar for wavelengths of 3.25 and 1.25 centimeters. Angles of arrival, measured simultaneously at the two wavelengths, agreed very well for times of single-path transmission; multiple-path transmission was observed on both wavelengths although the 3.25 centimeter antenna was too broad to resolve the com-

[11] W. M. Sharpless, "Measurement of the Angle of Arrival of Microwaves", *Proc. I. R. E.*, vol. 34, pp. 837–845; November 1946.

[12] A. B. Crawford and W. M. Sharpless, "Further Observations of the Angle of Arrival of Microwaves", *Proc. I. R. E.*, vol. 34, pp. 845–848; November, 1946.

ponents completely. This result suggested that the 1.25-centimeter scanning antenna might be a very useful tool for investigating the fading mechanism at 7 centimeters wavelength.

The 22.8-mile path between Crawford Hill and a hill on the Murray Hill Laboratory property was chosen for study as a representative link in a repeater circuit. (See Profile Map, Fig. I-4.) Transmitters for the 1.25-centimeter and 7 centimeter wavelengths were installed in the 100-foot tower at Murray Hill. At the Crawford Hill receiving site were the narrow-beam scanning antenna and a broad beam antenna for 1.25-centimeter operation; also two broad beam antennas, spaced vertically 15 feet, for 7-centimeter operation. In addition, a 1.25-centimeter radar could be operated with the scanning antenna. A corner reflector target, $5\frac{1}{2}$ feet on a side, was located at the Murray Hill tower. The signal reflected by this target was about 10 db stronger than the spurious reflections from other objects at the same range. By making use of this target and ground echoes at intermediate distances, the radar technique provided a considerable amount of useful information concerning the transmission phenomena.

Measurements were made on this path during the summer of 1946. As had been hoped, the observations showed that transmission on 1.25 centimeters and 7 centimeters was often affected by the same conditions except, of course, for atmospheric absorption effects at the 1.25-centimeter wavelength. While it was not possible to arrive at explanations for all the fading observed, the deep minima in the 7-centimeter signal, i.e., fades to levels of 15 to 20 db or more below the free space field, usually were the result of one of three types of propagation*:—

Type 1. The 7-centimeter fading was of the rapid, large amplitude type characteristic of wave interference. The 1.25-centimeter scanning records showed the presence of multiple-path transmission in which two or more readily separable wave paths were observed. While the signals on both of the vertically-spaced 7-centimeter antennas faded about the same in amplitude, their signal minima did not occur simultaneously. A space diversity system would be successful in reducing the effects of this type of fading.

Type 2. The 7-centimeter fading was somewhat slower than in Type 1, but still had the appearance of wave interference. The 1.25-centimeter scanning records appeared to be of the single path variety. However, close inspection showed that, in all probability, more than one transmission path was involved but the 0.12 degree beam of the antenna was not sharp enough to resolve them. The signals received on the vertically spaced 7-centimeter antennas faded together so that space diversity would not be expected to be successful unless an extremely large spacing of antennas were used.

* Recently, on a different overland path having barely one Fresnel zone clearance, an important fourth type has been observed when atmospheric refraction gives the ray path a curvature opposite to that of the earth, thus effectively reducing the path clearance.

Type 3. The 7-centimeter signal would fade to a low level and remain there for a considerable period of time; sometimes for an hour or so. The character of the fading was unlike that caused by wave interference. The 1.25-centimeter signal was simultaneously at a low level and the scanning records showed that only one path was involved. Reception was almost

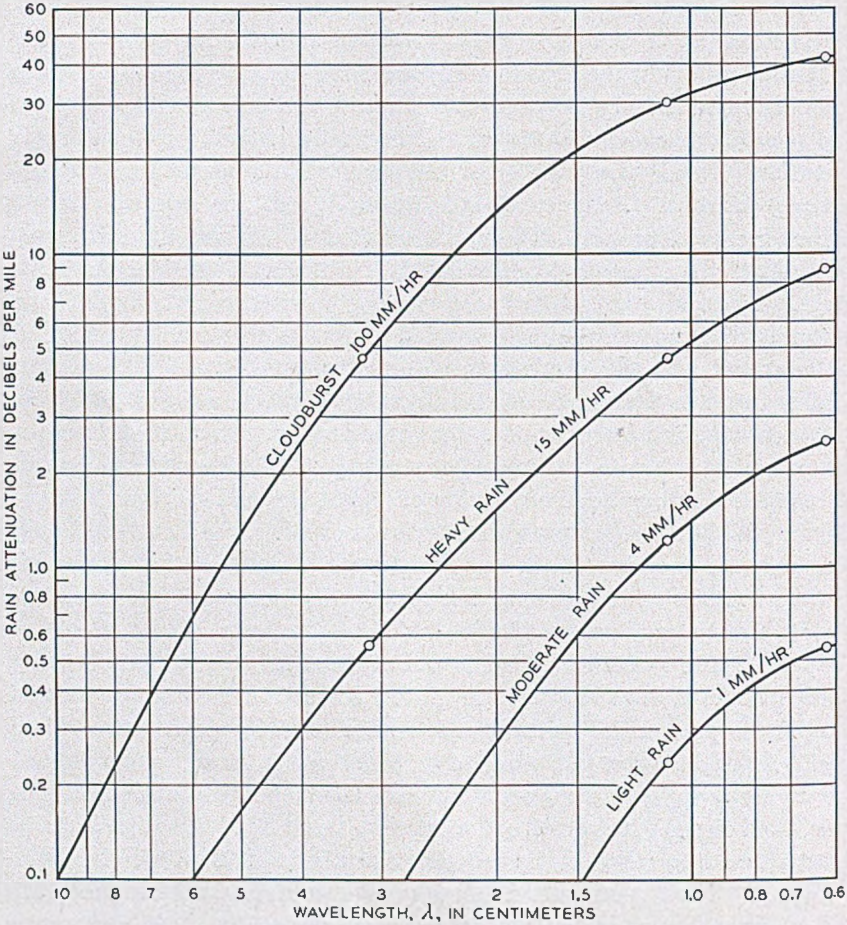


Fig. I-5.—Rain attenuation vs. wavelength.

identical on the two vertically-spaced 7-centimeter antennas. Radar observations suggested that this type of fading was due to attenuation by a reflecting layer in the atmosphere at a height intermediate to the heights of the transmitters and receivers. It was observed, for example, that while the echo from the Murray Hill corner reflector was absent, strong echoes were received from the hill directly in front of Murray Hill and some 250 feet lower in height; sometimes multiple paths were observed with this echo. Space diversity would fail to improve transmission under these propagation conditions, and no other means of improvement is apparent except, perhaps, an alternate path. Fortunately, this type of fading was the least frequent of the three types which were characterized by low signal levels.

Rain Attenuation and Atmospheric Absorption

Attenuation effects due to rainfall and absorption by atmospheric gases become increasingly important at the short-wave end of the microwave region. Measurements of rain attenuation have been made at the Holmdel Laboratory[13, 14]; the results are summarized in Fig. I-5. These curves show that for wavelengths above about 5 centimeters, rain attenuation is not very serious except for rains of cloudburst proportions. However, at wavelengths of one centimeter and less, even moderate rainfall will cause large attenuations on paths of the order of 10–20 miles in length.

Absorption by atmospheric gases, principally water vapor and oxygen, becomes important at wavelengths below about 1.5 centimeters. According to the theoretical work of Dr. J. H. Van Vleck, Harvard University, water vapor has an absorption band at 1.33 centimeters and oxygen has bands at 0.5 and 0.25 centimeters. Measurements made on the Deal-Holmdel path at 1.25 centimeters were in fair agreement with Van Vleck's results and indicated that a typical value of atmospheric absorption for this locality in summertime is about 0.4 db per mile.[14]

Summary

In the ultra-short-wave region, transmission has been found to be affected mainly by ground reflections and variable atmospheric refraction; only occasionally are atmospheric reflecting layers and trapping phenomena involved. These wavelengths ordinarily are not transmitted to great distances along the surface of the earth, but are diffracted around obstacles. They are used for local broadcasting and mobile radio communication.

Microwaves are attractive for radio repeater circuits since they permit the use of wide transmission bands. Ground reflections are apparently of small importance with terrain such as that of the Eastern seaboard and substantially free-space propagation is obtained during non-fading periods over optical paths which have approximately "first Fresnel region" clearance. Atmospheric reflecting layers and trapping phenomena are frequently observed and signal variations are considerably greater than in the ultra-

[13] Sloan D. Robertson and Archie P. King, "The Effect of Rain upon the Propagation of Waves in the 1- and 3-Centimeter Regions", *Proc. I. R. E.*, vol. 34, pp. 178P–180P; April 1946.

[14] G. E. Mueller, "Propagation of 6-Millimeter Waves", *Proc. I. R. E.*, vol. 34, pp. 181P–183P; April 1946.

short-wave region. Although fading becomes worse as the wavelength is decreased, the advantages of increased antenna gain and directivity possible at the shorter wavelengths suggest the use of a wavelength just above the region where rain attenuation becomes objectionable; i.e. above about 5 centimeters.

The use of two antennas, operated in space diversity, should reduce the effects of fading caused by multiple-path transmission. The use of space diversity may be essential in those localities where strong ground reflections are present. On the basis of the comparatively weak ground reflections measured on the New York-Boston path it was decided to avoid the complications that would result from the use of space diversity in this experimental system.

II. Repeater Circuit Planning

The diagram in Fig. II-1 shows schematically a repeater circuit. At the input terminal toward the left the signal, S, is fed to the terminal's transmitting antenna. The radiated signal is propagated as discussed in Section I and produces the signal power s_1 in the output of the receiving antenna of repeater 1. The signal is then amplified G_1 times and radiated toward repeater 2 and this process is repeated until the signal finally appears in the output terminal towards the right. In each repeater the signal is gain-controlled automatically for the same level of output powers, i.e., $S = S_1 = S_2 = \cdots = S_n$. It is assumed that the signals are amplitude or frequency-modulated C.W. carriers of substantially the same frequency in each link and that the repeaters have linear amplifiers. The diagram shows a West-to-East circuit only. A circuit for the opposite direction requires duplication of all the equipment with the exception of the antenna supporting towers.

Some simple formulas for the repeater gain and the signal-to-noise ratio at the terminal will be given in this section, without going into any details on propagation phenomena, antennas, amplifiers, etc. The formulas will orient the reader in regard to the importance of quantities such as:

$$\left.\begin{array}{l} d = \text{Repeater separation} \\ \lambda = \text{Wavelength of signal} \\ A = \text{Effective area of each antenna} \end{array}\right\}\text{same units of length}$$

F = Noise figure of each repeater amplifier
B = Bandwidth of circuit in cycles/sec.

The free space attenuation (S_{x-1}/s_x) of link "x" is[15]

$$L_x = \frac{d_x^2\lambda^2}{A^2}$$

[15] H. T. Friis, "A Note on a Simple Transmission Formula", *Proc. I. R. E.*, vol. 34, No. 5, pp. 254–256; May 1946.

Allowing the signal power to fade by a factor M below this value, the maximum gain of repeater "x" must be

$$G_x = \frac{d_x^2\lambda^2}{A^2} M_x \tag{II-1}$$

The total maximum gain in the circuit is $G_T = G_1 \times G_2 \times \cdot\cdot G_n$ which for the same repeater spacings and fading allowances becomes

$$G_T = G^n = \left(\frac{d^2\lambda^2 M}{A^2}\right)^n = \left(\frac{d^2\lambda^2 M}{A^2}\right)^{D/d} \tag{II-2}$$

where D is the total length of the circuit. Because of distortion, original costs, and maintenance costs, this total gain should be made as small as possible.

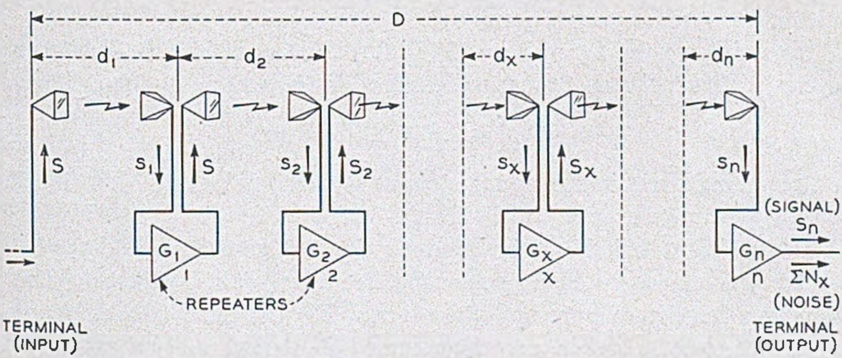


Fig. II-1.—Repeater circuit with n links.

The following example illustrates the maximum gain required of the amplifier in a repeater:

For $d = 4 \times 10^4$ meters (25 miles), $\lambda = 0.075$ meters ($f = 4000$ megacycles), $A = 4.6$ meter^2 (50 sq. ft.) and $M = 100$ (20 db), we have

$$G = 4.3 \times 10^7 \text{ (76 db)}$$

The noise output of a repeater due to noise sources in the repeater itself is approximately[16]

$$N_x = 4 \times 10^{-21}\ F\ B\ G_x \text{ Watts}$$

or from (II-1)

$$N_x = 4 \times 10^{-21}\ FB\ \frac{d_x^2\lambda^2}{A^2} M_x$$

[16] H. T. Friis, "Noise Figures of Radio Receivers", *Proc. I. R. E.*, vol. 32, pp. 419–22; July 1944.

This noise power is transmitted without gain or loss to the output terminals of the repeater circuit. The total noise power at the output terminals is therefore

$$\Sigma N_x = 4 \times 10^{-21} FB \frac{\lambda^2}{A^2} \Sigma(d_x^2 M_x)$$

Assuming the same repeater spacings and fading factors M in each link and substituting D for nd,

$$\Sigma N = 4 \times 10^{-21} FBD \frac{d\lambda^2}{A^2} M \text{ Watts} \tag{II-3}$$

The signal-to-noise ratio at the output terminals is $S/\Sigma N$. The circuit should be designed for a signal power, S, as low as possible. Therefore, it is very important to choose values of the several factors in (II-3) which give a low level of output noise.

Assuming a noise figure $F = 20$ (13 db) and bandwidth $B = 10$ megacycles, eight repeaters of the type described in the above example will have,

$$\Sigma N = 2.8 \times 10^{-4} \text{ Watts}$$

or, assuming a required minimum output signal to noise ratio of 30 db, the output power must be $S \geqq 0.28$ Watts.

In this example it has been assumed that the signals in all links have faded simultaneously 20 db below the free space value, which may only happen a fraction of a percent of the time. Most of the time the signal-to-noise ratio will be higher than the assumed 30 db and under normal transmission conditions it will be 50 db (fading allowance factor $M = 1$).

Assuming the same repeater spacings and fading allowances in each link, equation (II-1) and (II-3) give the following formula for the ratio of the output power to noise figure of the repeater amplifier

$$S/F = 4 \times 10^{-21} G B (S/\Sigma N)n \tag{II-4}$$

Equations (II-1) and (II-4) are the important design equations for the repeater amplifier.

The factors in (II-2) and (II-3) will now be discussed briefly. (II-3) shows that the noise figure F should be as small as possible. If by improving the equipment the noise figure could be halved, then the signal power S could also be halved (unless interference from other microwave circuits predominate). Later on the noise figure will be discussed further.

The bandwidth B is determined by the characteristics of the signal it is desired to transmit and by the method of transmission. Our aim has been to provide 10-megacycle bands which are sufficient for transmission of standard television signals by AM or low index FM.

An increase in the effective area, A, of the antennas reduces both the total gain required of the amplifier and the output noise power. Crosstalk between the several antennas in a repeater station and interference from outside sources also decrease as the antennas are increased in size because of the increased directivity. Therefore, the antennas should be as large as maintenance and initial costs will permit. Antennas will be discussed in detail in Section III.

Equations (II-2) and (II-3) show that the wavelength λ should be small. Also more frequency space or signal channels may be had at shorter wavelengths. On the other hand, the fading factor M increases somewhat as the wavelength is decreased and, besides, attenuation due to rain sets a lower limit for λ in the region of 5 centimeters. The status of the apparatus art has also been an important factor, but it now permits utilization of the wavelength range extending upward from 3 centimeters. Since the war, our work has been concentrated on a 10% band around 4000 megacycles or $7\frac{1}{2}$ centimeter wavelength. The manner in which this 4000-megacycle band may be divided up into separate channels is explained in Section IV.

The effects of varying the repeater separation d will now be discussed. d appears in the denominator of the exponent of (II-2) which indicates that large separations are favorable, while (II-3) shows that a decrease in separation cuts down the noise. Small separations are very costly, the cost being almost inversely proportional to the separation. We have concluded from propagation studies and site surveys that in the eastern part of the United States it is desirable to use separations of about 30 miles, which generally provide line-of-sight paths with reasonable tower heights. It should be mentioned that the fading allowance factor M is not independent of d; an increased d requires a larger fading factor.

III. Antenna Research*

There are three electrical characteristics which repeater antennas should possess. The first is high gain (large effective area), as this will reduce the path loss and accordingly the requirements on transmitter power. The second is good directional qualities so as to minimize interference from outside sources and also interference between adjacent antennas. The third is a good impedance match so that reflections between the antenna and the repeater equipment will not distort the transmitted signals. These characteristics should preferably be attainable without the imposition of severe mechanical or constructional requirements.

It was felt that a 10-foot round or square antenna would be the largest that maintenance and initial cost would permit. Propagation studies also

* This section prepared by W. E. Kock who performed the major part of the work on antennas.

showed that the variations in angle of arrival of the distant signals would be small compared to the beam width of 10-foot antennas. First experiments were therefore made with 10-foot diameter parabolic "dish" type antennas. The experimental models were made of wood with a metallized reflecting surface consisting of silver conducting paint. Fairly satisfactory tolerances were met in these first models, but it was anticipated that trouble would be experienced in constructing a permanent metal paraboloid of that size to the required tolerances without the use of a heavy and costly supporting means for the parabolic sheet. It was also found that an ice coating a quarter wavelength thick on the reflecting surface, when wet, acted as an effective absorber of power,† since the sheet of water is resistive and is backed up by the reflector. Such a condition could produce an intolerable drop in received signal and would have to be prevented by providing the dish with a plastic cover. As this cover should preferably house the feed also, it would have presented a difficult supporting problem.

Two electrical shortcomings of the paraboloid antenna also presented themselves. First, it was found extremely difficult to obtain a satisfactory impedance match between the antenna and the feed line. This was true partly because of energy reflected from the dish re-entering the feed horn, (this produced a constant 0.6 db standing wave ratio in the feed line), and partly because of the problem of matching the feed horn itself over the desired 400 megacycle band. Secondly, the mutual interference or "crosstalk" between two paraboloids was found to be only 50 to 60 db down when placed back to back** (Fig. III-1).

A type of reflector antenna was later investigated, which, although larger physically than a dish having the same aperture area, overcomes the above two objections.[17] It is shown sketched in Fig. III-2. The photograph of Fig. III-3 shows the antenna lying on its side. It can be seen that the feed is effectively "offset" and reflection back toward the feed is eliminated; the experimental model of Fig. III-3 showed only 0.1 db standing wave ratio in the feed line over a 10% band of frequencies. Furthermore, a horn or "shielded" type feed is used which confines the energy and minimizes stray radiation, and measurements indicate that the back-to-back crosstalk suppression of two such antennas will be high. This long horn is also partly responsible for the excellent impedance match. A horn having a large aperture "matches" free space quite well and the slight mismatch at the throat can be tuned out over a wide band of frequencies. This is not true

† A waveguide termination in common use today employs a resistive sheet placed one-quarter wavelength in front of a conducting plate; this device absorbs practically all the power falling on it.

** Back to back crosstalk suppression in the order of 125 db would be desirable for repeaters receiving and transmitting on the same frequency.

[17] *U. S. Patent* #2,236,393, H. T. Friis and A. C. Beck.

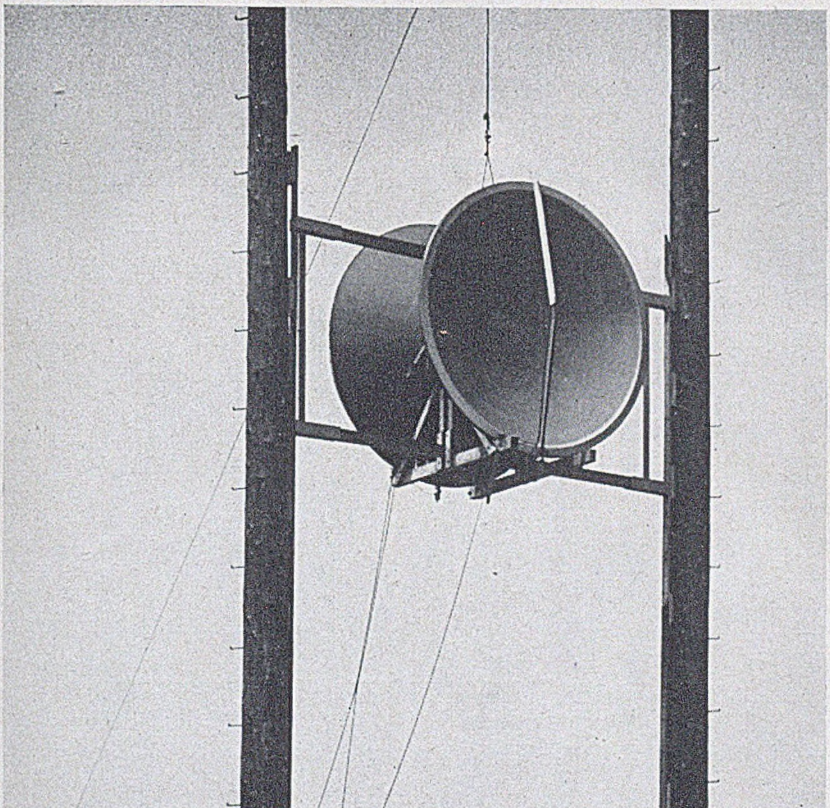


Fig. III-1.—Measuring the back to back crosstalk of two 10 ft. paraboloid antennas.

of the short, small aperture horn used in feeding the dish antenna. The antenna displayed an effective area which was 66% of its actual area which is only 0.9 db below the theoretical maximum.

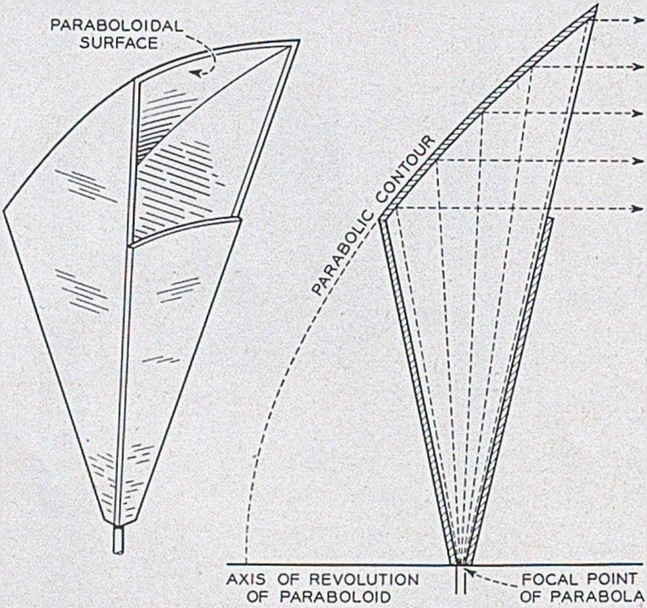


Fig. III-2.—Schematic of horn-reflector antenna.

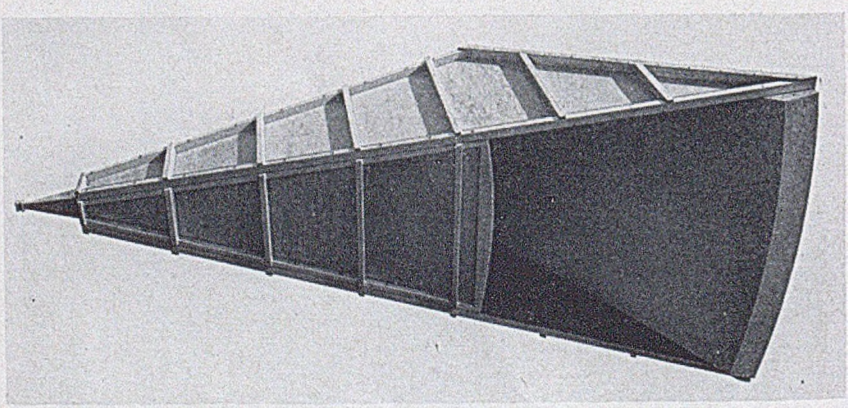


Fig. III-3.—Experimental model of horn reflector antenna.

The expected gain and directional characteristics of an antenna can be realized only if the emerging wave fronts are truly plane. Since deviations greater than $\pm \frac{1}{16}$ wavelength can materially impair the antenna performance,[18] reflector antennas have difficult tolerance requirements imposed upon them. For example, at 4000 megacycles, the 10-foot reflector must conform to parabolic shape to within $\pm \frac{1}{8}$ inches, and any twist or warp

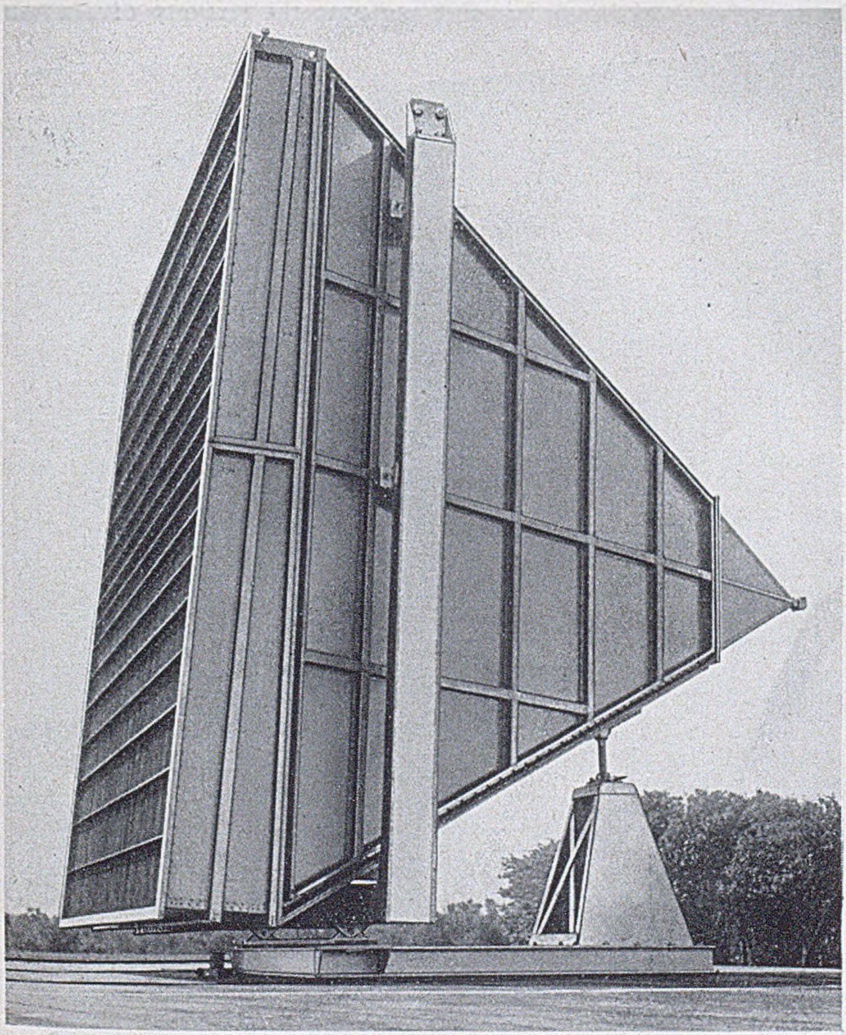


Fig. III-4.—Shielded metallic lens antenna.

of the reflector greater than this would be objectionable. Lenses, however, possess the property that a twist or warp in them does not impair their

[18] See, for example, "Radar Antennas", H. T. Friis and W. D. Lewis, *B. S. T. J.*, vol. 26, p. 219, April 1947, Figs. 17 and 28.

beam-forming properties, and, with the development of metal lenses for microwaves,[19] this type of antenna appeared to lend itself very well to repeater applications. The "shielded" type lens, which is a lens in the mouth of a short horn, is shown in Fig. III-4. This antenna, which was developed for the New York-Boston circuit,* possesses the property of excellent

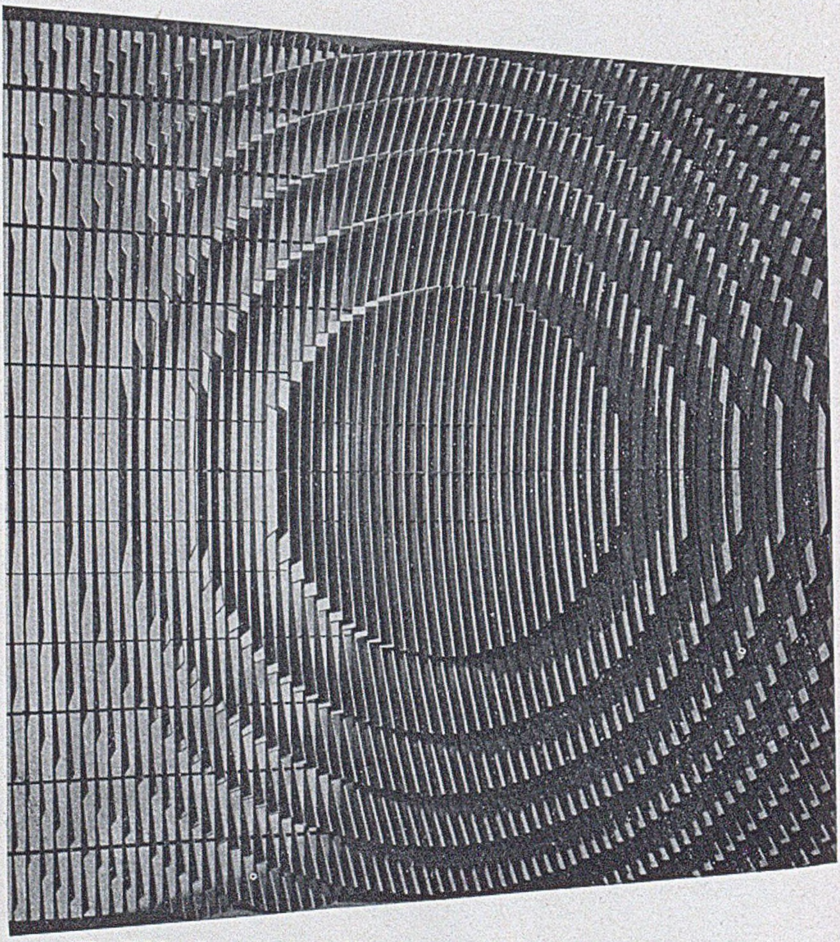


Fig. III-5.—Internal view of lens for the shielded metallic lens antenna.

crosstalk suppression both side to side (85 db) and back to back (125 db). Within the horn, the small amount of energy reflected back from the lens is directed away from the feed by tilting the lens, a procedure which does not noticeably affect the radiation characteristics, but which results in a fairly

[19] W. E. Kock, Metal Lens Antenna, *Proc. I. R. E.*, vol. 34, p. 828, November 1946.
* Developed by W. E. Kock and R. W. Friis.

good impedance match (under 0.8 db standing wave ratio) over the desired 400-megacycle band of frequencies. The lens in the mouth of the horn also provies a convenient support for a plastic impregnated Fiberglass sheet which acts as a weatherproof cover and protects the lens against ice forming between the plates.

The lens itself, Fig. III-5, is based on waveguide principles and causes the wave to be refracted by virtue of the fact that waves confined between plates parallel to the electric vector acquire a phase velocity higher than their free space velocity in accordance with the equation:

$$v_{\text{lens}} = v_{\text{free space}} \Big/ \sqrt{1 - \left(\frac{\lambda}{2a}\right)^2}, \tag{III-1}$$

where λ is the wavelength and, a, the distance between the plates. The index of refraction is thus less than one, and a converging lens must be made concave.

As seen in Fig. III-5, the lens is stepped to reduce its thickness. As a consequence of this stepping, the efficiency at midband of the antenna (50%), is a good deal less than the theoretical value of 81%. Furthermore, the index of refraction varies with wavelength, as seen from equation III-1, and this results in a defocussing of the lens, with a consequent drop in gain, at wavelengths different from the design wavelength. This amounts to a drop in gain of 1.5 db at the edges of a 400 megacycle band; however, its other characteristics of impedance, side lobe suppression (70 db in the two rear quadrants), crosstalk, and ease of construction, help to make up for the gain deficiency.

Measurements taken on the antenna when a thick coating of ice had formed on the plastic cover indicated that the impedance match was impaired (the maximum standing wave ratio increased from .8 db to 1.6 db), but that the gain was not appreciably affected (less than 1 db). Since propagation experiments indicate that severe atmospheric fades are not likely to occur during the winter months, some of the fading allowance can be applied against ice loss.

There was some doubt that the crosstalk figures quoted above could be relied upon during heavy rainfalls, as there was indication that the signal transmitted from one antenna might be reflected from the rain and thus caused to enter an adjacent side-by-side antenna. Measurements during a moderately heavy rain proved that this effect was small, but large enough so that the 85 db side-to-side figure was approaching a limit for the 4000 megacycle band.

The measurement of antenna characteristics involves microwave techniques whose development is an important part of a research program.

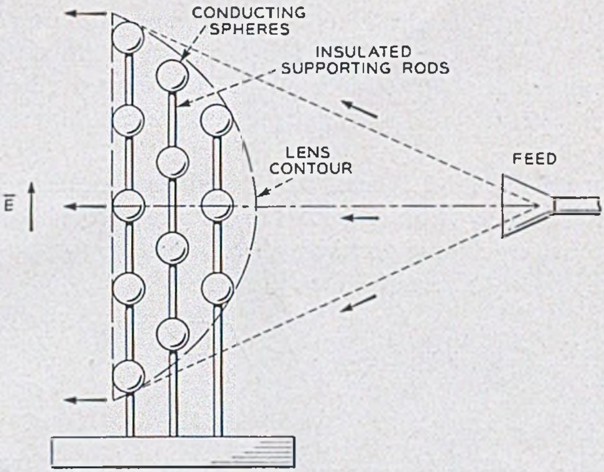


Fig. III-6.—Schematic of metallic delay lens using metal balls as the delay elements.

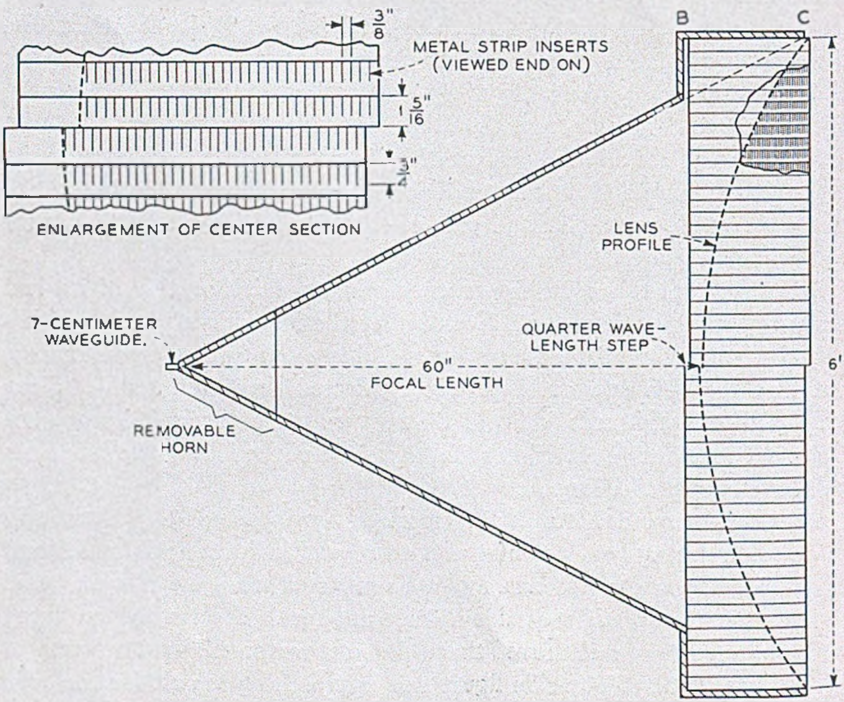


Fig. III-7.—Schematic of metallic delay lens for repeater applications using metal strips as the delay elements.

The antenna measuring methods which were employed in our repeater research follow along the lines of those described in a recent paper.[20] The very large signal ratios of 120 db or more necessitated double detection receivers and low noise figure crystal converters. Pattern and gain measurements of the antennas required measuring sites having large unob-

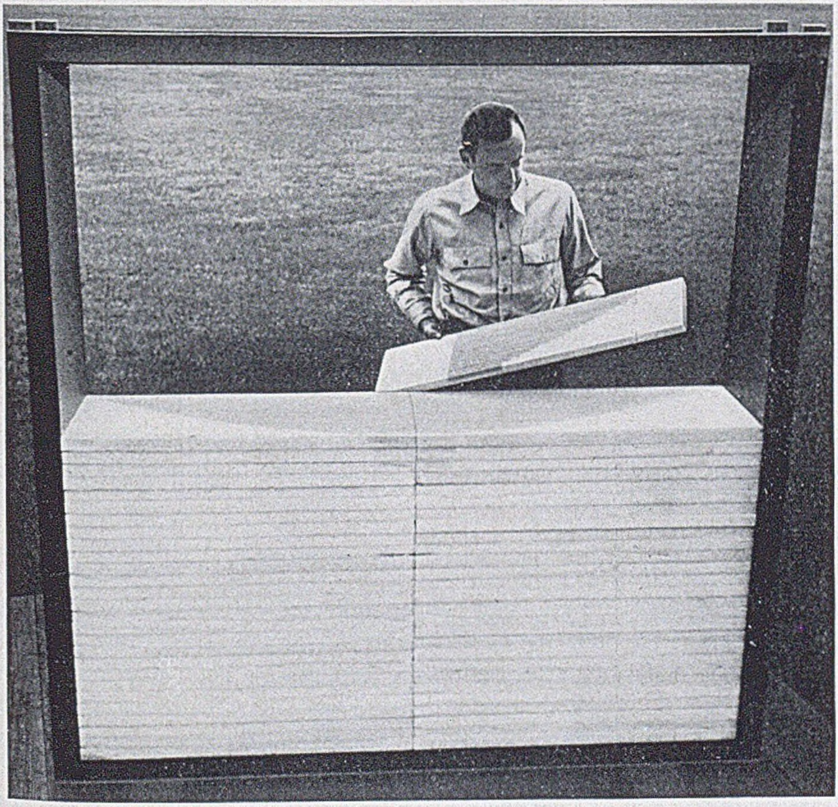


Fig. III-8.—A view of the partly assembled lens of Fig. III-7.

structed areas; these measurements were conveniently taken at the Holmdel Radio Research Laboratory. Impedance measurements involved the usual microwave equipment such as standing wave detectors in waveguide form, signal generators and calibrated receivers.

Research is now underway on an improved metal lens with gain and bandwidth properties which are superior to the lens of Fig. III-5. These lenses[21]

[20] C. C. Cutler, A. P. King, W. E. Kock, "Microwave Antenna Measurements", *Proc. I. R. E.*, vol. 35, No. 12, pp. 1462–1471, December 1947.

[21] Winston E. Kock, "Metallic Delay Lenses", *B. S. T. J.*, vol. 27, pp. 58–82, January 1948.

employ an artificial dielectric material which duplicates, on a much larger scale, processes occurring in a true dielectric. This involves arranging conducting elements in a three dimensional array or lattice structure to simulate the crystalline lattice of the true dielectric. Such an array responds to radio waves just as a molecular lattice responds to light waves, and if the spacing and size of the elements is small compared to the wavelength, the index of refraction is substantially constant, so that lenses made of this material are effective over large wavelength bands.

A lens employing conducting spheres as the lattice elements is sketched in Fig. III-6. A more convenient structure for large lenses is shown in Fig. III-7 and III-8; it uses thin metallic strips, with the width dimension parallel to the electric vector. Slotted polystyrene foam sheets support the strips and they are stacked up to form the lens. A quarter wavelength step in the lens causes the reflections from the lens surfaces to cancel at the feed point, which, in the drawing, is the apex of the horn shield.

Over a 10% wavelength band, a 6 foot square shielded lens antenna of this type exhibited an efficiency of better than 60% and the impedance mismatch due to the lens produces only a 0.2 db standing wave ratio in the feed line. This antenna thus retains the dimensional tolerance, weight, size and crosstalk advantages of the shielded lens over the shielded reflector, and has the advantage of higher gain and broader band performance over the shielded metal plate lens.

IV. Filter Research*

Frequency space for common carrier radio relay systems is available in blocks several hundred megacycles wide. Where heavy traffic is to be carried such bands must be efficiently exploited. This may in time be accomplished by using extremely wide band amplifiers, for example traveling wave tubes; however, more immediate success is offered by the possibility of operating a number of narrower band circuits of different frequencies. This could be done by using a separate transmitting and receiving antenna for each circuit. But each antenna, for sound technical reasons, must be large and expensive and in addition requires adequate tower support. Consequently there is a need for filters which can connect a number of individual radio channels to a common antenna.

The design[22] of these radio frequency branching filters must be coordinated with the design of the relay system as a whole. At lower frequencies where little or no antenna crosstalk protection can be counted on it is natural

* This section prepared by W. D. Lewis and L. C. Tillotson who were responsible for a large part of the research on filters.

[22] For more detailed discussion see, W. D. Lewis and L. C. Tillotson "A Constant Resistance Branching Filter for Microwaves," *B. S. T. J.*, vol. 27, pp. 83–95, Jan. 1948.

to lump transmitting frequencies in one group and receiving frequencies in another. When separate microwave shielded lens antennas are employed for transmitting and receiving in each direction it becomes practical to use a frequency plan in which transmitting and receiving frequencies are interleaved. Such a plan eases filter requirements considerably and has

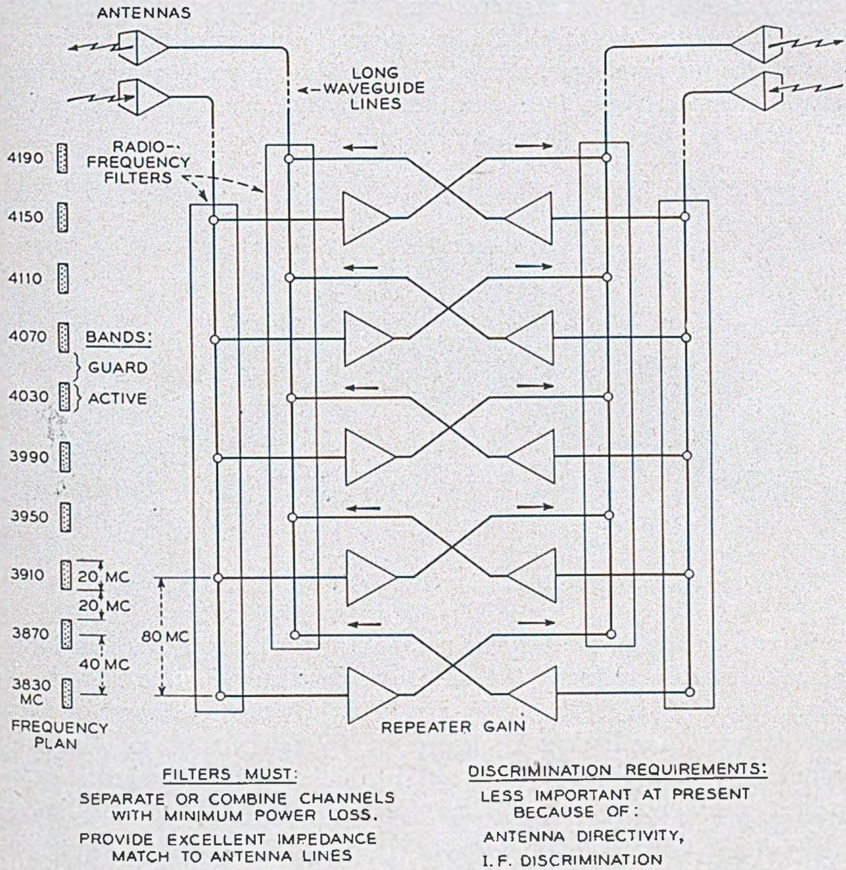


Fig. IV-1.—Schematic diagram of a possible five channel radio repeater station.

certain other advantages to be discussed in Section V. A possible repeater employing such a frequency scheme is illustrated schematically in Fig. IV-1.

If a radio frequency branching filter is to fit properly into a repeater it must separate or combine channels without excessive loss of signal. In addition it must provide an excellent match to the long line which leads to the antenna, otherwise troublesome echoes in this line may be caused.

Because of IF amplifier band-pass characteristics, suppression requirements on the filter, except possibly in the vicinity of receiver image bands, are not large.

Microwave band-pass filters consisting of one or more cavities arranged in sequence along a waveguide have been known for some time. The frequency, bandwidth and discrimination characteristics of such filters can all be chosen within wide limits by appropriate design of the cavities and the means for coupling to them. These filters are analogous to lumped-circuit channel-passing filters and can in principle, like them, be connected in groups to provide a branching network.

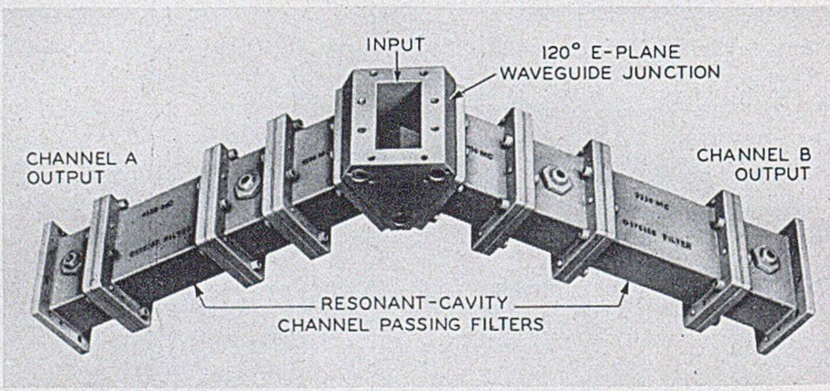


Fig. IV-2.—Photograph of a branching filter for an experimental radio relay system.

Several successful two-branch networks have been designed and constructed in this manner. One of these, developed for the New York-Boston circuit*, is illustrated in Fig. IV-2. Here two three-cavity filters are connected to an E plane Y junction, the waveguide analogue of a series connection. The two filters are tuned to different bands and each is connected to the junction through a line of length such that it causes no disturbance in the channel of the other. The electrical characteristics of this filter are plotted in Fig. IV-3.

Problems connected with the design of suitable microwave branching filters with more than two branches evidently differ considerably from previous filter problems. Channel passing networks which can be connected in series or parallel to form a channel branching filter can be designed at lower frequencies on the basis of lumped circuit theory and built of coils, condensers and resistances, but in the microwave region simple elements

* Developed by the group concerned with high-frequency filter design headed by A. R. D' heedene. A large part of the research underlying the design of these filters was performed by W. W. Mumford. Prior to the war a considerable amount of research on the band-pass type of waveguide filter had already been done by A. G. Fox.

and connections do not exist. Where more than two waveguide channel passing filters are to be connected to a common junction the design becomes complex, since in every channel the sum of the interactions of all the inactive filters on transmission through the active filter must be zero. It is evidently not easy to satisfy this condition, particularly since in doing so one must take account of the change with frequency of the effective length of all waveguide connecting lines. And even if such a solution is found it will be valid for only one set of channels, so that the problem must be solved all over again for every change in channel arrangement.

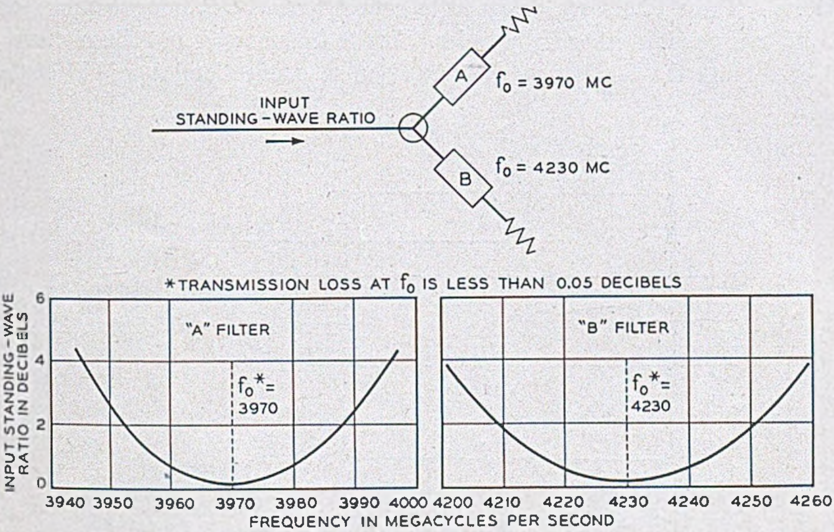


Fig. IV-3.—Input standing wave ratios of filter of Fig. IV-2.

As a result of these difficulties and after a few attempts to overcome them it became evident that a more flexible microwave branching filter technique should be found. Accordingly a solution on an iterative basis was developed. A channel dropping circuit was devised which, when inserted in a line, could extract or insert one channel while allowing others to pass through without disturbance. This circuit is of the constant resistance type; in other words it operates by diverting energy selectively but not by reflecting it back to the input. Consequently N such circuits placed in sequence do not interact reflectively; they, thus, form an N channel branching filter which is also of the constant resistance type.

An individual constant resistance channel dropping circuit is illustrated schematically in Fig. IV-4. It is made up of two hybrid[23] circuits, two

[23] For a general discussion of hybrid circuits see W. A. Tyrrell, "Hybrid Circuits for Microwaves", *Proc. I. R. E.*, vol. 35, No. 11, pp. 1294–1306, November 1947.

identical band reflection filters, and two quarter wavelengths of line. Each of the hybrids is analogous to a low-frequency hybrid coil and operates as follows. A wave in line C (See Fig. IV-4) incident on the hybrid is divided equally and with equal phase into A and B but does not appear in D or reappear in C. If waves in A and B are incident on the hybrid a wave proportional to their vector sum will appear in C, a wave proportional to their vector difference will appear in D but nothing will reappear in A or B. A wave in the input line incident on the channel dropping circuit will thus be divided by the input line into the lines leading to the two band reflection filters. These filters are designed to reflect frequencies lying within their band and pass all other frequencies. If the frequency is outside of the reflected band the two waves will travel on to connections A and B of the

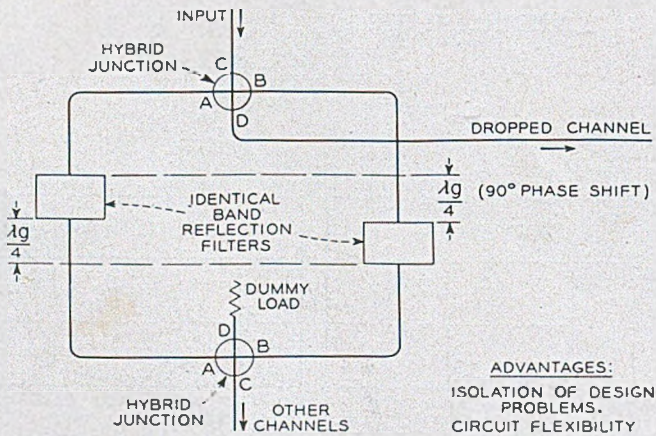


Fig. IV-4.—Schematic diagram of a constant impedance channel dropping filter using hybrid junctions and band reflection filters.

output hybrid. Here they will have equal phase and amplitude, their vector difference will be zero and the wave appearing in C of the output hybrid and consequently in the output line will contain all the power. If the frequency lies within the band of the reflection filters the two waves will be reflected by them and will travel back to the connections A and B of the input hybrid. The two waves strike these connections with opposite phase since one of them has traveled twice over an extra quarter wavelength of line. Their vector sum will consequently be zero and the wave which appears in terminal D of the input hybrid and consequently in the dropped channel line will contain all the power. The circuit of Fig. IV-4 is therefore a constant resistance channel dropping network which diverts energy lying within the band of the reflection filters but allows all other energy to pass through without disturbance. Conversely, by the law of reciprocity, this circuit can insert energy lying within the band of the reflection filters into the main line without disturbing any energy passing through it at other frequencies.

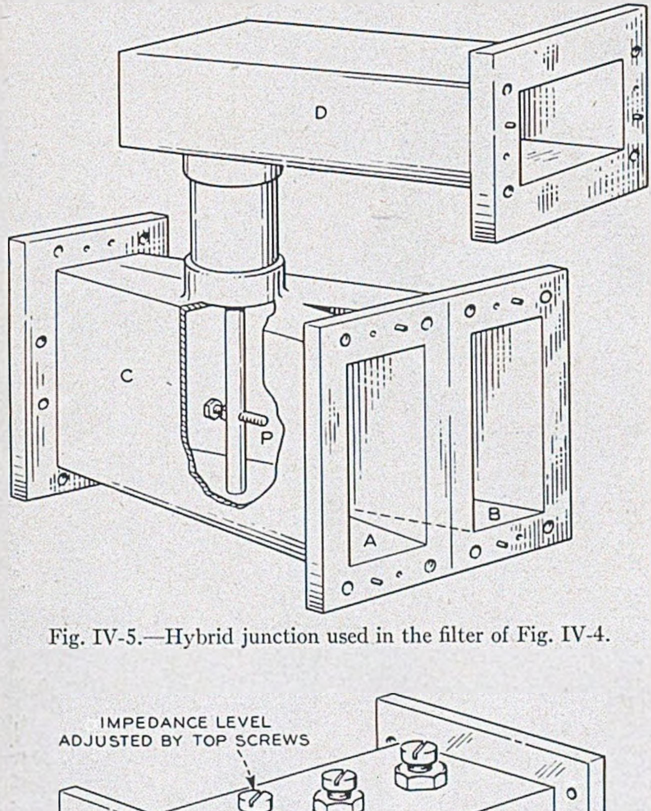


Fig. IV-5.—Hybrid junction used in the filter of Fig. IV-4.

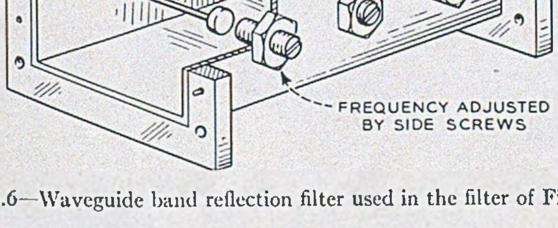


Fig. IV-.6—Waveguide band reflection filter used in the filter of Fig. IV-4.

An embodiment of the circuit of Fig. IV-4 suitable for use in the repeater arrangement of Fig. IV-1 has been constructed and tested. Figure IV-5 illustrates the waveguide hybrid employed. Here the waveguide opening

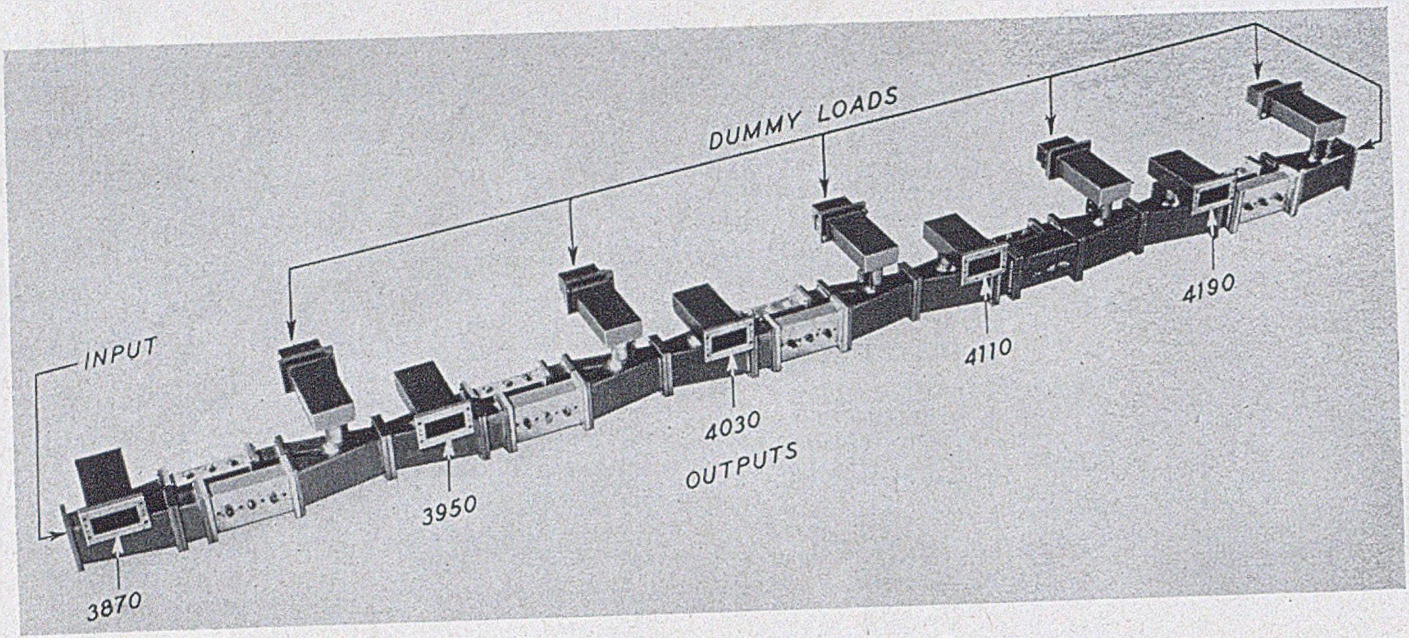


Fig. IV-7.—Photograph of a five channel branching filter.

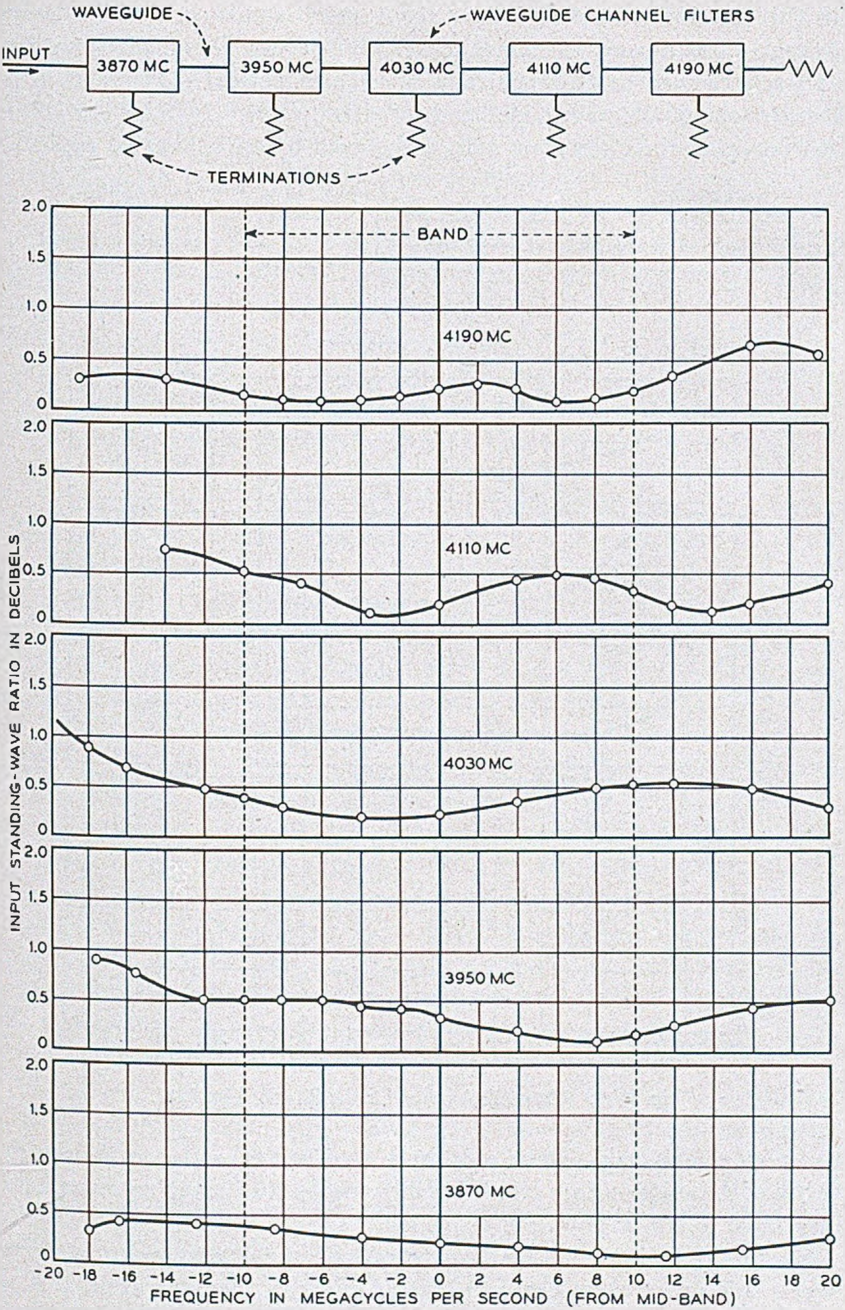
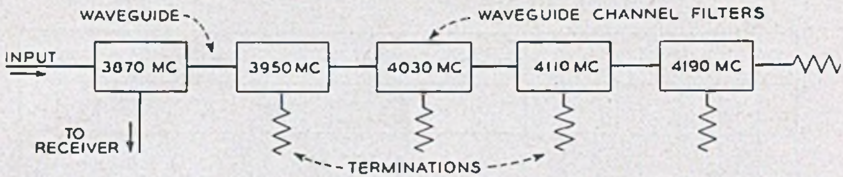


Fig. IV-8.—Input standing wave ratios for the filter of Fig. IV-7.

for C is physically parallel to those for A and B and is connected to them through a broad-band waveguide taper. Connection D is made through a relatively narrow band coaxial and probe arrangement. Figure IV-6 illustrates one of the waveguide band reflection filters. In this filter reflection occurs at three resonant rods, each tuned by an adjustable capacita-



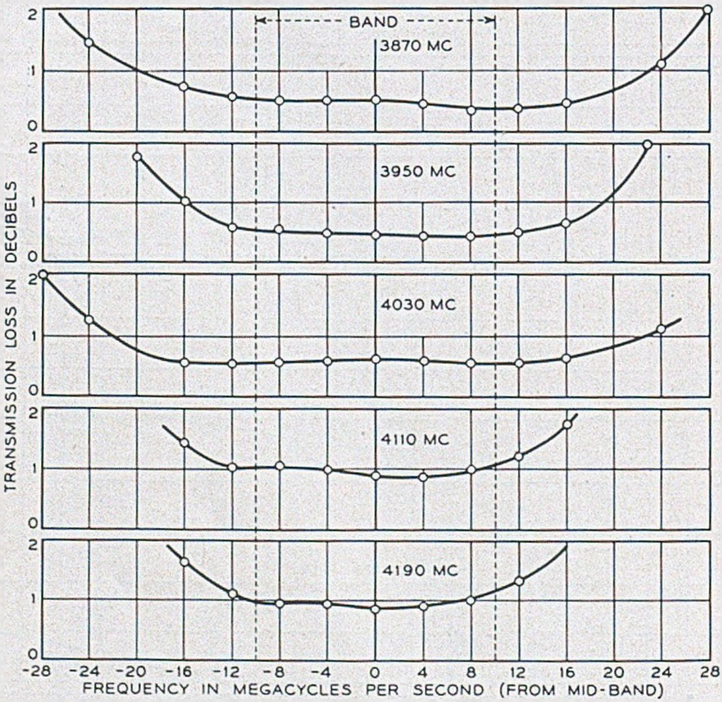


Fig. IV-9.—Transmission loss for the filter of Fig. IV-7.

tive plug at one end. These rods are placed perpendicular to the electric vector of the guide and are coupled to it by means of adjustable screws placed over them in the wide wall of the guide.

Channel dropping units for five channels in the 4,000 megacycle region were made up of these components and suitable quarter wavelengths of guide. These units were connected in sequence as shown in Fig. IV-7 and adjusted systematically. The electrical performance of the resulting five-channel branching filter was measured with a double detection measuring set and is plotted in Figs. IV-8 and IV-9.

These measured electrical characteristics serve as a check on the general theoretical ideas concerning constant resistance hybrid branching networks. They also indicate that when these ideas are embodied in the form illustrated in Fig. IV-7 the result is a branching filter which can be used in currently planned radio relay circuits.

The circuit of Fig. IV-7 provides a satisfactory channel splitting network. It does not, however, provide consistent high off-frequency discrimination between one of the channel output terminals and the other terminals of the filter. When systems requirements* are such that extra discrimination or special impedance behavior is required, this can be supplied by inserting suitably designed reflecting filters in the branch lines. These added filters will not interact reflectively with the branching filter.

V. The Repeater Amplifier†

In a microwave repeater circuit, Fig. II-1, the signal is amplified at each repeater to compensate for the transmission loss in the preceding radio path. Since we cannot build perfect amplifiers, the signal will not appear at the output terminals as a true replica of the input signal; the circuit will distort the shape of the signal and it will also add noise. Therefore, the main objectives in amplifier work have been to keep the signal distortion and the added noise within certain requirements.

To be more specific, the repeater amplifier in a relay system must be capable of supplying a maximum gain, G, as given by the equation II-1; it must have a ratio of output power capacity to noise figure which will meet the signal to noise ratio requirements of the system as given by equation II-4; since distortionless transmission is desired, it must have an amplitude characteristic as flat as possible and a phase characteristic as linear as possible over the essential range of frequencies of the signal it is desired to transmit;[21] and it must be equipped with an automatic gain regulating circuit to hold the output power constant over the expected range of input levels.

The simplest relay amplifier would be one which amplifies the signal and sends it on without a change in frequency. However, two major considerations indicated that early repeater amplifiers could not be so simple. No

* E.g., the converter may require a reflection in the input line at the image frequency, See Section V and Fig. V-4.

† Those parts of this section dealing with the general layout, the requirements, and the over-all testing of the repeater amplifier were prepared by D. H. Ring, who together with A. C. Beck did the work on this phase of the problem.

[21] Sallie Pero Mead, "Phase Distortion and Phase Distortion Correction", *B. S. T. J.*, vol. VII, No. 2, pp. 195–224, April 1928.

microwave amplifiers were known which gave promise of yielding an adequate ratio of output power capacity to noise figure, and there was considerable doubt of our ability to reduce, sufficiently, the feedback from the transmitting antenna to the receiving antenna.

These difficulties with straight-through amplification can be avoided by a repeater amplifier such as is shown schematically in Fig. V-1. The incoming signal is converted to an intermediate frequency, IF, where better amplifiers are available and where the major part of the required gain is supplied. The amplified IF signal is then converted back to the microwave frequency $f + \Delta f$, where Δf is relatively small. The difference Δf between the incoming and outgoing frequencies permits the use of circuit selectivity to counteract feedback troubles, and the radio frequency amplifier following the transmitting converter can have a relatively large noise figure.

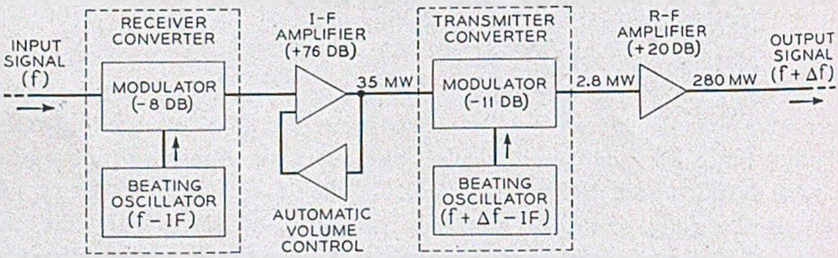


Fig. V-1.—Schematic of a repeater amplifier.

Our initial research on microwave repeaters was directed toward solving the problems associated with an amplifier of the type shown in Fig. V-1. The gain and level figures shown in this figure apply to the example of an eight-link relay system given in section II. They indicate approximate minimum objectives for the various components of the repeater amplifier to be discussed later.

Choice of I.F. Frequency—When selecting the intermediate frequency for a multichannel repeater circuit utilizing the interleaved radio frequency plan of Fig. IV-1 and the intermediate frequency type repeater amplifiers of Fig. V-1, the relative position of the various discrete frequencies and frequency bands shown in Fig. V-2 must be considered. In order to minimize the possibility of crosstalk from the image bands and interference from the beating oscillators caused by insufficient shielding, it is desirable to choose the intermediate frequency in such a way that the image bands fall midway between the active bands, and the oscillator frequencies fall midway between the image and active bands. These conditions are realized, as shown in Fig. V-2, if the intermediate frequency satisfies the relation

$$2\ \mathrm{IF} = n\Delta f + \frac{\Delta f}{2}$$

or

$$\mathrm{IF} = \frac{\Delta f}{4}(2n + 1)$$

where Δf is the frequency spacing and n is any integer greater than zero.

In general, better noise figures and circuit stability are obtained with low intermediate frequencies, while high intermediate frequencies lead to

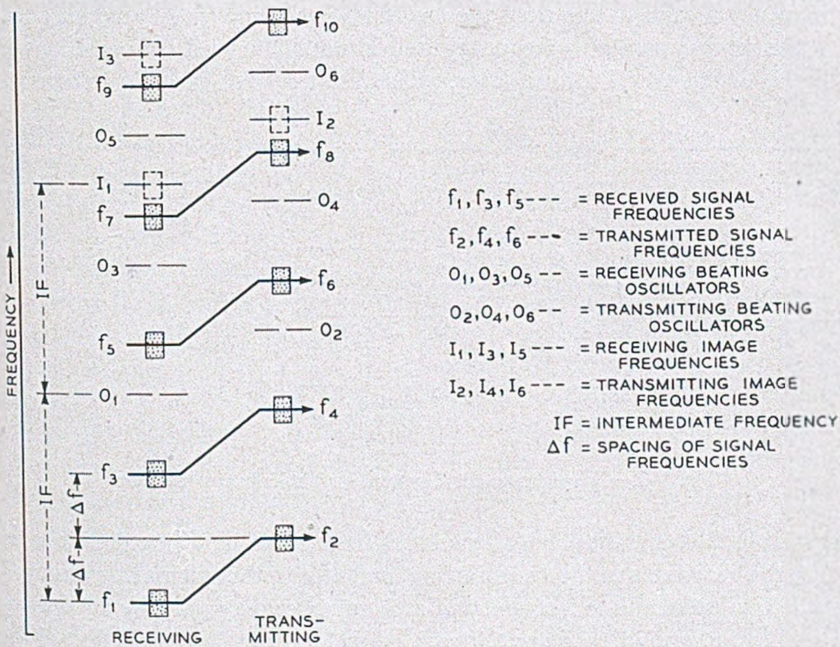


Fig. V-2.—Frequency spectrum of a multichannel microwave repeater circuit.

more symmetrical amplitude and phase characteristics. The research on the intermediate frequency components described below was conducted in the 60 to 70-megacycle range.

Frequency Stability—If all receiving and transmitting beating oscillators are independently controlled in a multichannel relay circuit of this kind, there is a possibility that small variations will add to produce large variations at the distant end of the system. This difficulty can be overcome by the frequency control system shown schematically in Fig. V-3. The transmitting and receiving beating frequencies are both derived from an oscillator operating at a frequency suitable for receiving the incoming signal. The frequency of this oscillator is controlled by a high Q cavity and a servo mechanism to 0.2 megacycles or better.[25] One portion of the output of the oscillator is used as the beat frequency in the receiving converter. A second portion is combined with the output from a crystal oscillator operating at a frequency equal to the difference, Δf, between the incoming and outgoing frequencies. In this way a beat frequency for the transmitting converter is obtained which has the same variations as that for the receiving converter except for negligibly small variations that may occur in the crystal oscillator. As a result of this method of deriving the beat frequencies the outgoing frequency always differs from the incoming frequency by an amount equal to the crystal oscillator frequency and is not influenced by variations in the high-frequency local oscillator. The result is that, except for the small

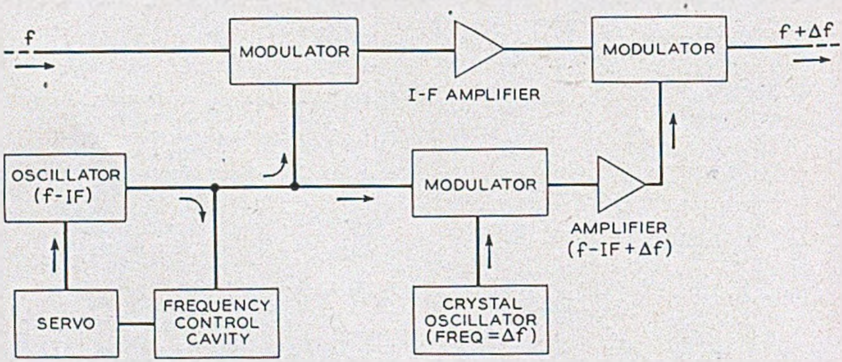


Fig. V-3.—Frequency control system for a microwave repeater.

crystal oscillator variations, all radiated frequencies in a long circuit carry only the variations in the transmitting oscillator of the originating terminal, while the intermediate frequency of each repeater may vary by an amount equal to the sum of the variation of its own local oscillator and that of the terminal transmitter frequency.

Automatic Gain Regulation—The function of the automatic gain control circuits is to hold the repeater output constant over the expected fading range. As already stated an allowance for fades 20 db down and 10 db up from free space have been made for 30-mile paths at a wavelength of about 7 centimeters. In addition to knowledge of the fading range to be compensated, it is necessary in the design of suitable circuits to know what the maximum fading rate is likely to be. Analysis of the records of a number of disturbed periods on the New York-Neshanic path indicate a maximum rate of 5 db per second.

[25] V. C. Rideout, "Automatic Frequency Control of Microwave Oscillators", *Proc. I. R. E.*, vol. 35, pp. 767–771, August 1947.

A conventional delayed automatic gain control circuit has been used in which a d.c. voltage, supplied by a peak rectifier in the output of the last IF stage, is amplified and fed back to bias several of the IF stages. Analysis of the transient response of the last repeater in response to a step function disturbance in the input to the first repeater, where the number of repeaters is of the order of 5 or more, shows that great care must be taken in shaping the frequency characteristic of the feedback circuit.

General Requirements of Components—The important factors bearing on the basic layout of the amplifier components have been discussed. Others, affecting the design of these various components will now be considered, after which a brief description of the research work on each component will be given.

Different repeater circuits will, in general, have different numbers of repeaters; also it may be necessary to feed signals into and extract signals from them at any point. Under these conditions it is impractical to specify only the over-all characteristics for a given number of repeaters. Each repeater itself should be individually good and should not depend upon any systematic compensation or equalization at any other point in the system. The repeater was designed in accordance with this philosophy and, in the interest of flexibility and ease of testing, the same line of reasoning was extended to cover the design of the components of the repeater as well.

In accordance with the above considerations major emphasis was placed on obtaining a minimum of amplitude variation and phase distortion over a 10-megacycle band for each repeater component. However, it was appreciated that even with the simplest circuits the inherent phase distortion would be excessive in long relay systems. Phase equalizers can be designed to equalize this distortion, but the difficulties of design and alignment increase with the magnitude of the distortion to be equalized. Accordingly, simple circuits were used wherever gain requirements would permit and at the same time parallel research was carried out on the problems of designing and testing appropriate phase equalizers.

While our aim was to provide a repeater 10 megacycles wide suitable for any type of modulation, it soon became apparent that it would be uneconomical to attempt to provide the extreme degree of linearity that would be required; for example, for a long relay circuit carrying an amplitude-modulated television signal. However, early tests indicated that very satisfactory transmission could be obtained with low-index frequency-modulated signals, for which reason the later stages of the work were aimed at providing an amplifier to be used for such signals. Nevertheless an attempt was made to limit the compression in each unit except the R.F. amplifier to 0.1 db at maximum rated load.

A further important requirement placed on all components was that their input and output impedances should match the corresponding impedances of the components to which they were to be connected with a minimum of reflected power over the 10-megacycle band. This requirement was imposed on the separate units to provide for flexibility in testing and to permit easy patching in of spare units in case of failure.

Receiving Converter*

The receiving converter, together with the input to the first stage of the intermediate-frequency amplifier, occupies a unique position in the repeater amplifier in that it is located at that point in the circuit where the signal level is lowest. As a consequence, research on receiving converters has been directed toward insuring that the least possible noise be added to the signal to be amplified. An extensive background of microwave converter design information was available from the work done on converters during the war[26,27] which led to the selection of a balanced converter using a waveguide hybrid junction and 1N23-B silicon point contact rectifiers. The information already available would probably have been adequate except for the two additional requirements imposed by the repeater amplifier: first, that the standing wave ratio at the input have a low value and, second, that uniform conversion efficiency be maintained over a band of at least 10 megacycles.

A receiving converter with its associated input and output circuits is shown in Fig. V-4. Frequency conversion is accomplished in a device of this kind by virtue of the fact that when two sinusoidal voltages (in this case the signal and the beating oscillator) are applied to a non-linear impedance such as a rectifier, new frequencies given by the sums and differences of the applied frequencies and their harmonics are generated. The difference frequency may thus be selected as the desired output frequency and passed through the IF transformers to the IF amplifier. The performance of a converter is, however, influenced by the impedances encountered by some of the other frequencies generated. For this reason, the separation S between the input filter, more properly a component of the channel selecting network, but here shown as part of the converter, and the converter must be given consideration, since it determines the phase of the reflection from the filter at the image frequency (the difference between the beating

* This section prepared by C. F. Edwards who was responsible for the research on the receiving converters.

[26] "Developments of Silicon Crystal Rectifiers for Microwave Radar Receivers", J. H. Scaff and R. S. Ohl, *B. S. T. J.*, vol. 26, pp. 1–30, January 1947.

[27] Descriptions of several converters developed prior to and during the war as well as a more complete description of the present converter are given in C. F. Edwards' paper, "Microwave Converters", *Proc. I. R. E.*, vol. 35, No. 11, pp. 1181–1191, November 1947.

oscillator second harmonic and the signal) which in turn affects the converter output impedance and hence the match between the rectifiers and the IF transformers. Failure to obtain a proper match results in non-uniform transmission over the channel band.

In addition, to maintain uniform conversion efficiency over a wide band, it is necessary to control the impedance encountered by frequencies near the beating oscillator second harmonic. This is done by means of the harmonic

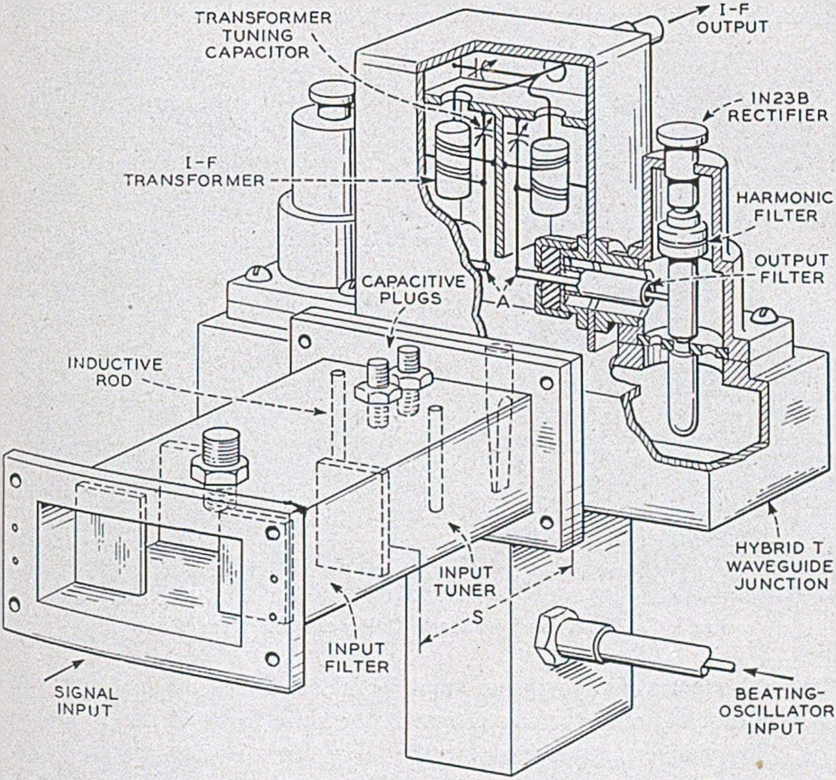


Fig. V-4.—Receiving converter with input filter.

filters shown which reflect these frequencies from a point close to the rectifier. Since the point of reflection is close, there is little opportunity for the phase of the reflection, and hence the conversion loss, to vary over the band of frequencies to be converted.

The converter in Fig. V-4 was designed to provide as good a termination as possible to the incoming waveguide over the band of channel frequencies so that a minimum of additional adjustment would be required to match the guide accurately at any particular channel frequency. This additional adjustment is provided by the input tuner shown. With it, the reflection coefficient may be reduced to zero at the desired channel band center, and when this is done the standing wave ratio at the input is less than 2 db over a 20-megacycle band.

The bandwidth of the converter is largely determined by the IF transformers shown which transform the balanced output impedance to the unbalanced 75-ohm coaxial line connecting to the IF amplifier. When transformers having a coupling coefficient of about 0.5 are used, transmission variations of less than 0.1 db over a 20-megacycle band are obtained.

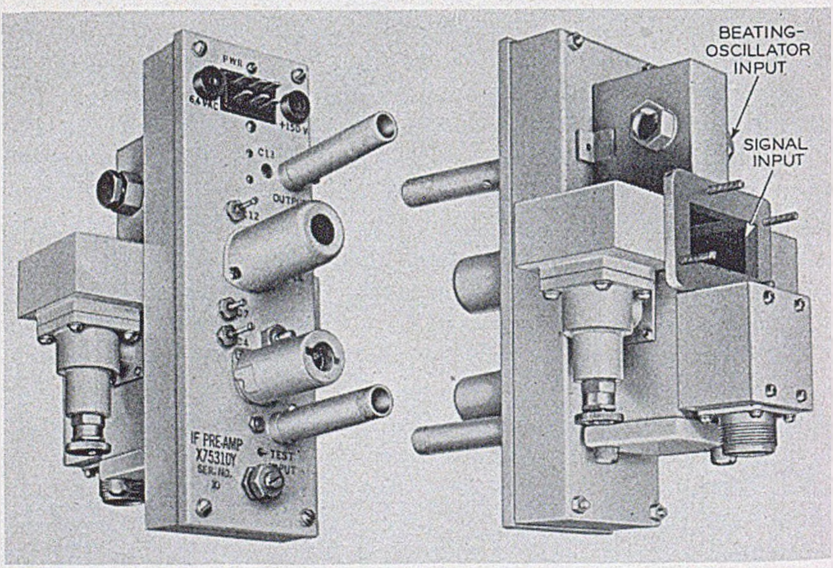


Fig. V-5.—Receiving converter with low noise IF preamplifier.

The noise figure of the repeater amplifier is determined by the conversion loss and noise figure of the converter and the noise figure of the IF amplifier.[16] The converter designed for the New York-Boston circuit has a conversion loss of about $5\frac{3}{4}$ db and its noise figure is 10 db. Thus when it is used with an IF amplifier having a noise figure of 7 db, a figure of 14 db is obtained for the over-all noise figure of the repeater amplifier. Figure V-5 shows this converter with a low noise preamplifier attached.*

I.F. Amplifier†

Most of the gain of the IF amplifier is obtained with stages using double tuned, symmetrically loaded (or 'matched") interstage transformers, de-

[16] Loc. cit.
* Developed by H. C. Foreman and B. C. Bellows, Jr.
† This section prepared by Karl G. Jansky who, together with V. C. Rideout, did the work on IF amplifiers.

signed in accordance with the formulas given by a former member of this laboratory.[28] In Fig. V-6, "a" shows a schematic diagram of such a transformer and "b" shows the equivalent π network generally used. The equivalent T network shown at "c" has also been used. In Fig. V-7 "a" shows the theoretical band-pass characteristic for this type of transformer with a coupling coefficient of 0.5 which was the value used in most stages. The circles indicate points measured on a typical transformer. This matched network is relatively insensitive to small changes in capacitance so that wherever it is used it is possible to change tubes without having to realign the amplifier. The delay distortion for this type of network is, as shown by

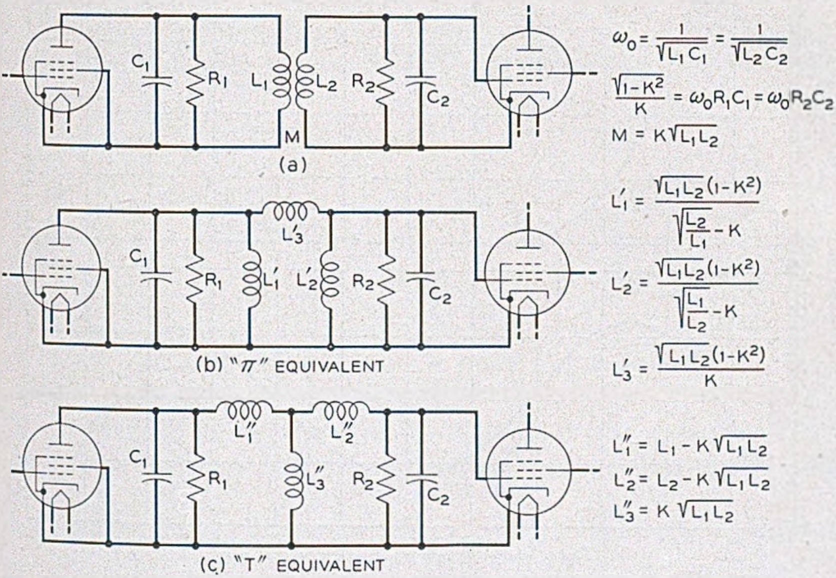


Fig. V-6.—The double tuned IF transformer.

"b" in this figure, also relatively small. The gain per stage for a coupling coefficient of 0.5 is approximately 6 db for 6AK5 vacuum tubes and about 12 db for the recently developed WE 404-A tubes. As shown by W. J. Albersheim[28] it is possible to design circuits which will give more gain than the "matched"transformer, but only at the cost of increased sensitivity to capacitance changes. For convenience the IF amplifier is usually divided into a preamplifier closely associated with the receiving converter and a main IF amplifier.

Pre-amplifier—At the time work was begun, noise figures of the order of

[28] V. C. Rideout, "Design of Parallel Tuned Transformers", *B. S. T. J.*, vol. 27, pp. 96–108, January 1948.

12 to 14 db were obtainable for 65 mc amplifiers of the required bandwidth with 6AK5 tubes and matched input transformers. This was much worse than was desired. By using a 6J4 close spaced grounded grid triode in the input stage, much lower noise figures were obtained, but the gain with matched input and output circuits was so low (approximately 3 db for a coupling coefficient of 0.5) that the noise from the following stage contributed considerably to the overall noise figure.

By removing the loading resistance on the output side of the first interstage transformer and reducing the coupling coefficient to 0.3, the gain was

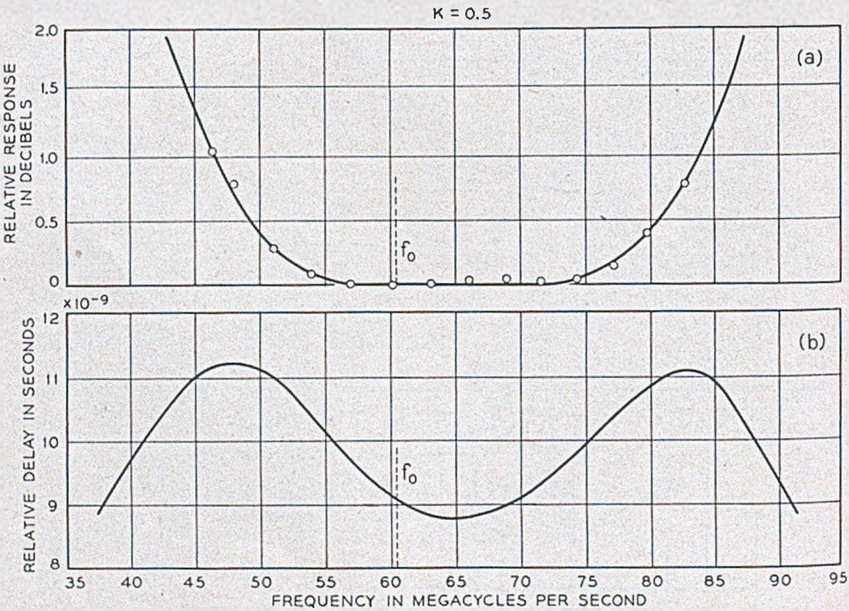


Fig. V-7.—Band pass and delay characteristics of the IF transformer.

raised sufficiently to give an overall noise figure of the order of 7 db. Figure V-5 shows the preamplifier developed for the New York-Boston circuit. This amplifier employs a 6J4 and a WE 404-A and provides a gain of 23 db.

The 6J4 tube, when used in a grounded grid circuit, has an input impedance of approximately 85 ohms which is close to the desired 75 ohm impedance. When a good match is required, it is necessary to use an input transformer, but it should be noted that an improvement in the noise figure can be obtained by deliberately producing a mismatch in the right direction at the input. Noise figures as low as 4 db have been obtained in this manner with recent experimental tubes.

Figure V-8 is a schematic diagram of an amplifier with a very low noise figure consisting of two of these experimental grounded grid tubes in tandem. The circuits were designed so that there is a mismatch at the input of each tube. The overall noise figure is 4.0 db with the amplifier connected between a 75-ohm generator and a 75-ohm load; the gain is $17\frac{3}{4}$ db; and the bandwidth at the 0.1 db down points is about 13.5 megacycles.

Main IF Amplifier—In order to obtain the required output power over the desired bandwidth it was necessary, at the beginning of this work, to use the WE 367-A tube in the output stage. Since the gain of a stage using this tube is very low, it must be driven by a tube similar to the 6AG7 and to prevent compression in this latter tube a special high gain, triple-tuned interstage transformer was designed, with a relatively low coupling coefficient.

Methods of paralleling tubes to get more power output or the same power output with a much wider bandwidth have been worked out, but recent

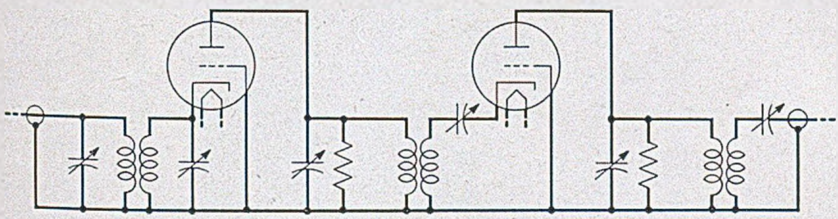


Fig. V-8.—Schematic of a "low noise figure" preamplifier.

developments in power output tubes and in transmitting converters have made them unnecessary, except for applications requiring very wide bandwidths.

Automatic gain control can be applied to the IF amplifiers by controlling the grid bias of the various stages. However, it is not advisable to apply the gain control bias to either of those stages which are preceded by the special high-gain transformers. In these stages the slight changes in input impedance of the tubes caused by variations in the grid bias would be sufficient to alter significantly the band-pass characteristics of the transformer.

Figure V-9 shows an amplifier with about 55 db gain developed for the New York-Boston circuit.* Automatic gain control bias is applied to the grids of the first three stages which employ wide band (K = 0.7) matched interstage transformers.

It will be noted that the first interstage transformer of the preamplifier and the last interstage transformer of the main IF amplifier require low coupling coefficients to obtain the gain desired at these points. For this

* Developed by A. L. Hopper and B. C. Bellows, Jr.

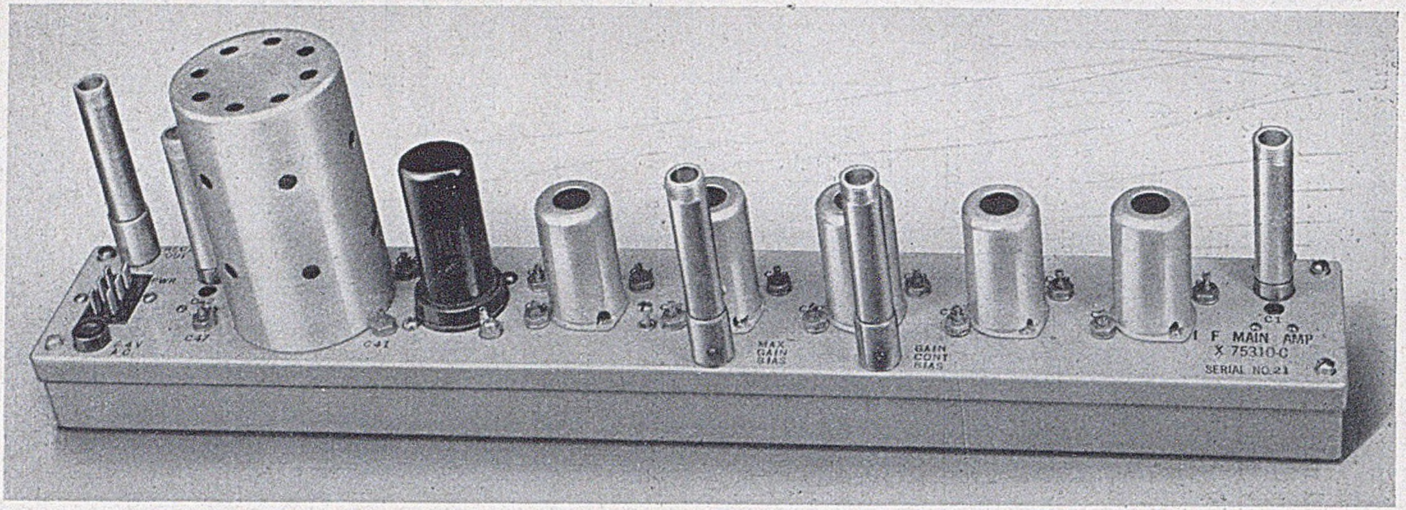


Fig. V-9.—The main If amplifier.

reason, these two transformers largely determine the band pass and delay distortion of the whole amplifier. Typical overall characteristics for a complete IF amplifier are shown in Fig. V-10.

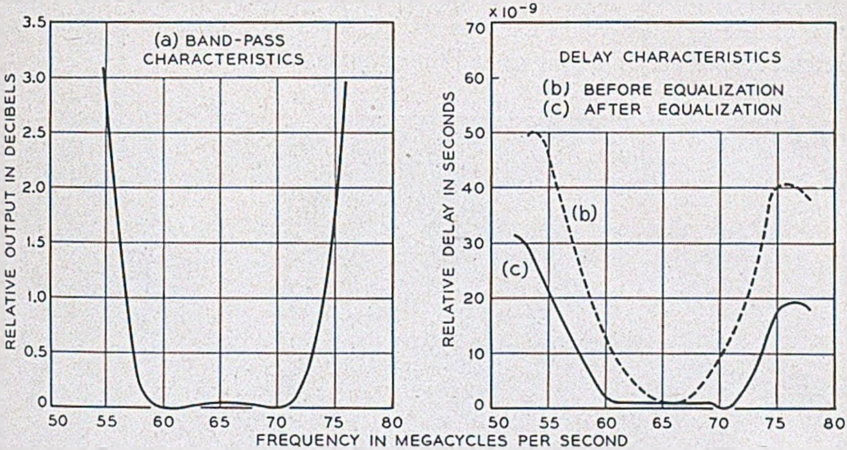


Fig. V-10.—Band pass and delay characteristics of complete IF amplifier.

Transmitting Converter, or Modulator*

The transmitting converter problem differed from the first converter problem in that high output power was the major consideration, rather than low conversion loss and noise figure.

There will normally be three frequencies present in the output of a converter. These are the desired output sideband frequency, f', the beating oscillator frequency, $f' + \text{IF}$ or $f' - \text{IF}$, and the unwanted sideband frequency, $f' + 2\ \text{IF}$ or $f' - 2\ \text{IF}$. The strongest of these is the beating oscillator frequency, but fortunately this can be suppressed by 20 to 30 db by balance in a balanced converter. Investigation showed that the input impedance at the IF terminals of the modulator was affected by the load impedance of the modulator at both the wanted and the unwanted sideband frequencies. The load impedance consists of the input impedance of the RF amplifier as seen through the length of transmission line which connects the two components. At the wanted sideband frequency, the RF amplifier presents a good termination and the load impedance is independent of the length of the connecting line. At the unwanted sideband frequency, however, the RF amplifier presents a short circuit and hence the load impedance is a function of the length of the connecting line. A waveguide filter was incorporated in the output circuit of the modulator so as to reflect the un-

* This section prepared by W. W. Mumford who did the work on the transmitting converters.

wanted sideband back into the modulator in proper phase, thereby making the length of the line going to the RF amplifier much less critical. The proper phase of reflection was obtained by adjusting the length of the waveguide between the modulator and the reflecting filter. The IF impedance of the crystals could thus be varied so as to obtain an impedance match between the IF amplifier and the crystals.

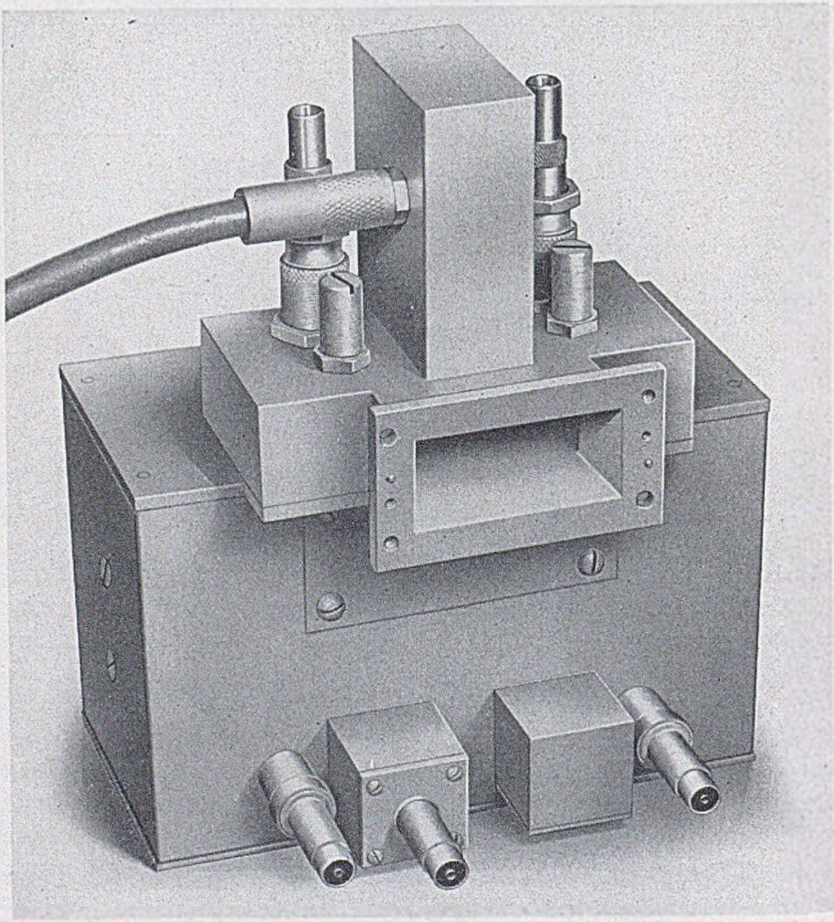


Fig. V-11.—Experimental model of transmitting converter.

An investigation of the power capacity of various types of standard and special silicon crystals indicated that one similar to the 1N28 crystal was the best available for this service, and life tests were made to check the stability of the crystal under high level conditions. Special IF input and crystal-to-waveguide matching circuits were developed to meet the stringent match requirements. The research on these major factors and many subsidiary details resulted in a converter which had an output of 6 milliwatts with less than 0.1 db compression, good input and output match, a flat amplitude response over more than 10 megacycles, and a conversion loss of about 11 db.

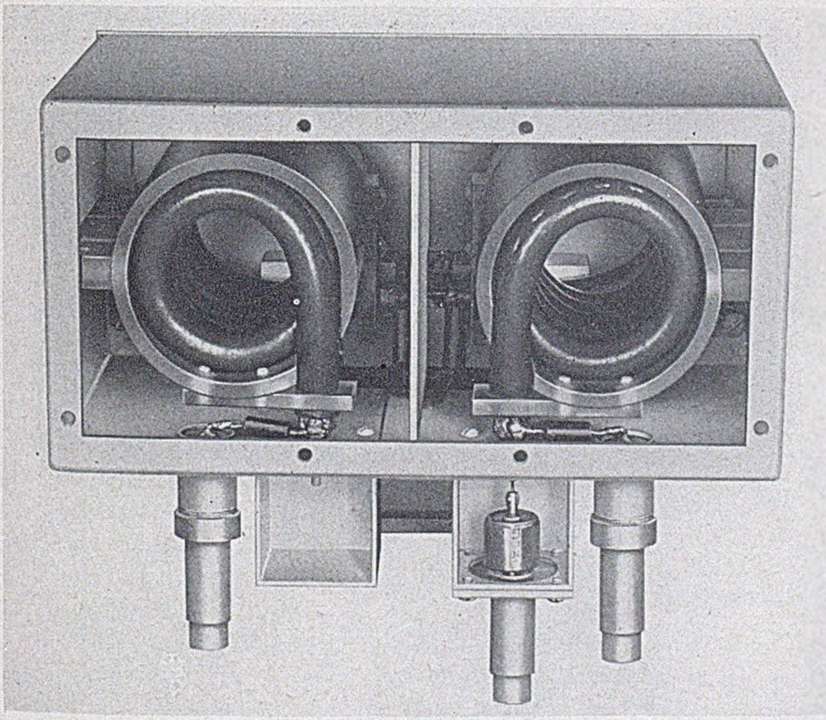


Fig. V-12.—Unbalanced to balanced coaxial transformer for feeding the 65 mc signal to the transmitting converter.

This converter, or transmitting modulator, had the 1N28 crystals mounted in the conjugate branches of a waveguide hybrid junction, as shown in Fig. V-11. Adjustable coaxial sleeves surrounded the crystal cartridges in order to effect an impedance match to the waveguide, and tuning studs were provided for trimming adjustments. The beating oscillator was injected into the hybrid junction through a broad-band coaxial-to-waveguide transducer in the upper branch of the hybrid junction, and the sidebands appeared in the conjugate waveguide branch. The 65-megacycle signal was fed onto the crystals in push-pull through the unbalanced to balanced coaxial transformer shown in Fig. V-12. Blocking condensers enabled the crystal currents to be monitored separately and RF filters kept the 4000-megacycle

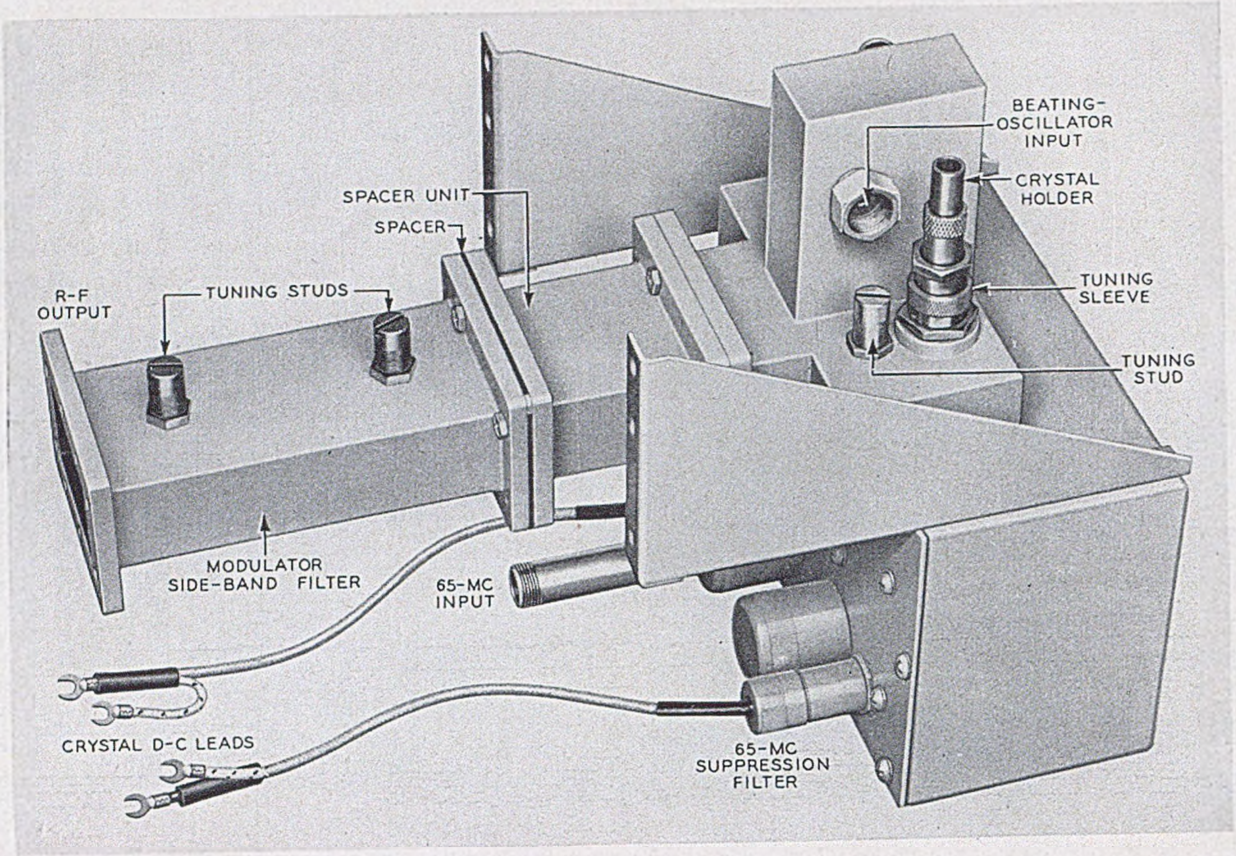


Fig. V-13.—Transmitting converter complete with output filter.

energy from entering the 65-megacycle transformer compartment. The converter developed for the New York-Boston circuit, complete with sideband filter, spacer, mounting brackets and d-c. leads, is shown in Fig. V-13.*

R. F. Amplifier**

Prior to the war, a considerable amount of research had been applied in the Bell Telephone Laboratories to electron tubes operating on the velocity modulation principle, with the expectation that such tubes would find applications in radio-relay systems. Although this work was interrupted by the war, enough information had been obtained to make it apparent, upon resumption of the radio relay work, that such tubes were the only ones then known which showed promise of meeting the stated requirements.

Velocity modulation tubes have been described by several authors,[29, 30, 31] and the theory of their operation has been discussed adequately in several places.[32, 33, 34, 35] However, a review in 1939 of the structures then known had led to the decision that a new type of construction would be necessary to obtain a satisfactory amplifier tube for radio relay purposes. To keep the tube voltages within reasonable limits it is desirable to make the input and output gaps as small as possible. Resonant circuits external to the evacuated envelope are desirable to enable coverage of as large a frequency range as possible with a single tube, and to facilitate addition of broad-banding circuits. Grids on the input and output gaps are undesirable because of the large interception of current and the difficulty and expense of construction.

With the above considerations in mind, decision was then made to explore the possibilities of focussed beams, and of gaps comprised of copper discs sealed through a cylindrical glass vacuum envelope. The latter technique had been developed at the Bell Telephone Laboratories in connection with

* Developed by H. C. Foreman and W. W. Halbrook.

** This section prepared by A. G. Fox and A. E. Bowen. Messrs. Fox and Bowen, in collaboration with A. L. Samuel, A. E. Anderson, and J. W. Clark of these Laboratories, did the major part of the research which resulted in this amplifier.

[29] Varian, R. H. and Varian, S. F., "A High Frequency Oscillator and Amplifier", *Jour. Applied Physics*, vol. 10, No. 5, pp. 321–327, May 1939.

[30] Hahn, W. C. and Metcalf, G. F., "Velocity Modulated Tubes", *Proc. I. R. E.*, vol. 27, No. 2, pp. 106–116, Feb. 1939.

[31] Harrison, A. E., "Klystron Tubes", McGraw-Hill Book Co., New York, 1947. (Book)

[32] Hansen, W. W. and Richtmyer, R. D., "On Resonators Suitable for Klystron Oscillators", *Jour. Applied Physics*, vol. 10, No. 3, pp. 189–199, March 1939.

[33] Hahn, W. C., "Small Signal Theory of Velocity Modulated Electron Beams", *G. E. Review*, vol. 42, No. 6, pp. 258–270, June 1939.

[34] Hahn, W. C., "Wave Energy and Transconductance of Velocity-Modulated Electron Beams", *G. E. Review*, vol. 42, No. 11, pp. 497–502, Nov. 1939.

[35] Ramo, S., "The Electronic-Wave Theory of Velocity Modulation Tubes", *Proc. I. R. E.*, vol. 27, No. 12, pp. 757–763, Dec. 1939.

the development of water cooled tubes, and showed promise of providing a very convenient means for construction of resonant circuits in which a part of the circuit was external to the vacuum envelope and part was internal.

Numerous forms of focussed beam-disc seal velocity modulation tubes were constructed for examination of the various factors affecting the performance of such tubes. Sizes and shapes of gaps, length of drift space, voltage and beam current, and other factors were adjusted to optimize the performance. Both magnetic and electrostatic focussing of the beam were explored. Triple gap tubes, from which gains of more than 30 db were obtained, were experimented with.

The end result of this work was the development of a four-stage amplifier employing velocity modulation tubes of the disc-seal type (Fig. V-14) designed especially for this application.* The electron gun is at the left, and the collector is at the right. These tubes employ external circuits in the form of resonant cavities assembled around the tube. The electron

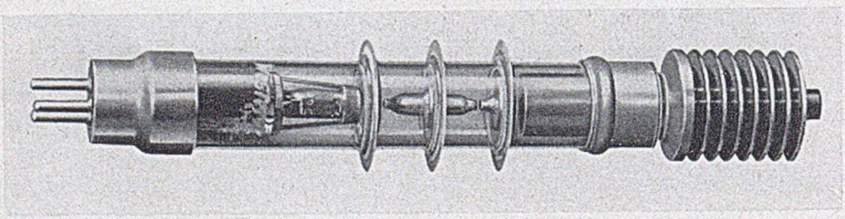


Fig. V-14.—Disc seal velocity modulation tube.

beams are accelerated by 1500 volts and focussed magnetically by means of small alnico permanent magnets placed around the cavity structure as shown in Fig. V-15. This shows a single stage of coaxial-coupled amplifier with half of each cavity removed to show internal details. The cavities are tunable over approximately a 250-megacycle range by means of metal screw plugs located around their periphery. Two different sets of cavities are sufficient to cover the entire frequency range of the radio repeater.

A single amplifier stage of the type shown in Fig. V-15 will exhibit a single-tuned type of gain characteristic which is only 5 megacycles wide at the 3 db points when operated under matched input and output impedance conditions. It was therefore necessary to widen the band by using double-tuned circuits throughout. In fact such a design would be difficult to avoid because each tube has its own resonant cavities. By running a short length of waveguide or coaxial transmission line from the output cavity of one stage to the input cavity of the next, a double-tuned structure results automatically. The coupling from cavity to transmission line is made by means

* These tubes were developed by A. L. Samuel and J. W. Clark.

of a window in the side of the cavity when a waveguide line is used, and by means of a wire loop projecting within the cavity when a coaxial line is used. Amplifiers employing both types of coupling were designed and tested. The coaxial coupling type with links approximately one-half wavelength long

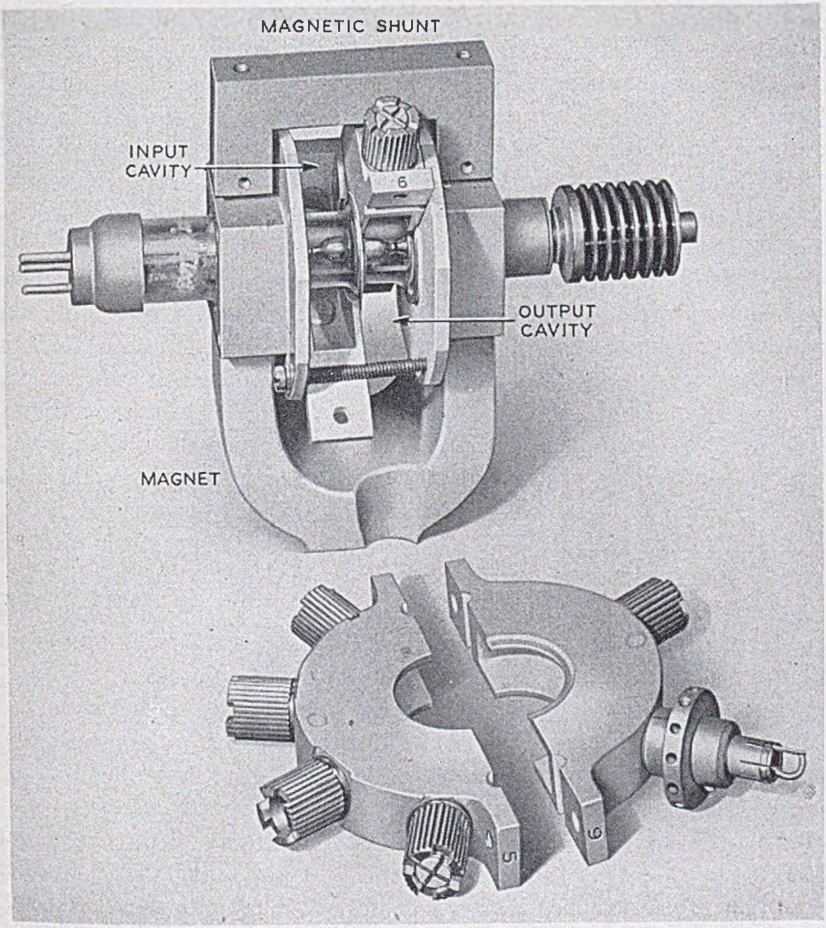


Fig. V-15.—A single stage of a coaxial-coupled amplifier using the velocity modulated tube of Fig. V-14.

proved the easier to adjust. Variation of the coupling from zero to a maximum is accomplished by rotating the plane of the coupling loop through 90°. Each cavity is provided with a small plate of resistance film projecting into the field in such a way that rotation of the plate about an axis lying in the plane of the plate will vary the resistive loading present in the cavity from zero to a maximum. Finally, the input and output cavities of the first and last stages respectively are coupled to separate tuned cavities in order to provide double-tuned terminals for the amplifier. This results in the overall structure shown schematically in Fig. V-16.

The required bandwidth is obtained by a process of stagger-coupling the circuits so that the individual responses are as indicated schematically at the top of Fig. V-16. Because there are available continuously variable adjustments of tuning, loading, and coupling for each circuit, it is possible to obtain a very smooth and symmetrical overall response characteristic as shown in Fig. V-17. The corresponding measured delay distortion characteristic is shown in Fig. V-18. For this type of tuning the output power is about .7 watt at ½ db of compression. More power is, however, obtainable if more compression is tolerable; and when used in the repeater for the transmission of FM signals, the amplifier is driven to an output of 1 watt.

Since the circuits are of fairly high Q, the frequency characteritics of the amplifier are markedly affected by changes in temperature of the cavities. In order to minimize such detuning effects, the amplifier has been placed in a temperature controlled compartment. Since about 45 watts are dissipated at the collector of each tube by the high-voltage electron beam, the collectors must be cooled by a forced air blast. In order to keep the cooling system separate from the temperature control system, the collector ends of the tubes project through a wall of the temperature controlled compartment into an external air duct. Figure V-19 shows a complete r.f. amplifier with the cover of the temperature control box removed. This is the amplifier developed for the New York-Boston circuit.* The electron-gun ends of the tubes face the reader. Short lengths of flexible coaxial cable couple the input and output of the amplifier to the associated equipment.

Testing of Components and Repeater Amplifier—Many special measuring and testing techniques had to be devised for the research work on each component. In addition, standard production tests had to be worked out and measuring equipment constructed.

Impedances were measured at RF with standing wave detectors, and at IF with three fixed taps on a coaxial line. It was found that input and output SW ratios could be held to 1.7 db or less over the 10 megacycle band. Both point by point and swept oscillator methods were used at RF and IF for measuring amplitude characteristics. Particularly as a result of the development of swept oscillator measuring techniques it was found practicable to adjust each unit to ±0.1 db amplitude variation over the 10 megacycle band. Noise figures[16] of IF equipment were measured by the noise diode method.[36]

* Developed by F. E. Radcliffe and R. C. Carlton.
[16] Loc. cit.
[36] H. Johnson, "A Coaxial Line Diode Noise Source for U.H.F.", *R. C. A. Review*, vol. VIII, No. 1, pp. 169–185, March 1947.

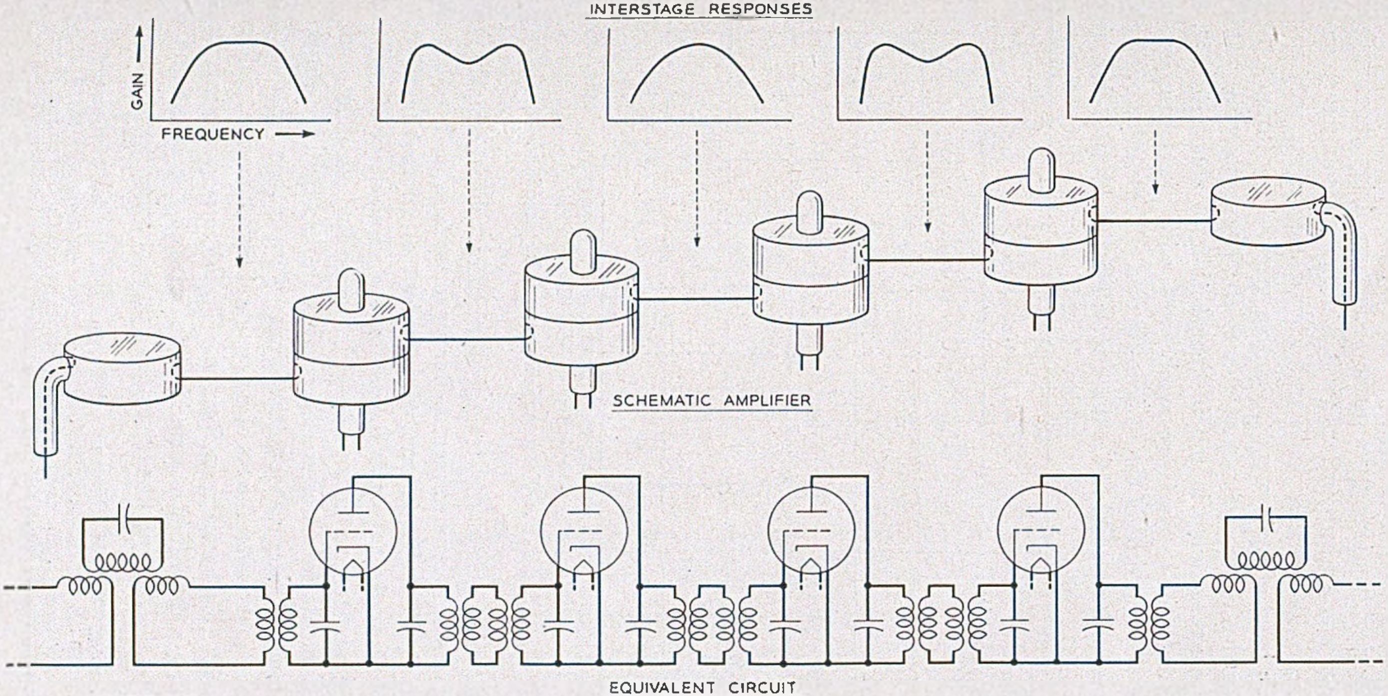


Fig. V-16.—Schematic diagram of four stage coaxial coupled amplifier using velocity modulated tubes.

A method[37] of measuring the phase and group delay characteristics of the repeater was worked out. This is a rather difficult measurement to make because the tolerable phase and delay distortion are very small due to the wide band of the system. Relative group delay through the band was measured with an accuracy of about ±0.001 microsecond. This corresponds to an accuracy in relative phase shift of about ±0.35°. The measured distortion agreed reasonably well with the distortion calculated from the constants of the various circuits. These measurements and calculations showed that while the characteristics of an 8-link repeater

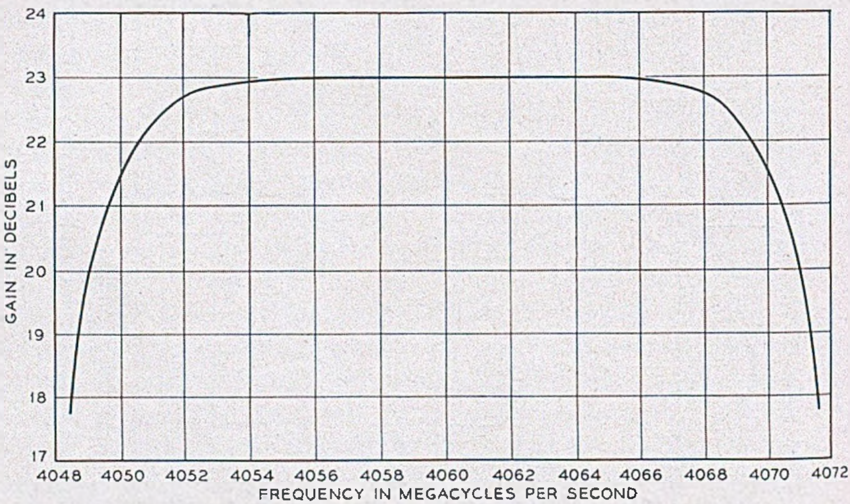


Fig. V-17.—Frequency response of RF amplifier.

circuit will give acceptable television pictures, phase equalization will provide a definite improvement. Longer circuits will therefore require phase equalization. Equalizers were devised for both the IF and RF circuits which made it possible to equalize the group delay of each component of the repeater amplifier to ±0.001 microsecond over the 10-mc band.

An experimental repeater amplifier was set up so that the output could be fed through an attenuator to the input. It was then possible to break the loop and make overall tests of delay, amplitude linearity, and transient response with IF measuring equipment. Transient response was measured by including a 2000-foot waveguide line between the input and output terminals and applying the circulated pulse testing technique.[38] This

[37] D. H. Ring, "The Measurement of Delay Distortion in Microwave Repeaters", elsewhere in this issue of the *B. S. T. J.*

[38] A. C. Beck and D. H. Ring, "Testing Repeaters with Circulated Pulses", *Proc. I. R. E.*, vol. 35, No. 11, pp. 1226–1230, November 1947.

testing method permits a study of the shapes of rectangular pulses which have passed through a repeater amplifier many times, and thus simulates transmission through a relay system with many repeaters. Figure V-20 shows an example of the results obtained from such tests made on the IF amplifier alone. The top row shows a 1.25-microsecond test pulse after 1,

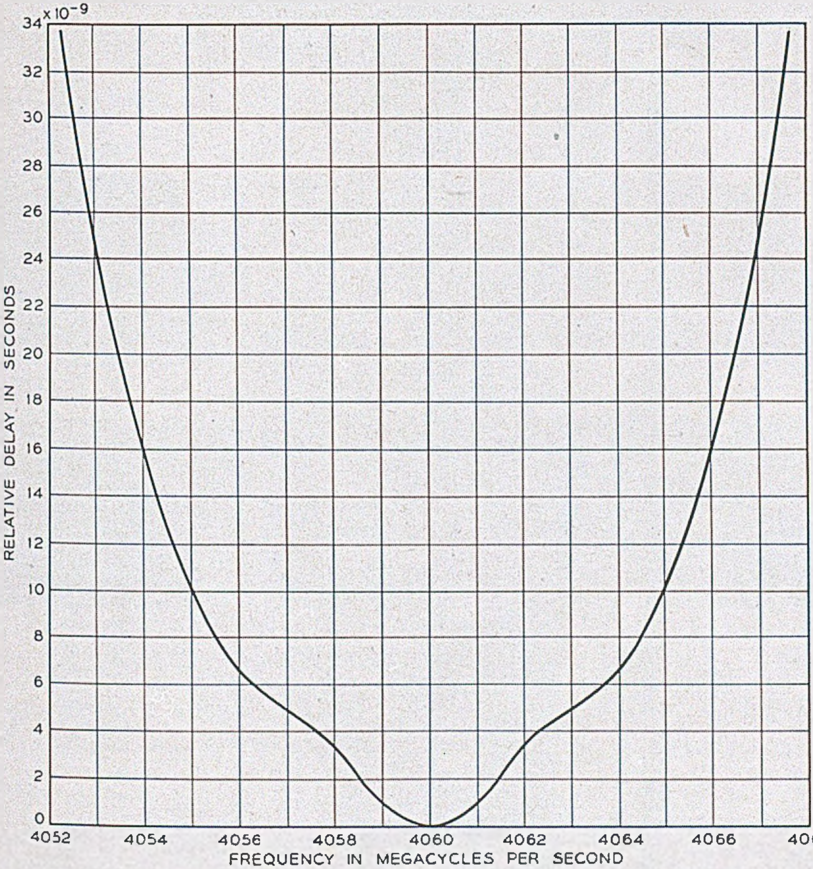


Fig. V-18.—Delay characteristic of RF amplifier.

10, and 30 trips through the IF amplifier. Part of the distortion appearing after one trip through the amplifier was in the viewing oscilloscope. It was difficult to detect distortion from a single trip through the amplifier, but Fig. V-20 demonstrates how it increases with successive trips. The second row of pulse pictures shows the improvement achieved by adding phase equalizers to the circuit. It was particularly noticeable in this test that without equalizers the details in the shape of the pulse after 10 or 30 trips were very sensitive to the exact value of the intermediate frequency.

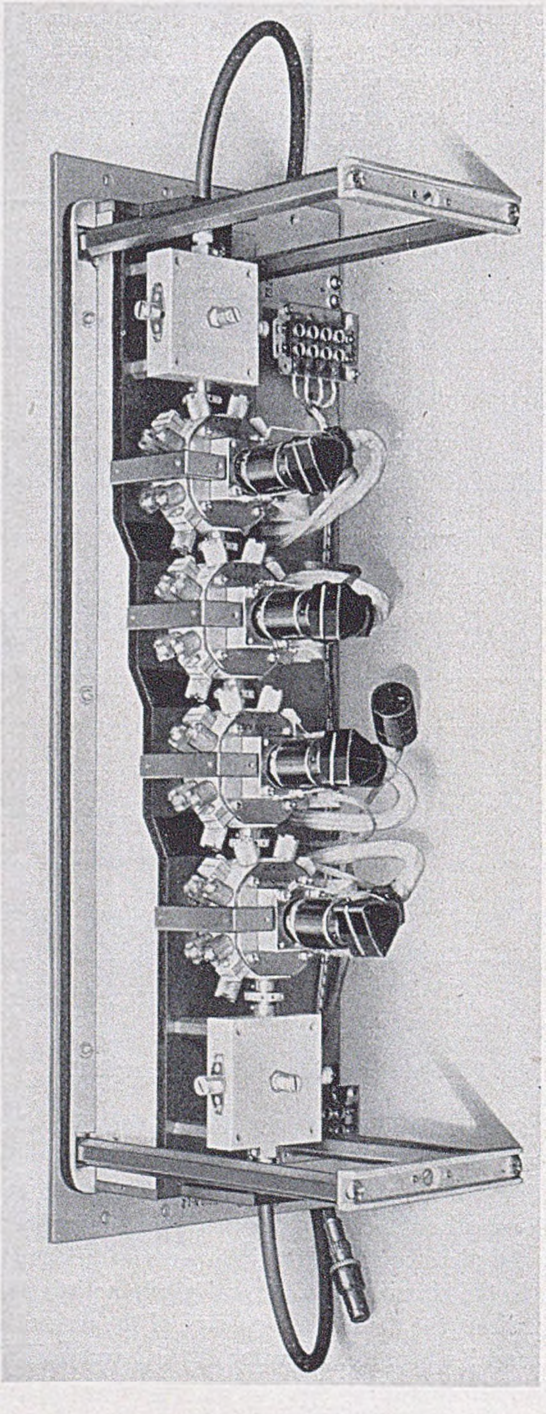


Fig. V-19.—The complete RF amplifier with cover removed.

When the equalizers were added, the pulse shapes shown in the second row were not greatly changed for variations of ± 3 megacycles or more in the intermediate frequency.

In concluding this section it should be noted that when the various components were connected together to form an IF-type repeater amplifier, it was found that the amplitude and phase characteristics added as expected and resulted in a satisfactory amplifier with only a few millimicroseconds variation in delay and only a few tenths of a db variation in amplitude over a 10-megacycle band. Nevertheless, the equipment is very complicated.

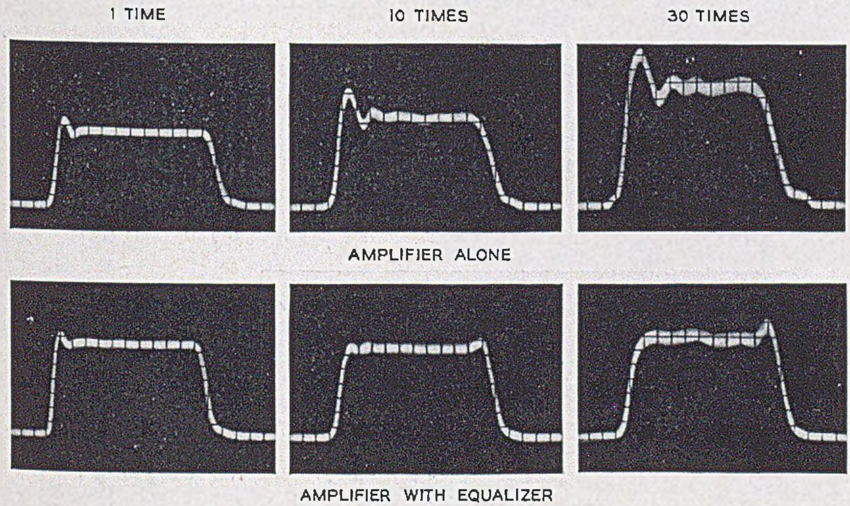


Fig. V-20.—Results of circulated pulse tests on IF amplifier.

A straight-through radio frequency amplifier repeater is still to be desired and, no doubt, further research will eventually produce such an amplifier.

VI. The Complete Repeater*

In the preceding section it has been pointed out how the various components that make up the repeater amplifier were added together without the introduction of additional distortion. The antennas, filters and amplifiers which go to make up one of the complete repeaters shown in Fig. II-1 were designed with the same ease of interconnection in mind. However, as will be discussed below, the length of the waveguide lines used for these connections has an important bearing on the distortion introduced.

The large amount of equipment in the repeater amplifier makes it desirable, from the maintenance standpoint, to locate the amplifier near the ground and to provide towers for the antenna where antenna elevation

* Prepared by D. H. Ring.

is necessary. This means relatively long transmission lines between the antenna and the rest of the repeater. If exact termination of the line by the antenna impedance or by the filter input impedance were possible no distortion would be produced, but in general there will be a slight mismatch and a corresponding reflection of energy at each of these junctions which

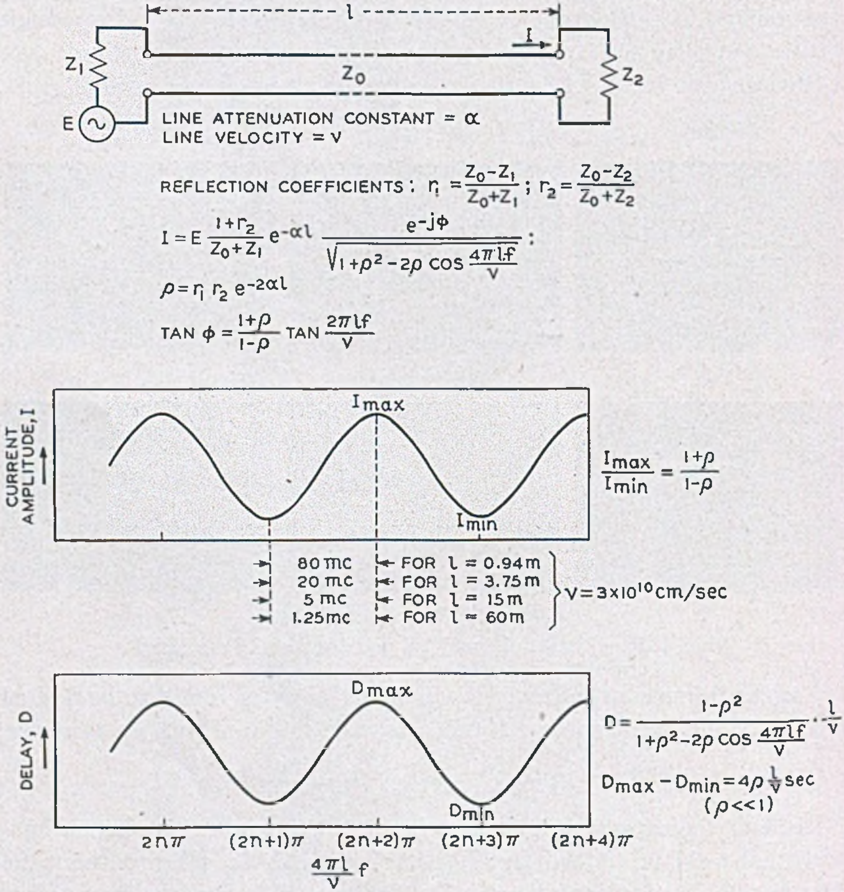


Fig. VI-1.—Effect of long lines on the amplitude and delay distortion of a repeater.

will produce variations in the amplitude and delay of the signal throughout the desired band. These variations may be greater than those produced anywhere else in the repeater.

The type of distortion originating in this way is illustrated by Fig. VI-1. In this figure there is represented an antenna of impedance Z_1 connected by a line of length ℓ and characteristic impedance Z_0, to a load, Z_2. In actual practice this load is the impedance presented by the filter to the line from the antenna. The variations produced in both the amplitude and delay of the signal currents are shown by the curves in the figure where relative change in the characteristics in question are shown as ordinates and values of the quantity $\frac{4\pi\ell f}{v}$ are shown as abscissas.

It will be seen from these curves that the amount of variation over a given band is a function of the line length. With short lines the amount of variation over a band 10 mc wide can be kept very small, but with lines about 150 feet in length there may be three or four full cycles of variations in a band 10 megacycles wide. Table A gives some typical values of the variations. Values are given for those degrees of mismatch which would produce standing wave ratios of 1, 1.4 and 2 db at the junctions. It will be seen that for lines between 100 and 200 feet in length the variations in some cases are greater than the limits achieved in the multistage amplifiers and other components of the system. Furthermore, it is usually impractical to compensate or tune out these variations because they are functions of the electrical length of the transmission lines and subject to change with temperature, frequency, and other small mechanical and electrical changes. It would appear that with present techniques this may be one of the most serious sources of distortion in a long relay system.

TABLE A

$SWR_1 = SWR_2$ in db $(\alpha = 0)$	1	1.4	2
$r_1 = r_2$ $(\alpha = 0)$	0.058	0.082	0.115
Max. Amplitude Variation in db	0.058	0.116	0.233
Max. Delay Variation in 10^{-9} seconds			
$\ell = 100$ feet	1.33	2.67	5.33
$\ell = 200$ feet	2.67	5.33	10.66

VII. CONCLUSION

Various phases of our work on microwave repeater circuits have been discussed. Such a circuit, made up of several repeaters as described in the last section, may be looked upon as a four-terminal network having specified amplitude and delay characteristics and should be suitable for the transmission of any signals for which the characteristics are adequate whether they be television signals or those of one of the various forms of multiplex telephony. It is outside the scope of this paper to discuss the uses to which such a repeater circuit might be put or the terminal equipment that any particular service might require.

The New York-Boston repeater circuit has been built to provide the experimental field trials necessary to answer many remaining questions which deal with the performance of an actual circuit. The results of these trials will be reported in a separate paper.

Acknowledgements

The work which has been discussed in this report of necessity involved the combined efforts of many individuals. In addition to those already mentioned, practically every other member of the staffs of the Deal and Holmdel Radio Laboratories has made valuable contributions to the work, as have also many in other departments of the Bell Telephone Laboratories.

The development of the components of the New York-Boston circuit has been under the direction of Mr. Gordon N. Thayer.

In particular, we wish to acknowledge the support and stimulating advice of Dr. R. Bown under whose general direction the work progressed.

The Measurement of Delay Distortion in Microwave Repeaters*

By D. H. RING

Measuring equipment is described which is capable of measuring delay distortion of the order of 10^{-9} seconds in a wide band microwave television relay repeater. Two measuring circuits are discussed. The first is a circuit for measuring the relative phase shift versus frequency from which the delay distortion may be computed. The second circuit gives the delay directly from a single measurement. The measuring equipment is designed to work in the intermediate frequency range from 50 to 80 megacycles, but by applying suitable conversion equipment measurements can be made at microwave frequencies.

THE successful transmission of broadband television and pulse signals over any communication circuit depends upon the preservation of the complex wave shapes of the original transmitted signals. Fourier analysis tells us that a complex signal wave can be resolved into a spectrum of frequencies with certain amplitude and phase relationships. It is well known that the amplitude relationships of all essential frequencies in this spectrum must be substantially preserved. It is equally important that the phase relationships of the essential frequencies should be preserved. The instantaneous value of the received signal is the vector sum of the instantaneous amplitudes of all the component frequencies. Therefore, if the relative phase of some frequency component is changed by 180° the sign of its contribution to the output is reversed, and it is clear that a closer approximation to the original signal could be obtained by suppressing this frequency component rather than permitting it to contribute negatively to the output.

It can be shown[1] that the relative phase relations of the component frequencies in a complex signal wave will be preserved if the phase shift in passing through a circuit is a linear function of the angular frequency. That is

$$\beta = T_0\omega + n\pi \tag{1}$$

where T_0 is a constant and n an integer. Distortion of the transmitted signal will occur if T_0 is not constant over the essential frequency band of the signal. We shall not be interested in distortion due to n not being an integer[2] since this case does not occur in carrier circuits where the phase shift at carrier frequency, rather than the phase shift at zero frequency, is

* Presented at National Convention of I. R. E., New York City, March 4, 1947.

[1] Phase Distortion and Phase Distortion Correction, S. P. Mead, *B.S.T.J.*, April 1928. 199–201.

[2] Phase Distortion in Telephone Apparatus, C. E. Lane, *B.S.T.J.*, July 1930, 494–496.

the reference point for phase. Departure of β from the linear relationship given by (1) is the phase distortion in the circuit.

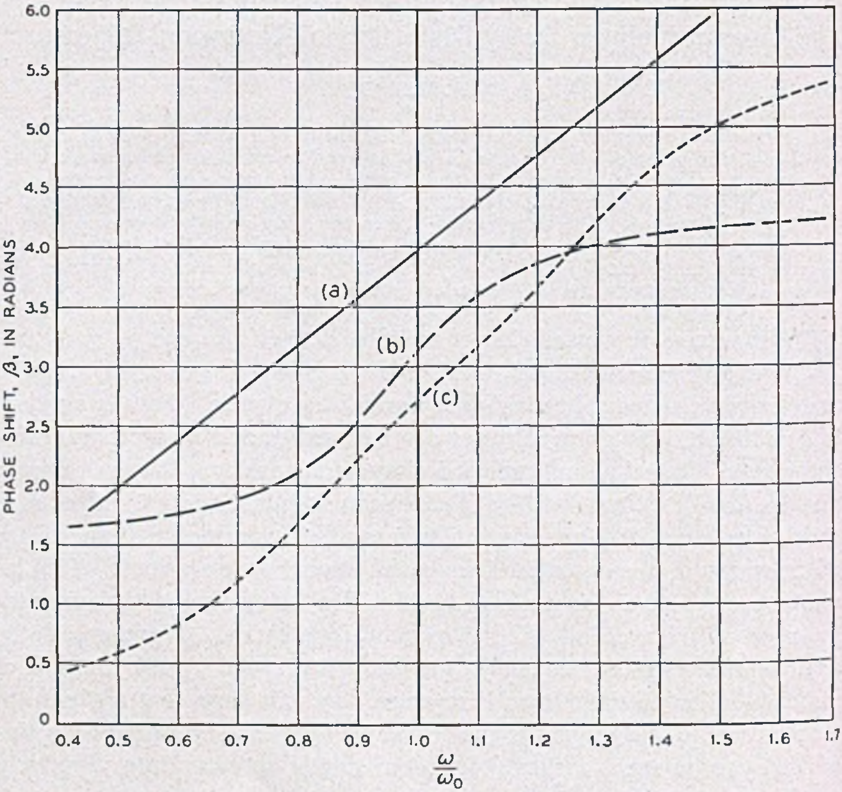


Fig. 1—Typical phase shift curves for various types of circuits
a. A Transmission line terminated in its characteristic impedance.
b. A single tuned circuit.
c. Two tuned circuits with approximately critical coupling.

The time of transmission[3] or the delay in passing through the circuit is obtained by differentiating (1):

$$\text{Delay} = \frac{d\beta}{d\omega} = T_0 \text{ seconds} \tag{2}$$

Variation in T_0 with frequency is the delay distortion in the circuit.

Figure 1 shows some typical phase curves, and Fig. 2 shows the corresponding delay curves. In each figure curve (a) represents an ideal distortionless circuit with linear phase and constant delay such as a simple transmission line. Curves (b) are obtained for single resonant circuits,

[3] T_0 has also been called the group delay and the envelope delay.

and curves (c) for coupled double tuned circuits. It is felt that the delay curves of Fig. 2 are easier to interpret and give a better physical picture of the distortion resulting from phase variations than the phase curves of Fig. 1. Therefore, most of the following discussion will be in terms of delay rather than phase.

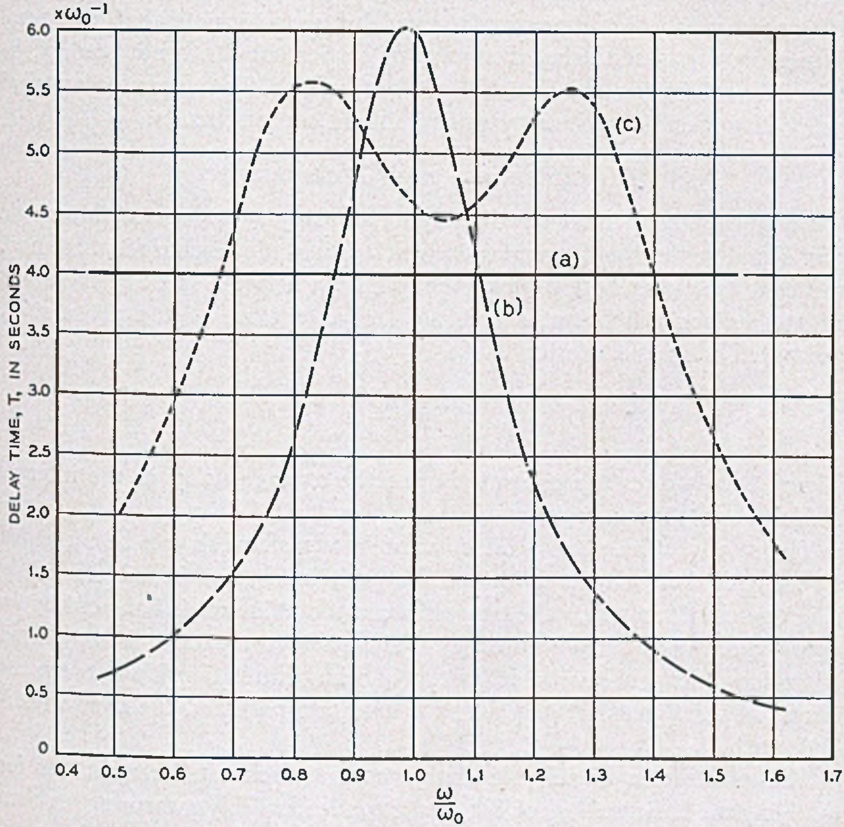


Fig. 2—Typical delay curves for various types of circuits
a. A transmission line terminated in its characteristic impedance.
b. A single tuned circuit.
c. Two tuned circuits with approximately critical coupling.

Thus far we have considered the general case and stated that the delay must be constant over the essential frequency band for distortionless transmission. It should be noted, however, that when delay distortion is present different types of signals may be affected differently. For instance, in the case of an amplitude modulated carrier of angular frequency ω_0, Fig. 2, we note that if the delay curves have arithmetic symmetry about ω_0, then the sidebands at $\omega_0 \pm \Delta\omega$ will suffer the same delay and therefore will add in phase upon demodulation. If the delay distortion is not symmetrical about ω_0 the sidebands will not add in phase upon demodulation, and the demodulated output will suffer both amplitude distortion and some delay distortion which differs from the delay distortion at both $\omega_0 + \Delta\omega$ and $\omega_0 - \Delta\omega$. In the case of frequency modulation dissymmetry introduces harmonics in the demodulated output. A detailed discussion of these effects is beyond the scope of this paper, but we may note that in general a true picture of the delay distortion in carrier circuits is not readily obtained by observing the demodulated output and that an unsymmetrical delay distortion is particularly undesirable in carrier circuits.

Principles of Delay Measurement

Delay cannot be measured directly on a steady state basis with a single test signal in the simple manner in which amplitude response is measured. Instead, the phase shift through the unknown network must be measured at two adjacent frequencies and the delay, or slope of the phase shift, computed from the relation

$$T = \frac{\Delta\beta}{\Delta\omega} \tag{3}$$

Figure 3 illustrates the computation of the average delay in the interval $\Delta\omega$ from two phase measurements.

The steady state phase shift of an unknown network can be measured by using the basic circuit shown schematically in Fig. 4.[4] This is essentially a bridge circuit in which the phase shift in the unknown is balanced by an equivalent calibrated phase shifter. The phase comparator is some kind of device which will give an indication of a known relationship between the phases of the signals arriving over the unknown path and the known reference path.

The exact form of suitable components and circuit arrangements for applying the basic method of Fig. 4 to a particular delay measuring problem is largely determined by the order of magnitude of the delay to be measured and accuracy desired. In the case of microwave television repeaters we are interested in video bands of the order of 5 mc wide. As a rough estimate we might say that the highest frequency in the band should not be shifted more than one quarter period from its normal phase position. In the case of linear delay distortion or a parabolic phase-frequency characteristic, one quarter period for 5 mc is 0.05 microseconds. In a repeater system with 50 repeaters this yields a tolerable systematic delay distortion of 10^{-9} seconds per repeater. Therefore we conclude that in developing

[4] Measurement of Phase Distortion, H. Nyquist and S. Brand, *B.S.T.J.*, July 1930, 526–527.

repeaters for this service an accuracy of better than 10^{-9} microseconds in measuring the relative delay over a band of frequencies will be desirable. The phase shift, $\Delta\beta$ in Fig. 3, which corresponds to a delay of 10^{-9} seconds is a function of the measuring interval $\Delta\omega = 2\pi\Delta f$. If Δf is small $\Delta\beta$ will

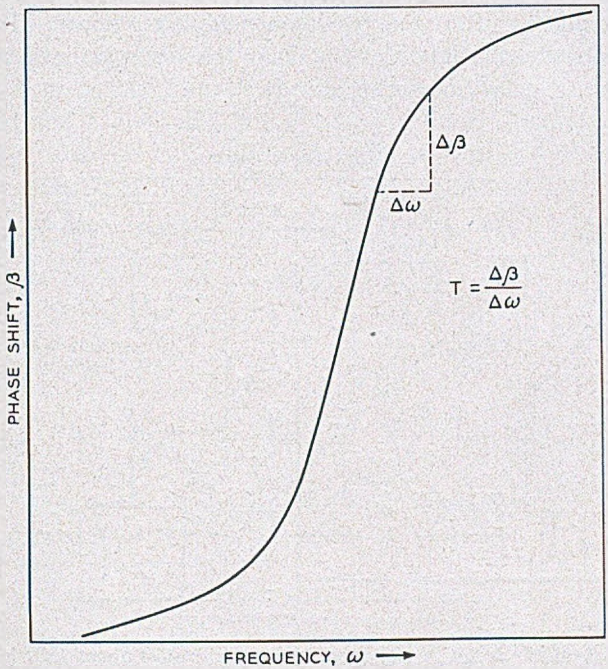


Fig. 3—Factors involved in calculating the delay of an electrical circuit

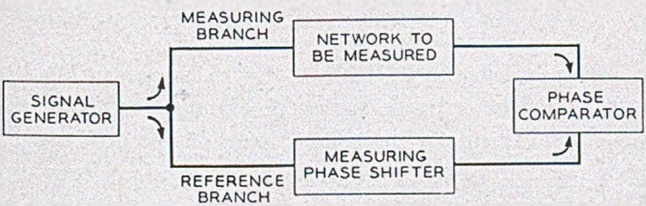


Fig. 4—Basic circuit for measuring the phase shift in a network

be small and difficult to measure. If Δf is large the average slope measured will not be the true slope in the center of the interval. For a 5-megacycle video signal the intermediate and radio frequency bands of interest will be in excess of 10 megacycles wide. These considerations led to a choice of 1 megacycle as a reasonable value for Δf. If $\Delta f = 1$ megacycle a phase shift $\Delta\beta = 0.36°$ will result from a delay of 10^{-9} seconds.

Two circuits for measuring delay distortion will be described. The first is a phase measuring circuit which, in principle, is an adaptation of Fig. 4 to practical operation at intermediate frequencies. The second circuit is a modification in which two frequencies differing by Δf are fed through the circuit under test simultaneously in such a way that the difference in the phase shift at the two frequencies is measured. This arrangement permits the calculation of the delay from a single measurement and is, therefore, a delay measuring circuit.

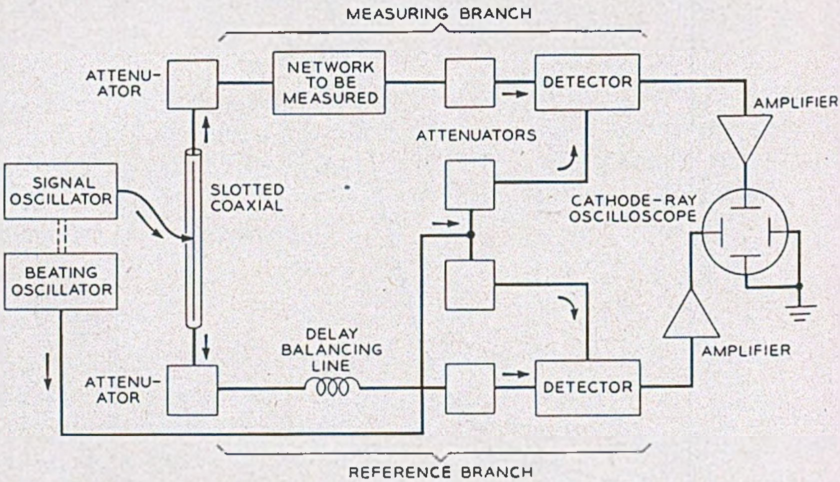


Fig. 5—Schematic circuit for the precise measurement of phase shift in the intermediate frequency range.

Phase Measuring Circuit

Figure 5 shows a schematic diagram of a phase measuring circuit which has been found to be suitable for precision measurements of wide-band circuits in the intermediate frequency range of 50 to 80 megacycles. It is a double detection system with an intermediate frequency of one megacycle. The test signal oscillator and the beating oscillator are ganged to a single control and adjusted to track so that they maintain a difference of approximately one megacycle throughout their tuning range. The test signal is fed to a sliding contact on a section of air dielectric coaxial transmission line. The signal divides at this point to feed the test branch and reference branch.

The sliding tap on the coaxial line provides a high precision measuring phase shifter if the coaxial line is well terminated at each end. The relative change $\Delta\beta$ in phase of the signals at the two ends of the line when the slider is moved a distance Δx is two times the phase shift corresponding to the movement of the slider at the working frequency or

$$\Delta\beta = \frac{720 f \Delta x}{c} \text{ degrees}$$

where c is the velocity of light.
For $\Delta x = 0.1$ centimeter and $f = 65$ megacycles, $\Delta\beta$ is 0.156 degrees.

The measuring branch and reference branch feed through the network to be measured and a balancing line respectively to identical detectors and one megacycle amplifiers. The amplifiers are connected to the plates of a cathode ray oscilloscope which is used as a phase comparator. A cathode

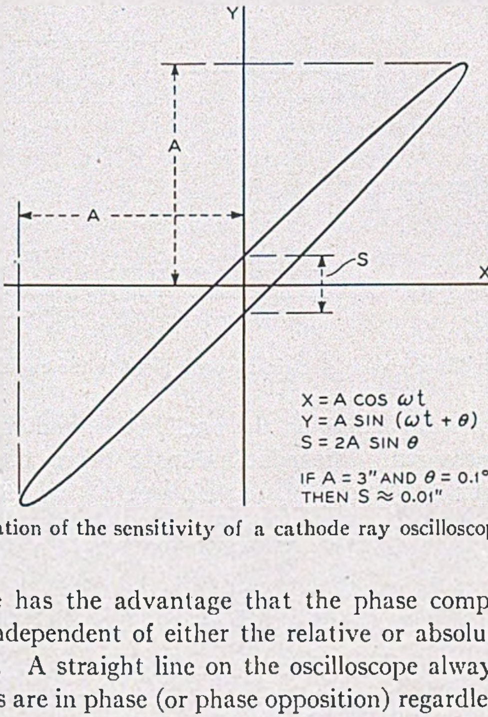


Fig. 6—Calculation of the sensitivity of a cathode ray oscilloscope as a phase comparator.

ray oscilloscope has the advantage that the phase comparison with this instrument is independent of either the relative or absolute amplitudes of the two signals. A straight line on the oscilloscope always indicates that the two voltages are in phase (or phase opposition) regardless of the relative amplitudes, which determine the slope of the line. The sensitivity, of course, is a function of the amplitudes. Figure 6 illustrates how the sensitivity of an oscilloscope phase comparator can be calculated. If each signal alone produces a 6-inch deflection, then 0.1 degree phase difference produces an opening of the pattern of 0.01 inches, which is sufficient to be detected on a "rocking" or in-out basis.

It is obvious that a circuit of this type measures the difference in the phase shifts of the two branches. The measured phase shift is, therefore, the absolute phase shift in the measuring branch less the phase shift in the reference branch. The balancing line shown in the reference branch of Fig. 5 can be adjusted in length so that it balances out the average delay in the measuring branch. The measured remainder will then be the distortion in the measuring branch, since a good transmission line does not have phase or delay distortion. The use of a balancing line in the reference branch simplifies measurements by reducing the range of movement of the slider, and it greatly decreases the errors in the calculation of the delay distortion due to inaccuracies in the measuring interval Δf.

The simplified diagram of Fig. 7 will be used to show how the delay of the unknown network can be calculated. With the signal oscillator set at frequency f_1 the slider is adjusted until the signals at C and E are in phase

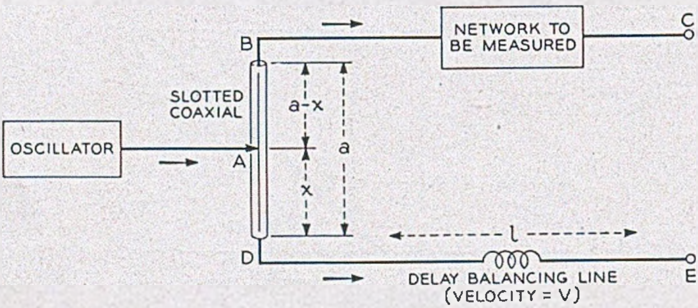


Fig. 7—Simplified circuit electrically equivalent to the circuit of Fig. 5.

as indicated by a straight line on the oscilloscope, and the corresponding distance x_1 is measured. The relationships between the phases of the signals at the various points in the circuit are:

$$\begin{aligned} \phi_C &= \phi_A - \phi_{AB} - \phi_{BC} \\ &= \phi_A - \frac{2\pi f_1}{c}(a - x) - \beta \qquad (4) \end{aligned}$$

$$\begin{aligned} \phi_E &= \phi_A - \phi_{AD} - \phi_{DE} \\ &= \phi_A - \frac{2\pi f_1 x_1}{c} - \frac{2\pi f_1 \ell}{v} \qquad (5) \end{aligned}$$

$$\phi_C = \phi_E$$

where β is the phase shift in the unknown network at frequency f_1. Solving for β,

$$\beta = \frac{2\pi f_1}{c}\left[\frac{\ell c}{v} + 2x_1 - a\right] \qquad (6)$$

The signal oscillator frequency is then changed to $f_2 = f_1 + \Delta f$ and the slider readiusted to the position x_2 that again makes $\phi_C = \phi_E$. Then as above

$$\beta' = \frac{2\pi f_2}{c}\left[\frac{\ell c}{v} + 2x_2 - a\right] \tag{7}$$

where β' is the phase shift in the unknown network at frequency f_2. The delav in the network is, from (3)

$$T = \frac{\beta' - \beta}{2\pi\Delta f} \tag{8}$$

Substituting (6) and (7) in (8) yields

$$T = \frac{1}{c}\left[\left(\frac{\ell c}{v} - a\right) + 2x_2 + \frac{2f_1}{\Delta f}(x_2 - x_1)\right] \tag{9}$$

The first term in T is a constant of the set-up and can be dropped when the delay distortion only is desired. The second term is small and can often be neglected. The third term gives the major part of the difference between the delay of the reference path and the delay of the network.

It has been found that the slider position can be easily reset to $\pm$ 0.05 centimeters for each frequency. This would mean a maximum error of $\pm$ 0.1 centimeter for the difference of two readings which corresponds to an error in T of about $\pm 0.4 \times 10^{-9}$ seconds. However, this reset accuracy will not be realized in overall accuracy unless a number of precautions are observed in setting up the circuit of Fig. 5. In order to avoid stray coupling the two branches of the system must be carefully shielded from each other. At least 60 db net attenuation must be provided between the detectors via the path through the balancing line, phase shifter and unknown network, and 40 db attenuation in the path via the BO supply lines. All attenuators, plugs, etc., must have voltage standing wave ratios of less than about 1.015 and the detectors and amplifiers in each branch must have identical phase shifts over the range of variation of the IF due to imperfect tracking of the oscillators.

Delay Measuring Circuit

The circuit of Fig. 5 is basically a phase measuring circuit. It can be rearranged as shown in Fig. 8 to yield a circuit that will read delay directly from a single setting of the slider. In Fig. 8 the signals from the two oscillators are both sent through the circuit to be measured, and both are sent through the reference branch. Any delay in either path will alter the relative phases of the two signals in that path and this change in relative phase shift will be passed on to the beat note formed in the detectors. If the delays in the two paths are different a relative phase shift proportional to the difference will appear in the two beat notes. This phase change can be measured with a phase shifter in the beat note circuit or with a phase shifter which varies the relative phase of the two signals fed to one branch

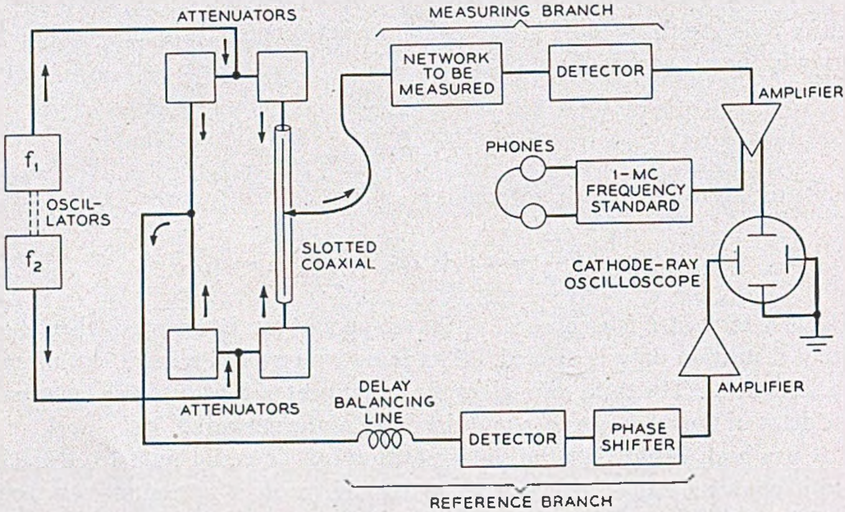


Fig. 8—Schematic circuit for the precise measurement of delay in the intermediate frequency range.

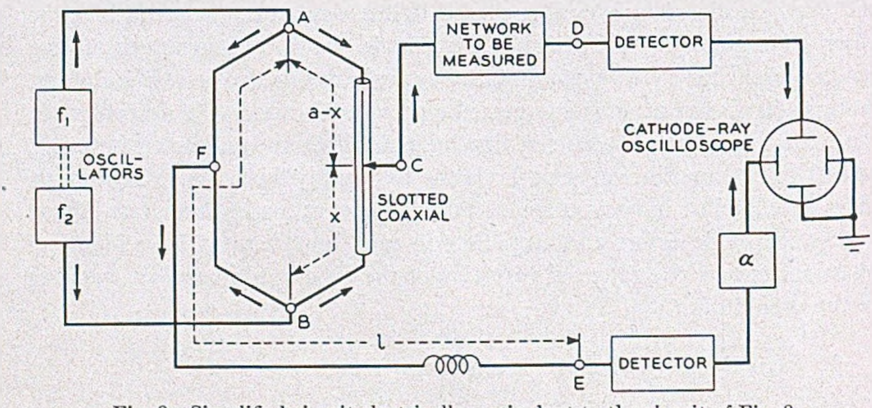


Fig. 9—Simplified circuit electrically equivalent to the circuit of Fig. 8

as compared with the relative phase of the two signals fed to the other branch. Figure 8 shows how the coaxial slider can be used for the latter type of measurement.

The simplified diagram of Fig. 9 will be used to show how the delay of the unknown network can be calculated in terms of this circuit. It will be assumed that the electrical lengths of the paths AF and BF are equal so that this length can be regarded as part of the balancing line ℓ. Then at point E we have

$$\phi_E = \phi_A - \phi_{AE} = \phi_A - \frac{2\pi f_1 \ell c}{v}$$

$$\phi'_E = \phi'_B - \phi'_{BE} = \phi'_B - \frac{2\pi f_2 \ell c}{v}$$

The primes indicate phase shifts at frequency f_2. The phase ϕ''_E of the beat note, $\Delta f = f_2 - f_1$, at E is

$$\phi''_E = \phi'_E - \phi_E = \phi'_B - \phi_A - \frac{2\pi \ell c}{v} \Delta f \tag{10}$$

Similarly, at point D

$$\phi''_D = \phi'_B - \phi_A - \frac{2\pi f_2 x}{c} + 2\pi f_1(a - x) - (\beta' - \beta) \tag{11}$$

If x is adjusted so that the Δf's in the two branches are in phase at the oscilloscope then

$$\phi''_D = \phi''_E - \alpha \tag{12}$$

where α is the difference in the phase shifts in the two beat note circuits. Substituting (10) and (11) in (12) and solving for $(\beta' - \beta)$ yields

$$(\beta' - \beta) = \frac{2\pi}{c}\left[\frac{\ell c \Delta f}{v} - x\Delta f - f_1(2x - a) + \frac{\alpha c}{2\pi}\right] \tag{13}$$

The delay in the network is

$$\begin{aligned} T &= \frac{(\beta' - \beta)}{2\pi\Delta f} \\ &= \frac{1}{c}\left[\frac{\ell c}{v} + \frac{\alpha c}{2\pi\Delta f} - f_1 \frac{(2x - a)}{\Delta f} - x\right] \end{aligned} \tag{14}$$

The first two terms in (14) are independent of x and f_1 and therefore are of interest only for absolute measurements. Relative measurements may be made without evaluating these constants. The last two terms are functions of x and yield the change in delay as f is varied. It is usually most convenient to adjust α and ℓ so that the average delay in the measuring branch is given by (14) with

$$(2x - a) = 0 \tag{15}$$

In general this will minimize the variation of the slider and make the delay distortion in the measuring branch roughly proportional to the slider movement. This is helpful in judging the effect of adjustments of the unknown network. The condition (15) corresponds to the optimum condition for the circuit of Fig. 5 which requires that the average delays of the two branches be equal. In the circuit of Fig. 8, the condition (15) can be realized by varying either ℓ or α. If α is made zero and (15) is fulfilled by adjusting ℓ, errors due to variations in Δf are minimized. However, if Δf is held sufficiently constant, the balancing line ℓ can be omitted entirely and a 360° variable phase shifter introduced in the beat note circuit of one branch to vary α.

The measuring interval Δf is determined by the difference in the two signal oscillators. Since Δf has an important influence on the measurement, oscillator tracking cannot be relied upon to maintain Δf with sufficient accuracy. A one-megacycle crystal frequency checker has therefore been included in the equipment as shown in Fig. 8. A sample of the signal from one of the amplifiers is compared with the crystal oscillator and a trimmer on one of the oscillators adjusted for zero beat before measuring each point. When T is so large that a balancing line is not practicable, as in the case of loop circuits including radio paths or long transmission lines, Δf must be held constant to about 1 part in 10^6. This has been accomplished in a modification of this equipment built by Messrs. W. J. Albersheim and J. P. Shafer of these laboratories, by deriving the two measuring frequencies from a crystal oscillator.

In order to obtain the absolute value of T, the constants ℓ, v, and α must be known. The value of ℓ may be found from the physical length of the line; α can be measured by measuring the movement of the slider required to rebalance the circuit when the connections to the two detectors are reversed. The absolute delay is particularly sensitive with respect to the difference between $2x$ and a. If a is measured as accurately as possible, then the exact setting of the slider corresponding to condition (15) can be found by reversing the connections of the slider at points A and B in Fig. 9. In this way a reference value of x may be found which will yield accurate results in spite of a small error in the value of a.

Successful operation of this circuit depends on low standing waves throughout and upon sufficient padding for satisfactory isolation of the oscillators. It will be noted that in the analysis it was assumed that f_1 reached the slider via the path AC, Fig. 9. There must be an attenuation for f_1 in the path AFBC sufficient to render the signal traversing this path negligible. Similar unwanted paths exist for f_2 from B to C and also for f_1 and f_2 to point F. Attenuation inserted as shown in Fig. 8 can be arranged to make these unwanted signals negligible.

Radio Frequency Measurements

The frequency range of a particular measuring equipment using the circuits of Fig. 5 or Fig. 8 is determined only by the range of the oscillators and the range over which the plugs and jacks and attenuators operate with sufficiently low reflection coefficients. While this range is greater than the IF range encountered in microwave repeaters, it does not include the actual microwave frequencies. However, it has been found that microwave components can be measured satisfactorily by using the circuit shown in Fig. 10. In this circuit the measuring equipment is operated at

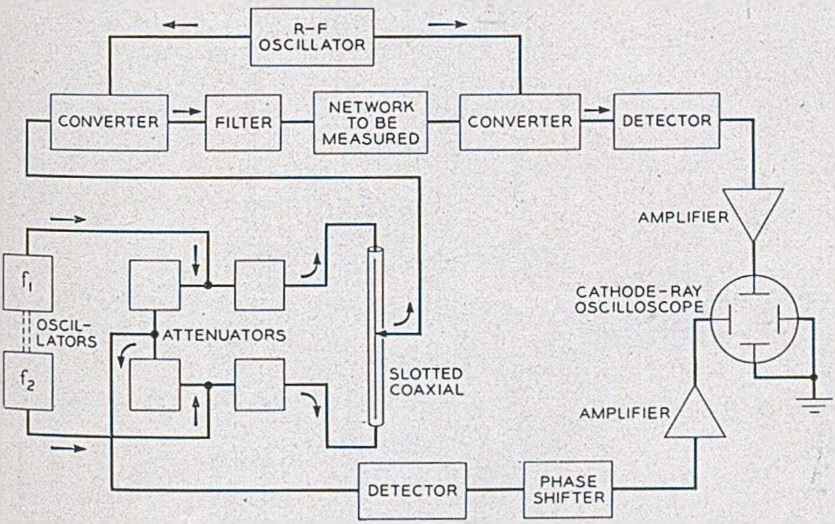


Fig. 10—Method of using the intermediate frequency measuring circuit of Fig. 8 for the measurement of radio frequency networks.

IF, and the reference branch is unaltered. The measuring branch signal is fed first to a converter where it is beat up to the desired microwave frequency. The converter is followed by a filter which eliminates the beating oscillator frequency and one of the beat frequencies. The filter output is then applied to the microwave component under test. The output of this is fed to another converter and converted back to IF by beating with the same oscillator that was used in the first converter. This process removes any variations in phase due to variations in the beating oscillator if the connections from the beating oscillator to each converter are of equal electrical length.

Since considerable extra equipment is included in the measuring branch in Fig. 10, it will usually be necessary to make a calibration run with the device to be measured removed in order to eliminate any possible delay distortion in the converters and filter. Figure 10 uses the measuring

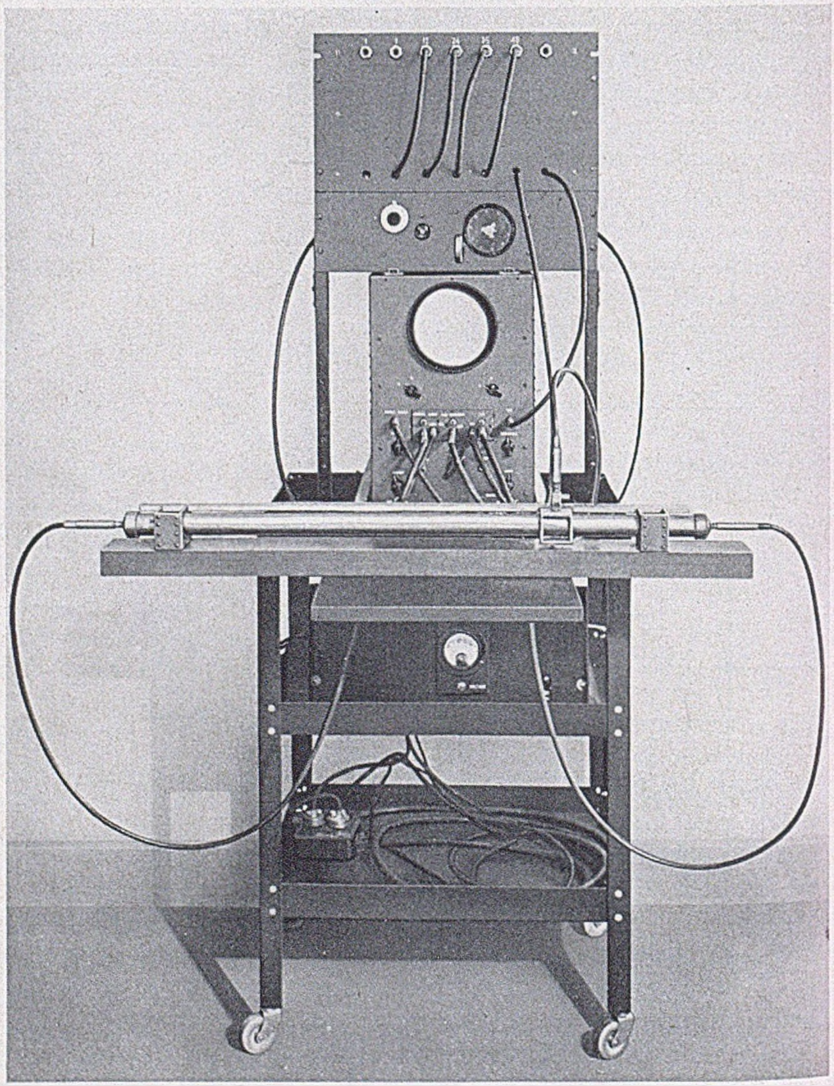


Fig. 11—Photograph of apparatus used in the measuring circuits shown in Figs. 5 and 8.

circuit of Fig. 8 rather than that of Fig. 5 because it was found that small variations in the transit time in microwave amplifiers cause small variations in the phase shift which are the same at all frequencies in the band. These variations cause changes in the relative phase of the two successive measurements required when using the circuit of Fig. 5 which do not represent changes in the delay. In effect the circuit of Fig. 8 makes the two phase measurements simultaneously, and thus eliminates effects due to variations of phase with time.

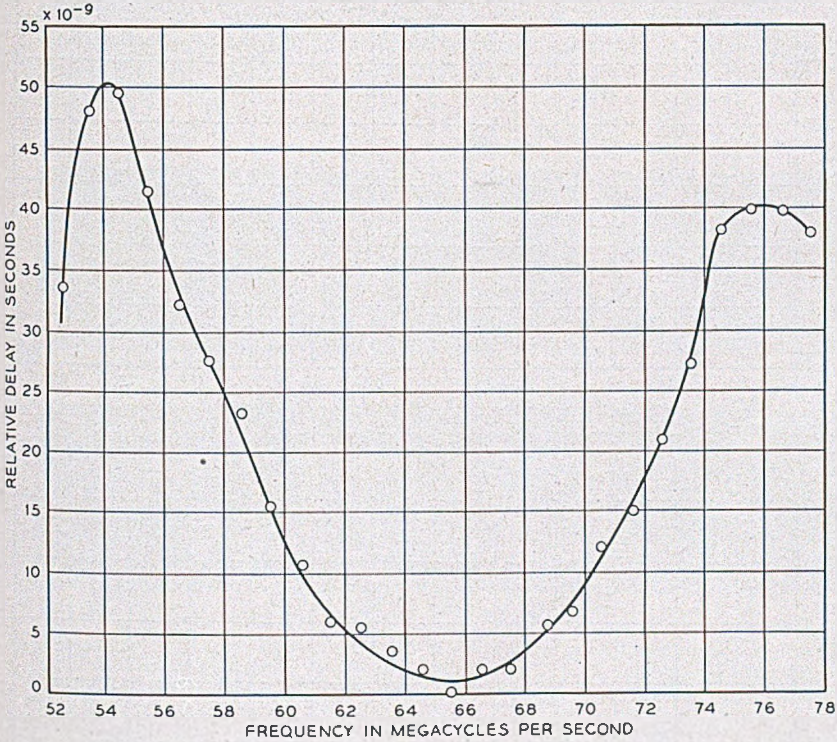


Fig. 12—Measured curve of the relative delay of an intermediate frequency amplifier

Results

Figure 11 shows a photograph of the delay measuring equipment which has been described. With the aid of patch cords and switches that have been included in the equipment this apparatus can be set up according to either Fig. 5 or Fig. 8. The ganged oscillators are on the panel above the oscilloscope box. The box contains the dividing attenuators, detectors, amplifiers and oscilloscope. A separate power supply is required for the oscillators and the output stages of the amplifiers. A number of different lengths of flexible coaxial cable are mounted on the panel above the oscillators and arranged so that various lengths of balancing line can be conveniently obtained by patching. The coaxial phase changer can be seen on the shelf in front of the oscilloscope.

Figure 12 shows the measured relative delay of a typical intermediate frequency amplifier. This amplifier has a substantially flat amplitude response over a band of about 12 mc centered on 65 mc. The delay distortion over the 10 megacycle band from 60 to 70 mc is about 10×10^{-9}

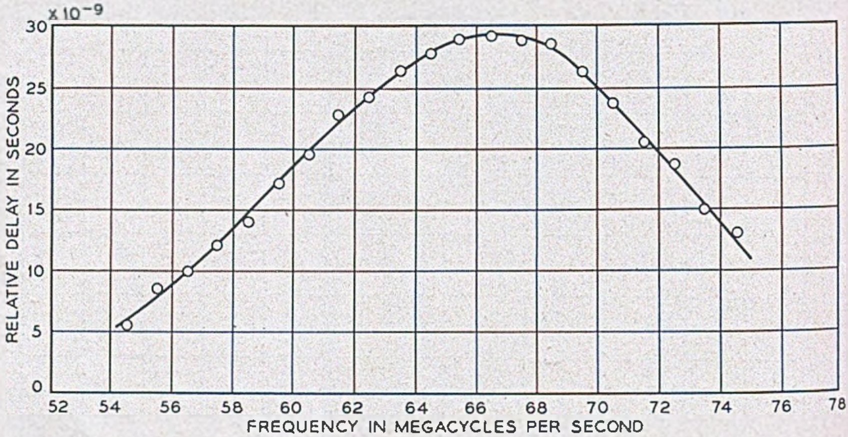


Fig. 13—Measured relative delay of an experimental delay equalizer

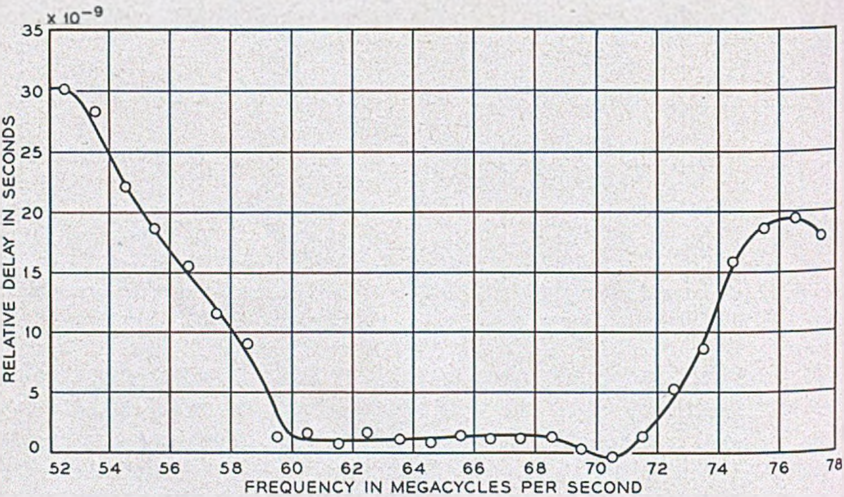


Fig. 14—Measured relative delay of the amplifier of Fig. 12 plus an equalizer

seconds. Figure 13 shows the measured relative delay of a bridged T phase equalizer which can be used to compensate for the distortion in the amplifier. Figure 14 shows the measured relative delay for the amplifier of Fig. 12 and an equalizer measured together. The equalizer has reduced the delay distortion over the 10 mc band from about 10^{-8} to 10^{-9} seconds. These measurements were made with an early model of the measuring equipment. Smoother curves are obtained with the apparatus shown in Fig. 11.

In Fig. 15 the top row shows the distortion of a square top pulse by the amplifier of Fig. 12 for 1, 10, and 30 trips through the amplifier without the equalizer. The lower row shows a similar set of pictures of the pulse when the distortion was reduced by a phase equalizer as shown in Fig. 14. The improvement due to the elimination of phase distortion is clearly illustrated. These pictures were obtained by the circulated pulse[5] technique which

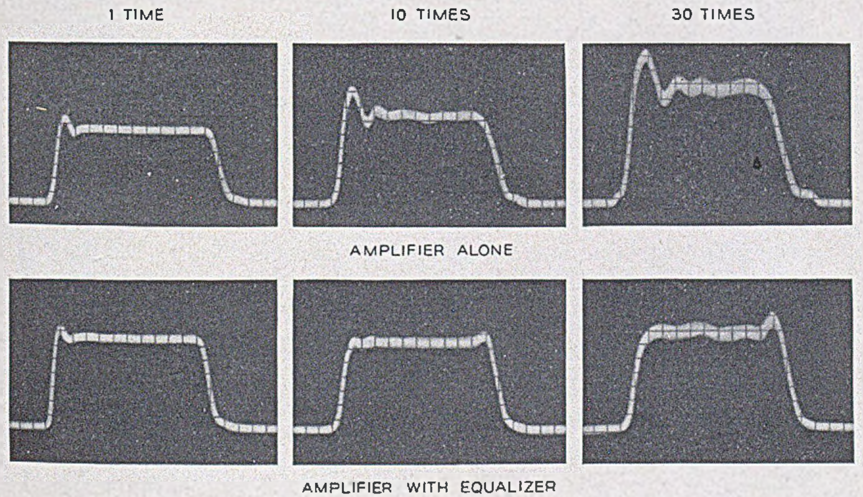


Fig. 15—Oscillograms showing the improvement in the square wave response of the amplifier of Fig. 12 obtained by delay equalization. The numbers above the traces indicate the number of times the test pulse has passed through the amplifier and equalizer.

permits the observation of a pulse after it has passed a number of times through the same amplifier.

Conclusions

Two measuring circuits have been described which are suitable for measuring the small variations in relative transmission time which are present in wide band microwave repeaters. If sufficient care is exercised in setting up these circuits an accuracy of better than $\pm 10^{-9}$ seconds can be realized in relative delay measurements. The circuit of Fig. 5 measures the relative phase shift as a function of frequency. It has the advantage of requiring less signal power and fewer important parameters for absolute

[5] Testing Repeaters with Recirculated Pulses, A. C. Beck and D. H. Ring. Proc. *I.R.E.* Nov. 1947, 1,226–1,230.

delay measurements. The circuit of Fig. 8 measures delay directly with a single measurement and has the advantage of ignoring uniform phase variations with time. It is useful for making relative measurements on circuits with long constant delay times. A significant improvement in the square wave response of carrier amplifiers has been obtained by applying delay equalizers based on measurements made with this equipment.

Frequency Shift Telegraphy—Radio and Wire Applications*

By J. R. DAVEY and A. L. MATTE

Frequency shift telegraphy is described and compared with amplitude modulation telegraphy under various conditions found in radio and wire transmission. Experimental data are given to demonstrate the influence of various design factors on the over-all performance under these conditions. It is shown that the most outstanding characteristic of the frequency shift method is its ability to accept large and rapid changes in signal amplitude. Frequency shift telegraphy thus proves to be of great advantage for use in the H.F. radio range. Frequency shift telegraphy also shows an advantage over amplitude modulation telegraphy with respect to noise. For applications where the level variations are small or slow the advantage of the frequency shift method over amplitude modulation is relatively small.

Introduction

DURING World War II, single-channel and multichannel frequency-shift radio telegraph systems proved of the utmost importance in providing the Allied Powers with a world-wide automatic printing telegraph network for handling with precision, secrecy and dispatch the unprecedented volume of traffic engendered by a war of global extent. It is expected that the next few years will witness a greatly expanded application of this method of operation by commercial telegraph companies and others interested in long distance telegraphy.

Frequency Shift carrier telegraphy (FS) may be applied to any carrier telegraph circuit, but, as will appear below, it provides particularly striking advantages in H.F. radio transmission. For some other radio frequency ranges and for wire line operation the conditions are such as to limit the advantages of the FS method. The main advantages of the FS over the AM method are a greater ability to accept rapid level changes, which results in better stability and lower distortion, and an improvement in signal-to-noise ratio, which permits a reduction in carrier amplitude. It is therefore of particular importance where automatic printing is desired over H.F. radio circuits. When it is necessary to transmit through very high noise levels, low speed AM signaling with aural reception of an audio beat note remains the superior method.

FS is a form of frequency modulation in which signaling is accomplished by shifting a constant amplitude carrier between two frequencies representing respectively the marking and spacing conditions of the telegraph code. Frequency variations in FS telegraphy correspond to amplitude variations in

*Published in *A.I.E.E. Transactions*, Vol. 66, p. 479, 1947.

AM telegraphy (CW); thus the signal transitions in FS are represented by frequency-time transients, while in the AM case they are amplitude-time transients. Since AM telegraph is the more common system, a discussion of the FS method involves numerous comparisons between the two systems. The merits of a telegraph system must be judged on its ability to combat the various adverse conditions encountered in the transmission medium and in the terminal apparatus. In general these adverse conditions involve variations in amplitude, frequency, and phase of the signals and the presence of extraneous signals and noise.

In the course of the development of a number of FS radio teletypewriter systems, a large amount of information concerning the characteristics and design parameters of such equipment has been obtained. It is the purpose of this paper to abstract therefrom selected data which will furnish a step-by-step comparison of the FS and AM methods. Typical terminal arrangements are described and the effects of varying certain design factors are illustrated by experimental data. Although the material presented applies largely to H.F. radio telegraph, much of it is of a general nature and with proper interpretation applies to other frequency ranges and transmission mediums and to cases in which the telegraph modulated carrier may be a sub-carrier or one of several sub-carriers.

General Discussion

Sideband Energy Distribution

The difference between FS and AM signals as regards distribution of sideband amplitudes is illustrated by the following two equations for a carrier of frequency $\omega/2\pi$ modulated with unbiased square wave dots of frequence $\rho/2\pi$.

For AM (On-off) keyed carrier of unity amplitude[1]:—

$$e = 0.5 \cos \omega t + \frac{1}{\pi}[\cos(\omega + \rho)t + \cos(\omega - \rho)t]$$
$$- \frac{1}{3\pi}[\cos(\omega + 3\rho)t + \cos(\omega - 3\rho)t]$$
$$+ \frac{1}{5\pi}[\cos(\omega + 5\rho)t + \cos(\omega - 5\rho)t]\ldots\text{etc.} \quad (1)$$

For FS keyed carrier of unity amplitude[2]:—

$$e = \frac{2m}{\pi}\left[\frac{1}{m^2}\sin\left(\frac{m\pi}{2}\right)\cos\omega t\right.$$

$$+ \frac{1}{m^2 - 1^2} \cos\left(\frac{m\pi}{2}\right) (\cos (\omega - p)t - \cos (\omega + p)t)$$

$$- \frac{1}{m^2 - 2^2} \sin\left(\frac{m\pi}{2}\right) (\cos (\omega - 2p)t + \cos (\omega + 2p)t)$$

$$- \frac{1}{m^2 - 3^2} \cos\left(\frac{m\pi}{2}\right) (\cos (\omega - 3p)t - \cos (\omega + 3p)t)$$

$$+ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \Big] \quad (2)$$

where m is the deviation ratio $= \dfrac{\text{frequency shift}}{2 \times \text{signaling speed}}$.

Typical sideband amplitudes calculated from these formulas are shown in graphical form in Fig. 1. In the case of FS keying, the relative amplitudes of the sidebands vary considerably as the amount of frequency shift is changed. For miscellaneous signals these line spectra do not exist but they do indicate the general distribution of energy over the band for a given signaling speed.

Methods of Modulating the Carrier

In AM telegraphy the carrier is usually modulated by simply interrupting it for the spacing condition. This is sometimes referred to as "on-off" keying. For low power and low frequencies the carrier may be keyed directly by electrical contacts. A more universally applicable method is to use vacuum tubes or other nonlinear elements to effectively interrupt the carrier. In some cases it is practical to start and stop an oscillator source of carrier.

The usual radio telegraph transmitter consists of an oscillator followed by a number of cascaded stages of amplification and frequency multiplication arranged to reach the desired output frequency and power. For on-off keying the carrier is usually interrupted by suitably varying the plate or grid voltage of one or more of the stages.

There are two general methods of obtaining frequency modulation: (a) The frequency of an oscillator may be modulated directly by suitably varying the frequency-determining circuit, (b) the output of a constant-frequency oscillator may be shifted in phase at such rates of change as to produce indirectly the desired frequency variations. In the latter case the marking and spacing intervals of an FS signal would be formed by periods of constant rate of phase change versus time. Square wave reversals would therefore require a triangular shaped wave of phase versus time. Since the transmission of long periods of steady mark or space would therefore involve huge phase swings, the modulator would have to be able to produce a steady phase rate of charge. A phase modulator capable of performing in this way while not producing undesirable phase discontinuities at the signal transitions becomes rather impractical. For this reason FS telegraphy usually utilizes the direct frequency modulation method. This may conveniently be accom-

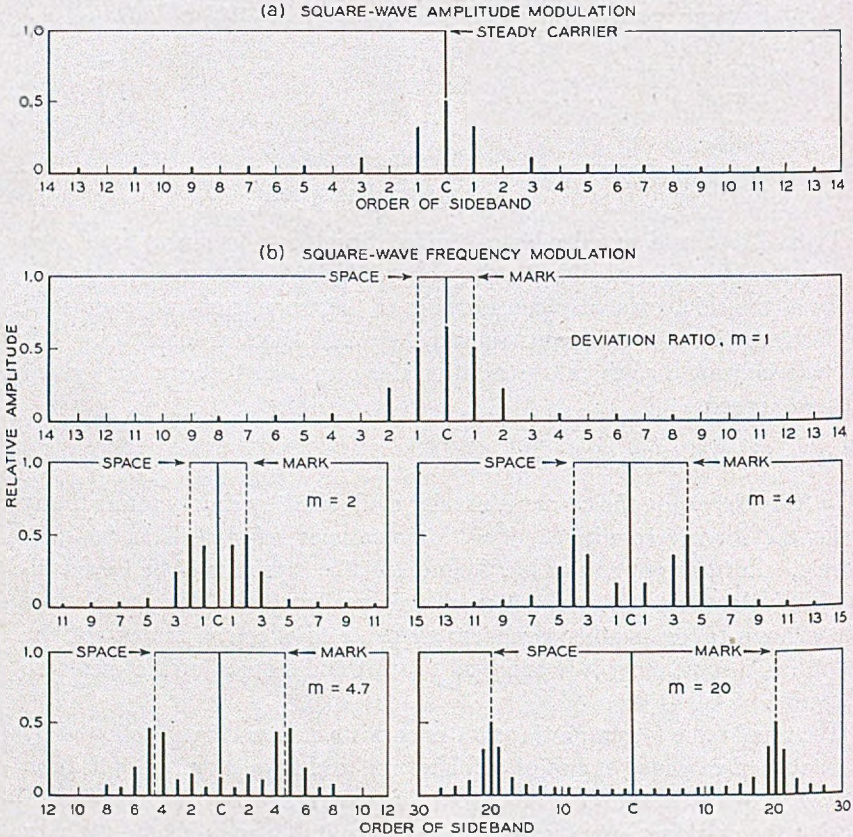


Fig. 1.—Amplitude of sideband components for (a) square-wave amplitude modulation (b) square-wave frequency modulation.

plished by the use of a reactance modulator which, by injecting a reactive component of current into the tuned circuit of the oscillator, varies the resonant frequency thereof. Such a modulator may be made linear so that a frequency shift proportional to the input voltage to the reactance modulator is obtained.

To apply FS telegraph signals to a radio transmitter the regular exciter oscillator is either replaced or modified by an arrangement providing a source of R.F. excitation that can be shifted in frequency in accordance with the telegraph signal. All the stages are operated with full R.F. excitation continuously to produce a constant amplitude carrier.

As a matter of expediency frequency shift keying has sometimes been provided by switching between two independent sources of carrier current separated in frequency by the desired shift. In such a case the frequency transitions involve sudden phase discontinuities of random values. This results in the instantaneous frequency swinging considerably outside the steady-state mark and space frequencies. If the band is wide, such as is the case in a radio transmitter, there results a very broad sideband radiation capable of causing severe interference to adjacent channels. If the band is narrow, as might be the case where sending filters are employed, the interference is eliminated but the amplitude transients resulting from the sudden phase shifts are capable of producing considerable distortion.

Restriction of Transmitted Band

As seen above, square-wave modulation results in a wide spread of sideband components which are of sufficient amplitude to interfere seriously with adjacent channels unless greatly attenuated. The transmitted band may be restricted either by the use of a band-pass filter centered about the carrier frequency or by a low-pass filter to suitably shape the modulating wave form. Band-pass filters are usually used if the power level is low and the frequency low enough to permit suitable filter construction. For multichannel systems the use of band-pass filters also permits efficient paralleling of the transmitting channels. For radio transmitters with a transmitted power measured in kilowatts, and with a frequency of several megacycles which is frequently changed to suit best the prevailing conditions, shaping of the modulating wave is the more practical method of restricting the transmitted band.

Insufficient attention has been given in the past to the envelope shape of the signals from on-off keyed radio transmitters. With the ever increasing crowding of frequency assignments it becomes more and more important to restrict the emission of unnecessary sidebands arising from keying. The envelope shape in on-off keying may be controlled by properly shaping the modulating grid or plate voltage wave. It is important that the stages following the keyed stage or stages be nearly linear, otherwise the wave shaping will be largely destroyed. In the case of frequency shift keying, on the other hand, the wave shaping is preserved after passage through class C amplifier or multiplier stages, and these may be operated for maximum efficiency. The greater ease of producing and maintaining the desired wave shaping, so necessary for close frequency spacing of channels, is one of the outstanding advantages of frequency shift keying.

Apparatus

Typical FS Exciter for Radio Telegraph

A typical FS exciter arrangement such as is often used with radio telegraph transmitters is shown in Fig. 2. A d-c. telegraph wave, after suitable shaping, is caused to frequency-modulate an intermediate frequency of 200 kc. which, in turn, amplitude-modulates a radio frequency from a crystal-controlled oscillator. The upper sideband of this latter modulation is an FS signal and is selected and amplified sufficiently to drive the first amplifier or multiplier stage of the transmitter. The 200-kc. oscillator is frequency-modulated by a reactance modulator which, by feeding a leading or lagging

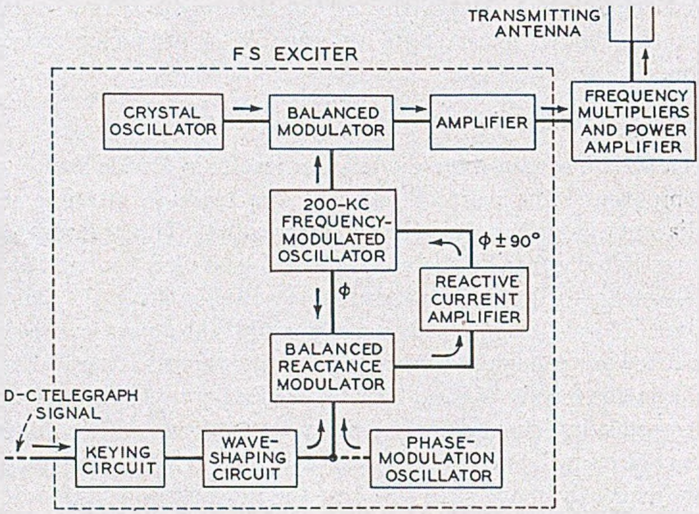


Fig. 2.—Block diagram of a typical FS transmitter.

quadrature component of current into the oscillator tuned circuit, decreases or increases the frequency. By operating the reactance modulator within its linear range the frequency shift wave form is made the same as the d-c. telegraph wave form into the modulator. A d-c. amplifier stage, designated "keying circuit", is provided to furnish a modulating wave effectively isolated from amplitude and wave front variations of the incoming telegraph signals. The d-c. telegraph signals may be polar or neutral and are often obtained from a tone demodulator unit which allows keying from a remote point by V.F. telegraph. The amount of frequency shift is adjusted by an amplitude control in the quadrature feed-back path to the 200 kc. oscillator. The shift may thus be varied continuously, or in definite steps to allow for subsequent frequency multiplications, by suitable attenuation controls. Controlling the shift in this manner keeps the instabilities of the reactance modulator a constant percentage of the frequency shift, which would not be the case if the shift were adjusted by varying the amplitude of the modulating wave. The use of a balanced instead of an unbalanced reactance modulator minimizes variation of the mean frequency and also allows the shift to be varied without affecting the mean frequency.

The frequency-shift signal transitions are wave-shaped, to restrict sideband radiation, by means of a low-pass filter in the d-c. telegraph signal path to the reactance modulator. The low-pass filtering is made adjustable to accommodate a range of signaling speeds. Frequency-versus-time wave

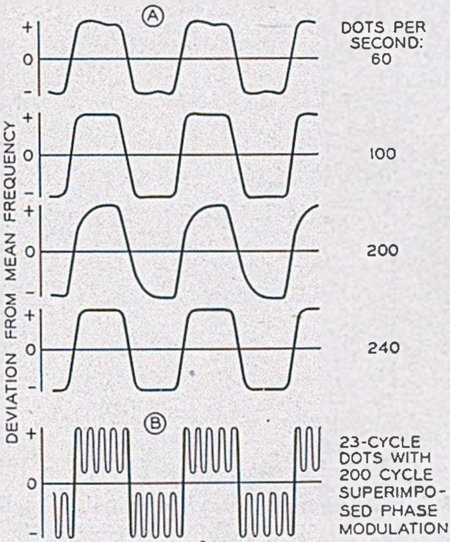


Fig. 3.—Frequency shift keyer output wave forms. (a) Low-pass filtering adjusted to produce similar wave shapes at dotting speeds of 60, 100, 200, and 240 cycles. (b) 200 cycle phase modulation superimposed on a 23 dots per second signal.

forms from an exciter of the type described are shown for several keying speeds in Fig. 3a. The effect of wave shaping on the sideband components in the R.F. output of such an exciter is shown in Fig. 4.

Phase modulation may readily be added to the signal in this type of exciter by superimposing the desired sine wave modulating frequency on the telegraph signal wave input to the reactance modulator as indicated in Fig. 2. Figure 3b shows the keyer output wave form with superimposed phase modulation. The use of this type of phase modulation is considered later.

To obtain optimum results in FS radio telegraph transmission and to allow close spacing of channels, a high degree of frequency stability is necessary. An over-all frequency stability of ± 100 cycles is desirable in a system using a value of frequency shift between 500 and 1000 cycles. A frequency shift exciter of the type described above, with the crystal oscillator and 200 kc. FS oscillator located in a temperature-controlled oven, usually has a frequency stability such that the mean R.F. carrier frequency may be held to within ±50 cycles up to frequencies of 20 mc over ordinary periods of operation on any one frequency. One of the advantages of this type of exciter is that small inaccuracies in crystal frequencies may be compensated for by adjusting the mean frequency of the 200 kc. oscillator.

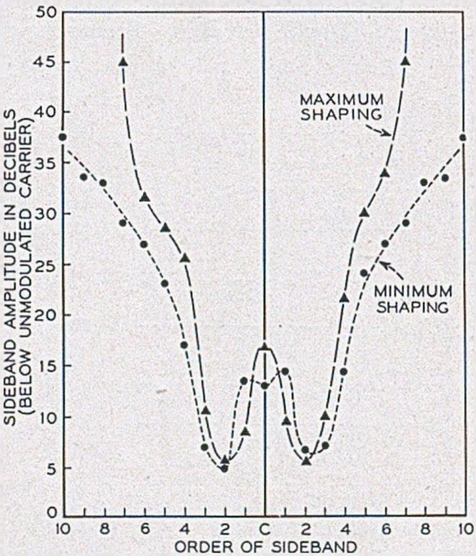


Fig. 4.—Effect of shaping FS transitions on amplitude of radiated sidebands. 100 dots per second and 500 cycle frequency shift. The maximum and minimum wave shaping conditions correspond to those used with 60 and 240 dots per second respectively in Fig. 3(a).

Receiving Terminal Arrangements

Typical receiving terminal arrangements are shown in the block diagrams of Fig. 5. Up to point "a" in the arrangements the FS and AM systems are identical, being of the usual H.F. superheterodyne type. The output from the second frequency-conversion stage may be either in the audio range or at a considerably higher frequency such as 50 kc. Following the second converter is a band-pass filter (shown at "b" in Fig. 5) which determines the final over-all band width before demodulation. The two systems differ only in the method of demodulation. The AM (on-off) signals are amplified and rectified to give a d-c.telegraph signal. The FS signals are amplitude limited and passed through a frequency discriminating network and then rectified to give a d-c.telegraph signal. Beyond this point the two systems are again identical. The d-c. signals pass through a low-pass filter to remove carrier ripple and higher frequency noise components and are then amplified to a suitable level to operate automatic recording or printing apparatus. The d-c.signals may also be used to modulate an audio frequency so as to pass the signals to a remote point by multichannel voice-frequency carrier telegraph methods.

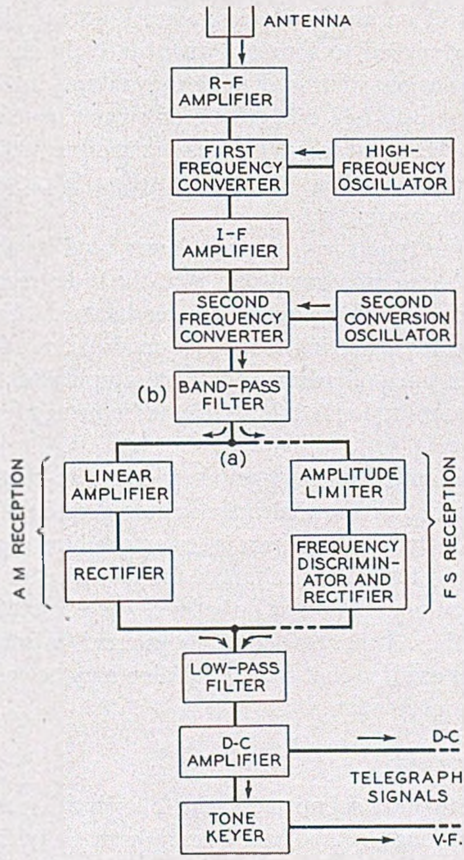


Fig. 5.—Block diagram of a typical receiving arrangement for either AM or FS telegraph signals.

The radio receiver portion of the terminals up to point "a" should be designed to have low noise and good selectivity. Extreme H.F. oscillator stability is necessary for either system if narrow band width operation is to be maintained without constant attention. An over-all frequency stability of ±50 cycles is desirable for the receiving terminal over a period of 6 to 8 hours. Sufficient selectivity and amplifier capacity should be provided at all points to prevent overloading by unwanted signals or loss of automatic gain control. In the following discussion those portions of the terminals beyond the second frequency converter will be given major attention.

Experimental Transmitting and Receiving Arrangements

For the laboratory transmission studies described in the following sections the transmitter and receiver were located nearby and connected together by means of an amplitude modulator and associated with various sources of noise designed to simulate quantitatively and under controlled conditions the variations which would be encountered in the actual medium.

Throughout the tests 7.42 unit start-stop signals were used unless otherwise stated, and the speed was 60 words per minute (23 dots per second). Their peak distortion and bias were measured on a cathode-ray tube telegraph distortion measuring set.

An exciter of the type shown in Fig. 2 was used as a source of signals. A frequency of 6.4 mc. was employed, with the radio receiver connected to the exciter output through an amplitude modulator. This modulator was an electronic circuit permitting amplitude modulation of a frequency-shift signal to produce unequal mark and space amplitudes. This modulator was also used to amplitude-modulate a single frequency for the AM portions of the measurements.

A temperature-limited diode together with a two-stage tuned amplifier was used as a source of thermal noise centered around 6.4 mc. A polar relay driven by 60-cycle a-c and arranged to produce sharp polar impulses from the discharge of small capacitances connected to its contacts was used as a source of impulse noise. The noise level was adjusted by an attenuator and mixed with the 6.4 mc. carrier of the exciter. A minimum amount of wave shaping was used, so that the modulation may be considered as having been essentially square-wave.

Receiving Arrangements

The experimental data submitted in the following discussion was obtained from reception through a laboratory setup essentially like that shown in Fig. 5. The radio receiver proper was a commercial type of H.F. superheterodyne. The output of the second frequency converter was in the audio-frequency range, which enabled the use of various band-pass filters at "b" of the type used in voice-frequency telegraph systems. The amplitude limiter was effective over an input range of −60 dbm* to above +20 dbm. The pass-band characteristics of the radio receiver and of the several band-pass filters used in position "b" are shown in Figs. 6 and 7 respectively. Unless otherwise stated, the frequency shift signals were centered about

* The symbol dbm signifies "db referred to one milliwatt".

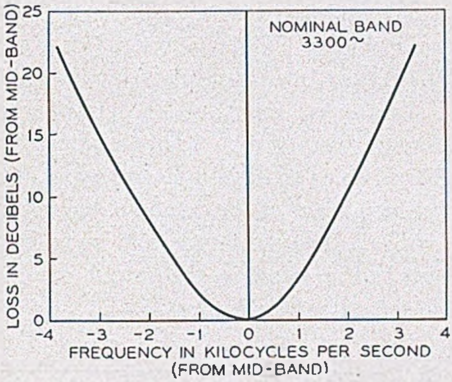


Fig. 6.—I.F. selectivity characteristic of radio receiver.

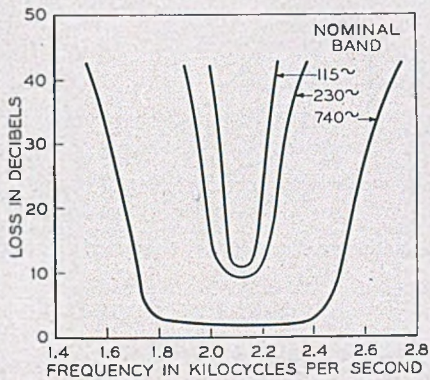


Fig. 7.—Attenuation versus frequency characteristics of bandpass filters shown at (b) in figure 5.

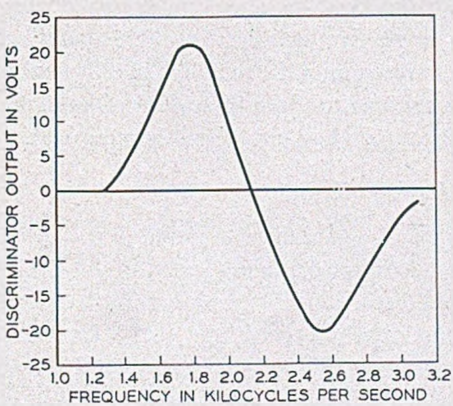


Fig. 8.—Linear discriminator characteristic.

2125 cycles and were demodulated by a linear discriminator centered about 2125 cycles as shown in Fig. 8. The characteristics of the low-pass filtering are shown in Fig. 9. These low-pass filters were adjusted by oscillographic observation of the signal wave form and had cut-off characteristics giving very little characteristic distortion[3]. The d-c. amplifier was a high-gain nonlinear type designed so as to have a square-wave output having transitions established by the passage of the demodulated voltage wave through a narrow amplitude range. The amplitude and wave front slope of the demodulated wave thus had no effect on the output wave form and could not affect the distortion measuring equipment.

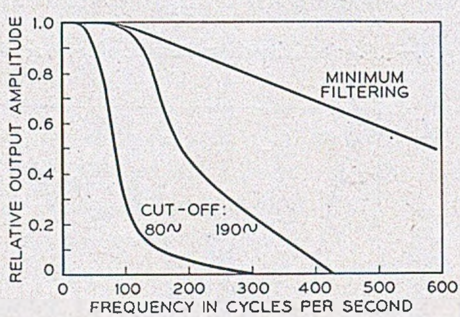


Fig. 9.—Attenuation versus frequency characteristics of low-pass filters.

Experimental Results

Band Width Before Demodulation

The band width before demodulation determines the amount of noise and interference which is to be accepted along with the desired signal and thus largely determines the signal-to-noise condition at the antenna at which the system fails to receive intelligence. The band width at this point (point "a" in Fig. 5) also limits the signaling speed capabilities of the system. In the following experimental data the values of band width were measured between the points of 6 db loss above that at midband.

Effect on Signaling Speed

For both methods of signaling considered here the band width must be at least twice the maximum signaling speed in dot-cycles per second but it is found that signal distortion rises rapidly for a band width less than three times the maximum signaling speed, and that a factor of at least four times is indicated for a system which is to have reasonably low distortion with any margin of safety. The signaling speed capability of a given band width is nearly the same for FS signals as for AM (on-off) signals. In Fig. 10 is shown the over-all signaling frequency response to FS and AM signals for a nominal band width at point "a" of about 740 cycles. It will be noted that the FS method is but slightly inferior and that both systems fail at a frequency of approximately one-half the band width.

Effect on Noise:

The effect of *thermal noise* on distortion for bandwidths of 115, 230, and 740 cycles is shown in Figs. 11 and 12. The rms. noise-to-carrier ratios

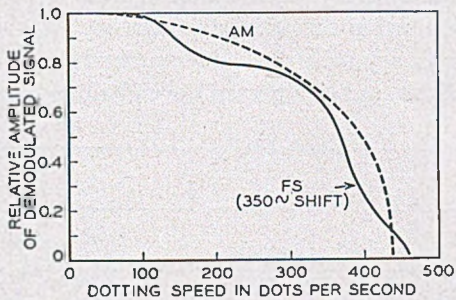


Fig. 10.—Overall frequency response of a 740-cycle band to AM and FS signals.

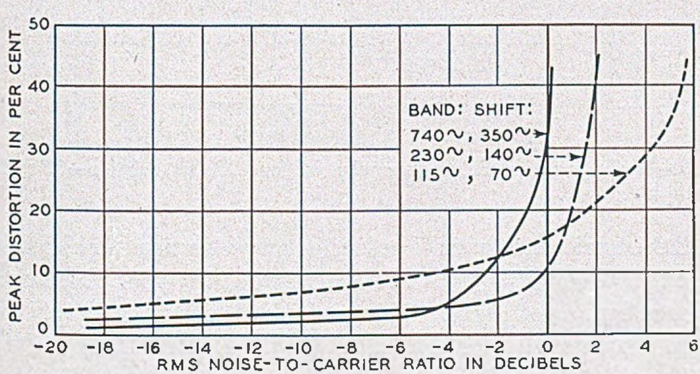


Fig. 11.—Peak distortion versus thermal noise for FS transmission—80-cycle cutoff low-pass filter.

indicated in the figures were measured in the 3300-cycle band of the radio receiver in all cases. The actual noise-to-carrier ratio existing in the transmission band used was lower and may be obtained from the following table. The values of correction are from rms. measurements.

Pass Band at Point "a"	Correction to be made
1920 cycles	−2.3 db
740	−6.5
230	−11.6
115	−14.3

For *AM signals*, Fig. 12, varying the bandwidth has little effect when the telegraph signal distortion is less than 15%. Although the wider bands accepted more noise power, this added noise merely produced high-frequency noise components which were removed by the low-pass filter. This added noise does, however, cause the peak noise to exceed the signal amplitude at a lower noise-to-carrier ratio and cause failure before that for a narrower band condition.

In the case of *FS signals*, Fig. 11, changing the bandwidth and the frequency shift simultaneously, and in approximately the same proportion, alters the whole distortion characteristic. At low noise levels a wider band with a greater frequency shift gives an improved signal-to-noise condition. However, as the noise level is increased the wider band causes the peak noise

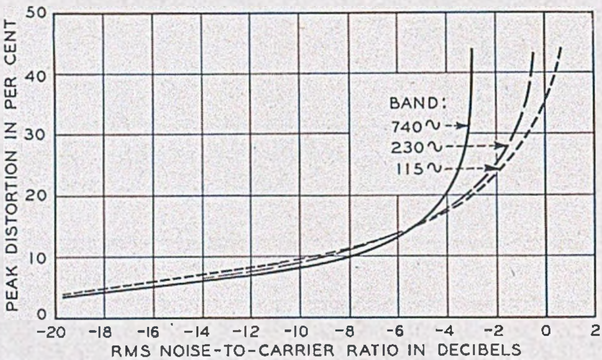


Fig. 12.—Peak distortion versus thermal noise for AM transmission—80-cycle cutoff low-pass filter.

to exceed the carrier at a lower noise level than with a narrower band. Thus a change to a wider band gives less distortion at low noise levels and more distortion at high noise levels. This results in a much more sharply breaking distortion characteristic for the wider band. This behavior is typical of frequency modulation systems in general.

Although the noise actually passed by the 740-cycle band filter was approximately 8 db above that passed by the 115-cycle band filter, the difference in the failure points (35–40% distortion) for the two bandwidths will be seen to be only about half this amount in db for both AM and FS. This phenomenon is typical of carrier telegraph systems when compared at the same signaling speed. This means that it is not particularly beneficial to decrease band width to obtain lower distortion under high noise conditions. The main reason for narrow bands is for more economical use of frequency space.

As to the *comparison between FS and AM*, the FS method has an advantage of 2.5 to 4.5 db at a distortion of 35% to 40%, corresponding to the selector failure point of the usual teletypewriter. From a signal-to-noise standpoint it is thus seen that the gain in changing to the FS method is approximately equal to the resulting increase in average transmitted power of about 3 db. A comparison at a lower distortion such as 15% shows an advantage of 4 to 6 db. At a still lower distortion the 740-cycle band, because of the higher deviation ratio, shows an improvement of over 10 db. In this region the slopes of the curves make accurate comparisons impossible due to the masking effect of other sources of distortion. These large improvements at low noise levels are similar to those associated with wide-band FM broadcast systems. However, in carrier telegraph transmission the criteria are so different that the difference between a nearly perfect

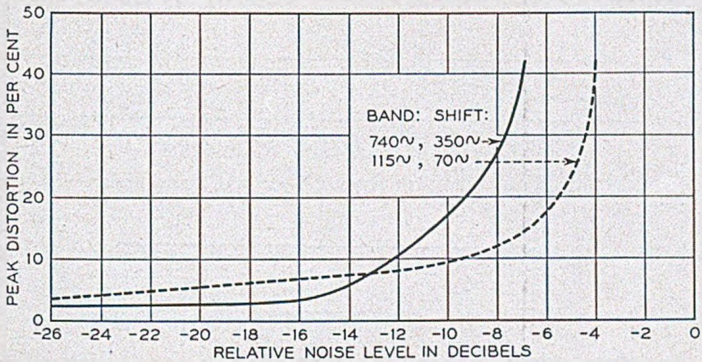


Fig. 13.—Peak distortion versus impulse noise for FS transmission—80-cycle cutoff low-pass filter.

circuit and one of small distortion is not of great importance except when a large number of telegraph sections are to be operated in tandem. From a practical standpoint the improvement in signal-to-noise is not more than about 6 db for equal band widths.

In Figs. 13 and 14 similar characteristics are shown for the case of *impulse noise.* The noise level values of these curves are purely relative since no attempt to measure the peak noise was made. The comparisons between AM and FS, and between different band widths, agree closely with those for thermal noise.

Effect of Limiter on Signal-to-Noise Ratio

The limiter in the FS system is a high-gain nonlinear amplifier which delivers to the frequency discriminating networks an essentially square wave having transitions coinciding with the passage of the instantaneous voltage of the input carrier signal through zero. The limiter thus passes only the frequency or phase changes of the signal. Noise voltages which are small compared to the signal cause approximately linear phase modulation of the signal and this is passed through the limiter. The amount of frequency deviation thus imparted to the carrier by a given component of noise is proportional not only to its amplitude but also to its frequency separation from the carrier. This gives rise to the so-called "triangular noise spec-

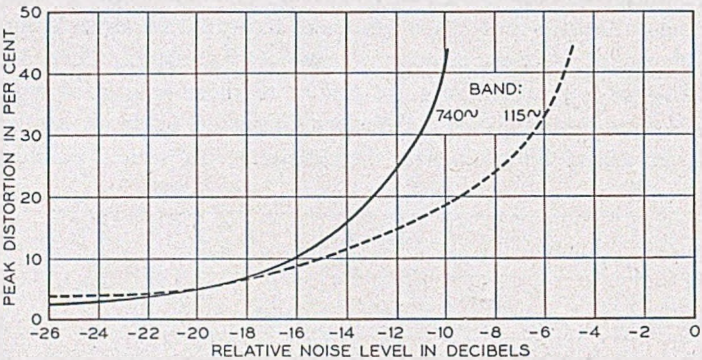


Fig. 14.—Peak distortion versus impulse noise for AM transmission—80-cycle cutoff low-pass filter.

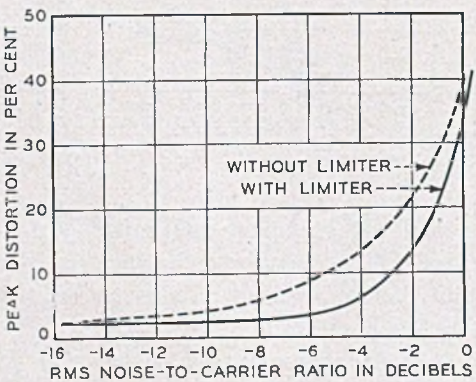


Fig. 15.—Effect of the limiter on distortion versus thermal noise for FS transmission—740-cycle band, 350-cycle frequency shift, 80-cycle cutoff low-pass filter.

trum"[4] when a linear frequency discriminator is used. If the limiter is removed from the frequency-shift terminal the noise components in phase with the carrier as well as those in quadrature therewith are allowed to reach the frequency discriminating network. For a balanced type discriminator this increases the demodulated noise for small amounts of noise about 3 db. In Fig. 15 is shown the effect of removing the limiter from the circuit. The limiter is seen to have little effect on the failure point, but an improvement of 2 to 4 db is shown in the 5 to 12% distortion region.

Other beneficial effects resulting from the use of a limiter are discussed below under "Level Variations".

Demodulation of Frequency Shift Signals

It is desirable that the frequency discriminating network be of the balanced type having two branches allowing differential combination of the two rectified outputs. This minimizes the response to amplitude modulation not eliminated by the limiter. Two general types of networks have been in common use for FS telegraph. One consists of two bandpass filters centered about the mark and space frequencies respectively and effectively dividing the total band into halves. The other consists of a two-branch network each branch of which has a varying amplitude characteristic extend-

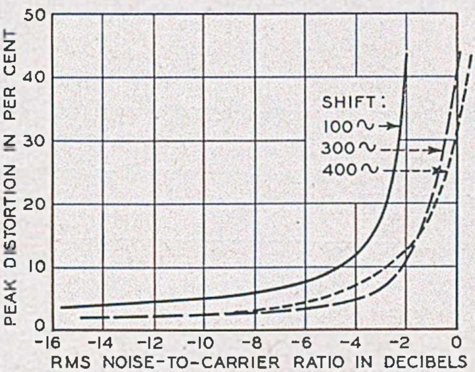


Fig. 16.—Effect of magnitude of frequency shift on distortion versus thermal noise in FS transmission—740-cycle band and 80-cycle cutoff low-pass filter.

ing over the complete transmission band and usually well beyond. The amplitude-versus-frequency characteristics of these two branches have opposite slopes and are of such shape that differential combination of their rectified outputs results in an approximately linear voltage-versus-frequency curve, passing through zero at midband. (Fig. 8)

In Fig. 17 is shown the characteristic of a two-bandpass-filter type of discriminator which was used in early frequency shift terminals. The characteristic is fairly flat near the mark and space frequencies so that this type of discriminator does not produce a triangular noise spectrum. In Fig. 18 is shown the type of discriminator characteristic obtained by the use of two narrow bandpass filters. In this case there is no broad flat region around the mark and space frequencies and an intermediate type of characteristic (approaching the linear type) is obtained.

With the linear type of discriminator the demodulated noise has the well known triangular spectrum and, as illustrated previously, the signaling speed capability is essentially the same as an AM system of equal band width.

To compare experimentally these two general types of discriminators a 740-cycle band system with linear discriminator and 350-cycle shift was

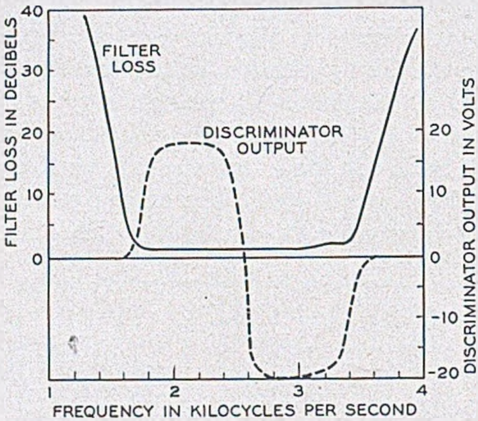


Fig. 17.—Characteristics of 1920-cycle bandpass filter and associated discriminator consisting of two bandpass filters.

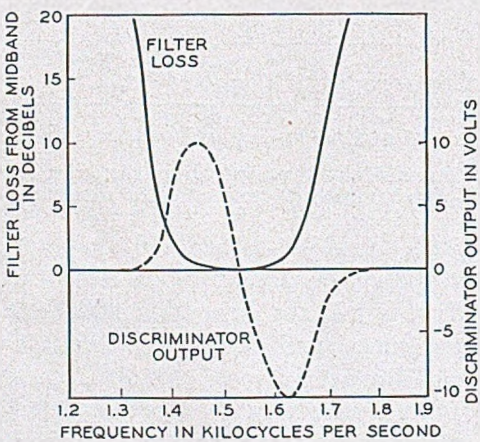


Fig. 18.—Characteristics of 295-cycle bandpass filter and associated discriminator consisting of two bandpass filters.

compared with a system with a bandwidth of about 1900 cycles, a discriminator consisting of two 740-cycle bandpass filters, and an 850-cycle shift. The results are shown in Fig. 19. The two systems are seen to reach failure distortion values at the same signal-to-noise point, with the linear discriminator becoming about 3 db superior at distortions around 5%. A second comparison was made using roughly equal bandwidths. A 295-cycle band with a discriminator consisting of two bandpass filters and 170-cycle shift was compared with a 230-cycle band with a linear discriminator and a shift of 140 cycles. The results are shown in Fig. 20. The linear discriminator in this case appears to fail slightly sooner but shows a superiority of about 2 db in the 5% distortion region. Due to the rounded character-

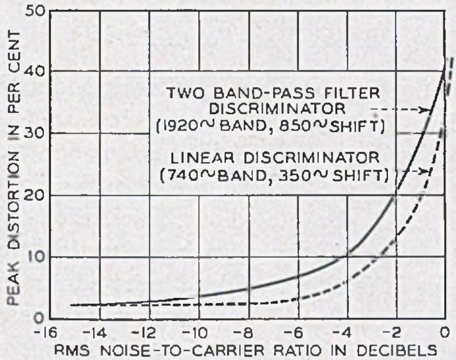


Fig. 19.—Linear discriminator versus two-bandpass-filter discriminator having the same signaling speed capability. Effect of thermal noise on peak distortion—80-cycle cutoff low-pass filter.

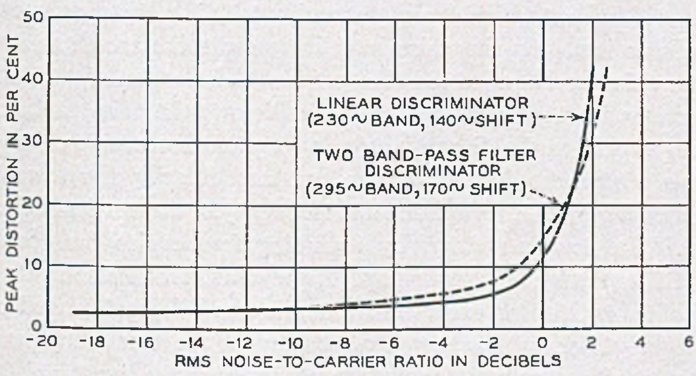


Fig. 20.—Linear discriminator versus two-bandpass-filter discriminator with equal bandwidths. Effect of thermal noise on peak distortion—80-cycle low-pass filter.

istic of the two bandpass filters used as a discriminator in the second comparison (Fig. 18) the difference in discriminators is less than in the previous test. It has been found[5] that for *a given signaling speed* capability almost twice the bandwidth is required if a two-bandpass-filter discriminator is used instead of the linear type. This added band width does not appear to cause any loss in signal-to-noise capabilities as to the failure point. The less sharp breaking point, however, makes the linear discriminator superior for moderate and low distortions. For a system occupying *a given bandwidth* the two-bandpass-filter discriminator provides some improvement at the failure point but is still somewhat inferior to the linear discriminator at moderate distortions. More importantly the two-bandpass-filter discriminator impairs the signaling speed capabilities to an extent which depends upon the shape of the cutoff of the filters used.

Bandwidth After Demodulation (*Low-Pass Filtering*)

In an *Am system* the low-pass filtering after demodulation can, to a large degree, make up for a greater than necessary bandwidth before demodulation. During *marking intervals* the added noise admitted by a wide band causes noise in the demodulated output at frequencies higher than the signaling frequency and this can be filtered out, unless the noise is so great as to over-modulate the carrier. During *spacing intervals* there is no carrier and hence only the noise is rectified. Added noise admitted by a wide band causes not only higher-frequency components in this rectified noise, which may be filtered out, but also an increase in the d-c. component. This tends to cause marking bias of the received signals as the noise level increases.

In an *FS system*, where the carrier is present continuously, the added noise from a wider band produces high-frequency noise components in the demodulated output which can be filtered out by the low-pass filter if the noise level is low. As the noise level increases there are short intervals when the noise envelope exceeds the carrier. The action of the limiter is to give preference to the greater signal, in this case the noise, and since the noise will appear to the discriminator as a carrier fluctuating around mid-band as a center, the demodulated output momentarily dips toward zero. As the noise increases, the duration and frequency of these holes in the signal increase. The low-pass filter, by excluding frequencies considerably in excess of the maximum signaling speed, prevents these holes in the signal from producing false or extra transitions in the telegraph signal output. The low-pass filtering, however, cannot prevent the true transitions from being displaced by this type of noise component since the signal is obliterated momentarily. Its most important function is to prevent a breakup in the signal output until a fairly high distortion is reached. For noise peaks exceeding the carrier the low-pass filter of an AM system also serves much the same purposes.

In Fig. 21 is shown the effect of changing the bandwidth of the low-pass filter in an FS and in an AM system in the presence of thermal noise. The effect of a narrower low-pass filter is seen to consist mainly in shifting the breaking point toward a higher noise level. Similar characteristics for the case of impulse noise are shown in Fig. 22.

Magnitude of Frequency Shift in Relation to Bandwidth

A frequency-shift transient in a band of given width has a wave shape much like that of an amplitude transient in the same band provided the shift is symmetrical and not over 50% of the bandwidth.[6] If the frequency shift approaches the total width of the band the transient is of such shape as to

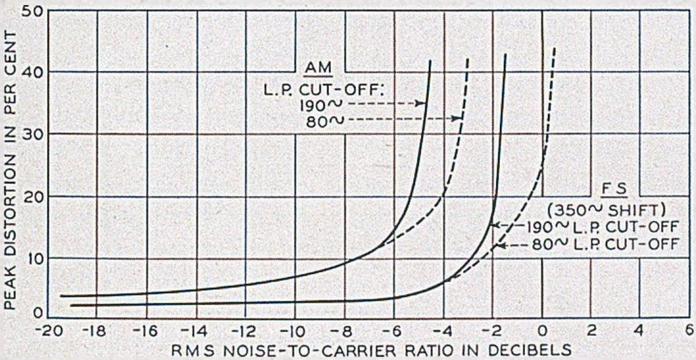


Fig. 21.—Effect of the low-pass filter cutoff frequency on distortion in the presence of thermal noise—740-cycle band.

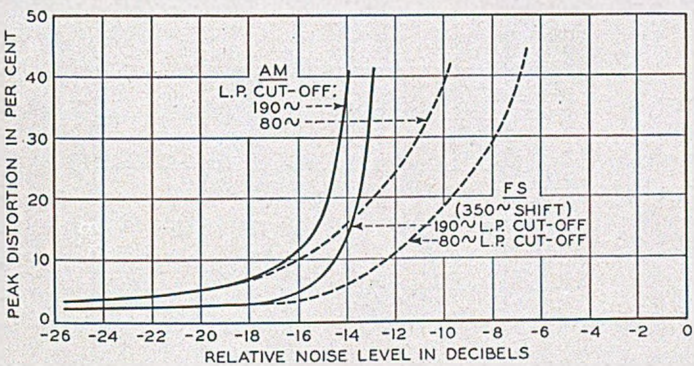


Fig. 22.—Effect of the low-pass filter cutoff frequency on distortion in the presence of impulse noise—740-cycle band.

cause distortion and to make the system more susceptible to noise. A very small shift results in a low amplitude of demodulated signal, which is more readily distorted by noise and biased by frequency drifts. It is of interest, however, that the signal-to-noise ratio of the demodulated signal is not proportional to frequency shift for high noise conditions. As described before, noise peaks tend to reduce momentarily to zero the output from a balanced discriminator. The amplitudes of the dips or holes are thus about one-half the demodulated signal amplitude for any value of shift. This tends to maintain a constant signal-to-noise condition and this characteristic is illustrated by Fig. 16 in which the breaking point with a 100-cycle shift occurs only 2 db before that with a 400-cycle shift although the difference in actual signal amplitude is 12 db. Figure 23 shows the effect on signal-to-noise ratio of progressively varying the frequency shift while the bandwidth

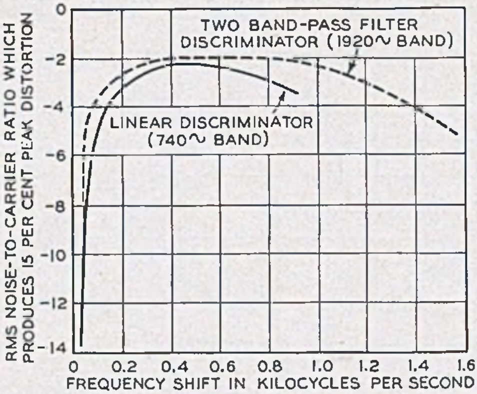


Fig. 23.—Effect of magnitude of frequency shift on distortion produced by thermal noise—80-cycle cutoff low-pass filter.

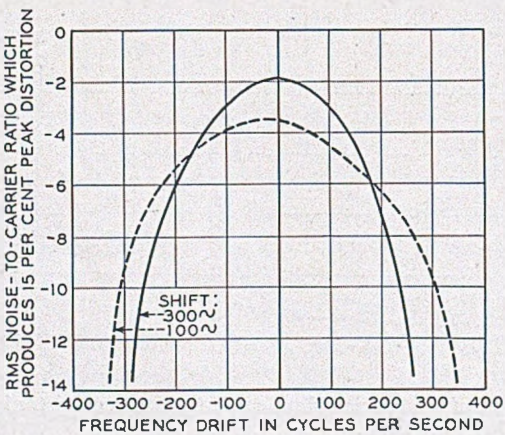


Fig. 24.—Effect of frequency drift on distortion produced by thermal noise in FS transmission—740-cycle band and 80-cycle cutoff low-pass filter.

is kept constant. There is a fairly broad region in which the signal-to-noise ratio is little affected by the amount of frequency shift. For optimum results a frequency shift of 50% to 60% of the band width is indicated, and thus has been used in most of the experimental results given herein.

Frequency Instabilities

Frequency drift of the carrier input to the receiving terminal in general causes biased signals and if severe enough results in failure of the system.

In Fig. 25 is shown the effect of frequency drift on signal bias in an *AM system.* In an AM system little bias is produced until the carrier reaches the cutoff region of the filter. The bias then becomes rapidly negative due to the increased loss and decreased amplitude of demodulated signal. When an automatic gain control arrangement is used the bias becomes positive due to a distorted envelope shape. The demodulated wave form determines the degree of sensitivity to frequency drift and depends on the bandwidth both before and after demodulation.

In an *FS System* using a linear discriminator, frequency drift changes the d-c. component of the signals and thus changes the operating point on the demodulated wave. The amount of bias depends upon the slope of the wave front and is thus affected by the amount of low-pass filtering. The effect of frequency drift on bias for a number of FS systems is shown in Figs. 26 and 27. If a two-bandpass filter type of discriminator is used the system

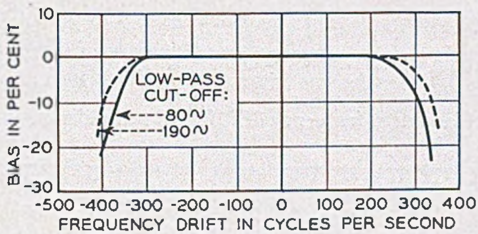


Fig. 25.—Signal bias versus frequency drift for AM transmission in a 740-cycle band.

is insensitive to moderate frequency drifts due to the flat pass bands as illustrated in Fig. 26. The relative shape and amplitude of the signal from a linear discriminator does not change appreciably with frequency drift; only a d-c. displacement occurs. This makes it desirable to have the low-pass filter coupled to the output amplifier by a network which passes only the useful signaling frequencies and blocks the d-c. and very slow drift components. When this is done the effect of frequency drift on bias is not greatly different from that for an AM system, as may be seen by comparing the dotted curves on Figs. 26 and 27 with Fig. 25.

The general method of performing this d-c. elimination is illustrated in the block diagram of Fig. 29. The output from the low-pass filter is passed through a coupling network which blocks the d-c. and passes the useful signaling frequencies. The output of the coupling network is passed through a positive feedback nonlinear amplifier which has but two output conditions representing the mark and space of the telegraph signal. The feedback network passes d-c. and low frequencies so as to just compensate for the loss of the coupling network. The time constant of the coupling network may be made large enough so that the signal wave form into the d-c. amplifier is practically the same as at the output of the low-pass filter. The operating point on the demodulated wave may be readily adjusted by

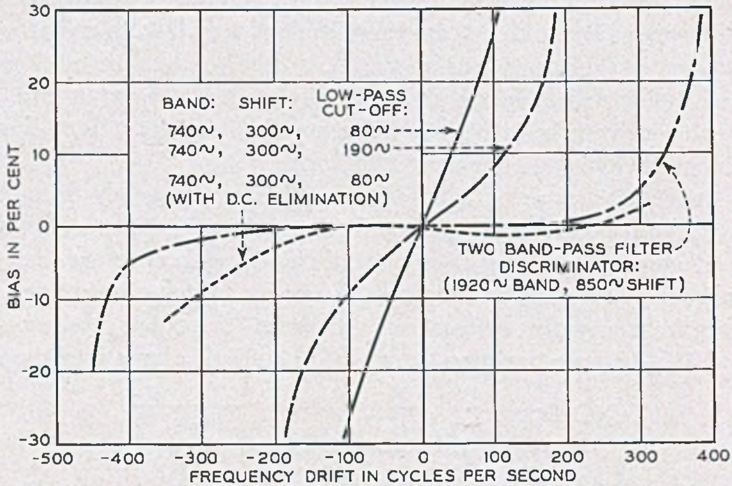


Fig. 26.—Signal bias versus frequency drift for FS transmission—wide filter bands.

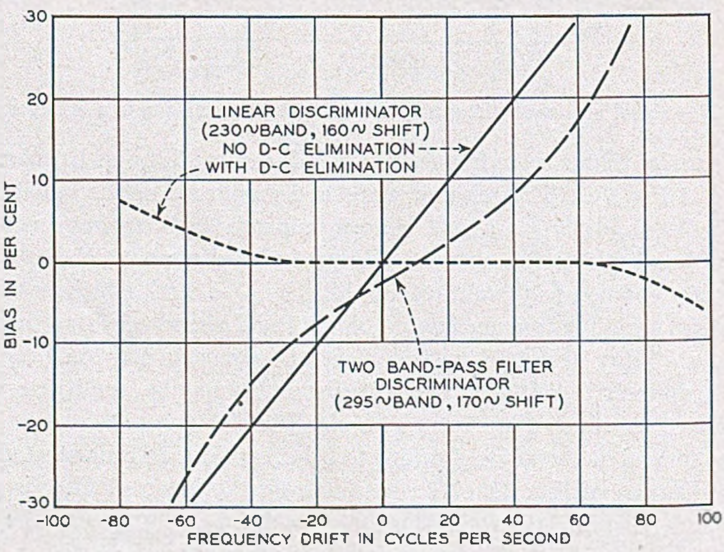


Fig. 27.—Signal bias versus frequency drift for FS transmission—narrow filter bands.

adjusting the bias voltage at the input to the d-c. amplifier. This arrangement differs from the impulse type of d-c. elimination, in which the demodulated wave form is effectively differentiated to form pulses but a fraction of a unit dot in length.

When d-c. elimination is used, operation with FS signals decentered in the pass band causes little bias but it does cause a loss in signal-to-noise ratio, especially at high noise levels. Due to the effect of noise peaks in causing a dip toward zero in the demodulator output, the effect of noise becomes exaggerated during the signal condition which is farther from midband. This change in signal-to-noise condition with frequency drift is shown in

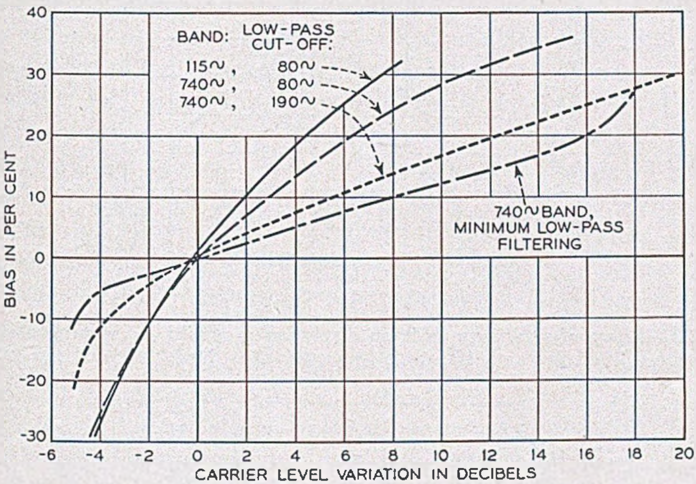


Fig. 28.—Signal bias versus carrier level variation for AM transmission.

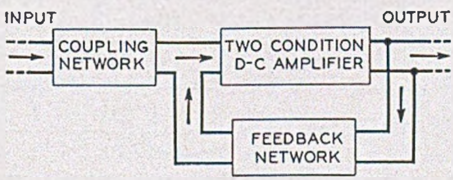


Fig. 29.—Block diagram of dc. elimination and restoration method used to minimize signal bias caused by frequency drift.

Figs. 24 and 30. A comparison between a two-bandpass-filter discriminator and a linear discriminator with d-c. elimination as regards frequency drifts in the presence of noise is shown in Fig. 31.

Radio telegraph systems operating in the H.F. region require a high degree of frequency stability if narrow bandwidths are to be used. Because of the several frequency conversions involved a number of different oscillators or frequency sources are involved but usually the major burden of frequency stability rests on the transmitter exciter and the high-frequency beating oscillator of the receiver.

Various methods of automatic frequency control may be used to hold the carrier input to the demodulator at the correct frequency. In the case of FS signals the control may be arranged to operate only on the marking frequency or to utilize both the mark and space conditions. It is preferable to have inherent frequency stability rather than to compensate for the drift at the receiving end since it is difficult if not impossible to provide an automatic frequency control which will not reduce the transmission capabilities

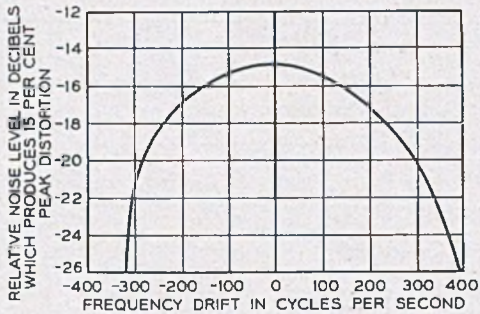


Fig. 30.—Effect of frequency drift on distortion produced by impulse noise in FS transmission—740-cycle band, 100 cycle frequency shift, 80-cycle cutoff low-pass filter.

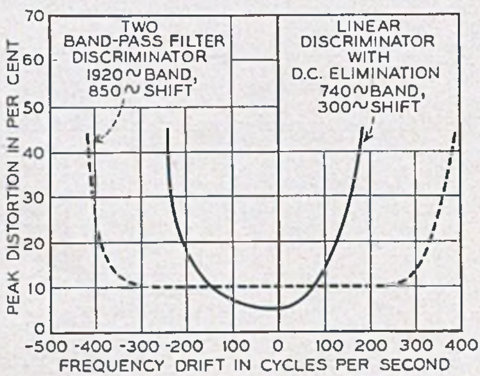


Fig. 31.—Linear discriminator versus two-bandpass-filter discriminator when frequency drift occurs in the presence of thermal noise—-4 db rms noise-to-carrier ratio.

of the system in the presence of noise and other interference. Best results are obtained if the frequency stability is high enough to require but a very slow correction, which usually dictates some mechanical rather than electronic tuning arrangement. Manual retuning may be found satisfactory where stability is reasonably good provided care is taken in making the adjustments. Suitable frequency stability with the retention of flexibility in frequency adjustment may be obtained by frequency sources making use of a combination of crystal oscillators and high-stability variable oscillators of lower frequency.

Level Variations

Extreme and rapid variations of received level exist in H.F. radio transmission. It is upon the ability to accommodate these level variations that the merits of an H.F. radio telegraph system must largely be judged. FS telegraph shows its outstanding advantage in this respect.

In an *AM system* in order to obtain zero-bias signals and optimum signal-to-noise conditions the operating point must be near the half amplitude point on the demodulated wave. This means that complete failure will result for a drop in level of 6 db unless some compensating arrangement is provided. The slope of the bias-versus-level characteristic depends upon the slope of the demodulated wave which in turn depends on the bandwidth of the system and upon the degree of low-pass filtering. The bias-versus-level characteristics of some AM systems are shown in Fig. 28. Where the level variations are relatively slow compared to the signaling speed, automatic gain control circuits can be used to maintain a nearly constant level into the demodulator. However, where large rapid level changes occur, as in the H.F. range, it is seen that a narrow band AM system would fail completely regardless of the amount of transmitted power. For printer operation over an AM system in the H.F. range a fairly wide band and little low-pass filtering should be used, so as to keep the wave shape of the signals as square as possible and thus obtain a fairly flat bias-versus-level characteristic. By adjusting the operating point low on the demodulated wave, approaching the spacing noise level, the greatest possible range of acceptable rapid level change will be obtained. The slower level change components may be handled by the usual automatic gain-control circuits. This will cause the bias of the signals to average somewhat marking but the peak distortions will be kept to a minimum.

In an *FS system* no bias is produced so long as both the marking and spacing frequencies are affected alike, with their received levels remaining equal. Such non-selective fading conditions cause no distortion even when they occur at quite rapid rates. If a balanced type of discriminator is used, amplitude limiting is not essential to obtaining this immunity from non-selective variations in attenuation. It is only when the mark and space levels are different that bias results. In Fig. 32 are shown bias versus mark-to-space level ratio characteristics both with and without a limiter. More bias exists when there is no limiter because the amplitude of the demodulated wave is directly affected and consequently the low-pass filtering also becomes a factor. With a limiter the amplitude of the demodulated wave is held constant and the amount of low-pass filtering has no effect on bias. Some bias is still produced, however, due to the differently shaped frequency transients in the passband of the receiving system when a level change occurs

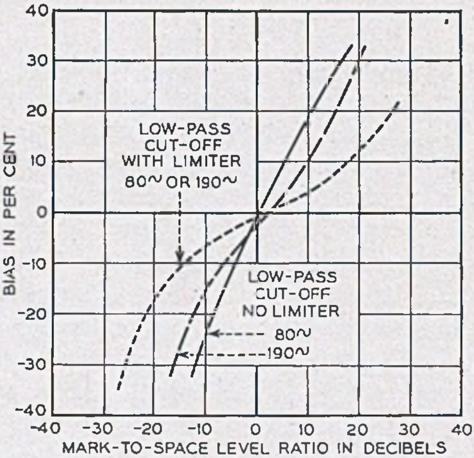


Fig. 32.—Signal bias versus mark-to-space level ratio in FS transmission—740-cycle band, 350-cycle frequency shift.

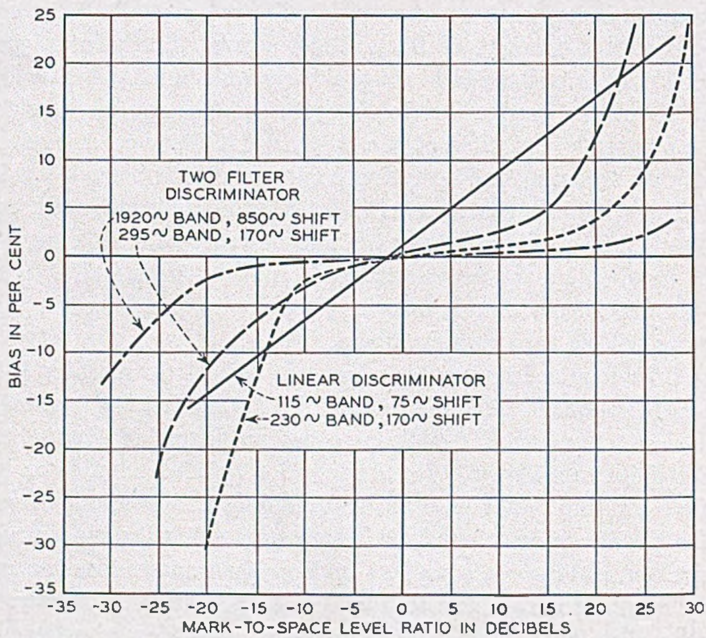


Fig. 33.—Signal bias versus mark-to-space level ratio in FS transmission—80-cycle cutoff low-pass filter.

at the moment when the frequency changes. In Fig. 33 this bias effect is demonstrated for various bandwidths. For moderate mark-to-space level ratios the bias effect is small and linear, with a slope which usually varies inversely with a change in bandwidth. At extreme level differentials the bias may rise very rapidly due to the amplitude and phase characteristics of the input passband; transients from the greater amplitude condition may severely interfere with the lower amplitude condition. For the types of bandpass filters used in the tests it appeared that the amount of level difference required to produce 20% bias did not change greatly with bandwidth. Severe wave shaping of the signal at the transmitter was found to be an aid in reducing the bias effect due to such transients but the characteristic distortion became too great to give any practical improvement.

The fading modulator used in obtaining the data for Fig. 33 caused no change in phase. Selective fading over an actual radio circuit would involve considerable phase shift and greater distortion might be expected. The data of Fig. 32 were obtained by use of the phase control associated with the crystal filter of the radio receiver to vary the loss-versus-frequency characteristics of the receiving pass band and thus cause unequal mark and space amplitudes. This method gave an amplitude and phase characteristic for the transmission band more like that over an actual radio circuit.

Multipath Propagation Effects

The rapid fading conditions prevailing in the H.F. range are brought about by multipath propagation. Under such conditions, the signal induced in a receiving antenna by a distant transmitter may be the resultant of two or three separate waves each propagated over a different path. If two waves arrive over paths differing in length by an odd number of half wavelengths the resulting 180° phase difference causes maximum cancellation. On the other hand if the paths differ in length by an integral multiple of whole wavelengths the waves arrive in-phase and maximum reinforcement results. The difference in path lengths may at times be as great as 500 to 1500 kilometers (delay times of 2 to 5 milliseconds) which in the H.F. region corresponds to thousands of wavelengths. Under these maximum conditions waves at one frequency may arrive in phase while waves at a frequency a few hundred cycles away may arrive in phase opposition. Since the path lengths are constantly changing, the transmission at a given frequency is subject to wide variations in amplitude and phase with time. When the difference in path lengths is not great enough to cause frequencies in one portion of a communication channel to fade differently from those in another portion the term "non-selective" or "flat" fading is applied. When the difference in path lengths becomes great enough to cause considerable amplitude or phase distortion over the transmission band the term "selective" fading is used. Since the propagation paths existing at a given moment vary for different antenna sites, the fading patterns obtained from two or three antennas separated by several wavelengths usually show a considerable phase difference so that a given frequency is not likely to fade into the noise level at all antennas simultaneously. By employing separate receivers for each antenna and suitably combining or selecting the demodulated outputs, a system is obtained which is much less susceptible to fading. Such a method is called *space diversity* reception. Inasmuch as fading over a given combination of paths is highly selective with respect to frequency much the same effect is obtained by *frequency diversity* reception. When this method is employed the intelligence is transmitted on two or more frequencies simultaneously and then received by separate receivers from a single antenna and the resulting demodulated signals combined or selected as for space diversity.

In telegraph transmission large differences in delay over two separate propagation paths cause the telegraph signal transitions to arrive at different instants over the two paths. Thus, there are intervals of overlap when a marking condition is received over one path and a spacing condition over a second path. When two components of nearly equal amplitude arrive at nearly 180° phase difference a signal transition may involve large and sudden amplitude and phase changes. The resulting transients in the bandpass networks of the receiving equipment may cause fortuitous distortions considerably greater than the difference in delay times over the two paths. The wider the pass band of the receiving system the shorter the duration of these fortuitous transients and hence the less the distortion. This phenomenon is one of the determining factors in the selection of bandwidth and frequency shift to be used in a given application of FS telegraphy. It becomes of increasing importance when the circuits are long and at higher signaling speeds such as are used in time-division multiplex methods.

In an AM system the effect of large differences in path lengths is usually a filling in of the spacing intervals with resulting marking bias. In an FS system the overlap time and associated transients may add to either marking or spacing intervals in a random fashion depending on the amplitude and phase conditions at each transition. The overlapping of the mark and space frequencies in FS transmission can sometimes be heard in an AM receiver as short pips of audio tone at each transition, the audio tone being the beat between the two frequencies.

Use of Superimposed Phase Modulation

Superimposed phase modulation has sometimes been employed as a simple means for achieving a certain amount of frequency diversity both in AM and FS telegraph systems. This consists in causing the radiated signal to oscillate continuously through a small phase angle at a rate relatively high compared to the dotting speed. Phase modulation spreads the energy of the signal over a wider frequency band so that the complete loss of the signal through selective fading becomes less probable. The spectra generated by sinusoidal phase modulation of 1.0, 1.4, and 2.0 radians are shown in Fig. 34. Most of the energy is seen to be concentrated in the carrier and first order sidebands. Less than 1.0 radian of modulation results in too little amplitude of the sidebands, while more than 1.5 radians results in too wide spread of energy outside the first order sidebands. The center three

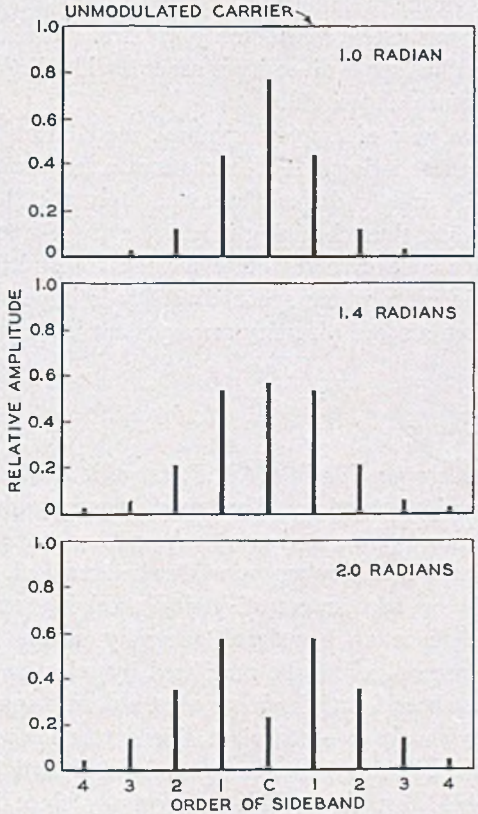


Fig. 34.—Frequency spectra for sinusoidal phase modulation.

components are equal at about 1.4 radians. In the case of FS the phase modulation frequency appears as a variation in amplitude of the signals from the discriminator. For AM no additional amplitude variation is caused by the phase modulation if there is no selective attenuation in the medium, but if such exists the phase modulation frequency or a multiple thereof appears in the rectified signal. To permit these unwanted amplitude variations to be removed by the low-pass filter so as not to break up the signals, the phase modulating frequency should preferably be ten or more times the maximum signaling frequency.

A test of superimposed phase modulation on FS signals was made over a radio circuit approximately 200 miles in length. A frequency shift of 850 cycles, and one radian of 200-cycle phase modulation, was used. A 60-word-per minute test sentence was transmitted and received without space diversity. It was found over a period of several hours that the phase modulation, on the average, gave a decrease in printed errors of about 50% when the error rate was in the proximity of 1 to 2%. For short intervals the reduction in errors was often considerably greater. The use of 200-cycle phase modulation when space diversity is used provides little or no improvement and is therefore undesirable.

A more effective way of employing phase modulation with FS signals would be to use a phase swing of ± 1.4 radians at a frequency of 2 to 3 times the frequency shift and to demodulate separately the three major components of the signal, thus obtaining in effect a triple-frequency diversity system. This of course involves quite a wide transmitted band, but it might be of use in cases where space diversity is impossible, such as on board ships. When a space diversity arrangement is feasible it is much to be preferred.

Diversity Operation

To obtain reliable operation in the H.F. range it is common practice to employ space diversity reception. The use of frequency diversity, with the increase of transmitted power and greater frequency space required, is seldom justified if space diversity reception can be arranged. For AM radio telegraph, double or triple-space diversity receiving arrangements are frequently used. Since an FS signal generally covers more frequency space, it is even more likely to be mutilated by selective fading than an AM signal. It has been found, however, that a double-space diversity system for FS signals usually gives sufficient diversity action provided it is of a type that permits switching between channels at signaling speed without causing appreciable distortion. This is necessary since it is a frequent occurrence that the mark of one channel may fade, leaving a good space, while the opposite may occur on the second channel. Since an FS system can accept rapid level changes, the main purpose of diversity methods is to insure that both the mark and space portions of the signal will be received above the noise level. In the case of AM telegraph, since it cannot accept rapid level changes, diversity operation is important not only in keeping the signal above the noise but also in averaging out some of the rapid level changes. For this reason AM systems usually show considerable improvement in going from double to triple diversity. It would be expected that a like change would show much less improvement in an FS system.

Diversity Channel Selection

The method employed to combine or select the channels of a diversity system is of great importance. For an AM system the relatively simple method of using a common load circuit for the diode detectors of the diversity channels is generally used. By deriving a common AVC voltage from the combined output and by properly adjusting the receiver sensitivities a fairly constant output is obtained. The parallel connection of the diode detectors causes the stronger signal to effectively block the weaker signal thus giving a fairly sharp diversity selection characteristic. The problem of combining the diversity channels of an FS system is more complicated mainly because of the amplitude limiting. If amplitude limiting is used in each diversity channel before demodulation, the resulting constant amplitude signals convey no information as to their relative amplitudes as received from the antennas. Any diversity selection must then be obtained by some indirect method. It is necessary to furnish some selecting device since the noise from a faded channel, if added directly to a good signal from another channel, will cause high distortion.

In an early frequency shift system employing a two-bandpass filter discriminator (shown previously in Fig. 17) it was found that for a poor signal-to-noise condition the sum of the outputs of the mark and space rectifiers increased above that for a good signal-to-noise condition. This increase was utilized to suppress the output of the poorer channel and emphasize that of the better channel. Although neither the degree of diversity selection nor the speed of response was as good as might be desired, fairly satisfactory results were obtained.

Another method which has been used involves the derivation of control currents or voltages proportional to the amplitudes of the incoming signals which in turn select the better diversity channel by some type of gate action. The time constants of the control circuits must be low enough to permit switching at signaling speed without introducing considerable distortion. The gate circuits must also be of a type which does not introduce interfering transients or otherwise allow the control voltages or currents to interfere with the signal. This method permits very sharp diversity selection and has the capability of approximating ideal results although it becomes somewhat involved in a practical form.

A considerably simpler method has been used in some of the more recent FS terminals. It is based on the use of a single-amplitude-limiter through which pass the signals of both diversity channels. This is made possible by arranging the two signals at the input to the limiter to be at different frequencies. At the output of the limiter the two signals are separately demodulated and then combined. When one of the signals is considerably greater in amplitude than the other at the input to the limiter the relative difference in level is increased by an additional amount of about 6 db at the limiter output. The limiter output may be considered as the stronger signal frequency-modulated by the weaker signal. For small modulation indices the amplitude of the first order sideband is approximately one-half the modulation index thus explaining the 6 db added difference in level at the limiter output. As the input levels approach equality the added level difference decreases to zero. A block diagram indicating the arrangement of such a diversity system is shown in Fig. 35. Tests were made of both parallel and series connections of the two discriminator outputs. With a parallel connection the discriminator having the greater output blocks the rectifier output of the other discriminator and thus gives a sharp diversity selection characteristic. However, the level ratio of the channels at which a switch

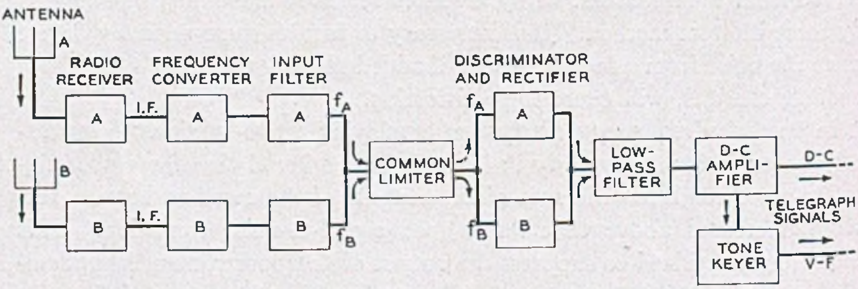


Fig. 35.—Block diagram of dual-diversity FS receiving system using a common limiter.

takes place is affected not only by the ratio at the limiter input but also by frequency drift and discriminator slope. With the series connection only the input level ratio at the limiter input affects the diversity switching; this arrangement was therefore selected as the preferred method although its selection characteristic is not sharp. The series and parallel combining characteristics are shown in Figs. 36 and 37.

Various tests were made on a terminal having a 1500-cycle bandwidth and using the series combining method to determine the signal-to-noise characteristics under different conditions of diversity fading. A frequency shift of 850 cycles was used and the midband frequencies of the two diversity channels at the common limiter input were 30 and 35 Kc. Figure 38 shows the distortion versus signal-to-noise ratio characteristics of each channel separately and in diversity combination for various relative level conditions of the two channels. During diversity operation equal noise levels were maintained in the two channels and various combinations of level differences of the two channels were preserved as the whole signal level combination was varied. The level differentials are indicated in the figure for each curve. The signal-to-noise level scales refer to the highest level portion of the diversity signal. Since the amplitude modulator which was used to simulate the selective fading did not produce phase shifts or

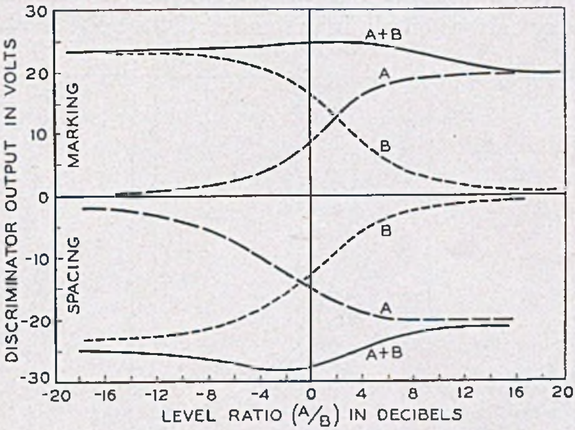


Fig. 36.—Diversity combination characteristic obtained by series addition of discriminator outputs—levels measured at output of 400 kc I.F. amplifier.

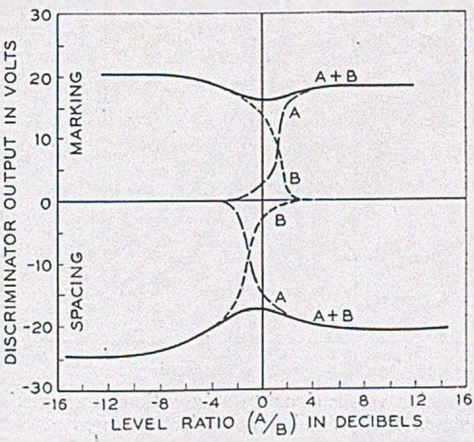


Fig. 37.—Diversity combination characteristic obtained by parallel addition of discriminator outputs—levels measured at output of 400 kc I.F. amplifier.

transmission delays as would actually exist over H.F. radio circuits the actual distortions shown by the curves are optimistic.

The ideal diversity selection circuit should theoretically give a signal-to-noise characteristic identical to that of a single channel under the signal-to-noise condition corresponding to the signal of the best momentary reception. It will be seen that the test results of Fig. 38 approach this limit within 2 or 3 db at a peak distortion of 20%. Part of this difference is due to the dissimilar bandpass characteristics of the two channels of the experimental unit used for the tests and part due to the lack of an extremely sharp diversity selection. It should be pointed out that the conditions under which the theoretical maximum diversity signal-to-noise condition may be reached are very hard to obtain in practice. If the noise levels in the two

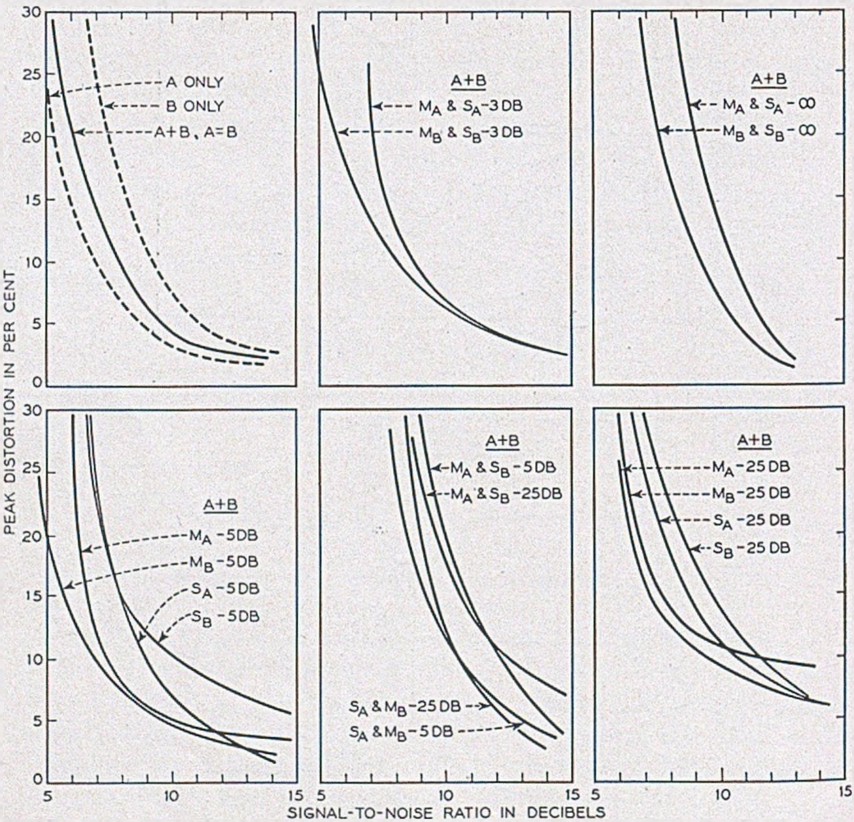


Fig. 38.—Peak distortion versus signal-to-noise ratio characteristics for dual-diversity operation with a common limiter. Signal-to-noise measured at output of 400 kc I.F. amplifier.

channels are not equal, and if the diversity selecting method does not at all times exclusively select the channel with the greater signal, the distortion characteristic will deteriorate accordingly. Because the AVC sensitivities of two radio receivers may differ considerably the noise levels cannot be maintained closely the same and usually no provision is made for determining the noise level except by ear. The slightly better diversity action which can theoretically be obtained is therefore felt to be of doubtful usefulness under actual operating conditions. The common limiter method has the advantage of being simple in that the diversity action is obtained in the transmission circuits directly and no added switching circuits or adjustments are required. Tests of this type of diversity selection in the field have indicated a marked superiority over earlier FS terminal equipment.

In connection with the use of a common limiter care must be taken in the selection of the two-channel frequencies. Frequencies having nearly integral ratios such as 3:5 and 5:7 produce disturbing amplitude modulation of the demodulated signal. The frequencies should be chosen so as to avoid low integral ratios; then all amplitude modulations are negligible or easily filtered out. Where frequency drift is to be allowed for, the frequencies should be chosen so as not to approach a low integral ratio at any place in the expected drift range.

If the radio receivers associated with a space diversity FS system have automatic gain control it must be a common control so the receivers will change gain equally. The use of common AVC prevents overloading of the receivers as the received signal strength varies. If no common AVC is available the receivers should be operated in the manual gain-control condition.

Conclusions

General Comparison of FS and AM Carrier Telegraphy

The foregoing sections have compared the characteristics of AM and FS carrier telegraph transmission under various conditions. Whether or not FS would prove to be the preferable method for a specific communication use depends largely on the transmission medium and the quality of transmission desired. As regards frequency space requirements, both methods provide essentially the same signaling speed capability for a given bandwidth.

As to the ability to transmit through noise, FS has an advantage of 3 to 4 db at distortions approaching the failure point when equal bandwidths are compared. At lower distortions the advantage of FS is 6 db or more so that it is attractive in this respect for tandem operation of several telegraph sections where regeneration of signals is not practiced. When frequency space permits wider bands, with correspondingly increased frequency shifts, the signal-to-noise advantage of FS over AM increases for low noise levels. Wide band FS therefore provides a means of obtaining higher quality circuits if the noise level is not too great.

The AM method is basically less susceptible to frequency variations than is the FS method. However, as has been illustrated, frequency drift can be compensated for by d-c. elimination so as to make FS comparable to AM in this respect.

FS transmission is essentially immune from effects of non-selective level variation, even when extremely rapid, and in this characteristic displays its most outstanding advantage over AM.

Operation Over Wire Circuits

A wire circuit usually provides a transmission medium having a low noise level with slow and relatively small variations in attenuation. Such circuits, when equipped with suitable automatic gain control, allow stable operation with AM telegraphy and but little improvement could probably be obtained by using FS. The choice between AM and FS under such relatively ideal conditions becomes one of economic considerations of the terminal equipment and carrier supply. However, when FS is applied to multichannel systems the problem of interchannel interference requires attention. For wire circuits having high noise levels or sudden changes in attenuation the use of FS instead of AM provides considerable improvement and in severe cases the FS method may be a necessity for satisfactory operation. Wide band FS operation with its sharper breaking distortion-versus-noise-level characteristic gives a low value of rms-to-peak distortion which would be especially advantageous for tandem operation. However, the necessary frequency space for wide-band operation is not usually economically justified for wire line operation.

Operation Over Radio Circuits

For operation over radio circuits providing stable conditions similar to those on wire circuits the FS method does not show a great advantage over the AM method. In the case of long distance telegraphy in the H.F. range, however, FS shows a marked advantage over AM. This is because of the rapid fading and high noise conditions which commonly prevail in the H.F. region. The amount of rapid variation in marking level that an AM system can accommodate is less than the difference between marking and spacing levels that an FS system can tolerate. In the worst case of selective fading the level differences between the mark and space frequencies might approach values equal to the short time level swings of a single frequency, but in general would be less. A given condition of selective fading thus causes less distortion in an FS system than in an AM system. FS allows the use of narrow bands without much loss in signal quality in the presence of fading, whereas AM does not. FS therefore is essential for satisfactory operation of closely spaced narrow band H.F. radio channels. Where frequency space is not restricted and wider bands are used to permit considerable frequency drift, the improvement afforded by FS over AM is materially less. To obtain optimum results from an AM system, however, requires more careful adjustment and more attention than does an FS system. This is partly due to the amplitude limiter in the FS system which results in a constant amplitude of signal from the discriminator and partly due to the fact that an FS signal is no more subject to noise interference during the spacing condition than during the marking condition. Therefore the operating point on the demodulated wave may be set and left for long intervals even though transmission conditions vary widely. This greater ease in maintaining good adjustment of the equipment probably accounts for some of the apparent improvement in changing from an AM to an FS system.

It should be noted that a system may fail either because of level variations well above the noise level or because of the signal becoming submerged in noise. If a system fails because it can accept only moderate level variations, an increase in transmitted power will provide no improvement since the level variations will remain the same as before. On the other hand, a system which can accept very wide variations in level will show improvement upon increasing transmitted power up to the point where no failures occur due to an unfavorable signal-to-noise ratio.

The over-all improvement obtained in changing from AM to FS radio telegraph is sometimes expressed as a ratio of transmitted powers required to give equivalent transmission results over the two systems. Such a ratio fluctuates widely depending upon the prevailing conditions. With little fading the improvement ratio will be mainly due to the better signal-to-noise obtained with FS and may be less than 5 db. Under severe fading conditions no amount of power may give good results with AM while FS may be satisfactory. Thus the power ratio would become infinite. By making a long-time comparison an average power ratio figure may be found which gives equal average error rates in the printed copy from each system. Such tests[7] between a triple space diversity AM system and a double space diversity FS system have indicated a power ratio of 11 db in favor of the latter when the error rate was 0.1 to 0.5 per cent.

When the two systems are thus made equal by adjustment of transmitted power, more errors due to the signal becoming submerged in noise occur in the FS system to compensate for a larger number of errors in the AM system due to rapid level changes. Often the reason for changing a radio telegraph system from AM to FS is to increase the reliability of the circuit and not just to save transmitted power. To insure a definite improvement in such cases the carrier level should not be decreased more than about 6 db.

References

1. Certain Topics in Telegraph Transmission Theory, H. Nyquist, *Trans. of A.I.E.E.* Vol. 47, April 1928, pp. 617–644.
2. Frequency Modulation, Balth. van der Pol, *Proc. of I.R.E.*, Vol. 18, July 1930, p. 1202.

3. Measurement of Telegraph Transmission, H. Nyquist, R. B. Shanck, and S. I. Cory, *Trans. of A.I.E.E.*, Vol. 46, Feb. 1927, pp. 367–376.
4. Frequency Modulation Noise Characteristics, M. G. Crosby, *Proc. of I.R.E.*, Vol. 25, April 1937, p. 472–514.
5. Performance Characteristics of Various Carrier Telegraph Methods, T. A. Jones, K. W. Pfleger, *Bell Sys. Tech. Jour.*, Vol. 25, pp. 483–531, July 1946.
6. Transients in Frequency Modulation, H. Salinger, *Proc. of I.R.E.*, August 1942, pp. 378–383.
7. Observations and Comparisons on Radio Telegraph Signaling by Frequency Shift and On-Off Keying, H. O. Peterson, J. B. Attwood, H. E. Goldstine, G. E. Hansell, R. E. Schock, *RCA Review*, March 1946.

Reflections from Circular Bends in Rectangular Wave Guides—Matrix Theory

By S. O. RICE

A method of computing reflections produced by circular bends in rectangular wave guides is presented. The procedure employs the theory of matrices. Although the matrix equations are quite simple, a considerable amount of calculation is necessary before quantitative results may be obtained. Fortunately, the approximate formulas pertaining to gentle bends hold surprisingly well for rather sharp bends. These formulas are obtained by a limiting process from the matrix equations. The approximate formula for reflection from an H-bend (in which the magnetic vector lies in the plane of the bend) generalizes an earlier result due to R. E. Marshak. The corresponding formula for the E-bend appears to be new.

Introduction

A NUMBER of investigators have studied the propagation of electromagnetic waves in a bent pipe of rectangular cross-section, the bend being along an arc of a circle. H. Buchholz[1], S. Morimoto[2], and W. J. Albersheim[3] have employed Bessel functions to express the field in the bend. The form assumed by the field when the radius of curvature of the bend becomes large has been obtained by K. Riess[4] and R. E. Marshak[5] who use approximations suited to this case. Marshak also obtains expressions for various reflection and transmission coefficients. A discussion of the subject using rather simple but approximate analysis is given on pages 324–330 of a text book[6] by S. A. Schelkunoff. The Bessel function approach is also sketched in the same section.

Here we study the disturbance produced when a wave goes around a circular bend (of some given angle) in a rectangular wave guide, the guide being straight on either side of the bend. Especial attention is paid to the dominant mode reflection coefficients g_{10}^{-} and d_{01}^{-} corresponding to H-bends and E-bends, respectively. As equations (4.2–6) and (4.4–4) show, these reflection coefficients (which are of the nature of voltage rather than power reflection coefficients) vary inversely as the square of the radius of curvature of the bend when the bend is gentle. The substance of (4.2–6) has been given by Marshak[5] for the important case in which only the dominant mode is propagated and the angle of the bend not too small.

When the bend is so sharp that the formulas mentioned above do not apply the reflection coefficients may be computed from the rather simple looking matrix expressions (2.3–3) together with (2.3–4). However, their appearance is deceptive and, as is shown by the numerical work in Part V, considerable labor is necessary to obtain an answer.

The gentle bend formulas were obtained from the matrix equations by the limiting process described in Part III. It seems likely that the matrix method, which is similar to the method used in an earlier paper[7] on transmission line equations, may be applied to other wave guide problems. With this thought in mind, the development of Parts II and III has been couched in general terms.

The matrices used in the present theory are of infinite order since the guide may support an infinite number of modes of propagation. This fact makes it difficult to justify all the steps in our analysis, and we do not attempt to do so.* Despite this lack of rigor, I believe that the procedures given here lead to the correct results since they yield, for gentle bends, expressions obtained by Buchholz and Marshak. Moreover, although numerical results tabulated in Part V were obtained by using matrices of only the second and third order, they indicate a rapid convergence as the matrix order is increased.

PART I

PROPAGATION OF WAVES IN GUIDE

1.1 *Propagation in a Straight Wave Guide*

Rather general expressions for the electric and magnetic intensities E and H in a field are (see pp. 127–128 of Reference[6])

$$\begin{aligned} E &= -i\omega\mu\vec{A} + \frac{1}{i\omega\epsilon}\text{ grad div }\vec{A} - \text{curl }\vec{B} \\ H &= \text{curl }\vec{A} + \frac{1}{i\omega\mu}\text{ grad div }\vec{B} - i\omega\epsilon\vec{B} \end{aligned} \tag{1.1-1}$$

The field is assumed to vary with the time t as $e^{i\omega t}$, ω is the radian frequency, μ the permeability and ϵ the dielectric constant (for free space $\mu = 1.257 \times 10^{-6}$ henries/meter, $\epsilon = 8.854 \times 10^{-12}$ farads/meter). The vector potentials $\vec{A}$ and $\vec{B}$ satisfy the wave equations

$$\begin{aligned} &\nabla^2\vec{A} = \sigma^2\vec{A}, \qquad \nabla^2\vec{B} = \sigma^2\vec{B} \\ &\nabla^2 \equiv \text{Laplacian operator} \\ &\sigma^2 = \omega^2\mu\epsilon \end{aligned} \tag{1.1-2}$$

In dealing with bends, it is convenient to choose $\vec{A}$ and $\vec{B}$ normal to the plane of the bend. In our notation, this plane is always taken to be the x, z plane so that $\vec{A}$ and $\vec{B}$ are parallel to the y axis. The z axis is parallel

* Similar questions arise in the rigorous treatment of an infinite set of linear equations. A discussion of this subject is given in Chap. III of Reference[8].

to the guide axis and, for the straight guide of the section, the guide walls are sections of the planes $x = 0, x = a, y = 0, y = b$.

Thus, a general wave traveling in the positive z direction may be described by the two functions (which represent the magnitudes of $\vec{A}$ and $\vec{B}$)

$$A = \sum_{m,n} g_{mn}^{+} e^{-\Gamma_{mn}z} \sin(\pi m x/a) \cos(\pi n y/b) \qquad m = 1, 2, 3, \cdots; \quad n = 0, 1, 2, \cdots \tag{1.1-3}$$

$$B = \sum_{m,n} d_{mn}^{+} e^{-\Gamma_{mn}z} \cos(\pi m x/a) \sin(\pi n y/b) \qquad m = 0, 1, 2, \cdots; \quad n = 1, 2, 3, \cdots \tag{1.1-4}$$

where the coefficients g_{mn}^{+} and d_{mn}^{+} are constants and the plus signs indicate propagation in the positive z direction.

The propagation constant Γ_{mn} is obtained from

$$\Gamma_{mn}^{2} = \sigma^2 + (\pi m/a)^2 + (\pi n/b)^2, \quad \sigma = i2\pi/\lambda_0, \qquad \lambda_0 = \text{wavelength in free space.} \tag{1.1-5}$$

Equation (1.1–5) arises when the typical term in (1.1–3) is substituted for A in the equation

$$\frac{\partial^2 A}{\partial x^2} + \frac{\partial^2 A}{\partial y^2} + \frac{\partial^2 A}{\partial z^2} = \sigma^2 A \tag{1.1-6}$$

This and a similar equation for B are the forms assumed by (1.1–2) for the rectangular coordinates of our straight guide.

The electric and magnetic intensities in the guide are given by

$$\begin{aligned} E_x &= \frac{1}{i\omega\epsilon}\frac{\partial^2 A}{\partial x \partial y} + \frac{\partial B}{\partial z} & \qquad H_x &= -\frac{\partial A}{\partial z} + \frac{1}{i\omega\mu}\frac{\partial^2 B}{\partial x \partial y} \\ E_y &= -i\omega\mu A + \frac{1}{i\omega\epsilon}\frac{\partial^2 A}{\partial y^2} & \qquad H_y &= -i\omega\epsilon B + \frac{1}{i\omega\mu}\frac{\partial^2 B}{\partial y^2} \\ E_z &= \frac{1}{i\omega\epsilon}\frac{\partial^2 A}{\partial z \partial y} - \frac{\partial B}{\partial x} & \qquad H_z &= \frac{\partial A}{\partial x} + \frac{1}{i\omega\mu}\frac{\partial^2 B}{\partial z \partial y} \end{aligned} \tag{1.1-7}$$

which follow from (1.1–1).

It is seen that the wave is completely specified by the g_{mn}^{+}'s and d_{mn}^{+}'s. These may be arranged as (infinite) column matrices in any convenient order. Thus in dealing with (1.1–3) and (1.1–4) we may write

$$g^{+} = \begin{bmatrix} g_{10}^{+} \\ g_{20}^{+} \\ g_{11}^{+} \\ g_{30}^{+} \\ g_{21}^{+} \\ g_{12}^{+} \end{bmatrix} \qquad d^{+} = \begin{bmatrix} d_{01}^{+} \\ d_{02}^{+} \\ d_{11}^{+} \\ d_{03}^{+} \\ \cdot \\ \cdot \end{bmatrix} \tag{1.1-8}$$

In our work we shall consider only those modes corresponding to a fixed value of m (or of n) and the order is almost automatically fixed.

The factors which determine the propagation of the typical terms in the summations (1.1–3) and (1.1–4) for A and B are

$$\alpha_{mn}(z) = g_{mn}^{+} e^{-z\Gamma_{mn}}, \qquad \beta_{mn}(z) = d_{mn}^{+} e^{-z\Gamma_{mn}} \tag{1.1–9}$$

The column matrices obtained by arranging these quantities in the same order as in (1.1–8) will be denoted by $\alpha(z)$ and $\beta(z)$. We may write

$$\alpha(z) = e^{-z\Gamma_\alpha} g^+, \qquad \beta(z) = e^{-z\Gamma_\beta} d^+ \tag{1.1–10}$$

where $\exp(-z\Gamma_\alpha)$ and $\exp(-z\Gamma_\beta)$ are square matrices defined by power series each term of which is a square matrix:

$$e^{-z\Gamma} = I - \frac{z\Gamma}{1!} + \frac{z^2\Gamma^2}{2!} - \frac{z^3\Gamma^3}{3!} + \cdots \tag{1.1–11}$$

I is the unit matrix and Γ_α is the diagonal matrix*

$$\Gamma_\alpha = \begin{bmatrix} \Gamma_{10} & 0 & 0 & \cdot \\ 0 & \Gamma_{20} & 0 & \cdot \\ 0 & 0 & \Gamma_{11} & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix} \tag{1.1–12}$$

in which the order of the diagonal elements is the same as the order of the elements in the column matrix g^+. Similarly Γ_β is a diagonal matrix whose elements are $\Gamma_{01}, \Gamma_{02}, \Gamma_{11}, \Gamma_{03}, \cdots$, the order being fixed by d^+. When Γ is replaced by Γ_α in (1.1–11) it is easy to obtain Γ_α^2, Γ_α^3, etc. and sum the resulting series to obtain

$$e^{-z\Gamma_\alpha} = \begin{bmatrix} e^{-z\Gamma_{10}} & 0 & 0 & \cdot \\ 0 & e^{-z\Gamma_{20}} & 0 & \cdot \\ 0 & 0 & e^{-z\Gamma_{11}} & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix} \tag{1.1–13}$$

A similar expression exists for $\exp(-z\Gamma_\beta)$. The expression (1.1–10) for $\alpha(z)$ is seen to be true when the square matrix (1.1–13) is multiplied, by matrix multiplication, into the column g^+.

It turns out that the field in a circular bend (in a rectangular guide) may be represented by a generalization of the foregoing expressions. In this generalization, which will be studied in the following sections, the square matrices Γ_α and Γ_β no longer have the simple form of diagonal matrices.

* That is, a square matrix in which all of the elements other than those in the principal diagonal are zero.

1.2 *Propagation in a Circular Bend*

In dealing with a circular bend we choose cylindrical coordinates (ρ, φ, y) as shown in Fig. 1. With these coordinates we associate new coordinates, shown in Figs. 1 and 2, (x, y, z) which have approximately the same significance as in the straight guide. z is the distance measured along the axis of

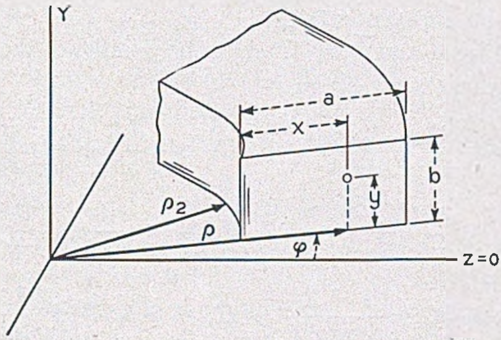


Fig. 1

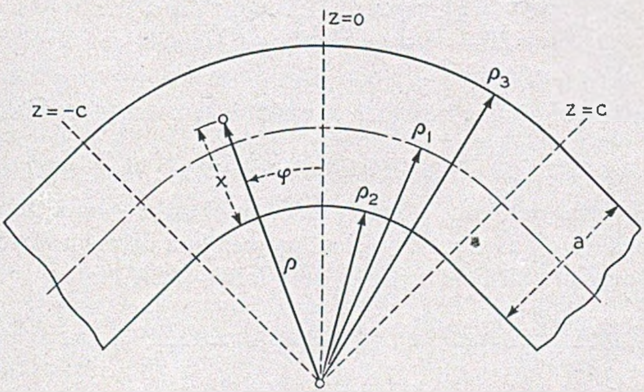


Fig. 2

the guide (defined as the locus of the centers of gravity of the transverse cross-sections of the guide), and x and y are the transverse coordinates.

Let $\rho = \rho_1 = (\rho_2 + \rho_3)/2 = \rho_2 + a/2$ be the radius of curvature of the guide axis, and let the origin of the polar coordinates be taken at the center of curvature. Then z is equal to $-\rho_1\varphi$ where the minus sign is necessary to make (x, y, z) a right-handed coordinate system. Since the vertical (in

Fig. 1) walls are to be specified by $x = 0$ and $x = a$ we set $x = \rho - \rho_1 + a/2$. Thus, the two sets of coordinates are related by

$$\begin{aligned} \rho &= x + \rho_1 - a/2 = x + \rho_2 \\ \varphi &= -z/\rho_1 \\ y &= y \end{aligned} \tag{1.2-1}$$

where ρ_1, ρ_2 and a are constants.

We again choose $\vec{A}$ and $\vec{B}$ in (1.1-1) to be parallel to the y axis. In the cylindrical coordinates,

$$\begin{aligned} E_\rho &= \frac{1}{i\omega\epsilon}\frac{\partial^2 A}{\partial\rho\partial y} - \frac{1}{\rho}\frac{\partial B}{\partial\varphi} & H_\rho &= \frac{1}{\rho}\frac{\partial A}{\partial\varphi} + \frac{1}{i\omega\mu}\frac{\partial^2 B}{\partial\rho\partial y} \\ E_\varphi &= \frac{1}{i\omega\epsilon\rho}\frac{\partial^2 A}{\partial\varphi\partial y} + \frac{\partial B}{\partial\rho} & H_\varphi &= -\frac{\partial A}{\partial\rho} + \frac{1}{i\omega\mu\rho}\frac{\partial^2 B}{\partial\varphi\partial y} \\ E_y &= -i\omega\mu A + \frac{1}{i\omega\epsilon}\frac{\partial^2 A}{\partial y^2} & H_y &= -i\omega\epsilon B + \frac{1}{i\omega\mu}\frac{\partial^2 B}{\partial y^2} \end{aligned} \tag{1.2-2}$$

where now, from (1.1-2), A satisfies the wave equation

$$\frac{1}{\rho}\frac{\partial}{\partial\rho}\left[\rho\frac{\partial A}{\partial\rho}\right] + \frac{1}{\rho^2}\frac{\partial^2 A}{\partial\varphi^2} + \frac{\partial^2 A}{\partial y^2} = \sigma^2 A \tag{1.2-3}$$

and likewise for B.

One method of dealing with (1.2-3) which is sometimes used is to assume

$$A = e^{ip\varphi} \times (\text{sine or cosine function of } y) \times f(\rho) \tag{1.2-4}$$

where $f(\rho)$ turns out to be a Bessel function of order p with its argument proportional to ρ. However, we shall proceed in a different direction.

The change of coordinates (1.2-1) transforms (1.2-2) into

$$\begin{aligned} E_x = E_\rho &= \frac{1}{i\omega\epsilon}\frac{\partial^2 A}{\partial x\partial y} + \frac{\rho_1}{\rho}\frac{\partial B}{\partial z} & H_x = H_\rho &= -\frac{\rho_1}{\rho}\frac{\partial A}{\partial z} + \frac{1}{i\omega\mu}\frac{\partial^2 B}{\partial x\partial y} \\ E_y &= -i\omega\mu A + \frac{1}{i\omega\epsilon}\frac{\partial^2 A}{\partial y^2} & H_y &= -i\omega\epsilon B + \frac{1}{i\omega\mu}\frac{\partial^2 B}{\partial y^2} \\ E_z = -E_\varphi &= \frac{\rho_1}{i\omega\epsilon\rho}\frac{\partial^2 A}{\partial z\partial y} - \frac{\partial B}{\partial x} & H_z = -H_\varphi &= \frac{\partial A}{\partial x} + \frac{\rho_1}{i\omega\mu\rho}\frac{\partial^2 B}{\partial z\partial y} \end{aligned} \tag{1.2-5}$$

and (1.2-3) into

$$\frac{\partial^2 A}{\partial x^2} + \frac{\partial^2 A}{\partial y^2} + \frac{\rho_1^2}{\rho^2}\frac{\partial^2 A}{\partial z^2} + \frac{1}{\rho}\frac{\partial A}{\partial x} - \sigma^2 A = 0, \tag{1.2-6}$$

where ρ_1 is a constant and $\rho = x + \rho_1 - a/2$ is to be considered a function of x. To solve (1.2–6) and the corresponding equation for B we assume

$$A = \sum_{m,n} \alpha_{mn}(z) \sin (\pi m x/a) \cos (\pi n y/b)$$
$$m = 1, 2, 3, \cdots; \qquad n = 0, 1, 2, \cdots \tag{1.2–7}$$

$$B = \sum_{m,n} \beta_{mn}(z) \cos (\pi m x/a) \sin (\pi n y/b)$$
$$m = 0, 1, 2, \cdots; \qquad n = 1, 2, 3, \cdots \tag{1.2–8}$$

these expressions being suggested by (1.1–3) and (1.1–4). The expressions (1.2–5) for the electric intensity show that this choice of A and B make its tangential component vanish at the walls of the guide. Thus the boundary conditions are satisfied.

In order to determine $\alpha_{mn}(z)$ so that the differential equation for A is satisfied, we substitute (1.2–7) in (1.2–6). The resulting left hand side of (1.2–6) may be regarded as a function, say $f(x, y)$, of x and y with the α's and their derivatives entering as parameters. We must choose the α's so as to make this function zero. Relations which must be satisfied by the α's may be obtained by expanding $f(x, y)$ in a double Fourier series for which the typical term is a coefficient times $\sin (\pi m x/a) \cos (\pi n y/b)$, and then setting the coefficient of each term to zero. This form of expansion is suggested by (1.2–7). However, it should be mentioned that such an expansion is best suited to a function which vanishes at $x = 0$ and $x = a$, a condition not fulfilled by $f(x, y)$ because of the term $\rho^{-1}\partial A/\partial x$ in (1.2–6). This causes no real trouble because our region of representation runs only from $x = 0$ to $x = a$ and hence our series is no worse than the Fourier sine series for the periodic function (of period $2a$) which is -1 for $-a < x < 0$ and $+1$ for $0 < x < a$.

To carry out the procedure outlined above, we multiply (1.2–6) (after putting in (1.2–7)) by $\sin(\pi p x/a) \cos(\pi \ell y/b)$ and integrate x from 0 to a and y from 0 to b. Using the expression (1.1–5) for Γ_{mn}^2 and reducing gives

$$-\Gamma_{p\ell}^2 \alpha_{p\ell}(z) + \sum_{m=1}^{\infty} [P_{pm} \alpha_{m\ell}''(z) - S_{pm} \alpha_{m\ell}(z)] = 0 \tag{1.2–9}$$

where p may have any one of the values $1, 2, 3, \cdots$ and the double prime on α denotes the second derivative with respect to z. The P's and S's are constants given by

$$P_{pm} = (2/a) \int_0^a (\rho_1^2/\rho^2) \sin (\pi p x/a) \sin (\pi m x/a)\, dx, \tag{1.2–10}$$

$$S_{pm} = -2\pi m a^{-2} \int_0^a \sin (\pi p x/a) \cos (\pi m x/a)\, dx/\rho \tag{1.2–11}$$

The evaluation of these integrals is discussed in Appendix I. Thus (1.2–9) is the p^{th} equation of a set of differential equations to be solved simultaneously for $\alpha_{1\ell}(z)$, $\alpha_{2\ell}(z)$, $\cdots$.

The customary method of solving a set of equations such as (1.2–9) is to assume that all the α's vary as $e^{\gamma z}$ so that for each $\alpha_{m\ell}(z)$ we may write $e^{\gamma z} g_{m\ell}$. This leads to a set of simultaneous homogeneous linear equations for the constants $g_{m\ell}$. In order that these equations may have a solution the determinant of the coefficients must vanish. Since the only derivative of $\alpha_{m\ell}(z)$ contained in (1.2–9) is the second, γ appears in the determinant only as γ^2. Let γ_1^2, γ_2^2, γ_3^2, $\cdots$ be the values of γ^2 which cause the determinant to vanish and let k_{1j}, k_{2j}, $\cdots$ be the values of $g_{1\ell}$, $g_{2\ell}$, $\cdots$ corresponding to $\gamma^2 = \gamma_j^2$. The k's are determined to within an arbitrary multiplying constant which, for the sake of convenience, is chosen so that $k_{jj} = 1$.

Thus one solution of the differential equation (1.2–6) is

$$A = e^{-\gamma_j z} \cos(\pi \ell y/b) \sum_{m=1}^{\infty} k_{mj} \sin(\pi m x/a). \tag{1.2–13}$$

This particular solution corresponds to the j^{th} one of the modes (traveling in the positive z direction) for which A is proportional to $\cos(\pi \ell y/b)$.

In much the same way it may be shown that the series (1.2–8) assumed for B is a solution of equation (1.2–6) (with A replaced by B) provided the coefficients $\beta_{mn}(z)$ satisfy the set of equations

$$-\Gamma_{p\ell}^2 \beta_{p\ell}(z) + \sum_{m=0}^{\infty} [Q_{pm} \beta_{m\ell}''(z) - U_{pm} \beta_{m\ell}(z)] = 0 \tag{1.2–14}$$

for $p = 0, 1, 2, \cdots$ and $\ell = 1, 2, 3, \cdots$. Here

$$Q_{pm} = (\epsilon_p/a) \int_0^a (\rho_1^2/\rho^2) \cos(\pi p x/a) \cos(\pi m x/a)\, dx \tag{1.2–15}$$

$$U_{pm} = \pi m \epsilon_p a^{-2} \int_0^a \cos(\pi p x/a) \sin(\pi m x/a)\, dx/\rho \tag{1.2–16}$$

where $\epsilon_0 = 1$ and $\epsilon_p = 2$ for $p > 0$. These integrals are discussed in Appendix I.

The problem of determining the reflection from a bend in a wave guide involves considerable manipulation of equations (1.2–9) and (1.2–14). The introduction of matrix notation in the manner suggested by the work of Section 1.1 for straight guides simplifies this work. Although $\alpha_{mn}(z)$ and $\beta_{mn}(z)$ are no longer the simple exponential functions given by (1.1–9), it turns out that the column matrices $\alpha(z)$ and $\beta(z)$ are still given by (for a wave traveling in the positive z direction) by the matrix expression (1.1–10). As mentioned earlier, Γ_α and Γ_β are no longer simple diagonal matrices.

We now turn to the task of expressing (1.2–9) and (1.2–14) in matrix form.

1.3 *Propagation Constant Matrix for Curved Rectangular Guide*

From this point onward in our investigation of propagation in the rectangular guide we shall assume A to be proportional to $\cos(\pi \ell y/b)$. Thus instead of the general expression (1.2–7) for A we shall deal with the more restricted form

$$A = \cos(\pi \ell y/b) \sum_{m=1}^{\infty} \alpha_{m\ell}(z) \sin(\pi m x/a) \qquad (1.3\text{–}1)$$

where ℓ has one of the values 0, 1, 2, 3, $\cdots$. Since the most general disturbance may be obtained by the superposition of disturbances of the form (1.3–1) no real generality will be lost.

The introduction of (1.3–1) is suggested by the fact that the set $\alpha_{1\ell}(z)$, $\alpha_{2\ell}(z)$, $\cdots$ may be determined from (1.2–9) (at least to within arbitrary constants of integration) without considering the other $\alpha_{mn}(z)$'s, $n \neq \ell$. The introduction of (1.3–1) is also suggested by physical reasons. The plane of the bend is the z, x plane and there is nothing in the system tending to change the field distribution in the y direction.

Equation (1.2–13) is a special case of (1.3–1). Furthermore the most general form of (1.3–1) (corresponding to a wave progressing in the positive z direction) may be obtained by multiplying (1.2–13) by an arbitrary constant c_j and summing on j.

In order to write the set of differential equations (1.2–9) for the $\alpha_{m\ell}(z)$'s in matrix form we introduce the infinite matrices

$$\Gamma_0 = \begin{bmatrix} \Gamma_{1\ell} & 0 & 0 & \cdot \\ 0 & \Gamma_{2\ell} & 0 & \cdot \\ 0 & 0 & \Gamma_{3\ell} & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix}, \qquad \alpha(z) = \begin{bmatrix} \alpha_{1\ell}(z) \\ \alpha_{2\ell}(z) \\ \alpha_{3\ell}(z) \\ \cdot \end{bmatrix} \qquad (1.3\text{–}2)$$

$$P = \begin{bmatrix} P_{11} & P_{12} & \cdot \\ P_{21} & P_{22} & \cdot \\ \cdot & \cdot & \cdot \end{bmatrix}, \qquad S = \begin{bmatrix} S_{11} & S_{12} & \cdot \\ S_{21} & S_{22} & \cdot \\ \cdot & \cdot & \cdot \end{bmatrix}$$

where the elements of Γ_0 are obtained by setting $n = \ell$ in equation (1.1–5) for Γ_{mn}, and the elements of P and Q are given by the integrals (1.2–10) and (1.2–11). The rules of matrix multiplication then show that (1.2–9) is the p^{th} element of the matrix equation

$$P\alpha''(z) - (\Gamma_0^2 + S)\alpha(z) = 0 \qquad (1.3\text{–}3)$$

Premultiplying by P^{-1} converts this equation into

$$\alpha''(z) - \Gamma_\alpha^2\, \alpha(z) = 0 \qquad (1.3\text{–}4)$$

where

$$\Gamma_\alpha^2 = P^{-1}(\Gamma_0^2 + S) \tag{1.3-5}$$

It may be verified by direct differentiation of the series (1.1–11) defining $\exp(-z\Gamma)$ that a solution of (1.3–4) is

$$\alpha(z) = e^{-z\Gamma_\alpha} g^+ \tag{1.3-6}$$

where, as in (1.1–10) for the straight guide, g^+ is a column of constants (of integration). However, now Γ_α is to be obtained by taking the square root of the right hand side of (1.3–5), a process which is not easy since it usually requires one to obtain the characteristic roots and modal columns of Γ_α^2 (see equation (1.3–10)).

As far as (1.3–6) being a solution of the differential equation is concerned, Γ_α may be any matrix whose square is given by (1.3–5). We shall restrict it as follows: When $\xi = a/\rho_1$ becomes small, as in the case of a gentle bend, it is seen from (1.2–10) and (1.2-11) that P approaches the unit matrix and S approaches zero. Hence, Γ_α^2 approaches the diagonal matrix Γ_0^2. Γ_α is chosen so that it approaches Γ_0, that is, all of the elements in the principal diagonal are either positive real or positive imaginary. This makes $\exp(-z\Gamma_\alpha)$ approach the diagonal matrix $\exp(-z\Gamma_0)$. With this choice of Γ_α the expression (1.3–6) for $\alpha(z)$ corresponds to a wave traveling in the positive z direction.

The various modes of propagation in the bend may be obtained from Γ_α^2 by expressing, in matrix notation, the steps leading to (1.2–13) (which gives A for the j^{th} mode). We assume $\alpha(z)$ to be the column matrix obtained by multiplying the column matrix g of constants by the scalar quantity $e^{\gamma z}$. Setting this in (1.3–4) gives

$$(\gamma^2 I - \Gamma_\alpha^2)g = 0 \tag{1.3-7}$$

where I is the unit matrix. In order that (1.3–7) may have a solution, the determinant of the coefficient of g must vanish. This leads to the characteristic equation* for γ^2:

$$|\gamma^2 I - \Gamma_\alpha^2| = 0 \tag{1.3-8}$$

The vertical bars denote the determinant of the inclosed matrix. The roots $\gamma_1^2, \gamma_2^2, \cdots$ are therefore the latent (or characteristic) roots of Γ_α^2. If we let k_j denote** the column g obtained when $\gamma = \gamma_j$ in (1.3–7) then

* See Section 3.6 of Reference[9].

** We choose this notation in order to adhere as closely as possible to that of Reference[9]. Incidentally, the column k_j is proportional to the j^{th} column of κ^{-1} where κ is the modal row matrix introduced in Section 5.1.

$$(\gamma_j^2 I - \Gamma_\alpha^2)\, k_j = 0 \tag{1.3-9}$$

and the elements $k_{1j}, k_{2j}, \cdots$ of the modal column k_j are the ones appearing as coefficients in (1.2–13).

Equation (1.3–9) and the methods of matrix analysis lead to

$$\Gamma_\alpha^2 = k[\gamma^2]_d\, k^{-1}, \qquad \Gamma_\alpha = k\,[\gamma]_d\, k^{-1} \tag{1.3-10}$$

where k is the square matrix whose j^{th} column is k_j and $[\gamma^2]_d$, $[\gamma]_d$ are diagonal matrices having γ_j^2, γ_j as the j^{th} elements in their principal diagonals. The representation (1.3–10) certainly holds for the rectangular guide since in this case no repeated roots occur.

In analogy with the expression (1.3–1) for A we shall henceforth deal with B in the form

$$B = \sin(\pi \ell y/b) \sum_{m=0}^{\infty} \beta_{m\ell}(z) \cos(\pi m x/a) \tag{1.3-11}$$

where ℓ has one of the values 1, 2, 3, $\cdots$. In much the same way as before it may be shown that for a wave traveling along the bend in the positive direction the $\beta_{m\ell}(z)$'s in (1.3–11) are given by

$$\beta(z) = e^{-z\Gamma_\beta}\, d^+ \tag{1.3-12}$$

where d^+ is a column of arbitrary constants and

$$\Gamma_\beta^2 = Q^{-1}(\Gamma_0^2 + U) \tag{1.3-13}$$

In (1.3–12) and (1.3–13)

$$\Gamma_0 = \begin{bmatrix} \Gamma_{0\ell} & 0 & 0 & \cdot \\ 0 & \Gamma_{1\ell} & 0 & \cdot \\ 0 & 0 & \Gamma_{2\ell} & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix} \qquad \beta(z) = \begin{bmatrix} \beta_{0\ell}(z) \\ \beta_{1\ell}(z) \\ \beta_{2\ell}(z) \\ \cdot \end{bmatrix}$$

$$Q = \begin{bmatrix} Q_{00} & Q_{01} & \cdot \\ Q_{10} & Q_{11} & \cdot \\ \cdot & \cdot & \cdot \end{bmatrix} \qquad U = \begin{bmatrix} 0 & U_{01} & U_{02} & \cdot \\ 0 & U_{11} & U_{12} & \cdot \\ 0 & U_{21} & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix} \tag{1.3-14}$$

where the elements of Γ_0, Q and U are given by equations (1.1–5), (1.2–15) and (1.2–16), respectively.

1.4 *Continuity Conditions at Junction of Straight and Curved Rectangular Guides*

Electromagnetic theory requires that E_x, E_y, H_x and H_y be continuous in crossing a plane z = constant which marks the junction of a straight and a curved wave guide (of the same cross-section). Comparison of the first equation in (1.1–7) with the first equation in (1.2–5) shows that E_x is continuous if (1)A is continuous and (2) if $\partial B/\partial z$ in the straight portion is equal (the equality being taken at the junction) to $(\rho_1/\rho)\ \partial B/\partial z$ in the curved portion. Examination of the expressions for the remaining field components shows that all the continuity conditions are satisfied if, at the junction,

$$\begin{aligned}[A \text{ in straight portion}] &= [A \text{ in bend}] \\ \left[\frac{\partial A}{\partial z} \quad \text{''} \quad \text{''} \quad \text{''}\right] &= \left[\frac{\rho_1}{\rho}\frac{\partial A}{\partial z} \text{ in bend}\right]\end{aligned} \tag{1.4–1}$$

and likewise for B.

Let A in the bend be given by (1.3–1) and let $\alpha(z)$ denote the column matrix of coefficients shown in (1.3–2). A in the straight portion may be represented in the same way except that $\alpha(z)$ has a simpler form as explained in Section 1.1. When these expressions for A are inserted in (1.4–1), both sides multiplied by $(2/a)\sin(\pi px/a)$ after cancelling out the cos $(\pi \ell y/b)$, and the results integrated with respect to x from 0 to a we obtain relations which may be expressed as the matrix equations

$$\begin{aligned}[\alpha(z) \text{ in straight portion}] &= [\alpha(z) \text{ in bend}] \\ \left[\frac{d\alpha(z)}{dz} \quad \text{''} \quad \text{''} \quad \text{''}\right] &= \left[V\frac{d\alpha(z)}{dz} \text{ in bend}\right]\end{aligned} \tag{1.4–2}$$

where V is the square matrix whose p^{th} row and m^{th} column $(p, m = 1, 2, 3, \cdots)$ is

$$V_{pm} = (2\rho_1/a)\int_0^a \sin(\pi px/a)\sin(\pi mx/a)\, dx/\rho, \tag{1.4–3}$$

ρ being equal to $\rho_1 + x - a/2$.

By using expression (1.3–11) for B in the continuity conditions, it may be shown in much the same way that the column matrix $\beta(z)$ given by (1.3–14) must satisfy the relations

$$\begin{aligned}[\beta(z) \text{ in straight portion}] &= [\beta(z) \text{ in bend}] \\ \left[\frac{d\beta(z)}{dz} \quad \text{''} \quad \text{''} \quad \text{''}\right] &= \left[W\frac{d\beta(z)}{dz} \text{ in bend}\right]\end{aligned} \tag{1.4–4}$$

where W is the infinite square matrix

$$W = \begin{bmatrix} W_{00} & W_{01} & \cdot \\ W_{10} & W_{11} & \cdot \\ \cdot & \cdot & \cdot \end{bmatrix} \tag{1.4–5}$$

whose elemenst are

$$W_{pm} = (\epsilon_p \rho_1/a) \int_0^a \cos(\pi p x/a) \cos(\pi m x/a)\, dx/\rho \tag{1.4-6}$$

$$\epsilon_0 = 1, \qquad \epsilon_p = 2 \quad \text{for} \quad p > 0$$

Both V_{pm} and W_{pm} are discussed in Appendix I.

PART II

THEORY FOR A GENERAL WAVE GUIDE

2.1 *Matrix Propagation Constant for a Curved Wave Guide of Arbitrary Cross-Section*

In Section 1.3 it has been shown that for a curved rectangular wave guide there exists a square matrix Γ_α (or Γ_β) which plays the same role in the propagation of a wave consisting of many modes as does the propagation constant in a simple transmission line. There Γ_α was obtained from a special form of the wave equation which is suited to bends in rectangular guides. Here we adopt a different approach with the idea of showing that a matrix propagation constant Γ exists under more general conditions.

The general theory of wave propagation in tubes shows that a wave traveling in the positive z direction may often be represented as

$$\Phi = \sum_{j=1}^{\infty} c_j e^{-z\gamma_j} \varphi_j(x, y) \tag{2.1-1}$$

where Φ is some quantity associated with the field and is analogous to the functions A and B of Part I. In (2.1-1) x and y are transverse coordinates and z a longitudinal coordinate. γ_j is the propagation constant for the j^{th} mode and $\varphi_j(x, y)$ the corresponding eigenfunction. For a circular bend in a rectangular wave guide $\varphi_j(x, y)$ is a combination of trigonometric and Bessel functions and γ_j is proportional to the order of the Bessel functions.

We assume that we may find a set of functions $\theta_m(x, y)$, $m = 1, 2, 3, \ldots$ such that every $\varphi_j(x, y)$ may be represented as

$$\varphi_j(x, y) = \sum_{m=1}^{\infty} k_{mj} \theta_m(x, y) \tag{2.1-2}$$

The usefulness of this procedure depends upon our ability to pick a system of $\theta_m(x, y)$'s which is appreciably simpler than the system of $\varphi_j(x, y)$'s. In the work of Part I $\theta_m(x, y)$ was taken to be the eigenfunction of the typical mode of propagation in a straight guide, i.e. the product of a sine and a cosine.

We assume further in (2.1-2) that the square matrix k^{-1} exists where k_{mj} is the element in the m^{th} row and j^{th} column of k; i.e. if a root γ_j is repeated, say, s times there are s linearly independent columns (k_j's) corresponding to γ_j. Substitution of (2.1–2) in (2.1–1) gives

$$\begin{aligned}\Phi &= \sum_{m=1}^{\infty} \theta_m(x, y) \sum_{j=1}^{\infty} k_{mj} c_j e^{-z\gamma_j} \\ &= \sum_{m=1}^{\infty} \mu_m(z)\theta_m(x, y)\end{aligned} \tag{2.1–3}$$

where

$$\mu_m(z) = \sum_{j=1}^{\infty} k_{mj} c_j e^{-z\gamma_j} \tag{2.1–4}$$

Since $\theta_m(x, y)$ is analogous to the product of the trigonometrical terms in (1.3–1) or (1.3–11) these equations show that $\mu_m(z)$ plays the same role as $\alpha_m\ell(z)$ or $\beta_m\ell(z)$. Therefore, in accordance with the discussion given at the beginning of this section, we wish to show that the column matrix $\mu(z)$ (which is similar to $\alpha(z)$ or $\beta(z)$) whose m^{th} element is $\mu_m(z)$ may be expressed as

$$\mu(z) = e^{-z\Gamma} f^+ \tag{2.1–5}$$

In this equation Γ is a square matrix to be determined and f^+ is a column matrix of constants similar to g^+ or d^+.

The rules of matrix multiplication and equation (2.1–4) show that

$$\mu(z) = k[e^{-z\gamma}]_d c \tag{2.1–6}$$

in which $[\exp(-z\gamma)]_d$ is a diagonal matrix having $\exp(-z\gamma_j)$ as the j^{th} element in its principal diagonal and c is the column matrix formed from the c_j's. We introduce the column f^+ by defining it as $\mu(0)$ whence

$$f^+ = kc, \quad c = k^{-1}f^+ \tag{2.1–7}$$

Incidentally, from (2.1–3), the value of Φ at $z = 0$ is

$$\Phi_{z=0} = \sum_{m=1}^{\infty} f_m^+ \theta_m(x, y) \tag{2.1–8}$$

where f_m^+ is the m^{th} element in f^+.

From (2.1–6) and (2.1–7)

$$\mu(z) = k[e^{-z\gamma}]_d\, k^{-1} f^+ \tag{2.1–9}$$

In this equation $k\,[\exp(-z\gamma)]_d\, k^{-1}$ is a square matrix which may be expressed as

$$\begin{aligned}\sum_{n=0}^{\infty} \frac{(-z)^n}{n!} k[\gamma]_d^n k^{-1} &= \sum_{n=0}^{\infty} \frac{(-z)^n}{n!} (k[\gamma]_d k^{-1})^n \\ &= e^{-z\Gamma}\end{aligned} \tag{2.1–10}$$

Here $[\gamma]_d$ represents the diagonal matrix having γ_j the j^{th} term in its principal diagonal and

$$\Gamma = k[\gamma]_d\, k^{-1} \tag{2.1-11}$$

Therefore we have shown that $\mu(z)$ is of the form (2.1–5) which is what we set out to do.

It is rather difficult to compute Γ from (2.1–11) using only the above definitions of k and γ_j for one must first obtain the functions $\varphi_j(x, y)$. In dealing with the rectangular guide it is easier to use equations (1.3–5) and (1.3–13) to determine Γ.

2.2 *Reflection at a Single Junction*

Let a straight wave guide extending from $z = -\infty$ to $z = 0$ be joined to a curved guide of the same cross-section which extends from $z = 0$ to $z = \infty$. Let an incident wave

$$\Phi_i = \sum_{m=1}^{\infty} h_m\, e^{-z\delta_m}\, \theta_m(x, y) \tag{2.2-1}$$

come in from the left along the straight guide. The h_m's are given constants, the δ_m's are the modal propagation constants for straight guides (for rectangular guides they are the Γ_{mn}'s given by (1.1–5)), and $\theta_m(x, y)$ is the m^{th} eigenfunction for the straight guide (the product of a sine and a cosine for a rectangular guide).

What are the reflected and transmitted waves set up by (2.2–1)? The reflected wave is of the form

$$\Phi_r = \sum_{m=1}^{\infty} f_m^-\, e^{z\delta_m}\, \theta_m(x, y) \tag{2.2-2}$$

where the f_m^-'s are to be determined.

If we assume the representation

$$\Phi = \sum_{m=1}^{\infty} \mu_m(z)\theta_m(x, y) \tag{2.2-3}$$

to hold for all real values of z then, since $\Phi = \Phi_i + \Phi_r$ for $z < 0$, equations (2.2–1) and (2.2–2) show that

$$\mu_m(z) = h_m e^{-z\delta_m} + f_m^-\, e^{z\delta_m}, \qquad m = 1, 2, 3, \ldots ; z < 0 \tag{2.2-4}$$

Introducing the column matrices $\mu(z)$, h, f^- and the diagonal matrices $\exp(\pm z\Gamma_0)$ where Γ_0 is a diagonal matrix having δ_m as the m^{th} term in its principal diagonal enables us to write (2.2–4) as

$$\mu(z) = e^{-z\Gamma_0}\, h + e^{z\Gamma_0} f^-, \qquad z < 0 \tag{2.2-5}$$

Equation (2.2–5) is more general than the expression (2.1–5) for $\mu(z)$ in that it contains waves going in both directions, but is more special in that Γ_0 is a diagonal matrix.

In the curved guide we take $\mu(z)$ to be given by (2.1–5), thus

$$\mu(z) = e^{-z\Gamma} f^+, \qquad z > 0 \tag{2.2-6}$$

where Γ is a square matrix whose elements are assumed to be known and f^+ is a column matrix whose elements are to be determined along with those of f^-.

The conditions that the transverse components of the electric and magnetic intensities be continuous at the junction of the two guides are assumed to lead to the requirements

$$\begin{aligned} [\mu(z) \text{ in straight portion}] &= [\mu(z) \text{ in curved portion}] \\ \left[\frac{d}{dz}\mu(z) \text{ in straight portion}\right] &= \left[V\frac{d}{dz}\mu(z) \text{ in curved portion}\right] \end{aligned} \tag{2.2-7}$$

where the quantities within the brackets are evaluated at the junction and V is a square matrix whose elements are constants. When the curvature of the curved portion becomes small V approaches the unit matrix. For the problem at hand (2.2–7) may be written as

$$[\mu(z)]_{z=-0} = [\mu(z)]_{z=+0} \tag{2.2-8}$$

$$\left[\frac{d}{dz}\mu(z)\right]_{z=-0} = V\left[\frac{d}{dz}\mu(z)\right]_{z=+0} \tag{2.2-9}$$

in which the subscripts $z = -0$, $z = +0$ refer to the straight and curved portions, respectively, of the guide at $z = 0$.

The requirements (2.2–7) have been established for the rectangular guide in Section 1.4. Their form is also suggested by the conditions that the voltage and current be continuous at the junction of two transmission lines. Thus if we let $\mu(z)$ play the role of the voltage, the current in the first line is $-Z_1^{-1}\, d\mu(z)/dz$ and the current in the second is $-Z_2^{-1}\, d\mu(z)/dz$ where Z_1 and Z_2 denote the distributed series impedances of the two lines. It is seen that this leads to scalar equations which look like (2.2–7), but now V denotes the scalar Z_1/Z_2 instead of a square matrix.

Setting the expressions (2.2–5) and (2.2–6) for $\mu(z)$ in the conditions (2.2–8) and (2.2–9) gives two equations which may be solved simultaneously to obtain f^- and f^+ in terms of h, Γ_0, Γ and V:

$$\begin{aligned} h + f^- &= f^+ \\ -\Gamma_0 h + \Gamma_0 f^- &= -V\Gamma f^+ \end{aligned} \tag{2.2-10}$$

$$f^- = (\Gamma_0 + V\Gamma)^{-1}(\Gamma_0 - V\Gamma)h \tag{2.2-11}$$

$$f^+ = (\Gamma_0 + V\Gamma)^{-1} 2\Gamma_0 h \tag{2.2-12}$$

Since f^- and f^+ specify the reflected and transmitted waves, respectively, they give the answer which we are seeking.

If the curved guide should extend from $z = -\infty$ to $z = 0$ and the straight guide from $z = 0$ to $z = \infty$ the response to an incident wave $e^{-z\Gamma} h$ coming in along the curved guide would be

$$\begin{aligned} \mu(z) &= e^{-z\Gamma} h + e^{z\Gamma} f^-, \qquad z < 0 \\ \mu(z) &= e^{-z\Gamma_0} f^+, \qquad z > 0 \end{aligned} \tag{2.2-13}$$

A procedure similar to that used above shows that

$$\begin{aligned} f^- &= -(\Gamma_0 + V\Gamma)^{-1} (\Gamma_0 - V\Gamma)h \\ f^+ &= (\Gamma_0 + V\Gamma)^{-1} 2V\Gamma h \end{aligned} \tag{2.2-14}$$

where, instead of condition (2.2-9), we have used

$$V \left[\frac{d}{dz} \mu(z) \right]_{z=-0} = \left[\frac{d}{dz} \mu(z) \right]_{z=+0} \tag{2.2-15}$$

2.3 *Reflection Due to a Bend*

Let the guide be straight for $-\infty < z < -c$ and for $c < z < \infty$, and let these two portions be connected by a curved portion in which the longitudinal coordinate z runs from $-c$ to $+c$. As in Section 2.2 we take the matrix propagation constants for the straight and curved portions to be the square matrices Γ_0 and Γ, respectively, and assume an incident wave, specified by the column matrix h, to come in from $z = -\infty$.

The column matrix $\mu(z)$ whose m^{th} element appears as the coefficient of $\theta_m(x, y)$ in the representation (2.2-3) for Φ is now given by

$$\begin{aligned} \mu(z) &= e^{-z\Gamma_0} h + e^{z\Gamma_0} f^-, \qquad z < -c \\ \mu(z) &= (\cosh z\Gamma)p + (\sinh z\Gamma)q, \qquad -c < z < c \\ \mu(z) &= e^{-z\Gamma_0} f^+, \qquad c < z \end{aligned} \tag{2.3-1}$$

In these expressions f^-, f^+, p, q are column matrices which may be determined as functions of the known matrices Γ_0, Γ, V and h by substituting (2.3-1) in the conditions (2.2-7) which must hold at the junctions $z = -c$ and $z = c$.

By straightforward algebra similar to that used for the analogous problem in transmission line theory we obtain

$$\begin{aligned} e^{-c\Gamma_0} f^- + e^{-c\Gamma_0} f^+ &= [-I + 2(V\Gamma \tanh c\Gamma + \Gamma_0)^{-1} \Gamma_0]\, e^{c\Gamma_0} h \\ e^{-c\Gamma_0} f^- - e^{-c\Gamma_0} f^+ &= [-I + 2(V\Gamma \coth c\Gamma + \Gamma_0)^{-1} \Gamma_0]\, e^{c\Gamma_0} h \end{aligned} \tag{2.3-2}$$

In these equations the infinite square matrix $\tanh c\Gamma$ is defined as $(\sinh c\Gamma)(\cosh c\Gamma)^{-1}$ and $\coth c\Gamma$ as its reciprocal. $\sinh c\Gamma$ and $\cosh c\Gamma$ may be defined as power series in $c\Gamma$ and may be expressed as combinations of $\exp(c\Gamma)$ and $\exp(-c\Gamma)$.

An expression for the column matrix f^- may be obtained by adding the equations in (2.3–2). Before doing this it is convenient to introduce the two column matrices x and y defined by

$$\begin{aligned}(Vc\Gamma \tanh c\Gamma + c\Gamma_0)\, x &= c\Gamma_0\, e^{c\Gamma_0}\, h \\ (Vc\Gamma \coth c\Gamma + c\Gamma_0)\, y &= c\Gamma_0\, e^{c\Gamma_0}\, h\end{aligned} \tag{2.3–3}$$

where the scalar length c has been introduced to make the various terms dimensionless. Each equation in (2.3–3) represents an infinite set of simultaneous linear equations to be solved for the elements of x or y.

Once x and y are known the reflected wave is given by

$$f^- = e^{c\Gamma_0}(x + y) - e^{2c\Gamma_0}\, h \tag{2.3-4}$$

and the transmitted wave by

$$f^+ = e^{c\Gamma_0}(x - y) \tag{2.3–5}$$

PART III

GENTLE BENDS—GENERAL THEORY

3.1 *Limiting Forms Assumed for* Γ *and* V

It will be shown in Part IV that for gentle circular bends in rectangular wave guides the matrix propagation constant Γ is such that

$$\Gamma^2 = \Gamma_0^2 + F \tag{3.1–1}$$

where Γ_0^2 is the square of the matrix propagation constant for the straight guide. Γ_0^2 is a diagonal matrix having δ_m^2 (which is one of the Γ_{mn}^2's, depending on the set of modes under consideration, given by (1.1–5)) for the m^{th} element in its principal diagonal. F is a square matrix of infinite order in which the elements F_{ii} in the principal diagonal are of order ξ^2 and the remaining elements F_{ij}, $i \neq j$ are of order ξ. Here $\xi = a/\rho_1$ is the ratio of the guide width to the radius of curvature of the bend. As the bend becomes more and more gentle, $\xi \rightarrow 0$.

The asymptotic expressions given in Appendix I show that, for gentle bends in rectangular guides, the square matrix V which appears in the junction conditions (2.2–7) approaches a unit matrix as $\xi \rightarrow 0$. In particular $V_{ii} = 1 + v_{ii}$ where v_{ii} is of order ξ^2, and V_{ij}, the element in the i^{th} row and j^{th} column, is of order ξ when $i \neq j$.

Throughout the remainder of Part III we shall assume that Γ^2, F and V behave as mentioned above. In addition we assume that there is no degeneracy, i.e. all of the δ_m's are unequal to each other and to zero.

3.2 *Propagation in a Gentle Bend*

Here we assume that the elements of Γ_0^2 and F in the expression (3.1–1) for Γ^2 are known. We wish to find the modal propagation constant γ_j and the corresponding eigenfunction $\varphi_j(x, y)$ for the j^{th} mode.

After squaring both sides of the collineatory transformation (2.1–11) connecting Γ and the diagonal matrix $[\gamma]_d$ we obtain a relation which may be written as $k[\gamma^2]_d - \Gamma^2 k = 0$. The left hand side is a square matrix having $(\gamma_j^2 I - \Gamma^2)k_j$ as its j^{th} column. Here I is the unit matrix and k_j is a column matrix having $k_{1j}, k_{2j}, \ldots$ as its elements (k_j is the j^{th} column of k). Thus we have a system of simultaneous linear equations in which the coefficients are furnished by the square matrix $\gamma_j^2 I - \Gamma^2$ and in which the unknowns are $k_{1j}, k_{2j}, \cdots$. Accordingly, γ_j^2 is the j^{th} latent root of Γ^2 and k_j is its corresponding modal column just as for the rectangular guide in Section 1.3.

In order to apply equations (A2–16) of Appendix II we set $\lambda_j = \gamma_j^2$ and $u = \Gamma^2$ so that, from (3.1–1),

$$u_{jj} = \delta_j^2 + F_{jj}, \; u_{ij} = F_{ij}, \qquad i \neq j \tag{3.2–1}$$

Therefore

$$\gamma_j^2 = \delta_j^2 + F_{jj} + \sum_{s=1}^{\infty}{}' F_{js}F_{sj}/(\delta_j^2 - \delta_s^2) \tag{3.2–2}$$

$$k_{jj} = 1, \qquad k_{sj} = F_{sj}/(\delta_j^2 - \delta_s^2), \qquad s \neq j \tag{3.2–3}$$

where we have neglected terms of order ξ^3 in (3.2–2) and of order ξ^2 in k_{sj}, $s \neq j$. The prime on the summation indicates that the term $s = j$ is to be omitted.

When $k_{1j}, k_{2j}, \ldots$ are known the eigenfunction $\varphi_j(x, y)$ may be written as a series in $\theta_m(x, y)$ by means of equation (2.1–2).

In Section 3.3 we shall need the form assumed by the square matrix $\Gamma \tanh c\Gamma$ in a gentle bend. This matrix is used in computing the reflection from such a bend, as might be inferred from equation (2.3–3). The formula to be used is (A2–18) with $u = \Gamma^2$, $\lambda_j = \gamma_j^2$ and with the elements of the square matrix k given by (3.2–3). In the diagonal matrix of (A2–18) we set

$$\begin{aligned} f(\lambda_j) &= \lambda_j^{1/2} \tanh c\lambda_j^{1/2} = \gamma_j \tanh c\gamma_j = \gamma_j t_j \\ t_j &= \tanh c\gamma_j \end{aligned} \tag{3.2–4}$$

and for the elements of k^{-1} we use (A2–19) together with the line above it. When the three matrices on the right of (A2–18) are multiplied out the element $(\Gamma \tanh c\Gamma)_{ij}$ in the i^{th} row and j^{th} column of $\Gamma \tanh c\Gamma$ is found to be

$$(\Gamma \tanh c\Gamma)_{ij} = (\gamma_j l_j - \gamma_i l_i) k_{ij}, \qquad i \neq j \tag{3.2-5}$$

$$(\Gamma \tanh c\Gamma)_{ii} = \gamma_i l_i + \sum_m{}' (\gamma_i l_i - \gamma_m l_m) k_{im} k_{mi} \tag{3.2-6}$$

where in (3.2–5) and (3.2–6) terms of $O(\xi^2)$ ("order of") and $O(\xi^3)$, respectively, have been neglected. This is in line with the fact that the terms in the principal diagonals of our matrices must be accurate to within $O(\xi^2)$ while the remaining terms need be accurate only to within terms of $O(\xi)$. The summation with respect to m runs from $m = 1$ to ∞ with $m = i$ omitted.

3.3 *Reflection from a Gentle Bend*

When the bend is gentle so that V and Γ behave according to the description given in Section 3.1, the matrix expressions for the reflection coefficients given in Section 2.3 may be evaluated. The results stated in Appendix II for "almost diagonal" matrices furnish the principal tools for this work.

It is assumed that the incident wave coming in along the straight guide from the left is $\Phi_i = \exp(-z\delta_p)\, \theta_p(x, y)$ and hence contanis only the p^{th} mode. Comparing this with the general expression (2.2–1) for Φ_i shows that $h_p = 1$, $h_m = 0$, $m \neq p$, and all the elements of the column matrix h are zero except the p^{th} which is unity.

We start by writing the first of equations (2.3–3) as

$$(\Gamma \tanh c\Gamma + V^{-1}\Gamma_0)x = V^{-1}\, \Gamma_0\, e^{c\Gamma_0}\, h \tag{3.3-1}$$

Since V approaches a unit matrix as $\xi \to 0$, the element $(V^{-1})_{ij}$ in the i^{th} row and j^{th} column of V^{-1} is

$$\begin{aligned} (V^{-1})_{ij} &= -V_{ij}, \qquad i \neq j \\ (V^{-1})_{ii} &= 1 - v_{ii} + \sum_m{}' V_{im}\, V_{mi} \end{aligned} \tag{3.3-2}$$

where $V_{ii} = 1 + v_{ii}$, $i, j = 1, 2, 3, \ldots$ and the summation with respect to m runs from 1 to ∞ with the term for $m = i$ omitted (as indicated by the prime on Σ). In omitting this term we are neglecting v_{ii}^2 because it is of order ξ^4. These results follow from equation (A2–2) of Appendix II. As usual, the elements in the principal diagonal are accurate to within $O(\xi^2)$ and the remaining elements to within $O(\xi)$.

It follows that $V^{-1}\, \Gamma_0\, e^{c\Gamma_0}\, h = \eta$ is a column matrix whose i^{th} element is

$$\begin{aligned} \eta_i &= -V_{ip}\, \delta_p\, e^{c\delta_p}, \qquad i \neq p \\ \eta_p &= (1 - v_{pp} + \sum_m{}' V_{pm}\, V_{mp})\, \delta_p\, e^{c\delta_p}, \qquad i = p \end{aligned} \tag{3.3-3}$$

Likewise, the element in the i^{th} row and j^{th} column of the square matrix $V^{-1}\Gamma_0$ is $-V_{ij}\delta_j$ when $i \neq j$ and is

$$(1 - v_{ii} + \sum_m{}' V_{im} V_{mi})\, \delta_i \tag{3.3-4}$$

when $i = j$.

By combining the approximate expressions for the elements of $\Gamma \tanh c\Gamma$ (given by (3.2–5) and (3.2–6)) and $V^{-1}\Gamma_0$ we find that if u_{ij} denotes the i^{th} row and j^{th} column of $\Gamma \tanh c\Gamma + V^{-1}\Gamma_0$ then

$$\begin{aligned} u_{ij} &= D_{ij} - V_{ij}\delta_j\,, \qquad i \neq j \\ u_{ii} &= \gamma_i t_i + \sum_m{}' D_{mi} k_{im} + \delta_i(1 - v_{ii} + \sum_m{}' C_{mi}) \\ &= \sigma_i - \delta_i v_{ii} + \sum_m{}' (D_{mi} k_{im} + \delta_i C_{mi}) \end{aligned} \tag{3.3-5}$$

In these equations we have set

$$\begin{aligned} \sigma_i &= \delta_i + \gamma_i t_i, \qquad C_{mi} = V_{im} V_{mi} \\ D_{mi} &= (\gamma_i t_i - \gamma_m t_m) k_{mi} = (\gamma_i t_i - \gamma_m t_m) F_{mi}/(\delta_i^2 - \delta_m^2) \end{aligned} \tag{3.3-6}$$

where γ_i and k_{mi} are given by (3.2–2) and (3.2–3).

We are now in a position to identify the matrix equation (3.3–1) for x with the set of equations (A2–20). The quantity η_i which appears on the right hand side of the i^{th} equation in (A2–20) is given by (3.3–3). The coefficients which appear on the left hand side are the u's defined by (3.3–5). Therefore, from (A2–21), when $i \neq p$,

$$\begin{aligned} x_i &= \frac{\delta_p e^{c\delta_p}}{\sigma_i \sigma_p} [-V_{ip}(\sigma_p - \delta_p) - D_{ip}] \\ &= -\frac{e^{c\delta_p}}{\delta_i(1 + t_i)(1 + t_p)} [V_{ip}\delta_p t_p + F_{ip}(\delta_i t_i - \delta_p t_p)(\delta_i^2 - \delta_p^2)^{-1}] \end{aligned} \tag{3.3-7}$$

where we have neglected higher order terms and in so doing have replaced γ_j by the simpler δ_j.

When $i = p$ (A2–21) yields

$$x_p = u_{pp}^{-1}[\eta_p + \sum_m{}' (u_{pm} u_{mp} \eta_p - u_{pm} u_{pp} \eta_m)/(u_{mm} u_{pp})] \tag{3.3-8}$$

In order to combine the second order terms in $1/u_{pp}$ with those in the rest of the expression for x_p we assume that σ_p is the major portion of u_{pp}. Then, approximately,

$$1/u_{pp} = \sigma_p^{-1}[1 + \delta_p v_{pp}/\sigma_p - \sum_m{}' (D_{mp} k_{pm} + \delta_p C_{mp})/\sigma_p] \tag{3.3-9}$$

This assumption, which is equivalent to assuming that tanh $c\gamma_p$ differs appreciably from -1, does not appear to restrict our results since tanh $c\gamma_p$ is either purely imaginary or else real and positive.

Substituting the appropriate values in (3.3–8), neglecting higher order terms, and using the definition (3.3–6) for C_{mp} leads to

$$x_p = \sigma_p^{-1} \delta_p e^{c\delta_p}[1 - (1 - \delta_p/\sigma_p)v_{pp} + \sum_m{}' (\sigma_p \sigma_m)^{-1}\{C_{mp}(\sigma_p - \delta_p)(\sigma_m - \delta_m)$$
$$+ D_{mp}(D_{pm} - \sigma_m k_{pm}) + (\sigma_p - \delta_p)D_{pm} V_{mp} - \delta_m D_{mp} V_{pm}\}] \quad (3.3\text{–}10)$$

A reduction similar to that used in going from the first to the second line of (3.3–7) gives our final expression for x_p

$$x_p = \frac{\delta_p e^{c\delta_p}}{\delta_p + \gamma_p t_p} - \frac{t_p e^{c\delta_p}}{(1 + t_p)^2}\Bigg[v_{pp} - \sum_m{}' V_{pm} V_{mp} t_m/(1 + t_m)$$
$$+ \sum_m{}' \frac{\delta_m t_m - \delta_p t_p}{\delta_m \delta_p(1 + t_m)(\delta_m^2 - \delta_p^2)} \{V_{pm} \delta_m F_{mp}/t_p \quad (3.3\text{–}11)$$
$$- V_{mp} \delta_p F_{pm} + F_{pm} F_{mp}(\delta_p + \delta_m/t_p)(\delta_m^2 - \delta_p^2)^{-1}\}\Bigg]$$

The above expressions for x_i and x_p have been derived from the first of equations (2.3–3). The second of equations (2.3–3) determines the column matrix y in the same way that the first equation determines x except that coth $c\Gamma$ now replaces tanh $c\Gamma$. Therefore, we may obtain expressions for the elements of y by replacing the t's (where t_i = tanh $c\gamma_i$) by their reciprocals in the expressions for the corresponding x's (i.e. in (3.3–7) and (3.3–11). The values obtained in this way lead to, when $i \neq p$,

$$\begin{aligned} f_i^- &= e^{c\delta_i}(x_i + y_i) \\ &= \delta_i^{-1} e^{c(\delta_i+\delta_p-\gamma_i-\gamma_p)} \sinh c(\gamma_i + \gamma_p)[-V_{ip}\delta_p - F_{ip}/(\delta_i + \delta_p)] \\ f_i^+ &= e^{c\delta_i}(x_i - y_i) \\ &= \delta_i^{-1} e^{c(\delta_i+\delta_p-\gamma_i-\gamma_p)} \sinh c(\gamma_i - \gamma_p)[V_{ip}\delta_p - F_{ip}/(\delta_i - \delta_p)] \end{aligned} \quad (3.3\text{–}12)$$

where we have used the expressions (2.3–4) and (2.3–5) for f^- and f^+. When $i = p$,

$$f_p^- = e^{c\delta_p}(x_p + y_p) - e^{2c\delta_p} = -e^{2c(\delta_p-\gamma_p)}[A_1 (\sinh 2c\gamma_p)/2 + A_2] \quad (3.3\text{–}13)$$

where

$$A_1 = 2v_{pp} + (\gamma_p^2 - \delta_p^2)\delta_p^{-2} - \sum_m{}' \left[V_{pm} V_{mp} + \frac{V_{pm} F_{mp} + V_{mp} F_{pm}}{\delta_m^2 - \delta_p^2}\right]$$

$$A_2 = \sum_m{}' \frac{(\cosh 2c\gamma_p - e^{-2c\gamma_m})}{2\delta_m \delta_p(\delta_m^2 - \delta_p^2)} [V_{pm} F_{mp} \delta_m^2 + V_{mp} F_{pm} \delta_p^2 + F_{pm} F_{mp}]$$

The expression for f_p^+ may be obtained in the same way but it is slightly more complicated.

$$f_p^+ = e^{c\delta_p}(x_p - y_p) = e^{2c(\delta_p - \gamma_p)}[1 - A_3 + A_4(\sinh 2c\gamma_p)/2 + A_5] \quad (3.3\text{–}14)$$

where

$$A_3 = (1 - e^{-4c\gamma_p})(\gamma_p - \delta_p)^2/(4\delta_p^2)$$

$$A_4 = \sum_m{}' e^{-2c\gamma_m}\left[-V_{pm}V_{mp} + \frac{V_{pm}F_{mp} - V_{mp}F_{pm}}{\delta_m^2 - \delta_p^2} + \frac{2F_{pm}F_{mp}}{(\delta_m^2 - \delta_p^2)^2}\right]$$

$$A_5 = \sum_m{}' \frac{(e^{-c\gamma_m}\cosh 2c\gamma_p - 1)}{2\delta_m\delta_p(\delta_m^2 - \delta_p^2)}$$

$$\cdot \left[V_{pm}F_{mp}\delta_m^2 - V_{mp}F_{pm}\delta_p^2 + \frac{(\delta_m^2 + \delta_p^2)F_{pm}F_{mp}}{\delta_m^2 - \delta_p^2}\right]$$

There are several points we should mention about these formulas for f_p^- and f_p^+: The summations with respect to m run from 1 to ∞ with the term $m = p$ omitted. γ_j and δ_j are the propagation constants of the jth mode in the bend and in the straight portion, respectively. The difference $\gamma_p^2 - \delta_p^2$ may be expressed in terms of the F's by equation (3.2–2). In the course of obtaining (3.3–13) and (3.3–14) relations of the following sort were used.

$$t_p(1 + t_p)^{-2} = e^{-2c\gamma_p}(\sinh 2c\gamma_p)/2$$

$$(t_m + t_p^2)(1 + t_p)^{-2}(1 + t_m)^{-1} = e^{-2c\gamma_p}(\cosh 2c\gamma_p - e^{-2c\gamma_m})$$

The term A_3 arises when we subtract $(1 - t_p^2)(1 + t_p)^{-2}$ from

$$\delta_p/(\delta_p + \gamma_p t_p) - \delta_p t_p/(\gamma_p + \delta_p t_p)$$

Since $\gamma_p - \delta_p$ is $O(\xi^2)$ for a circular bend in a rectangular guide $(\gamma_p - \delta_p)^2$ is $O(\xi^4)$ and hence A_3 is negligible in the cases we shall consider.

The reflected wave set up by an incident wave of unit amplitude and containing only the p^{th} mode (i.e. the incident wave described at the beginning of this section) is given by the column matrix f^- whose elements may be obtained from (3.3–12) and (3.3–13). Likewise, the transmitted wave is given by f^+.

PART IV

GENTLE CIRCULAR BENDS IN RECTANGULAR WAVE GUIDES

4.1 *Propagation of Dominant Mode in a Gentle Bend—H in Plane of Bend*

When the magnetic intensity H lies in the plane of the bend, $H_y = 0$, and equations (1.2–5) show that $B = 0$. Thus we have to deal only with

A. In order to study the dominant mode we set $\ell = 0$ in the $\cos(\pi \ell y/b)$ (A depends on y through this factor) in the formulas of Section 1.3 which involve A and assume the dimensions of the guide to be such that $b < a$.

We wish to determine γ_1^2, the first latent root of Γ_α^2 defined by (1.3–5), from the approximate formula (3.2–2). In our case the elements δ_m^2 of diagonal matrix Γ_0^2 are obtained by putting $n(= \ell)$ to zero in (1.1–5):

$$\delta_m^2 = \Gamma_{m0}^2 = \sigma^2 + (\pi m/a)^2, \qquad m = 1, 2, 3, \cdots \tag{4.1-1}$$

so that (3.2–2) becomes

$$\gamma_1^2 = \Gamma_{10}^2 + F_{11} - \sum_{m=2}^{\infty} F_{1m} F_{m1} a^2 \pi^{-2} (m^2 - 1)^{-1} \tag{4.1-2}$$

The first task is to find the elements of the matrix F where, from (3.1–1) and (1.3–5),

$$F = \Gamma_\alpha^2 - \Gamma_0^2 = (P^{-1} - I)\Gamma_0^2 + P^{-1}S \tag{4.1-3}$$

In the case under consideration $P = I + R$ where R is a square matrix whose elements are very small. In fact, the asymptotic expressions leading to (A1–18) show that R_{ii} and S_{ii} are $O(\xi^2)$, with $\xi = a/\rho_1$, while R_{ij} and S_{ij} are $O(\xi)$ if $i + j$ is odd and $O(\xi^2)$ if $i + j$ is even. When the approximate value of P^{-1} obtained from (A2–2) is set in (4.1–3) and the matrix multiplications carried out it is found that

$$\begin{aligned} F_{ij} &= -R_{ij}\Gamma_{j0}^2 + S_{ij} + O(\xi^2) \\ F_{ii} &= \left(-R_{ii} + \sum_{m=1}^{\infty} R_{im} R_{mi}\right) \Gamma_{i0}^2 + S_{ii} - \sum_{m=1}^{\infty} R_{im} S_{mi} + O(\xi^3) \end{aligned} \tag{4.1-4}$$

The "order of" symbol $O(\ )$ will be omitted in the following equations, it being understood that the terms in the principal diagonal are correct to within $O(\xi^2)$ and the others to within $O(\xi)$.

The values of the F's which enter (4.1–2) may be computed from the asymptotic expressions (A1–18) for the R's and S's. They turn out to be

$$\begin{aligned} F_{1m} &= -4\xi m[4\Gamma_{10}^2 \pi^{-2}(m^2 - 1)^{-2} + 3a^{-2}(m^2 - 1)^{-1}] \\ F_{m1} &= -4\xi m[4\Gamma_{10}^2 \pi^{-2}(m^2 - 1)^{-2} + a^{-2}(m^2 - 1)^{-1}] \\ F_{11} &= \xi^2[\Gamma_{10}^2(1 - 6\pi^{-2}) + 6a^{-2}]/12 \end{aligned} \tag{4.1-5}$$

In the expressions for F_{1m} and F_{m1} m is supposed to have the values 2, 4, 6, $\cdots$. For odd values of m F_{1m} and F_{m1} are $O(\xi^2)$. When $i = 1$ in the expression (4.1–4) for F_{11}, the two series therein reduce to S_3 and S_4 where

$$S_p = \sum_{m=2,4,6\cdots} m^2(m^2 - 1)^{-p} \tag{4.1-6}$$

By expanding the typical terms in partial fractions and using the fact that the sums of (see, for example, page 238 of Reference[10])

$$
\begin{aligned}
U_q &= 1 + 3^{-q} + 5^{-q} + \cdots \\
&= (-)^{q/2-1}\tfrac{1}{2}(2^q - 1)B_q\pi^q/q!
\end{aligned}
\tag{4.1-7}
$$

for $q = 2, 4, 6$, are $\pi^2/8$, $\pi^4/96$, $\pi^6/960$, it may be shown that

$$
\begin{aligned}
S_3 &= \pi^2/64, \qquad S_4 = \pi^4/768 - \pi^2/128, \\
S_5 &= (15\pi^2 - \pi^4)/3072.
\end{aligned}
\tag{4.1-8}
$$

In (4.1-7) B_q denotes the q^{th} Bernoulli number. The values of S_p may also be computed in succession from the two relations*

$$
U_{2p} = \sum_{i=1}^{p} 2^{2i-1} C_{p+i-1,2i-1} S_{p+i} = \sum_{i=1}^{p+1} 2^{2i-2} C_{p+i-1,2i-2} S_{p+i}
$$

where $C_{m,n}$ is a binomial coefficient. Still another method is to make use of the generating function

$$
\sum_{p=0}^{\infty} t^p S_{p+1} = (1 + t) \sum_{m=2,4,6\cdots} (m^2 - 1 - t)^{-1} = \tfrac{1}{2} - \tfrac{1}{2}\pi x \cot \pi x
$$

where $4x^2 = 1 + t$. Note that by this definition S_1 is $\frac{1}{2}$ in contrast to the non-convergent series obtained by putting $p = 1$ in (4.1-6).

Substituting the values for the F's given by (4.1-5) in the expression (4.1-2) for γ_1^2 and using the sums (4.1-8) of the series which occur gives

$$
\gamma_1^2 = \Gamma_{10}^2 - \frac{\xi^2}{4a^2}\,[1 + a^2\Gamma_{10}^2(1 - 6\pi^{-2}) + (a\Gamma_{10}/\pi)^4(5 - \pi^2/3)] \tag{4.1-9}
$$

When the dominant mode is propagated without attenuation both γ_1^2 and Γ_{10}^2 are negative.

The general form of (4.1-9) has been obtained by both Buchholz[1] and Marshak[5] by different methods. In our notation their result is

$$
\gamma_{mn}^2 = \Gamma_{mn}^2 - \frac{\xi^2}{4a^2}\left[1 + a^2\Gamma_{mn}^2(1 - 6\pi^{-2}m^{-2}) + \left(\frac{a\Gamma_{mn}}{\pi m}\right)^4 (5 - \pi^2 m^2/3)\right] \tag{4.1-10}
$$

where γ_{mn} is the propagation constant for the m, n^{th} mode when the magnetic vector is in the plane of the bend.

* I am indebted to John Riordan for these relations.

4.2 *Reflection Due to Dominant Mode Incident upon Gentle Bend—H in Plane of Bend*

Let the system be the one described in the first paragraph of Section 2.3 and let the incident wave contain only the dominant mode. Then the matrix propagation constant is the Γ_α of Section 4.1 and the column matrix h specifying the incident wave has unity for its top element and zero for its remaining elements, i.e., $p = 1$ in the formulas of Section 3.3.

We shall be interested only in the reflection coefficient, $f_{\bar{1}}$ of the dominant mode. Here we shall denote it by $g_{\bar{1}0}$, in line with the notation of equation (1.1–3), in order to distinguish it from the corresponding coefficient (which will be denoted by $d_{0\bar{1}}$) when E lies in the plane of the bend.

Setting $p = 1$ in the expression (3.3–13) for the reflection coefficient and using equation (4.1–1) for δ_m gives

$$f_{\bar{1}} = g_{\bar{1}0} = -e^{2c(\Gamma_{10}-\gamma_1)}[A_1(\sinh 2c\gamma_1)/2 + A_2] \tag{4.2–1}$$

where γ_1 has just been obtained in (4.1–9) and

$$\begin{aligned} A_1 &= 2v_{11} + (\gamma_1^2 - \Gamma_{10}^2)\Gamma_{10}^{-2} \\ &\quad - \sum_{m=2}^{\infty}\left[V_{1m}V_{m1} + \frac{V_{1m}F_{m1} + V_{m1}F_{1m}}{\pi^2 a^{-2}(m^2 - 1)}\right], \\ A_2 &= \sum_{m=2}^{\infty} \frac{\cosh 2c\gamma_1 - e^{-2c\gamma_m}}{2\Gamma_{m0}\Gamma_{10}\pi^2 a^{-2}(m^2 - 1)} \\ &\quad \cdot [V_{1m}F_{m1}\Gamma_{m0}^2 + V_{m1}F_{1m}\Gamma_{10}^2 + F_{1m}F_{m1}] \end{aligned} \tag{4.2–2}$$

From (A1–18) and $V_{11} = 1 + v_{11}$ it follows that

$$\begin{aligned} v_{11} &= \xi^2(1 - 6\pi^{-2})/12 \\ V_{1m} &= V_{m1} = 8\pi^{-2}\xi m(m^2 - 1)^{-2} \end{aligned} \tag{4.2–3}$$

where $m = 2, 4, 6, \cdots$. For odd values of m, V_{1m} and V_{m1} are $O(\xi^2)$. Substituting these values together with those for the F's given by (4.1–5), using the sums (4.1–8) and the expression (4.1–9) for $\gamma_1^2 - \Gamma_{10}^2$ finally leads to (after considerable cancellation)

$$A_1 = -\xi^2\Gamma_{10}^{-2}a^{-2}/4 \tag{4.2–4}$$

Likewise, for even values of m,

$$V_{1m}F_{m1}\Gamma_{m0}^2 + V_{m1}F_{1m}\Gamma_{10}^2 + F_{1m}F_{m1} = 16\xi^2 m^2 a^{-4}(m^2 - 1)^{-2} \tag{4.2–5}$$

All of the terms in the expression (4.2–1) for $g_{\bar{1}0}$ are now known (the values of γ_m may be obtained by setting $n = 0$ in (4.1–10)). We shall make the further approximation of putting Γ_{m0} for γ_m. Since $\Gamma_{m0} - \gamma_m$ is $O(\xi^2)$ no serious error is introduced and we have

$$g_{\overline{10}} = \frac{\xi^2 \sinh 2c\Gamma_{10}}{8\Gamma_{10}^2 a^2} - \frac{\xi^2 8}{\pi^2} \sum_{m=2,4,6\cdots} \frac{(\cosh 2c\Gamma_{10} - e^{-2c\Gamma_{m0}})}{\Gamma_{10}\,\Gamma_{m0}\,a^2} \frac{m^2}{(m^2-1)^3} \tag{4.2–6}$$

in which

$$a\Gamma_{m0} = [\pi^2(m^2 - 1) + a^2\Gamma_{10}^2]^{\frac{1}{2}}, \qquad \xi = a/\rho_1 . \tag{4.2–7}$$

For frequencies such that only the dominant mode is propagated the ratio of the power in the reflected wave to the power in the incident wave is $| g_{\overline{10}} |^2$. Marshak has given an expression for this ratio which is the same as that obtained from (4.2–6) when the negligible (for his case) terms $e^{-2c\Gamma_{m0}}$ are omitted.

The corresponding expression for the transmission coefficient derived from (3.3–14) for f_1^+ is not as simple as (4.2–6).

4.3 *Propagation of Dominant Mode in a Gentle Bend—E in Plane of Bend*

When the electric intensity E lies in the plane of the bend, $E_y = 0$, and equations (1.2–5) show that $A = 0$. Here we deal with B in much the same way as we dealt with A in Section 4.1. The dominant mode is obtained by setting $\ell = 1$ in the $\sin(\pi \ell y/b)$ in the formulas pertaining to B in Section 1.3. It is assumed that $b > a$.

Examination of the matrices (1.3–14) indicates that, for the sake of convenience, we should call the top row of our matrices the 0^{th} row and the left-most column the 0^{th} column. In line with this we call γ_0 the propagation constant of the dominant mode in the bend. The elements δ_m^2 of the diagonal matrix Γ_0^2 are obtained by putting $n(= \ell) = 1$ in (1.1–5):

$$\delta_m^2 = \Gamma_{m1}^2 = \sigma^2 + (\pi m/a)^2 + \pi^2/b^2, \qquad m = 0, 1, 2, \cdots \tag{4.3–1}$$

When we make the appropriate shift in the subscripts, equation (3.2–2) yields

$$\gamma_0^2 = \Gamma_{01}^2 + F_{00} + \sum_{m=1}^{\infty} F_{0m} F_{m0} a^2 \pi^{-2} m^{-2} \tag{4.3–2}$$

in which the elements of the matrix F are to be determined from (1.3–13):

$$F = \Gamma_\beta^2 - \Gamma_0^2 = (Q^{-1} - I)\Gamma_0^2 + Q^{-1}U \tag{4.3–3}$$

As in (4.1–4) we have, with $Q = I + T$,

$$\begin{aligned} F_{ij} &= -T_{ij}\Gamma_{j1}^2 + U_{ij} \\ F_{ii} &= \left(-T_{ii} + \sum_{m=0}^{\infty} T_{im} T_{mi}\right) \Gamma_{i1}^2 + U_{ii} - \sum_{m=0}^{\infty} T_{im} U_{mi} . \end{aligned} \tag{4.3–4}$$

By using the asymptotic expressions (A1–19) for the T's and U's and summing the series with the value of (4.1–7) for $q = 4$ given in Section 4.1 we obtain

$$
\begin{aligned}
F_{0m} &= -2\xi[2\Gamma_{01}^2\pi^{-2}m^{-2} + a^{-2}] \\
F_{m0} &= -8\xi\Gamma_{01}^2\pi^{-2}m^{-2} \qquad (4.3\text{–}5) \\
F_{00} &= \xi^2\Gamma_{01}^2/12
\end{aligned}
$$

In these expressions m is supposed to have the values 1, 3, 5, $\cdots$. When m is even F_{0m} and F_{m0} are $O(\xi^2)$.

Substituting (4.3–5) in (4.3–2) and summing the series with the help of the values of (4.1–7) given in Section 4.1 gives

$$\gamma_0^2 = \Gamma_{01}^2 - \xi^2\Gamma_{01}^2(5 + 2a^2\Gamma_{01}^2)/60 \qquad (4.3\text{–}6)$$

A result equivalent to (4.3–6) has been given by Buchholz who also gives the approximation to the propagation constant when $m > 0$ (and the electric vector in the plane of the bend). In our notation his approximation is

$$
\begin{aligned}
\gamma_{mn}^2 = \Gamma_{mn}^2 + \frac{\xi^2}{4a^2}\Bigg[3 &- \left(\frac{a\Gamma_{mn}}{\pi m}\right)^2 (10 + \pi^2 m^2) \\
&+ \tfrac{1}{3}\left(\frac{a\Gamma_{mn}}{\pi m}\right)^4 (21 + \pi^2 m^2)\Bigg] \qquad (4.3\text{–}7)
\end{aligned}
$$

In writing (4.3–7) we have corrected a misprint in Buchholz's expression. In order to agree with Buchholz's equation (5.30a) the leading term within the square brackets would have to be changed from 3 to -3. This change was indicated by the results obtained when our equation (3.2–2) was used to obtain special cases of (4.3–7). Probably the best way of obtaining (4.3–7) is furnished by Marshak's method (WKB approximation, out to second order terms, applied to Bessel's differential equation). If one wishes to verify (4.3–7) by using Marshak's report[5] as a guide, he should correct the misprint in Marshak's equation (12a).

4.4 *Reflection Due to Dominant Mode Incident upon Gentle Bend—E in Plane of Bend*

The problem here is the same as that treated in Section 4.2 except that now the electric vector lies in the plane of the bend. In line with equation (1.1–4), the reflection coefficient f_1^- for the dominant mode will be denoted by d_{01}^-. As in Section 4.3 the subscripts indicating the position of matrix elements will be adjusted so as to start with 0 instead of 1. The square matrix W given by (1.4–5), and associated with the junction conditions for

B in the same manner as V is associated with A, now replaces V. Thus our expression (3.3–13) for the reflection coefficient becomes

$$f_{01} = \bar{d_1} = -[A_1(\sinh 2c\Gamma_{01})/2 + A_2] \tag{4.4-1}$$

where we have neglected the difference in Γ_{01} and γ_0 and where

$$A_1 = 2w_{00} + (\gamma_0^2 - \Gamma_{01}^2)\Gamma_{01}^{-2} - \sum_{m=1}^{\infty} \cdot \left[W_{0m} W_{m0} + \frac{W_{0m} F_{m0} + W_{m0} F_{0m}}{\pi^2 a^{-2} m^2} \right] \tag{4.4-2}$$

$$A_2 = \sum_{m=1}^{\infty} \frac{\cosh 2c\Gamma_{01} - e^{-2c\Gamma_{m1}}}{2\Gamma_{m1}\Gamma_{01}\pi^2 a^{-2} m^2} \cdot [W_{0m} F_{m0} \Gamma_{m1}^2 + W_{m0} F_{0m} \Gamma_{01}^2 + F_{0m} F_{m0}]$$

From $W_{00} = 1 + w_{00}$ and the asymptotic expressions (A1–19) it follows that, for $m = 1, 3, 5, \cdots$

$$\begin{aligned} w_{00} &= \xi^2/12 \\ W_{0m} &= 2\xi m^{-2}\pi^{-2}, \qquad W_{m0} = 4\xi m^{-2}\pi^{-2} \end{aligned} \tag{4.4-3}$$

For even values of m, W_{0m} and W_{m0} are $O(\xi^2)$. Substitution of these values together with those for the F's given by (4.3–5), using the sums (4.1–7) and expression (4.3–6) for $\gamma_0^2 - \Gamma_{01}^2$ leads to

$$A_1 = \xi^2/12$$

$$W_{0m}F_{m0}\Gamma_{m1}^2 + W_{m0}F_{0m}\Gamma_{01}^2 + F_{0m}F_{m0} = -8\xi^2\Gamma_{01}^2 a^{-2}\pi^{-2}m^{-2}$$

for m odd.

Thus the reflection coefficient for the dominant mode when E lies in the plane of a gentle bend of length $2c$ is approximately

$$\bar{d}_{01} = -\frac{\xi^2 \sinh 2c\Gamma_{01}}{24} + \frac{\xi^2 4\Gamma_{01}}{\pi^4} \sum_{m=1,3,5\cdots}^{\infty} \frac{\cosh 2c\Gamma_{01} - e^{-2c\Gamma_{m1}}}{m^4\Gamma_{m1}} \tag{4.4-4}$$

where Γ_{m1} is given by (4.3–1) and $b > a$.

PART V

NUMERICAL CALCULATIONS

5.1 *Bend in Plane of Magnetic Vector*

Let $a/b = 2.25$ and $\lambda_0/a = 1.400$ where λ_0 is the free-space wavelength of the dominant wave striking the bend. The propagation constant

Γ_{10} of the dominant wave is obtained by setting $m = 1$, $n = 0$ in (1.1–5). The Γ's corresponding to the higher modes may be obtained from the same formula:

$$
\begin{aligned}
a^2\Gamma_{10}^2 &= -\left(\frac{2\pi a}{\lambda_0}\right)^2 + \pi^2 = -10.272 \qquad & a\Gamma_{10} &= i\,3.205 \\
a^2\Gamma_{20}^2 &= a^2\Gamma_{10}^2 + 3\pi^2 = 19.336 & a\Gamma_{20} &= 4.397 \\
a^2\Gamma_{30}^2 &= a^2\Gamma_{10}^2 + 8\pi^2 = 68.684 & a\Gamma_{30} &= 8.288
\end{aligned}
\tag{5.1-1}
$$

We shall consider a 90° bend. The approximation (4.2–6) appropriate to gentle bends becomes

$$g_{10}^- = i\xi^2[-\ .0122 \sin (5.03/\xi) + .0087 \cos (5.03/\xi)] \tag{5.1-2}$$

where the exponential terms have been omitted since they are generally negligible. In (5.1–2), $\xi = a/\rho_1$ and the arguments of the sine and cosine terms arise from $2c\Gamma_{10} = \pi a\Gamma_{10}/(2\xi)$. From (4.1–9) the approximate change in the propagation constant produced by the curvature is obtainable from

$$\gamma_1^2 - \Gamma_{10}^2 = .294\xi^2/a^2 \tag{5.1-3}$$

where γ_1 is the propagation constant of the dominant mode in the bend.

The determination of $g_{10}^{\pm}$ by matrix methods will be illustrated for a 90° bend in which $\rho_1/a = 0.6$. This makes $c/a = \rho_1\pi/(4a) = .4712$, $c\Gamma_{10} = i1.510$ and the appropriate equations in (2.3-5) and (2.3–4) become, upon setting $f_1^- = g_{10}^-$ and $f_1^+ = g_{10}^+$,

$$
\begin{aligned}
g_{10}^+ &= e^{c\Gamma_{10}}(x_1 - y_1) = (.061 + i.998)(x_1 - y_1) \\
g_{10}^- &= e^{c\Gamma_{10}}(x_1 + y_1) - e^{2c\Gamma_{10}} = (.061 + i.998)(x_1 + y_1) \\
&\qquad + .993 - i.121
\end{aligned}
\tag{5.1-4}
$$

Here x_1, y_1 are the top elements in the column matrices x, y. The problem is to compute x and y from the matrix equations (2.3–3) with Γ replaced by Γ_α, Γ_0 defined by (1.3–2) with $\ell = 0$, and h a column matrix whose elements are zero except the top one which is unity. Since the order of the matrices is infinite, an exact solution calls for an infinite amount of work. A compromise must be made between the accuracy desired and amount of work one is willing to do. The following numerical work uses third order matrices.

The first step is to compute the square matrix, obtained from (1.3–5),

$$a^2\Gamma_\alpha^2 = P^{-1}(a^2\Gamma_0^2 + a^2S) \tag{5.1-5}$$

The elements of the diagonal matrix $a^2\Gamma_0^2$ are given by (5.1–1) and those of P and S by the equations and tables of Appendix I.

$$a^2\Gamma_\alpha^2 = \begin{bmatrix} 1.4429 & .8812 & .6821 \\ .8812 & 2.1250 & 1.3745 \\ .6821 & 1.3745 & 2.4879 \end{bmatrix}^{-1} \begin{bmatrix} -12.292 & 3.3447 & 1.6087 \\ -6.3039 & 16.369 & 6.9738 \\ -3.5034 & -11.3031 & 65.213 \end{bmatrix}$$
$$= \begin{bmatrix} -9.086 & -1.785 & -5.178 \\ .157 & 17.218 & -19.362 \\ .996 & -13.566 & 38.329 \end{bmatrix} \tag{5.1–6}$$

The next step is to use (5.1–6) to evaluate the coefficients of x and y in (2.3–3). The square matrices $\Gamma_\alpha c \tanh \Gamma_\alpha c$ and $\Gamma_\alpha c \coth \Gamma_\alpha c$ cause most of the computational difficulties. We shall deal with these matrices by using Sylvester's theorem (an account of this theorem is given in Section 3.9 of Reference[9]). This requires the determination of the latent roots and modal rows of $a^2\Gamma_\alpha^2$. However, it is interesting to note that the matrices in question may also be computed from $c^2\Gamma_\alpha^2$ (which is easily obtained from $a^2\Gamma_\alpha^2$) by processes which employ only matrix multiplication, addition, and inversion.

Thus, setting A^2 for $c^2\Gamma_\alpha^2$,

$$A^{-1}\sinh A = I + \frac{A^2}{3!} + \frac{A^4}{5!} + \cdots$$

$$\cosh A = I + \frac{A^2}{2!} + \frac{A^4}{4!} + \cdots$$

$$A \coth A = (\cosh A)(A^{-1} \sinh A)^{-1}$$

$$A \tanh A = A^2(A \coth A)^{-1}.$$

Although the series always converge, they do so too slowly to be of use in our computations. The same is true of the series

$$A \tanh A = \sum_{m=1}^{\infty} 8A^2 [(2m - 1)^2 \pi^2 I + 4A^2]^{-1}.$$

For the matrices we shall encounter it appears best to use Sylvester's theorem even though this requires the determination of the latent roots and modal rows of $a^2\Gamma_\alpha^2$. The square matrix formed from the modal rows* will be denoted by κ.

* As has already been mentioned in the footnote associated with equation (1.3–9), we shall use the notation and theory set forth in Sections 3.5 and 3.6 of Reference[9].

We shall use the relations*

$$\begin{aligned} \Gamma_\alpha c \tanh \Gamma_\alpha c &= \kappa^{-1} [\gamma c \tanh \gamma c]_d \, \kappa \\ \Gamma_\alpha c \cosh \Gamma_\alpha c &= \kappa^{-1} [\gamma c \coth \gamma c]_d \, \kappa \end{aligned} \tag{5.1-7}$$

where the subscript d on the brackets stands for "diagonal" matrix, the i^{th} element in the principal diagonal of $[\gamma c \tanh \gamma c]_d$ being $\gamma_i c \tanh \gamma_i c$ where

$$\gamma_i c = c\lambda_i^{1/2}/a \tag{5.1-8}$$

and λ_i is the ith latent root of $a^2\Gamma_\alpha^2$. In our applications γ_i is either positive real or positive imaginary.

From (5.1-6) the λ_i's are the roots of

$$\begin{vmatrix} \lambda + 9.086 & 1.785 & 5.178 \\ -.157 & \lambda - 17.218 & 19.362 \\ -.996 & 13.566 & \lambda - 38.329 \end{vmatrix} \tag{5.1-9}$$

$$= \lambda^3 - 46.461\,\lambda^2 - 101.96\lambda + 3464.5 = 0$$

and have the values

$$\lambda_1 = -8.886, \qquad \lambda_2 = 8.284, \qquad \lambda_3 = 47.06 \tag{5.1-10}$$

The elements κ_{21}, κ_{31} of the modal row $[1, \kappa_{21}, \kappa_{31}]$ corresponding to λ_1 may be obtained by solving the two equations derived from the last two elements of

$$[1, \kappa_{21}, \kappa_{31}](\lambda_1 I - a^2\Gamma_\alpha^2) = 0 \tag{5.1-11}$$

namely,

$$1.785 + (\lambda_1 - 17.218)\,\kappa_{21} + 13.566\,\kappa_{31} = 0$$

$$5.178 + 19.362\,\kappa_{21} + (\lambda_1 - 38.329)\,\kappa_{31} = 0$$

When the value of λ_1 from (5.1-10) is used these equations yield

$$\kappa_{21} = .1593, \qquad \kappa_{31} = .1750$$

Likewise, the first and third elements of

$$[\kappa_{12}, 1, \kappa_{32}](\lambda_2 I - a^2\Gamma_\alpha^2) = 0$$

and the first and second elements of

$$[\kappa_{13}, \kappa_{23}, 1](\lambda_3 I - a^2\Gamma_\alpha^2) = 0$$

* This is the modal row matrix analogue of equation (11) in Section 3.6 of the Reference[9]. The modal rows of Γ_α are equal to the modal rows of $a^2\Gamma_\alpha^2$.

give

$$\kappa_{12} = .0465 \qquad \kappa_{32} = .6524$$

$$\kappa_{13} = .0165 \qquad \kappa_{23} = -.4555$$

Thus, the numbers entering (5.1–7) are

$$\gamma_1 c = .4712\,(-8.886)^{1/2} = i\,1.404, \qquad \gamma_2 c = .4712\,(8.284)^{1/2} = 1.356$$

$$\gamma_1 c \tanh \gamma_1 c = -8.382 \qquad \gamma_2 c \tanh \gamma_2 c = 1.187$$

$$\gamma_1 c \coth \gamma_1 c = .2354 \qquad \gamma_2 c \coth \gamma_2 c = 1.549$$

$$\begin{aligned} \gamma_3 c &= .4712\,(47.06)^{1/2} = 3.233 \\ \gamma_3 c \tanh \gamma_3 c &= 3.228 \\ \gamma_3 c \coth \gamma_3 c &= 3.243 \end{aligned} \qquad \kappa = \begin{bmatrix} 1 & .1593 & .1750 \\ .0465 & 1 & .6524 \\ .0165 & -.4555 & 1 \end{bmatrix}$$

For the purpose of calculation it is convenient to transform (2.3–3) by inserting (5.1–7) and premultiplying by κV^{-1}. We obtain

$$\begin{aligned} ([\gamma c \tanh \gamma c]_d \kappa + \kappa V^{-1}\Gamma_0 c)\, x &= \kappa V^{-1} c \Gamma_0 e^{c\Gamma_0} h \\ ([\gamma c \coth \gamma c]_d \kappa + \kappa V^{-1}\Gamma_0 c)\, y &= \kappa V^{-1} c \Gamma_0 e^{c\Gamma_0} h \end{aligned} \qquad (5.1\text{–}12)$$

in which

$$\kappa V^{-1} = \begin{bmatrix} 1 & .1593 & .1750 \\ .0465 & 1 & .6524 \\ .0165 & -.4555 & 1 \end{bmatrix} \begin{bmatrix} 1.1204 & .3911 & .1629 \\ .3911 & 1.2833 & .4946 \\ .1629 & .4946 & 1.3460 \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} .9492 & -.1992 & .0883 \\ -.2608 & .7686 & .2339 \\ .1427 & -.7900 & 1.0160 \end{bmatrix}, \qquad \begin{aligned} \Gamma_{10} c &= i\,1.510 \\ \Gamma_{20} c &= 2.072 \\ \Gamma_{30} c &= 3.905 \end{aligned}$$

where the elements of V are obtained from the formulas and tables of Appendix I.

The i^{th} equation of the set obtained by writing out the first of equations (5.1–12) is

$$\sum_{j=1}^{3} [\kappa_{ji}\gamma_i c \tanh \gamma_i c + (\kappa V^{-1})_{ij}\Gamma_{j0} c]x_j = (\kappa V^{-1})_{i1} c\Gamma_{01} e^{c\Gamma_{01}} \qquad (5.1\text{–}13)$$

where $(\kappa V^{-1})_{ij}$ denotes the element in the i^{th} row and j^{th} column of κV^{-1}, κ_{ji} is the element in the i^{th} row and j^{th} column (note the reversal of the usual convention regarding the order of subscripts) of κ, $\kappa_{ii} = 1$, and h has disappeared because it is a column matrix whose top element is unity while the remaining elements are zero. It will be noted that the only imaginary terms in (5.1–13) occur in the coefficients of x_0 and arise from the imaginary quantity $\Gamma_{10}c$. By making the substitution

$$x_j = \frac{u_j c\Gamma_{10} e^{c\Gamma_{10}}}{1 + u_1 c\Gamma_{10}} \tag{5.1–14}$$

the set (5.1–13) may be reduced to

$$(\kappa_{1i}\gamma_i c \tanh \gamma_i c)u_1 + \sum_{j=2}^{3} [\kappa_{ji}\gamma_i c \tanh \gamma_i c + (\kappa V^{-1})_{ij}\Gamma_{j0} c]u_j = (\kappa V^{-1})_{i1} \tag{5.1–15}$$

in which the coefficients are all real. It should be noticed, however, that nothing is gained by making the substitution (5.1–14) when the frequency is so high that other modes in addition to the dominant are propagated.

The equation for y corresponding to (5.1–15) may be obtained by replacing tanh by coth and u by v where now

$$y_j = \frac{v_j c\Gamma_{10} e^{c\Gamma_{10}}}{1 + v_1 c\Gamma_{10}} \tag{5.1–16}$$

Incidentally, if we set $j = 1$ in (5.1–14) and (5.1–16) and substitute in the expressions (5.1–4) for $g_{10}^{\pm}$ we may show that, since u_1 and v_1 are real,

$$| g_{10}^{+} |^2 + | g_{10}^{-} |^2 = 1 \tag{5.1–17}$$

Equation (5.1–17) may be obtained at once from the fact that the energy of the waves leaving the bend must equal the energy of the incident wave. It may also be shown that g_{10}^{-} vanishes when $u_1 v_1 c^2 \Gamma_{10}^2 = 1$.

When the above numbers are set in the three equations obtained from (5.1–15) we get

$$\begin{aligned}
-8.382\, u_1 - 1.748\, u_2 - 1.122\, u_3 &= .9492 \\
.055\, u_1 + 2.780\, u_2 + 1.688\, u_3 &= -.2608 \\
.053\, u_1 - 3.105\, u_2 + 7.191\, u_3 &= .1427
\end{aligned}$$

from which

$$u_1 = -.0940, \qquad x_1 = .1400 + i.0113$$

The equations for v_1 obtained by substituting coth for tanh are

$$\begin{aligned}
.2354\, v_1 - .3732\, v_2 + .3861\, v_3 &= .9492 \\
.0717\, v_1 + 3.1417\, v_2 + 1.924\, v_3 &= -.2608 \\
.0534\, v_1 - 3.1143\, v_2 + 7.211\, v_3 &= .1427
\end{aligned}$$

from which

$$v_1 = 3.930, \qquad y_1 = -.1045 + i.9803$$

When these values are set in (5.1–4) we finally obtain

$$g_{10}^{+} = .9822 + i.1858$$
$$g_{10}^{-} = .0048 - i.0255$$

The following table lists values of $g_{10}^{\pm}$ obtained by the methods of this section*. Here the bend is in the plane of H, $a/b = 2.25$, $\lambda_0/a = 1.400$, where λ_0 is the free space wavelength of the incident wave. ρ_1 is the radius of curvature of the axis of the guide. The smallest possible value of ρ_1/a is 0.5. The term "approx." refers to equations (5.1–2) while "1st order", "2nd order", etc. refers to the order of the matrices used in the computations. The amplitude of the reflected wave is g_{10}^{-} and the amplitude of the wave sent forward is g_{10}^{+} when the incident wave is of unit amplitude.

g_{10}^{+}

ρ_1/a	Approx.	1st order	2nd order	3rd order
.6		.964 $+i.267$	.980 $+i.197$	.982 $+i.186$
.7		.974 $+i.224$	.994 $+i.105$	.994 $+i.111$
.8		.984 $+i.178$	.997 $+i.082$	.997 $+i.082$
.9		.988 $+i.153$	.997 $+i.074$	.997 $+i.073$
1.0		.991 $+i.135$	.998 $+i.066$	.998 $+i.066$
1.2		.994 $+i.110$	.998 $+i.056$	.998 $+i.056$
1.5		.996 $+i.084$	.999 $+i.043$	.999 $+i.044$

g_{10}^{-}

ρ_1/a	Approx.	1st order	2nd order	3rd order
.6	$-i.0280$	.0020 $-i.0074$	.0056 $-i.0280$	.0048 $-i.0255$
.7	$-i.0068$	$-.0005$ $+i.0023$	.0013 $-i.0131$	.0007 $-i.0066$
.8	$+i.0062$	$-.0013$ $+i.0074$	$-.0003$ $+i.0039$	$-.0004$ $+i.0051$
.9	$+i.0128$	$-.0014$ $+i.0087$	$-.0009$ $+i.0123$	$-.0009$ $+i.0123$
1.0	$+i.0143$	$-.0010$ $+i.0075$	$-.0010$ $+i.0148$	$-.0010$ $+i.0147$
1.2	$+i.0079$	$-.0002$ $+i.0018$	$-.0005$ $+i.0086$	$-.0005$ $+i.0085$
1.5	$-i.0040$	$+.0003$ $-i.0034$	$+.0002$ $-i.0041$	$+.0002$ $-i.0042$

It appears that the values obtained from the first order matrices are quite far from the true values. On the other hand there is considerable agreement between the approximation and the second and third order values, especially at the larger values of ρ_1/a.

5.2 *Bend in Plane of Electric Vector*

The calculations for this case are quite similar to those presented in Section 5.1. If we are to deal with the same waveguide it is necessary to

* The computations were performed by Miss M. Darville. I am also indebted to her for the values given in the tables in Section 5.2 and Appendix I.

interchange the dimensions a and b so that now $b/a = 2.25$ and, for the same frequency, $\lambda_0/b = 1.400$.

For a 90° bend the approximation (4.4–4) for the reflection coefficient d_{01}^- for gentle bends, i.e. for $\xi = a/\rho_1$ small, becomes

$$d_{01}^- = i\xi^2[-.0417 \sin (2.23/\xi) + .0209 \cos (2.23/\xi)]$$

where the negligible exponential terms have been neglected just as in the analogue (5.1–2) for g_{10}^-.

The following table, which is similar to the table at the end of Section 5.1, gives the results of computations for bends in the plane of the electric vector.

d_{10}^+

ρ_1/a	Approx.	1st order	2nd order
.6		$.823 + i.547$	$.975 + i.223$
.7		$.887 + i.447$	$.994 + i.051$
.8		$.921 + i.380$	$.996 + i.042$
.9		$.941 + i.332$	$.997 + i.035$
1.0		$.954 + i.295$	$.998 + i.031$
1.2		$.970 + i.242$	$.999 + i.023$
1.5		$.982 + i.190$	$1.000 + i.017$

d_{10}^-

ρ_1/a	Approx.	1st order	2nd order
.6	$-i.0996$	$-.0855 + i.1284$	$-.0020 + i.0086$
.7	$-i.0848$	$-.0520 + i.1031$	$+.0050 - i.0975$
.8	$-i.0706$	$-.0330 + i.0800$	$+.0033 - i.0792$
.9	$-i.0575$	$-.0214 + i.0605$	$+.0022 - i.0635$
1.0	$-i.0457$	$-.0137 + i.0443$	$.0021 - i.0507$
1.2	$-i.0258$	$-.0051 + i.0204$	$.0007 - i.0282$
1.5	$-i.0051$	$+.0001 - i.0004$	$.0001 - i.0062$

The agreement between the approximation for d_{10}^- and its second order matrix value is fairly good from $\rho_1/a = .7$ onward.

APPENDIX I

Calculation of P_{pm} etc. for Circular Bend

It is convenient to write P_{pm} and Q_{pm} as given by (1.2–10) and (1.2–15) in the form

$$\begin{aligned} P_{pm} &= \delta_m^p + R_{pm}, \qquad p, m = 1, 2, 3, \cdots \\ Q_{km} &= \delta_m^p + T_{pm}, \qquad p, m = 0, 1, 2, \cdots \end{aligned} \tag{A1-1}$$

where δ_m^p is unity if $p = m$ and is zero otherwise and

$$R_{pm} = (2/a) \int_0^a (\rho_1^2 \rho^{-2} - 1) \sin(\pi p x/a) \sin(\pi m x/a)\, dx$$

$$T_{pm} = (\epsilon_p/a) \int_0^a (\rho_1^2 \rho^{-2} - 1) \cos(\pi p x/a) \cos(\pi m x/a)\, dx \tag{A1-2}$$

in which $\epsilon_0 = 1, \epsilon_p = 2, p = 1, 2, \cdots$

In (A1–2), (1.2–11), (1.2–16), (1.4–3), (1.4–6) we make the substitutions

$$\begin{aligned} p - m &= r, & u &= \rho_1 \pi/a - \pi/2 = \rho_2\pi/a \\ p + m &= s, & v &= \rho_1 \pi/a + \pi/2 = \rho_3\pi/a \\ y &= \pi x/a, & w &= \rho_1\pi/a, \qquad \rho = x + \rho_1 - a/2 = a(y + u)/\pi \end{aligned} \tag{A1-3}$$

Introduction of the integrals

$$\begin{aligned} I_s &= (1/\pi) \int_0^\pi [w^2(y + u)^{-2} - 1] \cos sy\, dy \\ &= (1/a) \int_0^a (\rho_1^2 \rho^{-2} - 1) \cos(\pi s x/a)\, dx \\ J_s &= \pi \int_0^\pi \frac{\sin sy}{y + u}\, dy = \pi \int_0^a \sin(\pi s x/a)\, dx/\rho, \\ K_s &= \frac{\rho_1}{a} \int_0^\pi \frac{\cos sy}{y + u}\, dy \end{aligned} \tag{A1-4}$$

enables us to write

$$\begin{aligned} R_{pm} &= I_r - I_s, & S_{pm} &= -ma^{-2}(J_s + J_r) \\ T_{pm} &= \epsilon_p(I_r + I_s)/2, & U_{pm} &= m\epsilon_p a^{-2}(J_s - J_r)/2 \\ V_{pm} &= K_r - K_s, & W_{pm} &= \epsilon_p(K_r + K_s)/2 \end{aligned} \tag{A1-5}$$

where I_s and K_s are even functions of s and J_s is an odd function of s. $\epsilon_0 = 1$ and $\epsilon_p = 2, p = 1, 2, 3, \cdots$. Since w and u depend only upon the ratio ρ_1/a, the values of I_s, K_s and J_s depend only upon ρ_1/a and the integer s. These quantities are tabulated at the end of this appendix.

Setting $y + u$ equal to t gives

$$\begin{aligned} J_s &= \pi \int_u^v \sin s(t - u)\, dt/t \\ &= \pi[Si(sv) - Si(su)] \cos su - \pi[Ci(sv) - Ci(su)] \sin su \end{aligned} \tag{A1-6}$$

where Si and Ci denote the integral sine and cosine functions. Integrating by parts enables us to express I_s in terms of J_s. Thus

$$\int_0^\pi (y + u)^{-2} \cos sy\, dy = u^{-1} - v^{-1} \cos s\pi - \pi^{-1} s J_s \tag{A1-7}$$

and

$$I_s = \pi^{-1}w^2[u^{-1} - (-)^s v^{-1} - \pi^{-1}sJ_s] \tag{A1-8}$$

except when $s = 0$ in which case

$$\begin{aligned} I_0 &= \pi^{-1}w^2(u^{-1} - v^{-1}) - 1 = w^2/(uv) - 1 \\ &= \pi^2/(4uv) = [(2\rho_1/a)^2 - 1]^{-1} \end{aligned} \tag{A1-9}$$

When ρ_1/a is large, u and v are large, and the asymptotic expansion of (A1-6) gives

$$\pi^{-1}J_s \sim s^{-1}[u^{-1} - (-)^s v^{-1}] - 2!s^{-3}[u^{-3} - (-)^s v^{-3}] + \cdots \tag{A1-10}$$

When (A1-10) is placed in (A1-8)

$$I_s \sim \pi^{-1}2!w^2 s^{-2}[u^{-3} - (-)^s v^{-3}] - \cdots \tag{A1-11}$$

Formulas for K_s may be obtained in much the same way.

$$\begin{aligned} K_s &= (\rho_1/a) \int_u^v \cos s(t - u)\, dt/t \\ &= (\rho_1/a)\{[Ci(sv) - Ci(su)] \cos su + [Si(sv) - Si(su)]\sin su\} \end{aligned} \tag{A1-12}$$

and when $s = 0$

$$K_0 = (\rho_1/a) \log (1 + \pi/u) \tag{A1-13}$$

The asymptotic expression is

$$a\rho_1^{-1}K_s \sim s^{-2}[u^{-2} - (-)^s v^{-2}] - 3!s^{-4}[u^{-4} - (-)^s v^{-4}] + \cdots$$

It is convenient to write the asymptotic expressions in terms of the new variable

$$\xi = a/\rho_1 \tag{A1-14}$$

When s is even and greater than zero

$$J_s \sim \xi^2/s, \qquad I_s \sim 6\xi^2\pi^{-2}s^{-2}, \qquad K_s \sim 2\xi^2\pi^{-2}s^{-2} \tag{A1-15}$$

and when s is odd

$$J_s \sim 2\xi/s, \qquad I_s \sim 4\xi\pi^{-2}s^{-2}, \qquad K_s \sim 2\xi\pi^{-2}s^{-2} \tag{A1-16}$$

When $s = 0$

$$I_0 \sim \xi^2/4, \qquad K_0 \sim 1 + \xi^2/12 \tag{A1-17}$$

We shall need the following asymptotic expressions which may be obtained from the above work

$$R_{11} \sim \frac{\xi^2}{4}\left(1 - \frac{6}{\pi^2}\right), \qquad S_{11} \sim -a^{-2}\xi^2/2, \qquad V_{11} = 1 + \xi^2\left(\frac{1}{12} - \frac{1}{2\pi^2}\right)$$

$$R_{1m} = R_{m1} \sim 16\xi m\pi^{-2}(m^2 - 1)^{-2}$$
$$S_{1m} \sim -S_{m1}, \qquad S_{1m} \sim 4a^{-2}\xi m(m^2 - 1)^{-1} \tag{A1-18}$$
$$V_{1m} = V_{m1} \sim 8\xi m\pi^{-2}(m^2 - 1)^{-2} \sim R_{1m}/2$$

where $m = 2, 4, 6, \cdots$. Also, if now $m = 1, 3, 5, \cdots$,

$$T_{00} \sim \xi^2/4, \qquad U_{00} = 0, \qquad W_{00} = 1 + \xi^2/12$$
$$T_{m0} \sim 8\xi m^{-2}\pi^{-2}, \qquad U_{m0} = 0$$
$$T_{0m} \sim 4\xi m^{-2}\pi^{-2}, \qquad U_{0m} \sim 2a^{-2}\xi \tag{A1-19}$$
$$W_{0m} \sim 2\xi m^{-2}\pi^{-2}, \qquad W_{m0} \sim 4\xi m^{-2}\pi^{-2}$$

Values of I_s, J_s and K_s

ρ_1/a	I_0	I_1	I_2	I_3	I_4	I_5	I_6
.5	∞	∞	∞	∞	∞	∞	∞
.6	2.2723	2.31879	1.82979	1.43755	1.14772	.94432	.78458
.7	1.04166	1.28232	.76232	.53315	.37692	.28832	.22090
.8	.64103	.88256	.44628	.29206	.18995	.14101	.09986
.9	.44643	.68200	.30111	.18992	.11546	.08517	.05908
1.0	.33333	.56052	.22008	.13637	.07818	.05814	.03865
1.1	.26041	.47844	.16916	.10459	.05772	.04298	.02759
1.2	.21008	.41905	.13486	.08413	.04344	.03361	.02056
1.3	.17361	.37361	.11059	.06961	.03500	.02732	.01633
1.4	.14620	.33780	.09259	.05926	.03168	.02293	.01297
1.5	.12500	.30876	.07872	.05147	.02315	.01969	.01068
2.0	.06667	.21803	.04133	.03080	.01135		
2.5	.04167	.16980	.02564	.02217	.00675		

	J_0	J_1	J_2	J_3	J_4	J_5	J_6
.5	0						
.6	0	3.99809	2.01979	2.31576	1.48355	1.66348	1.15716
.7	0	3.21624	1.3054	1.58176	.84936	1.04899	.61931
.8	0	2.72356	.73339	1.21541	.56683	.77644	.40134
.9	0	2.37231	.70698	.99327	.41079	.62183	.28546
1.0	0	2.10615	.55663	.84343	.31379	.52170	.21578
1.1	0	1.89624	.45091	.73508	.24849	.45123	.16981
1.2	0	1.72581	.37335	.65279	.20254	.39869	.13768
1.3	0	1.58448	.31450	.58812	.16843	.35788	.11413
1.4	0	1.46507	.26878	.53573	.14216	.32515	.09636
1.5	0	1.3628	.23251	.49238	.12243	.29825	.08254
2.0	0	1.0122	.12817	.35299	.06596		
2.5	0	.8062	.08128	.27660	.04140		

ρ_1/a	K_0	K_1	K_2	K_3	K_4	K_5	K_6
.5	∞	∞	∞	∞	∞	∞	∞
.6	1.43874	.61659	.31832	.22547	.15539	.12199	.09275
.7	1.25423	.42303	.17355	.11508	.06971	.05353	.03704
.8	1.17306	.33141	.11439	.07463	.04093	.03195	.02038
.9	1.12748	.27575	.08261	.05431	.02737	.02214	.01325
1.0	1.09861	.23761	.06305	.04244	.01976	.01675	.00938
1.1	1.07891	.20953	.04994	.03471	.01509	.01341	.00705
1.2	1.06476	.18785	.04075	.02935	.01192	.01126	.00548
1.3	1.05421	.17050	.03391	.02539	.00977	.00956	.00443
1.4	1.04610	.15626	.02871	.02243	.00807	.00838	.00365
1.5	1.03972	.14434	.02467	.02013	.00682	.00745	.00315
2.0	1.02165	.10508	.01332	.01330	.00358		
2.5	1.01366	.08295	.00838	.01010	.00217		

APPENDIX II

Functions of Almost Diagonal Matrices

Let E be a matrix whose elements are small in comparison with unity. It is then often possible to approximate a matrix defined as some function of the matrix $I + E$, where I is the unit matrix, by the expansion

$$f(I + E) = If(1) + \frac{E}{1!}f'(1) + \frac{E^2}{2!}f''(1) + \cdots . \tag{A2-1}$$

Thus, for example, when we take $f(z)$ to be z^{-1} we obtain

$$(I + E)^{-1} = I - E + E^2 - \cdots . \tag{A2-2}$$

Here we shall give similar formal results for $f(D + E)$ where now D is a diagonal matrix

$$D = \begin{bmatrix} d_1 & 0 & \cdot & 0 \\ 0 & d_2 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & d_N \end{bmatrix} \tag{A2-3}$$

whose diagonal elements are unequal and the elements E_{ij} and E_{ji} are small in comparison with the absolute value of $| d_i - d_j |$. We shall restrict ourselves to a first approximation of the non-diagonal terms of $f(D + E)$ and to a second approximation of the diagonal terms. The results are closely related to the ones obtained from the perturbation theory used in wave mechanics.

We assume that $f(D + E)$ may be defined by the series

$$f(D + E) = a_0 I + a_1(D + E) + a_2(D + E)^2 + \cdots \tag{A2-4}$$

where a_n is a scalar and

$$(D + E)^2 = (D + E)(D + E) = D^2 + DE + ED + E^2$$

and so on. The sum of the terms independent of E is $f(D)$. The terms of order E are

$$\begin{aligned} &E \text{ in } D + E \\ &DE + ED \text{ in } (D + E)^2 \\ &D^2E + DED + ED^2 \text{ in } (D + E)^3 \\ &\Sigma\, D^{\ell} E D^{m} \text{ in } (D + E)^n \end{aligned} \tag{A2-5}$$

where the summation extends over the non-negative integer values of ℓ and m for which $\ell + m = n - 1$. The element in the ith row and jth column of $D^{\ell}ED^{m}$ is $d_i^{\ell}E_{ij}d_j^{m}$ and hence the corresponding element in the summation in (A2–5) is

$$E_{ij}\Sigma d_i^{\ell}\, d_j^{m} = \begin{cases} (d_i^n - d_j^n)/(d_i - d_j), & i \neq j \\ n\, d_i^{n-1}, & i = j. \end{cases} \tag{A2-6}$$

Thus the terms of order E in the ith row and jth column of $f(D + E)$ are, from (A2–6) and (A2–4),

$$\begin{aligned} &E_{ij}\frac{f(d_i) - f(d_j)}{d_i - d_j}, \qquad i \neq j \\ &E_{ii}f'(d_i), \qquad i = j \end{aligned} \tag{A2-7}$$

where the prime on f denotes its first derivative.

The terms of order E^2 in $(D + E)^n$ are

$$\sum_{k,\ell,m} D^k E D^{\ell} E D^m = \sum_{k,\ell,m} [d_i^k E_{ij}]_M [d_i^{\ell} E_{ij} d_j^m]_M \tag{A2-8}$$

where the summations extend over all the non-negative integer values of k, ℓ, m for which $k + \ell + m = n - 2$. On the right $[d_i^k E_{ij}]_M$ denotes a square matrix whose element in the ith row and jth column is $d_i^k E_{ij}$. Likewise the second factor in brackets is a matrix having $d_i^{\ell} E_{ij} d_j^m$ in the ith row and jth column. The element in the ith row and jth column of (A2–8) is, from the rule for the product of two matrices,

$$\sum_{k,\ell,m} \sum_{s=1}^{N} (d_i^k E_{is})(d_s^{\ell} E_{sj} d_j^m) = \sum_{s=1}^{N} E_{is} E_{sj} \sum_{k,\ell,m} d_i^k d_s^{\ell} d_j^m.$$

If i, s, and j are unequal the sum in k, ℓ, m is

$$\frac{1}{d_i - d_j}\left[\frac{d_i^n - d_s^n}{d_i - d_s} - \frac{d_j^n - d_s^n}{d_j - d_s}\right]$$

and in case of equality the sum may be found by a limiting process. Since we are interested in this order of approximation only for the diagonal terms we set $i = j$ and obtain for the sum

$$\frac{nd_i^{n-1}}{d_i - d_s} - \frac{d_i^n - d_s^n}{(d_i - d_s)^2}, \qquad i \neq s$$

$$\frac{n(n-1)}{2!} d_i^{n-2}, \qquad i = s.$$

Thus the contribution to the ith diagonal element of $f(D + E)$ from terms of type (A2-8) is

$$\frac{E_{ii}^2}{2!} f''(d_i) + \sum_{s=1}^{N}{}' E_{is} E_{si} \left[\frac{f'(d_i)}{d_i - d_s} - \frac{f(d_i) - f(d_s)}{(d_i - d_s)^2} \right] \tag{A2-9}$$

where the prime on Σ indicates that the term $s = i$ is to be omitted.

Thus, to summarize, we may say that the first approximation to the non-diagonal term in the ith row and jth column $(i \neq j)$ of $f(D + E)$ is

$$E_{ij} \frac{f(d_i) - f(d_j)}{d_i - d_j} \tag{A2-10}$$

and the second approximation to the diagonal term in the ith row and ith column of $f(D + E)$ is

$$\begin{aligned} f(d_i) &+ E_{ii} f'(d) + \frac{E_{ii}^2}{2!} f''(d_i) \\ &+ \sum_{s=1}^{N}{}' E_{is} E_{si} \left[\frac{f'(d_i)}{d_i - d_s} - \frac{f(d_i) - f(d_s)}{(d_i - d_s)^2} \right] \end{aligned} \tag{A2-11}$$

where the primes on f denote derivatives and the prime on Σ indicates that the term $s = i$ is to be omitted.

Two results obtained from (A2-10) and (A2-11) are of interest. For the first result we set $f(z) = z^{-1}$ and get the following approximations to the elements of $(D + E)^{-1}$:

$$\begin{aligned} &- E_{ij}(d_i d_j)^{-1}, \qquad i \neq j \\ &d_i^{-1} - d_i^{-2} \left[E_{ii} - \sum_{s=1}^{N} E_{is} E_{si} d_s^{-1} \right], \qquad i = j. \end{aligned} \tag{A2-12}$$

For the second result we set $f(z) = z^{1/2}$ and obtain the following approximations to the elements of $(D + E)^{1/2}$:

$$\begin{aligned} &E_{ij}(d_i^{1/2} + d_j^{1/2})^{-1}, \qquad i \neq j \\ &d_i^{1/2} + \tfrac{1}{2} d_i^{-1/2} \left[E_{ii} - \sum_{s=1}^{N} E_{is} E_{si} (d_i^{1/2} + d_s^{1/2})^{-2} \right], \qquad i = j. \end{aligned} \tag{A2-13}$$

In (A2-12) and (A2-13) the summations include the term $s = i$.

We shall now state several results related to the above formulas. Let u denote the matrix $D + E$ so that the typical element $u_{ij} = E_{ij}$, $i \neq j$, and $u_{ii} = d_i + E_{ii}$. Then the latent roots $\lambda_1, \lambda_2, \cdots \lambda_N$ of u are the roots of the equation obtained by setting the determinant of $\lambda I - u$ to zero:

$$|\lambda I - u| = \begin{vmatrix} \lambda - u_{11} & -u_{12} & \cdot \\ -u_{21} & \lambda - u_{22} & \cdot \\ \cdot & \cdot & \cdot \end{vmatrix} = 0. \tag{A2-14}$$

The modal column k_j corresponding to the jth root λ_j satisfies the matrix equation

$$(\lambda_j I - u)k_j = 0. \tag{A2-15}$$

Since the non-diagonal elements of u are small, we see from (A2-14) that we may label the roots so as to make λ_j nearly equal to u_{jj}, and this together with (A2-15) shows that all the elements of k_j are nearly zero except the jth which we may choose to be unity. When these approximate values are taken as a first approximation in the process of solving (A2-15) by successive approximations, the second approximation is found to be

$$\lambda_j = u_{jj} + \sum_{s=1}^{N}{}' \frac{u_{js} u_{sj}}{u_{jj} - u_{ss}} = u_{jj} + \sum_{s=1}^{N}{}' u_{js} k_{sj}$$

$$k_j = \begin{bmatrix} k_{1j} \\ k_{2j} \\ \cdot \\ 1 \\ \cdot \\ k_{Nj} \end{bmatrix} \qquad k_{sj} = \frac{u_{sj}}{u_{jj} - u_{ss}}, \qquad s \neq j \tag{A2-16}$$

where the 1 in the column for k_j occurs as the jth element. This expression for λ_j occurs in the perturbation method often used in wave mechanics.

For the modal row κ_j corresponding to λ_j we have in much the same way

$$\kappa_j(\lambda_j I - u) = 0$$

$$\kappa_j = [\kappa_{1j}, \kappa_{2j}, \cdots \kappa_{jj}, \cdots \kappa_{Nj}] \tag{A2-17}$$

$$\kappa_{sj} = \frac{u_{js}\kappa_{jj}}{u_{jj} - u_{ss}}$$

where the last expression is an approximation and where κ_{jj} may be chosen at our convenience.

The results (A2-10) and (A2-11) may also be obtained from (A2-2), (A2-16) and the relation*

$$f(u) = k \begin{bmatrix} f(\lambda_1) & 0 & \cdot & 0 \\ 0 & f(\lambda_2) & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & & f(\lambda_N) \end{bmatrix} k^{-1} \tag{A2-18}$$

* This is equation (12) in Section 3.6 of Reference[9]. Although proved only for polynomials it may be verified to be true for the applications which we shall make.

where k is the square matrix $[k_1, k_s, \cdots k_N]$ composed of the modal columns k_j, and the latent roots are assumed to be distinct. The element in the ith row and j column of k^{-1} is, to our order of approximation, equal to $-k_{ij}$ when $i \neq j$ and to

$$1 + \sum_{s=1}^{N}{}' k_{js} k_{sj} \tag{A2-19}$$

when $i = j$. Here, as usual, the prime on Σ indicates that the term $s = j$ is to be omitted.

Another result which is sometimes useful is the approximate solution of

$$\begin{array}{l} u_{11}x_1 + u_{12}x_2 + \cdots + u_{1p}x_p + \cdots \quad u_{1N}x_N = \eta_1 \\ u_{21}x_1 + u_{22}x_2 + \cdots + u_{2p}x_p + \cdots + u_{2N}x_N = \eta_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ u_{p1}x_1 + u_{p2}x_2 + \cdots + u_{pp}x_p + \cdots + u_{pN}x_N = \eta_p \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{array} \tag{A2-20}$$

where $|\eta_j| \ll |\eta_p|, j \neq p$, and the u's are of the type assumed in (A2-14):

$$x_j = \frac{\eta_j}{u_{jj}} - \frac{u_{jp}\eta_p}{u_{jj}u_{pp}}, \qquad j \neq p \tag{A2-21}$$

$$x_p = \frac{\eta_p}{u_{pp}} + \sum_{s=1}^{N}{}' \frac{u_{ps}}{u_{ss}u_{pp}^2}(\eta_p u_{sp} - \eta_s u_{pp}).$$

The prime on Σ indicates that the term $s = p$ (which happens to be zero here) is to be omitted. A somewhat more general result which states that the ith element in the column matrix $(D + E)^{-1}\eta$ is

$$\eta_i d_i^{-1} + \sum_{s=1}^{N} E_{is} d_i^{-2} d_s^{-1}(E_{si}\eta_i - d_i \eta_s) \tag{A2-22}$$

may be derived from (A2-12). Here the elements of the column matrix need not be restricted by $|\eta_j| \ll |\eta_p|, j \neq p$.

References

1. Der Einfluss der Krümmung von rechteckigen Hohlleitern auf das Phasenmass ultrakurzer Wellen, *E. N. T.*, *16*, 73–85 (1939).
2. Research on Wave Guides and Electromagnetic Horns (Report III), *E. T. J.*, *4*, 64–67 (1940).
3. Unpublished work concerned with elbows of small radius of curvature and minimum reflection.

4. Electromagnetic Waves in a Bent Pipe of Rectangular Cross-Section, *Quart. App. Math.*, *1*, 328–333 (1944). See also a note on this paper by S. A. Schelkunoff in the same journal *2*, 171 (1944).
5. Theory of Circular Bends in Rectangular Wave Guides, *Rad. Lab. Report* 43–45, June 24, 1943. An abstract under the number PB2869 appears in *Bibl: Sci. and Indus. Repl.*, Vol. 1, No. 7, p. 253.
6. Electromagnetic Waves, D. Van Nostrand, New York (1943).
7. Steady State Solutions of Transmission Line Equations, *B. S. T. J.*, *20*, 131–178 (1941).
8. H. T. Davis, The Theory of Linear Operators, Principia Press (1936).
9. Frazer, Duncan and Collar, Elementary Matrices, Camb. Univ. Press (1938).
10. K. Knopp, Theory and Application of Infinite Series, Blackie and Son (1928).

The Approximate Solution of Linear Differential Equations

By MARION C. GRAY and S. A. SCHELKUNOFF

Linear differential equations with variable coefficients occur in many fields of applied mathematics: in the theories of acoustics, elastic waves, electromagnetic waves in stratified media, nonuniform transmission lines, wave guides, antennas, wave mechanics. The "Wave Perturbation" method described in greater detail elsewhere[1] is particularly useful in those ranges of the independent variable in which the "WKB Approximation" is not sufficiently accurate. The present paper endeavors to illustrate the remarkable accuracy of this method, particularly when compared with Picard's method.

I. Introduction

In a recent paper[1] the approximate solution of linear differential equations by a wave perturbation method was described. When the method was applied to equations whose exact solutions were known we were greatly impressed by the rapidity of convergence of the successive approximations. Hence the purpose of this note is to present some illustrations in the hope that others may be interested and may find the proposed method an improvement on those now in use.

In essence the wave perturbation method dates back to Liouville[2], but in his mémoires he was interested in a problem of heat conduction involving a non-homogeneous differential equation with homogeneous boundary conditions, whereas we consider a homogeneous equation

$$y'' = F(x)y \tag{1}$$

with non-homogeneous initial conditions

$$y(a) = 1, y'(a) = 0 \tag{2a}$$

or

$$y(a) = 0 \qquad y'(a) = 1, \tag{2b}$$

the solution being desired in an interval $a \leqq x \leqq b$. Since the solution for any assigned initial or boundary conditions can be expressed as a linear combination of the solutions satisfying (2a) and (2b) we have not imposed any real limitation.

II. Theory

Comparison of the wave perturbation method with Picard's method (which is essentially a linear perturbation method) is particularly instructive. It will be recalled that in Picard's formulation the differential equation (1) is replaced by an integral equation

$$y(x) = y(a) + (x - a)y'(a) + \int_a^x F(u)y(u)(x - u)\,du \tag{3}$$

where $y(a)$ and $y'(a)$ are assigned initial values*. Writing

$$L_0(x) = y(a) + (x - a)y'(a), \tag{4}$$

$$L_n(x) = \int_a^x F(u)L_{n-1}(u)(x - u)\,du, \qquad n = 1, 2, 3, \cdots,$$

the series

$$y(x) = L_0(x) + L_1(x) + L_2(x) + \cdots \tag{5}$$

is shown to converge to a solution of the original equation. In practical applications, unfortunately, it is usually found that the successive approximations converge rather slowly unless the interval (a, b) is small.

In the wave perturbation method we first rewrite equation (1) in the form

$$y'' = -\beta^2 y + [\beta^2 + F(x)]y = -\beta^2 y + f(x)y, \tag{6}$$

and instead of the integral equation (3) we use

$$\begin{aligned} y(x) = y(a) \cos \beta(x - a) &+ \frac{1}{\beta}\, y'(a) \sin \beta(x - a) \\ &+ \frac{1}{\beta} \int_a^x f(u)y(u) \sin \beta(x - u)\,du. \end{aligned} \tag{7}$$

The parameter β is arbitrary and might be defined in various ways. We have found it convenient to use the definition

$$\beta^2 = -\frac{1}{b - a} \int_a^b F(x)\,dx, \tag{8}$$

so that if $F(x)$ is negative β is real and our first approximation

$$W_0(x) = y(a) \cos \beta(x - a) + \frac{1}{\beta}\, y'(a) \sin \beta(x - a) \tag{9}$$

is sinusoidal. If F is positive β is imaginary and we start with an exponential approximation. If F changes sign in (a, b) the best procedure is to

* This is not quite the usual form of the integral equation but it is substantially equivalent.

subdivide the interval and obtain separate approximations, though this is not necessary if F is predominantly of one sign throughout (a, b). To (9) we now add the sequence

$$W_n(x) = \frac{1}{\beta} \int_a^x f(u) W_{n-1}(u) \sin \beta(x - u)\, du, \tag{10}$$

and the series

$$y(x) = W_0(x) + W_1(x) + W_2(x) + \cdots \tag{11}$$

is the desired solution.

The flexibility of the wave perturbation method as compared with Picard's linear method lies essentially in the introduction of the variable parameter β. Since we make β depend on the length of the interval (a, b) in which a solution is desired the approximations may be extended over much longer intervals than is feasible in Picard's method. If $F(x)$ is a slowly varying function throughout (a, b), so that $f(x)$ is small, it will be found that the first approximation $W_0(x)$ is good, and the second, $W_0 + W_1$ is generally adequate.

Another choice for β is

$$\beta = \frac{1}{b - a} \int_a^b \sqrt{-F(x)}\, dx. \tag{12}$$

However, the integration in (8) will often be simpler than in (12).

Picard's method is a special case of the wave perturbation method, with $\beta = 0$. In fact, if $F(x)$ changes sign in (a, b), then in some cases β as defined by (8) will reduce to zero.

If $F(x)$ is a rapidly varying function, or if the solution is desired over an infinite interval, it is usually advantageous to transform equation (1) by first introducing a new independent variable

$$\theta = \int^x \sqrt{-F(x)}\, dx, \tag{13}$$

and then removing the first order term in the new equation by an appropriate transformation of the dependent variable.

III. Examples

For our illustrations we have used mainly the simple equation

$$y'' = -xy \tag{14}$$

whose exact solution can be expressed in terms of Bessel functions of order $\pm 1/3$. Since the Bessel functions are oscillatory in nature it might be suggested that comparison with Picard's method is weighted in our favor. This does not seem to be the case, as will be illustrated in example 4 where the exact solution is a monotonically increasing function. It has also been suggested that Picard's solution might be improved by starting from a better initial approximation, say W_0, rather than from the linear approximation L_0, but we have not found any marked improvement in the succeeding approximations (see examples 1 and 2). The various points of interest will be brought out in our examples, with the accompanying figures, which we shall now briefly describe. In each figure the heavy curve is the accurate solution while the approximations are indicated by self-explanatory letters.

Example 1, Fig. 1

$$y'' = -xy, 0 \leqq x \leqq 2$$

$$y(0) = 1, y'(0) = 0$$

Exact solution: $$y(x) = \Gamma(\tfrac{2}{3})3^{-1/3} x^{1/2} J_{-\frac{1}{3}} (\tfrac{2}{3}x^{3/2})$$

(a) Wave perturbation

$$W_0 = \cos x$$

$$W_1 = -\tfrac{1}{4}x \cos x + \tfrac{1}{4}(1 + 2x - x^2) \text{ si}$$

(b) Linear perturbation

$$L_0 = 1$$

$$L_1 = -\frac{x^3}{6}$$

(c) Linear perturbation using initial sinusoidal approximation

$$\bar{L}_0 = \cos x = W_0$$

$$\bar{L}_1 = x + x \cos x - 2 \sin x$$

Example 2, Figs. 2, 3 and 4

$$y'' = -xy, 2 \leqq x \leqq 6$$

$$y(2) = 0, y'(2) = 1$$

Exact solution:

$$y(x) = -.84423x^{1/2} J_{-1/3}(\tfrac{2}{3}x^{3/2}) - .019291x^{1/2} J_{1/3}(\tfrac{2}{3}x^{3/2})$$

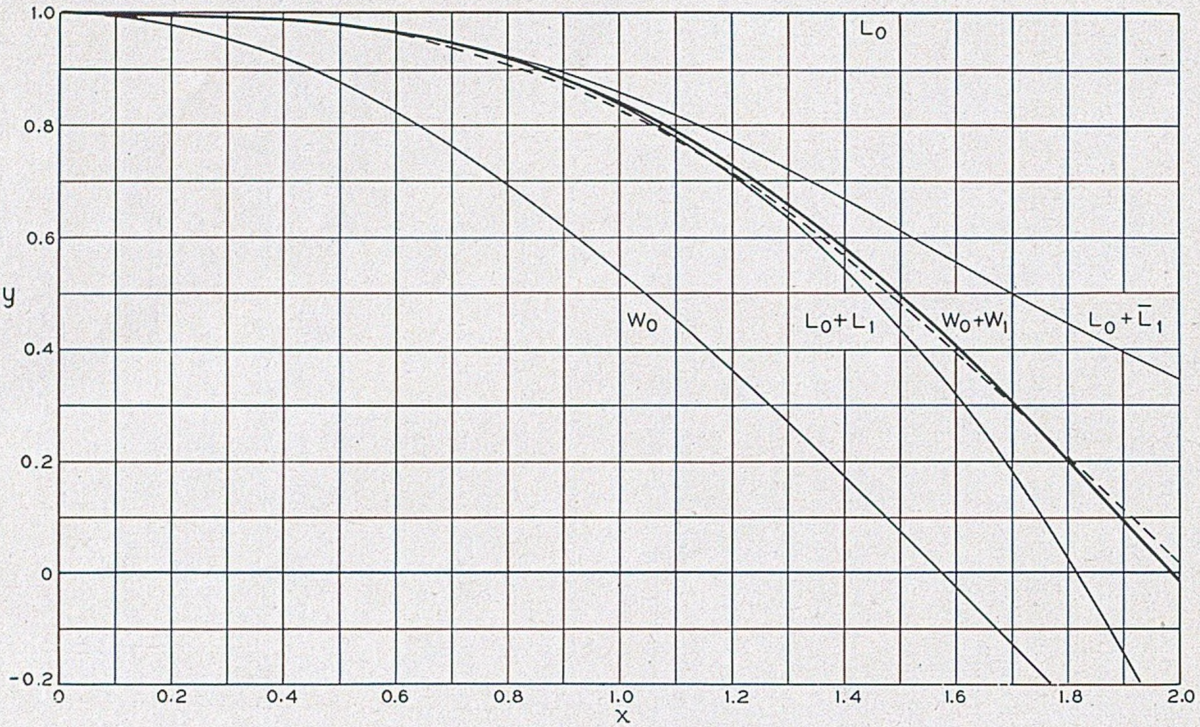

1.0
0.8
0.6
0.4
0.2
0
-0.2
y
0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
2.0
x
L_0
W_0
L_0+L_1
W_0+W_1
$L_0+\bar{L}_1$

Fig. 1

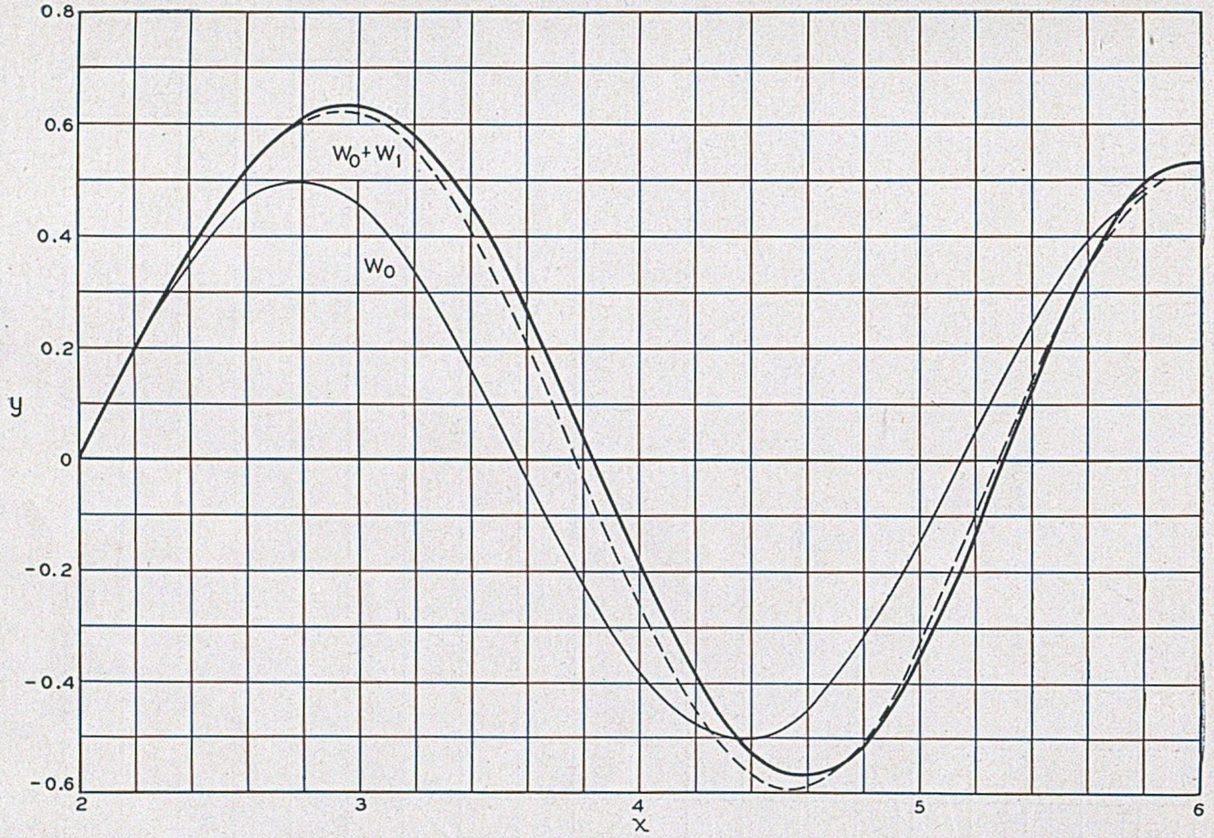


Fig. 2

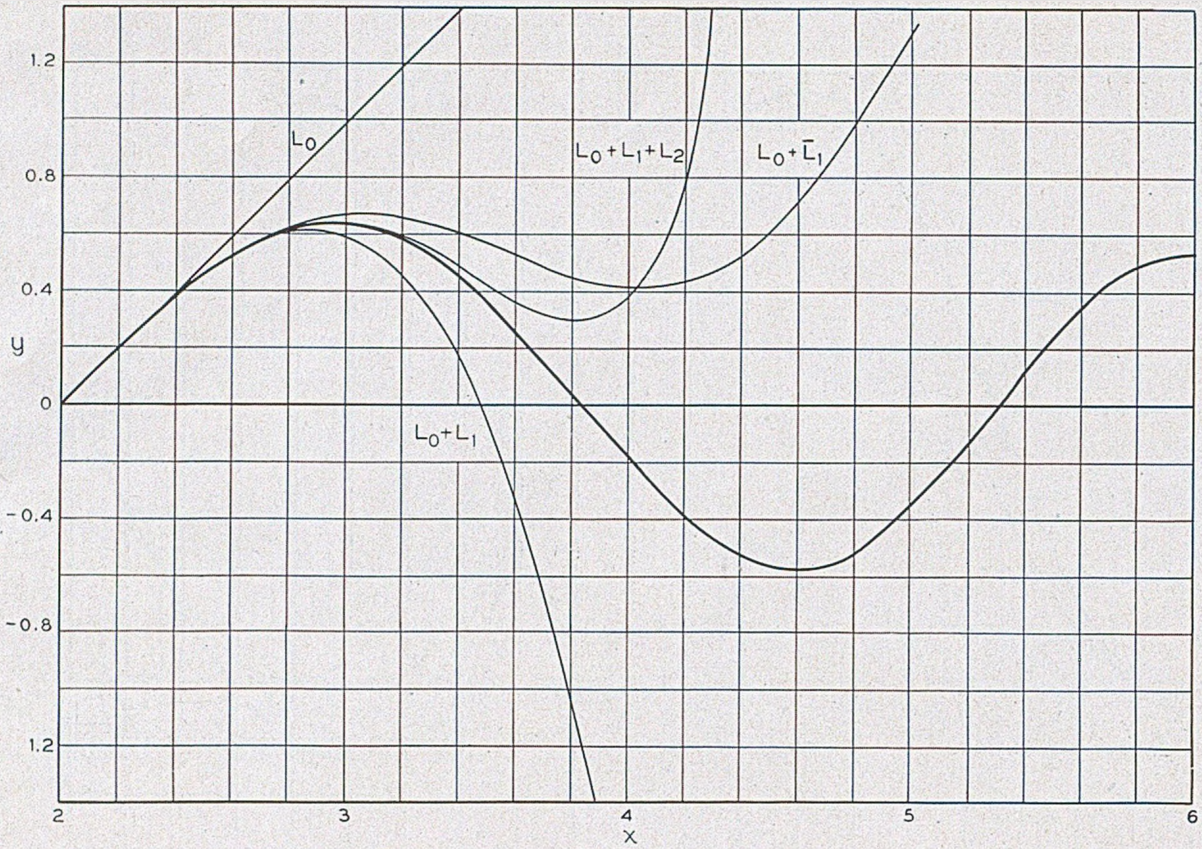


Fig. 3

(a) Wave perturbation, Fig. 2

$$W_0 = \tfrac{1}{2}\sin 2(x-2)$$

$$W_1 = \frac{6-x}{32}\sin 2(x-2) + \frac{12-8x+x^2}{16}\cos 2(x-2).$$

Figure 2 exhibits rapid pulling of the successive approximate waves toward the exact even though the interval has been chosen deliberately unfavorable to the straight wave perturbation method [see example (c) and Fig. 4 for the improved treatment].

(b) Linear perturbation, Fig. 3

$$L_0 = x - 2$$

$$L_1 = \tfrac{1}{12}(16 - 16x + 4x^3 - x^4)$$

$$L_2 = -\frac{16}{63} + \frac{16x}{45} - \frac{2x^3}{9} + \frac{x^4}{9} - \frac{x^6}{90} + \frac{x^7}{504}.$$

Using W_0 instead of L_0

$$\bar{L}_1 = \tfrac{7}{8} - \frac{x}{2} + \frac{x}{8}\sin 2(x-2) + \tfrac{1}{8}\cos 2(x-2).$$

(c) Preliminary transformation of variables, Fig. 4
Introduce $\theta = \frac{2}{3}x^{3/2}$, $y = \theta^{-1/6}v$
and the modified equation is

$$v'' = -\left(1 + \frac{5}{36\theta^2}\right)v \qquad \frac{4\sqrt{2}}{3} \leqq \theta \leqq 4\sqrt{6}.$$

Then, using for simplicity $\beta = 1$

$$v_0 = 2^{-1/2}\theta_0^{1/6}\sin(\theta - \theta_0),\ \theta_0 = 4\sqrt{2}/3$$

or

$$W_0 = (2x)^{-1/4}\sin\tfrac{2}{3}(x^{3/2} - 2^{3/2})$$

It will be seen that W_0 is a very good approximation throughout the range (2, 6). Adding W_1 obtained from

$$v_1 = \frac{5\theta_0^{1/6}}{36\sqrt{2}}[\cos(\theta+\theta_0)(Si\,2\theta - Si\,2\theta_0) - \sin(\theta+\theta_0)(Ci\,2\theta - Ci\,2\theta_0)]$$

the accurate curve y is reproduced.

In Fig. 2 the third approximation could not be distinguished from the accurate curve though numerically the values are not identical. For purposes of comparison the table of numerical values (Table A) may be found interesting.

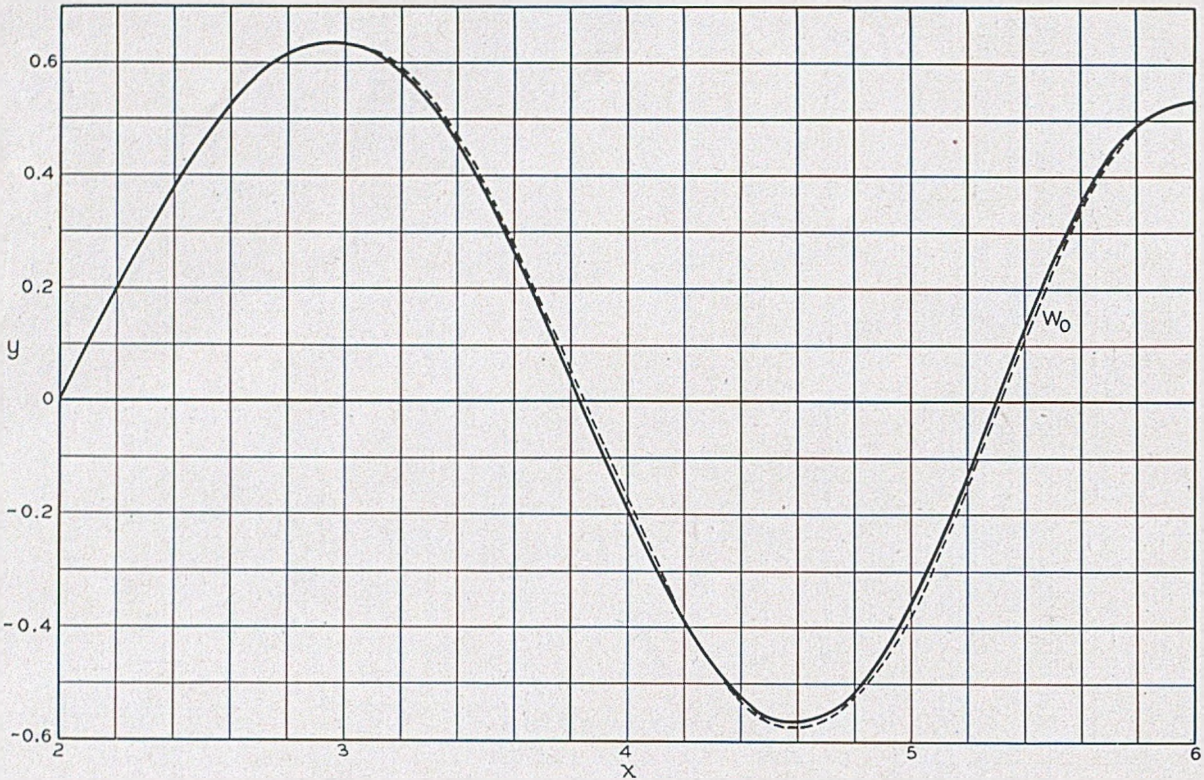


Fig. 4

TABLE A

x	y	W_0	$\sum_0^1 W$	$\sum_0^2 W$	$\sum_0^3 W$
2.0	0.	0.	0.	0.	0.
2.2	.19721	.19471	.19720	.19721	.19721
2.4	.37694	.35868	.37668	.37694	.37694
2.6	.52056	.46602	.51885	.52053	.52056
2.8	.61035	.49979	.60442	.61020	.61034
3.0	.63236	.45465	.61792	.63180	.63232
3.2	.57922	.33773	.55169	.57781	.57918
3.4	.45287	.16749	.40907	.44995	.45726
3.6	.26584	−.02919	.20603	.26085	.26561
3.8	.04126	−.22126	−.02974	.03408	.04087
4.0	−.18921	−.37840	−.26229	−.19800	−.18974
4.2	−.38951	−.47580	−.45326	−.39852	−.39011
4.4	−.52506	−.49808	−.56889	−.53240	−.52557
4.6	−.56943	−.44173	−.58697	−.57328	−.56964
4.8	−.51062	−.31563	−.50217	−.51000	−.51044
5.0	−.35548	−.13971	−.32847	−.35076	−.35494
5.2	−.13068	.05827	−.09772	−.12364	−.13006
5.4	−.12052	.24706	.14547	.12725	.12080
5.6	.34582	.39683	.35200	.34988	.34536
5.8	.49485	.48396	.47807	.49525	.49347
6.0	.53114	.49467	.49467	.52903	.52903

Example 3, Fig. 5

$$y'' = -y + \frac{2}{x^2}y, \quad 1 \leqq x \leqq \infty$$

$$y(1) = 1, y'(1) = 0$$

Exact solution: $y(x) = \sin(x-1) + \dfrac{1}{x}\cos(x-1)$

(a) Wave perturbation, with the initial conditions satisfied exactly

$$W_0 = \cos(x-1)$$

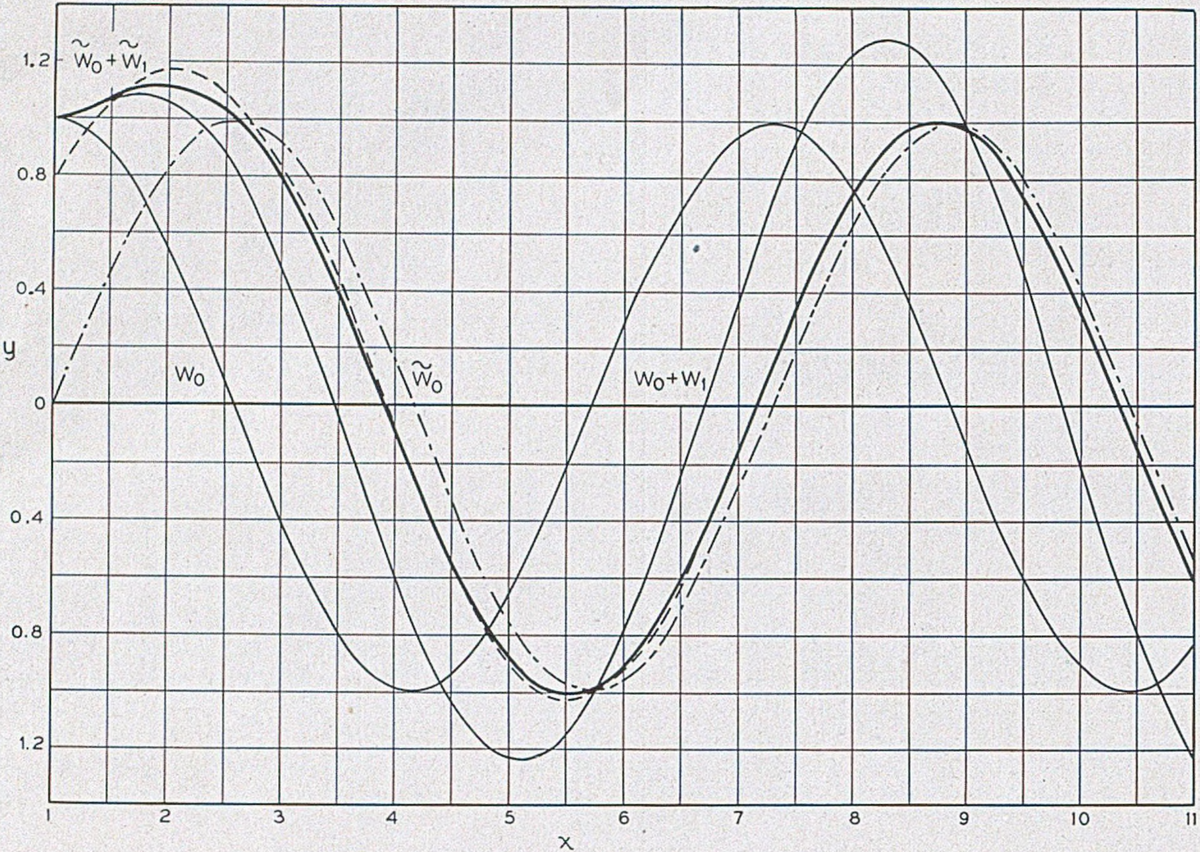
$$W_1 = 2\sin(x-1) - 2\cos(x+1)(Ci\,2x - Ci\,2) - 2\sin(x+1)(Si\,2x - Si\,2)$$

(b) Wave perturbation, matching the exact solution at infinity

$$\tilde{W}_0 = \sin(x-1)$$

$$\tilde{W}_1 = 2\sin(x+1)\,Ci\,2x - 2\cos(x+1)(Si\,2x - \pi/2)$$

This is an example of a solution in an infinite interval, where the perturbation term is not small throughout. It is interesting to note that the second form gives good agreement with the accurate solution in most of the range of integration.



$\tilde{W}_0 + \tilde{W}_1$
W_0
$\tilde{W}_0$
$W_0 + W_1$
y
x
1.2
0.8
0.4
0
0.4
0.8
1.2
1
2
3
4
5
6
7
8
9
10
11

Fig. 5

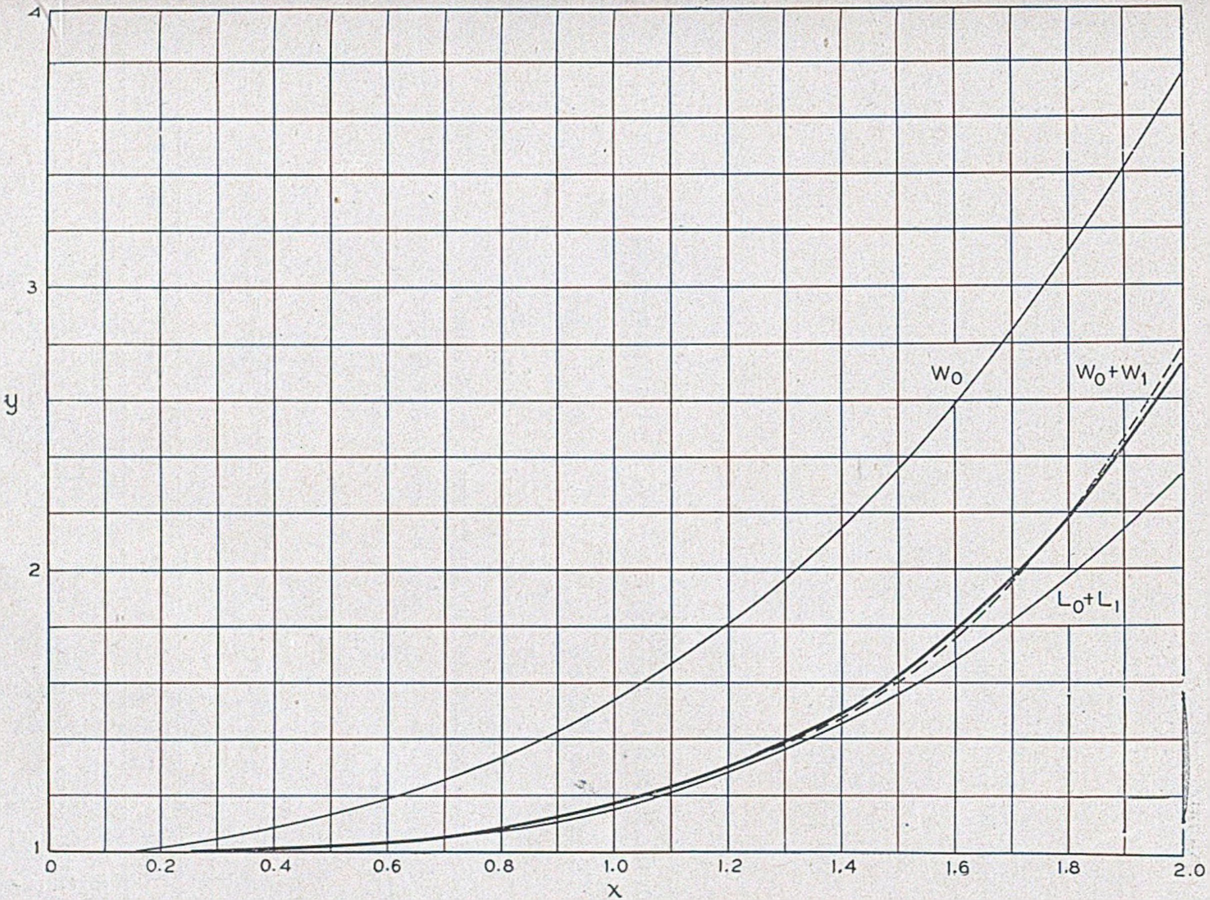


Fig. 6

Example 4, Fig. 6

$$y'' = +xy, \qquad 0 \leqq x \leqq 2$$

$$y(0) = 1, y'(0) = 0$$

Exact solution:

$$y(x) = \Gamma(\tfrac{2}{3})3^{-1/3}x^{1/2}I_{-1/3}(\tfrac{2}{3}x^{3/2}).$$

(a) Wave perturbation

$$W_0 = \cosh x$$

$$W_1 = -\frac{x}{4}\cosh x + \tfrac{1}{4}(x^2 - 2x + 1)\sinh x$$

(b) Linear perturbation

$$L_0 = 1$$

$$L_1 = \frac{x^3}{6}.$$

This is an example in which the exact solution is non-oscillatory yet even in the short interval (0, 2) $W_0 + W_1$ is a better approximation than $L_0 + L_1$.

Example 5, Table I

$$y'' + \frac{1}{x}y' + y = 0, \quad 0 < x \leqq \infty$$

Solution required to match the accurate solution

$$y(x) = J_0(x) - i\,N_0(x)$$

at infinity:

$$\bar{W}_0 = \frac{1+i}{\sqrt{\pi x}}e^{-ix}$$

$$\bar{W}_1 = \frac{1+i}{4\sqrt{\pi x}}e^{ix}\left[Ci2x - i\left(Si2x - \frac{\pi}{2}\right)\right].$$

TABLE I

x	$J_0 - iN_0$	$\bar{W}_0$	$\bar{W}_0 + \bar{W}_1$	$\bar{W}_0 + \bar{W}_1 + \bar{W}_2$
10	$-.2459 - i\,.0557$	$-.2468 - i\,.0526$	$-.2460 - i\,.0557$	
9	$-.0903 - i\,.2499$	$-.0938 - i\,.2489$	$-.0903 - i\,.2500$	
8	$.1717 - i\,.2235$	$.1683 - i\,.2264$	$.1717 - i\,.2235$	
7	$.3001 + i\,.0259$	$.3009 + i\,.0207$	$.3001 + i\,.0260$	
6	$.1506 + i\,.2882$	$.1568 + i\,.2855$	$.1507 + i\,.2883$	
5	$-.1776 + i\,.3085$	$-.1704 + i\,.3135$	$-.1777 + i\,.3086$	
4	$-.3975 + i\,.0169$	$-.3979 + i\,.0291$	$-.3973 + i\,.0169$	
3	$-.2601 - i\,.3769$	$-.2765 - i\,.3684$	$-.2601 - i\,.3772$	
2	$.2239 - i\,.5104$	$.1967 - i\,.5288$	$.2246 - i\,.5109$	
1	$.7652 - i\,.0883$	$.7796 - i\,.1699$	$.7683 - i\,.0860$	$.7651 - i\,.0882$
0.8	$.8463 + i\,.0868$	$.8920 - i\,.0130$	$.8499 + i\,.0916$	$.8461 + i\,.0868$
0.6	$.9120 + i\,.3086$	$1.0124 + i\,.1899$	$.9152 + i\,.3183$	$.9116 + i\,.3084$

For values of x less than 1 $\bar{W}_2$ was evaluated numerically.

Example 6, Table II

$$y'' + \frac{1}{x}y' = \left(\frac{1}{x^2} - 1\right)y, \qquad 1 \leqq x \leqq 3.$$

$$y(1) = 1, \qquad y'(1) = 0.$$

The solution of this equation using Picard's method and the integraph has been described by Thornton C. Fry.[3] We compare his results with those obtained by the wave perturbation method. The equation is first reduced to normal form by the substitution $y = x^{-1/2} u$, so that

$$u'' = \left(-1 + \frac{3}{4x^2}\right)u$$

and we have $\beta = \frac{1}{2}\sqrt{3}$. Then

$$x^{1/2} W_0 = \cos \beta\,(x - 1) + \frac{1}{2\beta} \sin \beta(x - 1)$$

$$x^{1/2} W_1 = \frac{1}{4\beta}\int_1^x \left(\frac{3}{u^2} - 1\right)\left[\cos \beta(u - 1) + \frac{1}{2\beta}\sin \beta(u - 1)\right] \sin \beta(x - u)\, du.$$

While W_1 may be evaluated in terms of Ci and Si functions the values tabulated below were obtained by numerical integration. The values of the accurate solution

$$y = 1.4034\, J_1(x) - 0.3251\, N_1(x),$$

and of the third and eighth Picard approximations, are copied from Fry's paper.

TABLE II

x	y	y_3	y_8	W_0	$W_0 + W_1$
1.0	1.000	1.000	1.000	1.000	1.000
1.2	.998	.998	.998	.990	.998
1.4	.985	.984	.986	.961	.985
1.6	.956	.951	.955	.913	.956
1.8	.908	.894	.910	.848	.908
2.0	.842	.809	.844	.769	.842
2.2	.759	.694	.760	.677	.758
2.4	.659	.548	.661	.575	.659
2.6	.547	.370	.549	.466	.547
2.8	.425	.156	.427	.352	.427
3.0	.297	−.096	.300	.236	.300

References

1. S. A. Schelkunoff, Solution of linear and slightly nonlinear equations, *Quart. App. Math.*, vol. 3, p. 348, Jan., 1946.
2. J. Liouville, Mémoires sur le développement des fonctions ou parties de fonctions en séries dont les divers termes sont assujétis à satisfaire à une même équation différentielle du second ordre, contenant un paramètre variable, *Jl. de Math: Pures et Appl.*, v. *1*, p. 253–265, 1836, v. *2*, p. 16–35 and p. 418–436, 1837. A brief discussion of the method will be found in "Numerical studies in differential equations" by H. Levy and E. A. Baggott, London, 1934; but these authors apply it to the numerical solution of non-linear equations and do not seem to have appreciated that its real potentialities lie in the field of linear equations, where they use only the better known methods.
3. Thornton C. Fry, The use of the integraph in the practical solution of differential equations by Picard's method of successive approximations, *Proc. Int. Math. Congress of Toronto*, pp. 405–428, 1924.

Potential Coefficients for Ground Return Circuits

By W. HOWARD WISE

This paper is concerned with the effect of the finite conductivity and dielectric constant of the earth on the potential coefficient for a 1-wire ground return circuit. It has been customary to say that the potential coefficient V/Q is

$$p_{12} = c^2 2 \log \rho''/\rho', \text{ elm units per cm.}$$

It is generally realized of course that this is just a good approximation to the true p_{12}. To see that it is just an approximation one has only to imagine the earth turning into air, in which case the distance ρ'' will eventually cease to have significance. The object of this paper is to derive the complete expression for p_{12}.

It turns out that

$$p_{12} = c^2[2 \log \rho''/\rho' + 4(M + iN)] \tag{7}$$

where

$$M + iN = \int_0^\infty \frac{e^{-(h+z)\zeta\sqrt{\alpha}\,t} \cos y\zeta\sqrt{\alpha}t}{\sqrt{t^2 + ie^{i2\eta}} + (\epsilon - i2c\lambda\sigma)t}\,dt \tag{8}$$

$\alpha = 4\pi\sigma\omega$, as in Carson's work on Z_{12} and in mine on Z_{12} at high frequencies[1,2]

$$\zeta e^{i\eta} = \sqrt{1 + i(\epsilon - 1)/2c\lambda\sigma} = s$$

ϵ = dielectric constant in electrostatic units
σ = conductivity in electromagnetic units
= 10^{-13} to 10^{-14} in ordinary soil
λ = wavelength in centimeters
c = velocity of light, in cm per sec.

$$M + iN \text{ vanishes as } f \to 0, f \to \infty, \epsilon \to \infty \text{ or } \sigma \to \infty.$$

Ordinarily $4(M + iN)$ will not be an important correction to $2 \log \rho''/\rho'$; but if the frequency is high and h or z is small it can be a worthwhile correction. For example, if a .02535 inch wire be thrown out on the ground to be a 2 mc antenna and we assume that $\sigma = 10^{-13}$, $\epsilon = 15$ and $h = 3$ cm. then, with a for wire radius,

$$\begin{aligned} p_{11} &= c^2[2 \log 2h/a + 4(M + iN)] \\ &= c^2[10.455 + .152 + i\,.319]. \end{aligned}$$

$1/p_{11}$ is the capacity to ground. If there were two parallel wires the scalar potential at the first wire would be $V_1 = p_{11}Q_1 + P_{12}Q_2$.

Derivation of the Formula

We begin with the wave-function for an exponentially propagated current in a straight wire parallel to a flat earth. The wave-function for a horizontal current-element dipole has been formulated as an infinite

[1] John R. Carson: "Wave Propagation in Overhead Wires with Ground Return," *Bell Sys. Tech. Jour.* 5, pp. 539–554, 1926.
[2] W. Howard Wise: "Propagation of High-Frequency Currents in Ground Return Circuits," *Proc. I. R. E.* 22, pp. 522–527, 1934.

integral by H. von Hoerschelmann[3]. The wave-function for the current in the wire is obtained by integrating the wave-functions of the current-element dipoles along the wire from minus infinity to plus infinity. It is

$$\Pi = a \times e^{-\gamma x} \int_{-\infty}^{\infty} I_0 e^{-\gamma x} \left(\frac{e^{-ikR_1}}{R_1} - \frac{e^{-ikR_2}}{R_2} \right.$$
$$\left. + \int_0^{\infty} \frac{2J_0(\nu\rho)}{l + m} e^{-wl} \nu \cdot d\nu \right) dx + b \times 0 \qquad (1)$$
$$- c \times 2e^{-\gamma x} \int_{-\infty}^{\infty} I_0 e^{-\gamma x} \frac{\partial}{\partial x} \int_0^{\infty} \frac{(1 - \tau^2) J_0(\nu\rho)\nu}{(l + m)(l + \tau^2 m)} e^{-wl} d\nu \cdot dx.$$

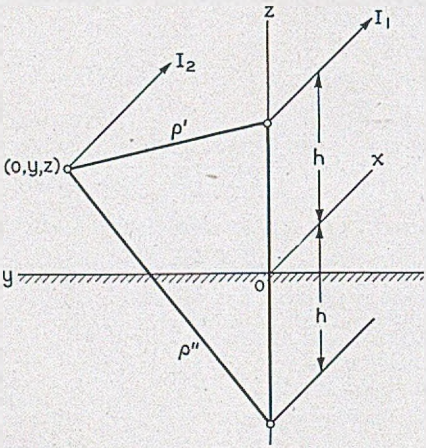


Fig. 1

The time factor is $e^{i\omega t}$. a, b and c are unit vectors pointing in the x, y and z directions.

$$\begin{aligned} R_1 &= (x^2 + y^2 + (h - z)^2)^{1/2} \\ R_2 &= (x^2 + y^2 + (h + z)^2)^{1/2} \\ \rho &= (x^2 + y^2)^{1/2} \\ k &= 2\pi/\lambda \\ k_2^2 &= \epsilon\mu\omega^2 - i4\pi\sigma\mu\omega \text{ in electromagnetic units} \end{aligned}$$

By supposing ϵ to be measured in electrostatic units we can write

$$k_2^2 = k^2(\epsilon - i2c\lambda\sigma)\mu.$$

[3] H. Von Hoerschelmann, Jahrb. der draht. Teleg. 5, pp. 14–188, 1912.

It is assumed that μ is everywhere unity in electromagnetic units.

$$l = (\nu^2 - k^2)^{1/2}, \qquad m = (\nu^2 - k_2^2)^{1/2}$$
$$w = h + z, \qquad \tau^2 = k^2/k_2^2$$
$$\gamma = \alpha + i\beta \text{ is the desired propagation constant}$$

The electric field parallel to the wire is

$$E_x = -i\omega\left[\Pi_x + k^{-2}\frac{\partial}{\partial x}\left(\frac{\partial}{\partial x}\Pi_x + \frac{\partial}{\partial y}\Pi_y + \frac{\partial}{\partial z}\Pi_z\right)\right]$$
$$= -I_0\, e^{-\gamma x}\, Z_{12} - \frac{\partial V}{\partial x} \tag{2}$$

It has previously been shown that[2]

$$Z_{12} = i\omega[2 \log \rho''/\rho' + 4(Q - iP)] \tag{3}$$

where

$$Q - iP = \frac{1}{is^2}\int_0^\infty (\sqrt{\nu^2 + is^2} - \nu)e^{-w'\nu} \cos y'\nu \cdot d\nu,$$

$$w' = w\sqrt{\alpha} \text{ and } y' = y\sqrt{\alpha}.$$

To get the potential coefficient for a ground return circuit it is necessary to compute the scalar potential.

$$V = i\omega k^{-2}\left(\frac{\partial}{\partial x}\Pi_x + \frac{\partial}{\partial y}\Pi y + \frac{\partial}{\partial z}\Pi_z\right)$$
$$= Q\, p_{12} \tag{4}$$

As in previous work the propagation constant γ is assigned the value ik as a first approximation. This is an ideal value for γ but the following considerations make it an imperative choice: (1) to assume that the current is propagated down the line with a velocity less than that of light makes the integrals very hard to evaluate, (2) to assume that the attenuation is not zero on an infinite line amounts to assuming an infinite source of energy and makes the integrals diverge.

It should not be inferred that the resulting formulas are necessarily poor if the physical system does not closely approximate the ideal one in which γ is ik. ik is employed as a convenient first approximation in evaluating the correction terms in Z_{12} and p_{12}. Eventually, if there were but one wire, one would compute $\gamma = \sqrt{(z + Z_{11})(G + i\omega/p_{11})}$, wherein Z_{11} and p_{11} have been evaluated with ik for γ, and this would be a second approximation to γ. Past experience with the second approximation so obtained has justified the expectation that it would be a satisfactory final result. Since the integrals diverge if the attenuation is not zero the use of an infinite line formula presupposes reasonably efficient transmission.

Since $\Pi \propto e^{-\gamma x}$ we have

$$i\omega k^{-2}\frac{\partial}{\partial x}\Pi_x = \frac{\omega}{k}\Pi_x = \frac{I}{ik}Z_{12} = \frac{cI}{i\omega}Z_{12}\,.$$

Since $-\dfrac{\partial I}{\partial x} = \dfrac{\partial Q}{\partial t}$ or $ikI = i\omega Q$ or $I = cQ$ this is

$$i\omega k^{-2}\frac{\partial}{\partial x}\Pi_x = Qc^2[2\log \rho''/\rho' + 4(Q - iP)]. \tag{5}$$

We have next to consider

$$\frac{i\omega}{k^2}\frac{\partial}{\partial z}\Pi_z = \frac{2\omega(1-\tau^2)}{ik^2}I\frac{\partial}{\partial z}\int_{-\infty}^{\infty}e^{-\gamma x}\frac{\partial}{\partial x}\int_0^{\infty}\frac{J_0(\nu\rho)e^{-wl}\,\nu\cdot d\nu}{(l+m)(l+\tau^2 m)}\,dx$$

$$= Qc^2\,\frac{2(1-\tau^2)}{ik}\frac{\partial}{\partial z}\int_0^{\infty}\frac{e^{-wl}\,\nu\cdot d\nu}{(l+m)(l+\tau^2 m)}\int_{-\infty}^{\infty}e^{-\gamma x}\frac{\partial}{\partial x}J_0(\nu\rho)\,dx.$$

The infinite integral is

$$\int_0^{\infty}\frac{e^{-wl}\,\nu\cdot d\nu}{(l+m)(l+\tau^2 m)}\left[e^{-\gamma x}J_0(\nu\rho)\Big|_{-\infty}^{\infty} + \gamma\int_{-\infty}^{\infty}e^{-\gamma x}J_0(\nu\rho)\,dx\right].$$

Since $J_0(\nu\sqrt{x^2+y^2})$ and $\cos kx$ are even functions of x and $\sin kx$ is an odd function of x

$$\int_{-\infty}^{\infty}J_0(\nu\sqrt{x^2+y^2})e^{-ikx}\,dx = 2\int_0^{\infty}J_0(\nu\sqrt{x^2+y^2})\cos kx\cdot dx$$

$$= 0 \quad \text{if} \quad \nu < k$$

$$= 2\,\frac{\cos y\sqrt{\nu^2-k^2}}{\sqrt{\nu^2-k^2}} \quad \text{if} \quad k \leq \nu$$

and so our integral is

$$2ik\int_k^{\infty}\frac{e^{-wl}\,\nu}{(l+m)(l+\tau^2 m)}\cdot\frac{\cos yl}{l}\,d\nu$$

or, since $l^2 = \nu^2 - k^2$,

$$2ik\int_0^{\infty}\frac{e^{-wl}\cos yl\cdot dl}{(l+\sqrt{l^2+i^2(k_2^2-k^2)})(l+\tau^2\sqrt{l^2+i^2(k_2^2-k^2)})}$$

or, if we put $l = \nu\sqrt{\alpha}$

and $i(k_2^2 - k^2) = 4\pi\sigma\omega\left(1 + i\dfrac{\epsilon - 1}{2c\lambda\sigma}\right) = \alpha s^2,$

$$\frac{2ik}{\sqrt{\alpha}}\int_0^\infty \frac{e^{-w'\nu}\cos y'\nu\cdot d\nu}{(\nu + \sqrt{\nu^2 + is^2})(\nu + \tau^2\sqrt{\nu^2 + is^2})}$$

where, as in $Q - iP$, $w' = w\sqrt{\alpha}$ and $y' = y\sqrt{\alpha}$.

Noting next that $\dfrac{\partial}{\partial z}e^{-w'\nu} = -\sqrt{\alpha}\nu e^{-w'\nu}$

we have

$$\frac{i\omega}{k^2}\frac{\partial}{\partial z}\Pi_z = -Qc^2 4\int_0^\infty \frac{(1 - \tau^2)e^{-w'\nu}\cos y'\nu\cdot\nu\cdot d\nu}{(\nu + \sqrt{\nu^2 + is^2})(\nu + \tau^2\sqrt{\nu^2 + is^2})}$$

$$= -Qc^2\frac{4}{is^2}\int_0^\infty \frac{\sqrt{\nu^2 + is^2} - \nu}{\nu + \tau^2\sqrt{\nu^2 + is^2}}e^{-w'\nu}\cos y'\nu\cdot(1 - \tau^2)\nu\cdot d\nu.$$

Since $(1 - \tau^2)\nu = \nu + \tau^2\sqrt{\nu^2 + is^2} - \tau^2(\sqrt{\nu^2 + is^2} + \nu)$
this is

$$-Qc^2\frac{4}{is^2}\int_0^\infty\left[\sqrt{\nu^2 + is^2} - \nu - \frac{\tau^2 is^2}{\nu + \tau^2\sqrt{\nu^2 + is^2}}\right]e^{-w'\nu}\cos y'\nu\cdot d\nu$$

$$= Qc^2\left[-4(Q - iP) + 4\int_0^\infty \frac{e^{-w'\nu}\cos y'\nu\cdot d\nu}{\sqrt{\nu^2 + is^2} + \nu/\tau^2}\right]. \tag{6}$$

On adding (6) to (5) we have

$$Qp_{12} = Qc^2[2\log\rho''/\rho' + 4(M + iN)] \tag{7}$$

where

$$M + iN = \int_0^\infty \frac{e^{-w'\nu}\cos y'\nu}{\sqrt{\nu^2 + is^2} + \nu/\tau^2}d\nu$$

$$= \int_0^\infty \frac{e^{-(h+z)\zeta\sqrt{\alpha}t}\cos y\sqrt{\alpha}\zeta t}{\sqrt{t^2 + ie^{i2\eta}} + t/\tau^2}dt. \tag{8}$$

$M + iN$ vanishes as $f \to 0, f \to \infty, \epsilon \to \infty$ or $\sigma \to \infty$.

When $k_2^2 - k^2$ is minute the leading terms in the approximation (9)

for $M + iN$ are $\dfrac{\pi}{8} - \dfrac{C}{2} - \dfrac{1}{2}\log\left(\rho''\sqrt{(k_2^2 - k^2)/2}\right) - i\dfrac{\pi}{4}$.

An Approximation for $M + iN$

It is possible to get series expansions for $M + iN$ but those which have been obtained do not facilitate computation. A fairly good approximation to $M + iN$ is arrived at as follows.

Let $ie^{i2\eta} = u^2, \qquad u = e^{i(\eta+\pi/4)}$

$$\epsilon - i2c\lambda\sigma = a = 1/\tau^2$$

$$e^{-(h+z)\zeta\sqrt{\alpha}t} \cos y\zeta\sqrt{\alpha}t = \tfrac{1}{2}(e^{-g't} + e^{-g''t})$$

$$g' = (h + z - iy)\zeta\sqrt{\alpha}$$

$$g'' = (h + z + iy)\zeta\sqrt{\alpha}$$

$$g = (h + z)\zeta\sqrt{\alpha}$$

Since $(t^2 + u^2)^{1/2} = (t^2 + 2tu + u^2 - 2tu)^{1/2}$

$$= t + u - tu/(t + u) + \ldots$$

we put

$$1/(\sqrt{t^2 + u^2} + at) = 1/(t + u - tu/(t + u) + at)$$

$$= (t + u)/[(a + 1)t^2 + (a + 1)tu + u^2]$$

$$= (t + r_1 + r_2)/(a + 1)(t + r_1)(t + r_2)$$

where $r_1 = u(1 - \sqrt{1 - 4/(a + 1)})/2.$

and $\quad r_2 = u(1 + \sqrt{1 - 4/(a + 1)})/2.$

Then

$$M + iN \approx \frac{1}{(a + 1)(r_2 - r_1)} \int_0^\infty \left(\frac{r_2}{t + r_1} - \frac{r_1}{t + r_2}\right) \frac{e^{-g't} + e^{-g''t}}{2} dt$$

$$= \frac{1}{2(a + 1)(r_2 - r_1)} [-r_2 e^{g'r_1} \operatorname{li}(e^{-g'r_1}) + r_1 e^{-g'r_2} \operatorname{li}(e^{-g'r_2})$$

$$- r_2 e^{g''r_1} \operatorname{li}(e^{-g''r_1}) + r_1 e^{g''r_2} \operatorname{li}(e^{-g''r_2})]. \qquad (9)$$

When $y = 0$ this reduces to

$$M + iN \approx \frac{1}{(a + 1)(r_2 - r_1)} [-r_2 e^{gr_1} \operatorname{li}(e^{-gr_1}) + r_1 e^{gr_2} \operatorname{li}(e^{-gr_2})]. \qquad (10)$$

$$\operatorname{li}(e^{-z}) = -\int_z^\infty \frac{e^{-t}}{t} dt = C + \log z + \sum_{\iota=1}^\infty \frac{(-z)^\iota}{\iota!\iota},$$

where $C = .577215665$ and, if $z = re^{i\theta}, \quad -\pi \leq \theta \leq \pi$.

$$\operatorname{li}(e^{-z}) \sim \frac{e^{-z}}{-z} \sum_{\iota=0}^{n} \frac{\iota!}{(-z)^\iota} + R_n.$$

The accompanying charts give M and N for $y = 0$ and $\epsilon = 15$. With $z = h$ they give M and N for p_{11}. The computed points are indicated by solid dots on the chart for M; they were obtained by numerical integration.

The approximation (10) was checked against the values obtained by numerical integration at a number of points. The discrepancy in each case amounted to less than one per cent for both M and N. This approximation is a much easier way to evaluate the integral than is numerical integration but it is a tedious computation with many chances for error. Conse-

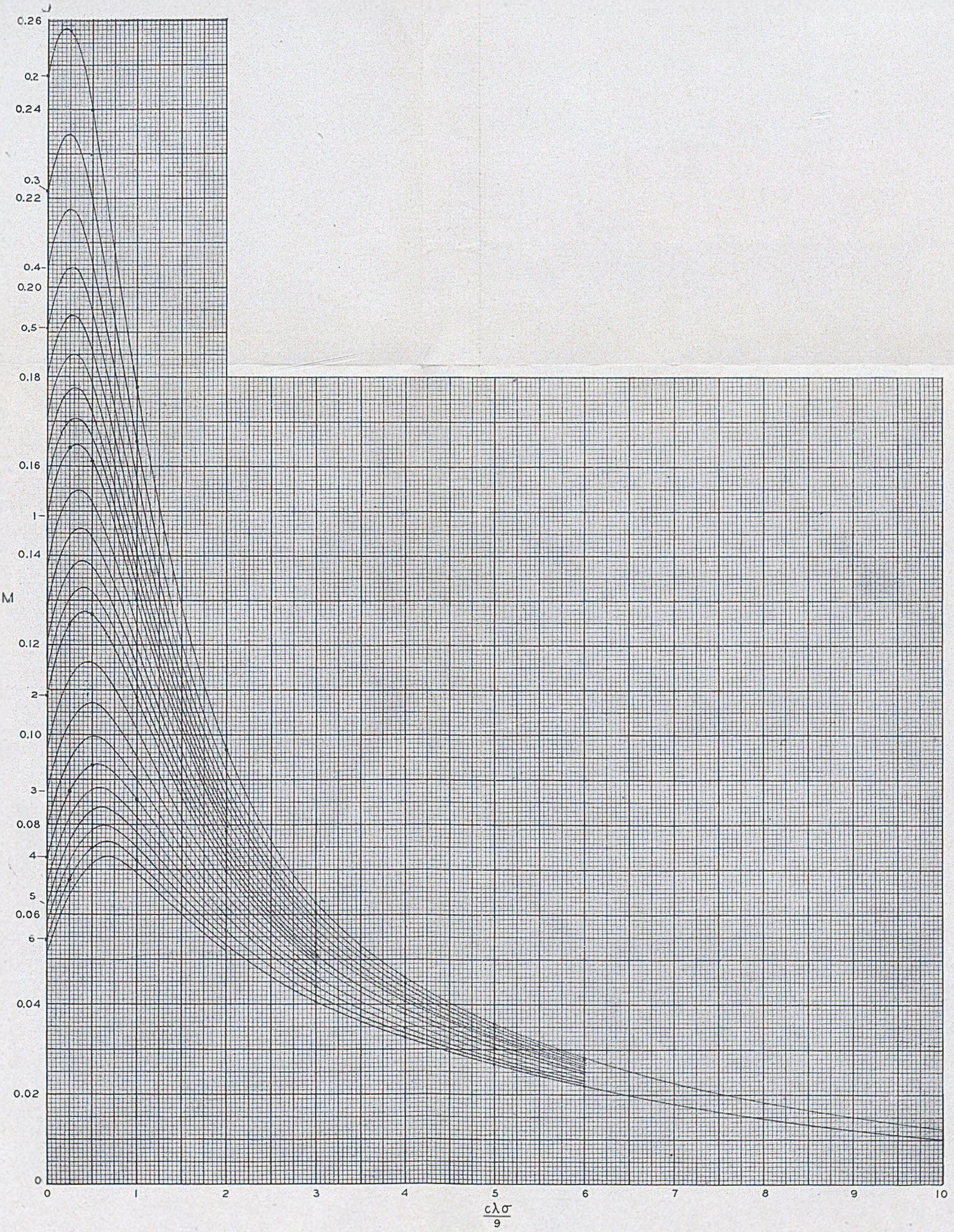


Fig. 2—M, computed with $\epsilon = 15$ and $y = 0$. The number associated with a curve is the value of $(h + z)\zeta\sqrt{a}$.

quently it is important to observe that a coarser kind of approximation may often be good enough. Thus, taking y to be zero,

$$M + iN \approx \int_0^\infty \frac{e^{-gt}}{u + at}\,dt = -\frac{1}{a}\,e^{gu/a}\,\mathrm{li}\,(e^{-gu/a}).$$

If g is very small one might use

$$\begin{aligned} M + iN &\approx \int_0^1 \frac{dt}{u + at} + \int_1^\infty \frac{e^{-gt}}{(1 + a)t}\,dt \\ &= \frac{1}{a}\log\left(1 + \frac{a}{u}\right) - \frac{1}{1 + a}\,\mathrm{li}\,(e^{-g}). \end{aligned}$$

Ordinarily precision is not required in $M + iN$ because $4(M + iN)$ is a small term in p_{12}.

Abstracts of Technical Articles by Bell System Authors

Electrochemical Factors in Underground Corrosion of Lead Cable Sheath.[1] V. J. Albano. Stray current is the principle cause of corrosion failures on underground telephone cables in most cities where trolleys are operated. To mitigate this condition, the cable sheaths are "drained" to the negative return system of the traction system. By this means not only is the stray current anodic area largely eliminated, but the cables automatically become negative to earth, and therefore are cathodically protected. The cathodic protection afforded in this manner prevents other types of corrosion from occurring. With the gradual abandonment of trolley systems, and with the extension of underground cables into non-trolley areas, the percentage of underground telephone plant receiving this protection is decreasing. As a result, the problems of lead corrosion due to such causes as galvanic and local cell action of various types, and chemical action by substances in the soil are becoming more prevalent. It is the purpose of this article to review some of the basic principles of corrosion not involving stray currents, and show how they apply to the problems of lead cable sheath corrosion.

PCM Equipment.[2] H. S. Black and J. O. Edson. PCM, pulse code modulation, is a new solution to the problem of overcrowded frequency spectrum. It appears to have exceptional possibilities from the standpoint of freedom from interference, and seems to have inherent advantages over other types of multiplexing.

Coaxial-Cable Networks.[3] Frank A. Cowan. This paper discusses the general features of the coaxial system, its application for both telephone and television, and the future prospects for very-broadband transmission facilities in the communication network.

Parabolic-Antenna Design for Microwaves.[4] C. C. Cutler. This paper is intended to give fundamental relations and design criteria for parabolic radiators at microwave frequencies (i.e., wavelengths between 1 and 10 centimeters). The first part of the paper discusses the properties of the parabola which make it useful as a directional antenna, and the relation of phase polarization and amplitude of primary illumination to the over-all radiation characteristics. In the second part, the characteristics of practical feed systems for parabolic antennas are discussed.

[1] *Corrosion*, October 1947.
[2] *Electrical Engineering*, November 1947.
[3] *Proc. I. R. E.—Waves and Electrons Section*, November 1947.
[4] *Proc. I. R. E.*, November 1947.

Microwave Antenna Measurements.[5] C. C. CUTLER, A. P. KING and W. E. KOCK. A description is given of the techniques involved in measuring the properties of microwave antennas. The measuring methods which are peculiar to these frequencies are discussed, and include the measurement of gain, beam width, minor lobes, wide-angle radiation, mutual coupling between antennas, phase, and polarization. The requirements of the antenna testing site are taken up, and components of a complete measuring system are briefly described.

Microwave Converters.[6] C. F. EDWARDS. Microwave converters using point-contact silicon rectifiers as the nonlinear element are discussed, with particular emphasis on the design of the networks connecting the rectifier to the input and output terminals. Several converters which have been developed during recent years for use at wavelengths between 1 and 30 centimeters are described, and some of the effects of the impedance-versus-frequency characteristics of the networks on the converter performance are discussed.

Recent Developments in Relays:[7] *Glass-Enclosed Reed Relay,* W. B. ELLWOOD; *Mercury Contact Relays,* J. T. L. BROWN and C. E. POLLARD. Relays which combine high-speed and great uniformity of performance over long periods of time are required for some uses in the telephone plant. The relays described possess these qualities to an unusual degree. Detailed description is limited to two types, each typical of a generic family in which the principles involved apply to all.

These relays are based on the philosophy that a motor element (any device for conversion of electromagnetic to mechanical energy), which is efficient and magnetically and elastically stable and operates contacts sealed in a proper atmosphere free from dirt and film, will give reliable performance if the contact load is engineered to the capacity of the contact. The relays require no maintenance beyond unit replacement, for there is no possibility of a change in adjustment after assembly is completed.

In one form the contact is provided for by metal in solid form, while in the other a mercury film supported on solid metal surfaces provides the contacting medium. The mercury at the contacting surfaces is replenished continuously through a capillary path from a mercury reservoir below the contact.

An Adjustable Wave-Guide Phase Changer.[8] A. GARDNER FOX. A very interesting and useful component of the wave-guide art is the differential phase-shift section, wherein dominant waves of one polarization are caused to travel through a section of wave guide at a different velocity than waves

[5] *Proc. I. R. E.*, December 1947.
[6] *Proc. I. R. E.*, November 1947.
[7] *Elec. Engg.*, November 1947.
[8] *Proc. I. R. E.*, December 1947.

polarized at right angles to the first. Particularly useful are the Δ90-degree and Δ180-degree differential phase-shift sections which produce differential delays between the two polarizations of 90 degrees and 180 degrees, respectively. The properties of these sections are discussed, and it is shown how they may be combined to form a phase changer which will transmit substantially 100 per cent of the incident power with a phase which is readily adjustable. Several different methods of building these sections are finally described.

Considerations in the Design of a Radar Intermediate-Frequency Amplifier.[9] ANDREW L. HOPPER and STEWART E. MILLER. The intermediate-frequency amplifier of a microwave radar receiver is commonly required to provide approximately 100 decibels amplification in a bandwidth of 1 to 10 megacycles, centered at frequencies in the 30- and 60-megacycle regions. Meeting such requirements involves the use of five to ten amplifier stages of the highest efficiency that can be suited to production methods. In addition, the noise figure of the radar intermediate-frequency amplifier is a significant contributor to the over-all radar receiver noise figure, and must therefore be maintained at an absolute minimum. By examining a particular intermediate-frequency-amplifier design (one providing an over-all bandwidth of 10 megacycles centered at 60 or 100 megacycles), this paper discusses qualitatively the theoretical problems involved in such a design and gives data of practical importance to the engineer attempting to build a similar amplifier. Measured characteristics of approximately fifty amplifiers are summarized to illustrate the end results achieved.

Historical Note on the Rate of a Moving Atomic Clock.[10] HERBERT E. IVES. The history of the idea of variation of frequency with velocity is followed through Goigt, Larmor, Lorentz, and Einstein. The Michelson-Morley experiment is explainable by any contraction of dimensions in the ratio $(1 - v^2/c^2)^{\frac{1}{2}}:1$ along and transverse to the direction of motion. To each contraction corresponds a different value of frequency change. The theoretical speculations pointing to the relation $v_m = v_0(1 - v^2/c^2)^{\frac{1}{2}}$ are discussed, together with the significance of the experimental test by means of canal rays.

New Low-Coefficient Synthetic Piezoelectric Crystals for Use in Filters and Oscillators.[11] W. P. MASON. Two crystals of the monoclinic sphenoidal class have been found which have modes of vibration with zero temperature coefficients of frequency, high electromechanical coupling constants, and high Q's or low dissipation. These properties make it appear probable that such crystals may have a considerable use in filters and oscillators as a sub-

[9] *Proc. I. R. E.*, November 1947.
[10] *Jour. Opt. Soc. Amer.*, October 1947.
[11] *Proc. I. R. E.*, October 1947.

stitute for quartz, which is difficult to obtain in large sizes. These crystals are ethylene diamine tartrate (EDT) having the chemical formula $C_6H_{14}N_2O_6$, and di-potassium tartrate (DKT) having the formula $K_2C_4H_4O_6 - \frac{1}{2} H_2O$.

The paper describes the properties of EDT, since this crystal has been found more advantageous than DKT. The 13 elastic constants, the 8 piezoelectric constants, and the 4 dielectric constants have been measured over a temperature range, and from these measurements the regions of low temperature coefficients and high electromechanical coupling have been located. Six low-temperature-coefficient cuts have been discovered and the properties of these cuts are given. These cuts are being applied in the crystal channel filters of the long-distance telephone system, and may be applied to the control of oscillators.

Multi-Channel Carrier Telegraph.[12] A. L. Matte. Discussion of a carrier telegraph system, adapted specifically to railway requirements, to meet the needs for high-quality line transmission.

Reflex Oscillators for Radar System.[13] J. O. McNally and W. G. Shepherd. The advantages to be gained in the operation of radar systems at very high frequencies have led to the use of frequencies of several thousand megacycles. Operation at these frequencies has imposed serious problems in obtaining suitable tube behavior. Because of the difficulty in obtaining amplification at the transmission frequency, the r.f. section of the usual radar receiver consists of a crystal converter driven by a beating oscillator and operating directly into an i.f. amplifier. Since the midband frequency of the latter has commonly been either 30 or 60 Mc., it has been necessary to provide beating oscillators operating at frequencies differing from those of the transmitter by only a few per cent.

For radar systems intended to operate at approximately 3000 Mc., which were under development in the early days of the war, it was found that triodes then available gave unsatisfactory performance. Attention shifted to the possibility of using velocity-modulated tubes, and the particular form known as the reflex oscillator came into general use.

In this paper the requirements on beating-oscillator tubes for radar systems are discussed, and the design features which have made the reflex oscillator eminently satisfactory in this application are pointed out. Problems encountered in such oscillators are outlined, and the solution in a number of cases is indicated. In some instances military requirements and expediency were in conflict with the optimum performance, and hence certain compromises were necessary.

[12] *Railway Signaling*, December 1947.
[13] *Proc. I. R. E.*, December 1947.

Space-Charge and Transit-Time Effects on Signal and Noise in Microwave Tetrodes.[14] L. C. PETERSON. Signal and noise in microwave tetrodes are discussed with particular emphasis on their behavior as space-charge conditions are varied in the grid-screen, or drift, region. The analysis assumes that the electron-stream velocity is single-valued. For particular conditions the noise figure may be substantially improved by increasing the space-charge density in the grid-screen region until an entering electron encounters a field of a certain magnitude. The noise reduction is largely due to the cancellation in the output of the noise produced by the random cathode emission. The method of noise reduction described is applicable only when the transit angles of both input and drift regions are fairly long.

In a forthcoming paper, H. V. Neher describes experimental results which broadly agree with the theory.

"Cloverleaf" Antenna for F. M. Broadcasting.[15] PHILLIP H. SMITH. The radiation requirements and general design considerations for transmitting antennas suitable for f.m. broadcoasting are briefly discussed, and an explanation of the design and operation of the arrangement of radiating elements and associated feed system employed in the "cloverleaf" antenna is given. Both calculated and measured data are included, showing field-intensity distribution, gain, impedance-frequency characteristics, etc. Design features which are discussed include a simple coaxial impedance-matching transformer developed initially for microwave application, and the method and facilities provided for the removal of sleet.

Hybrid Circuits for Microwaves.[16] W. A. TYRRELL. The fundamental behavior of hybrid circuits is reviewed and discussed, largely in terms of reciprocity relationships. The phase properties of simple wave-guide tee junctions are briefly considered. Two kinds of hybrid circuits are then described, the one involving a ring or loop of transmission line, the other relying upon the symmetry properties of certain four-arm junctions. The description is centered about wave-guide structures for microwaves, but the principles may also be applied to other kinds of transmission lines for other frequency ranges. Experimental verification is provided, and some of the important applications are outlined.

[14] *Proc. I. R. E.*, November 1947.
[15] *Proc. I. R. E.—Waves and Electrons Section*, December 1947.
[16] *Proc. I. R. E.*, November 1947.

Contributors to this Issue

J. R. Davey, B.S. in Electrical Engineering, University of Michigan, 1936. During the war Mr. Davey was engaged in the development of H. F. radio teletype systems, and has since been concerned with the development of electronic telegraph circuits.

H. T. Friis, E.E., Royal Technical College, Copenhagen, 1916; Sc.D., 1938; Assistant to Professor P. D. Pedersen, 1916; Technical Advisor at the Royal Gun Factory, Copenhagen, 1917–18; Fellow of the American Scandinavian Foundation, 1919; Columbia University, 1919. Western Electric Company, 1920–25; Bell Telephone Laboratories, 1925–. Formerly as Radio Research Engineer and since January 1946 as Director of Radio Research, Dr. Friis has long been engaged in work concerned with fundamental radio problems. He is a Fellow of the Institute of Radio Engineers.

Marion C. Gray, Edinburgh University, M.A., 1922; Bryn Mawr College, Ph.D., 1926. Instructor in physics, Edinburgh University, 1926–27; Research Assistant in Mathematics, Imperial College, London, 1927–30. American Telephone and Telegraph Company, Department of Development and Research, 1930–34; Bell Telephone Laboratories, 1934–. Dr. Gray has been engaged mainly in mathematical work in the field of electromagnetic theory.

A. L. Matte, B.S. in Electrical Engineering, Massachusetts Institute of Technology, 1909; Graduate Studies, M.I.T., 1912–13. New England Investment and Securities Company, 1910–12; Detroit United Railways, 1913–18. American Telephone and Telegraph Company, Department of Development and Research, 1918–34; Bell Telephone Laboratories, 1934–. Mr. Matte has been engaged principally in transmission studies relating to telegraphy.

S. O. Rice, B.S. in Electrical Engineering, Oregon State College, 1929; California Institute of Technology, 1929–30, 1934–35. Bell Telephone Laboratories, 1930–. Mr. Rice has been concerned with various theoretical investigations relating to telephone transmission theory.

D. H. Ring, A.B., Stanford, 1929; Engineer, Stanford, 1930. Bell Telephone Laboratories, 1930–. Mr. Ring has been engaged in radio research.

S. A. Schelkunoff, B.A., M.A. in Mathematics, The State College of Washington, 1923; Ph.D. in Mathematics, Columbia University, 1928. Engineering Department, Western Electric Company, 1923–25; Bell Telephone Laboratories, 1925–26. Department of Mathematics, State College of Washington, 1926–29. Bell Telephone Laboratories, 1929–. Dr. Schelkunoff has been engaged in mathematical research, especially in the field of electromagnetic theory.

W. H. Wise, B.S., Montana State College, 1921; M.A., University of Oregon, 1923; Ph.D., California Institute of Technology, 1926. American Telephone and Telegraph Company, Department of Development and Research, 1926–34; Bell Telephone Laboratories, 1934–. Dr. Wise has been engaged in various theoretical investigations relating to transmission theory.