The Bell System Technical Journal

Vol. XVI *July, 1937* *No. 3*

Scientific Research Applied to the Telephone Transmitter and Receiver *

By EDWIN H. COLPITTS

LET us recall a scene at the Centennial Exhibition in Philadelphia in 1876. Across a room had been strung wires connecting crude instruments, at one end of the room a transmitter and at the other end of the room a receiver. Dom Pedro, Emperor of Brazil, takes up the receiver and listens while Alexander Graham Bell speaks into the transmitter. The Emperor, astonished at hearing Mr. Bell's voice in the receiver, exclaims in amazement, "My God, it talks."

When at the same place, Sir William Thomson (later Lord Kelvin) took up the receiver and listened to Mr. Bell, the words of this distinguished scientist were, "It does speak," and continuing, "it is the most wonderful thing I have seen in America."

Sixty years have passed and, as a result of continued effort, the use of the telephone has become such an everyday matter that even the ability to talk from Tokyo in your country to New York in my country scarcely excites comment or wonder. It is not surprising that, to the layman, the element of distance seems the most striking factor in the technical development of the telephone art. As a matter of fact, while the conquest of distance has involved much scientific effort, and very ingenious and highly developed methods for the transmission of speech currents, the magic of the telephone still resides in the instruments which provide for the conversion of mechanical energy, namely speech sounds of highly complex wave form, into electrical currents of corresponding wave form, and the reverse process of converting these electrical currents into speech sounds. These instruments, the transmitter and the receiver, are basic to the whole telephone art. As they have been improved by development and design, it has become possible not only to render a higher grade of service but to effect economies in other portions of the plant. For example, the

* Another of three Iwadare Foundation lectures delivered during this past spring in Japan by Dr. Colpitts. One lecture was published in the April 1937 issue of this *Journal*.

very extensive use of fine-gauge cables in the plant of the Bell System was, to a large extent, made possible by the development of more efficient transmitters and receivers. Further perfecting of these instruments promises additional improvements in service and some further economies.

Telephony, restricting the term to ordinary two-way talking between individuals, involves an element not present in any other service. It does not greatly concern one customer of an electric light or power company whether another customer chooses to use inadequate or inefficient or poorly located lamps or other equipment. That is, each user of the service is, under any ordinary conditions, independent of all other users. In the case of telephony, however, the problem is entirely different; for one user of the telephone is greatly concerned with not only the apparatus furnished to any one with whom he has occasion to talk but also with other factors affecting the use of this apparatus, such as the amount of noise in the room where the apparatus is located, the user's habits of speech, and whether his ability to hear is normal. Telephone instrumentalities must therefore be so designed and the plant so engineered as to meet reasonably wide variations from what may be termed normal conditions, and ratings of performance should be similarly established.

I believe telephony in your country as in ours will find an increasingly wide field of service, and there is no single factor more important to a sound development of this art than the subscriber apparatus. With your permission, therefore, I will broadly outline certain work of the Bell Telephone Laboratories which has had a very direct bearing on these telephone instrumentalities and the form they are likely to assume. I will first discuss the research program which has been carried on in these laboratories, and then indicate to you the general trend which development and design have taken.

The research program basic to the development and design of transmission instruments has itself been a matter of development as a better understanding of the problems unfolded and as the need for research in this or that direction became apparent. The research problem basic to the development and design of transmission instruments may be described as having the following very broad scope: an understanding of their physical operation viewed as electromechanical structures; an understanding of speech mechanism and an accurate physical definition of speech air waves; an understanding of the hearing processes and a determination of how hearing is affected by factors present in telephony. Also, our research program may be said to have included research upon certain materials, the results of

which have an important bearing either upon an understanding of the operation of these instruments or upon their practical design. In addition to the development of many methods of measurement and testing applicable to laboratory research and development, of very great importance has been a development of the testing methods which permit of a better final evaluation of the developments based upon the results of this activity.

Instruments as Electromechanical Structures

The telephone transmitter itself is a complex mechanical and electrical structure. Its general method of operation can be described qualitatively in relatively simple terms, but the operation of few structures is more difficult to define in definite quantitative terms and relationships. For example, we are concerned with acoustical problems such as those involved in the air connection between the lips of the speaker and the diaphragm of the instrument. This air connection may involve a short column of air as in those instruments which have a telephone mouthpiece. Connection between the column of air and the working parts of the transmitter may be partially closed by a perforated section. When we come to consider the operation of the instrument itself, there is involved the mechanical vibration of the diaphragm as it operates on the carbon, and further, the whole question of electric conduction in the small mass of granular carbon itself.

In the case of the receiver which converts telephonic currents into speech sounds, we have very similar acoustical, mechanical and electrical problems with the exception, of course, of the mechanical and electrical problems introduced by the carbon of the transmitter.

A large amount of research work has been carried on in the Laboratories relating broadly to the transmitter and the receiver as electromechanical physical structures. The theory of these devices as vibrating systems has been developed so that their overall performance can be related to the various structural features. Consequently, our development and design engineers are now enabled to predetermine by calculation how certain modifications in structure will affect the physical performance of the instrument. In other words, the design process has become very much less "cut and try."

Research has been undertaken and substantial progress has been made on a study of microphonic action in carbon. In order to develop a complete theory of the operation of the transmitter, it is necessary to understand fully what takes place between each carbon granule in the carbon chamber.

Speech Sounds

Let me outline briefly some of the results of these studies on speech. The source of any voiced sound is in the larynx. On both sides of this larynx there are two muscular ledges called the vocal cords. When we breathe, these two ledges are widely separated, but when a voiced sound is produced, they come close together, forming a long narrow slit. As they come close together, the air passing through the resulting slit is set into vibration producing a sound. It has been generally supposed that the pitch of the tone thus produced was determined by the natural frequency of vibration of the two vocal cords, and that by changing the tension of these cords, the pitch of the tone can be raised or lowered at will. As most of you know, their natural frequency of vibration is the rate that they would vibrate to and fro if they were plucked and set into vibration like a banjo string or an elastic band. Our studies revealed that the natural pitch of these cords while a tone is being produced is considerably below that of the pitch of the tone. It is true that the pitch of the tone produced is affected, somewhat, by the elasticity of the vocal cords, but it is principally controlled by the size of the air opening between them. The little plug of air between the two vocal cords vibrates through a very much larger amplitude than the amplitude of the cords themselves and is the real source of the sound. The mass of this small plug is controlled by the size of the opening and by the elastic forces pushing it to and fro—namely, the air pressures on either side of it. It is evident that these oscillating pressures will be influenced by the size and shape of the trachea leading into the lungs on one side and by the size and shape of the tongue, mouth, and nasal cavities on the other. The mechanical action involved is analogous to the electrical action in a vacuum-tube oscillator. The sound which is generated at the vocal cords is modified as it passes through the throat, mouth, and nasal passages. The real character of the sound which enables us to identify words is wholly dependent upon the manner in which this cord tone is modified by the changing sizes, shapes, and characters of these passages and the outlet to the outside air.

After the various speech sounds leave the mouth, they are transmitted to the ear of the listener by means of air vibration. As an example of the type of disturbance created in the air, consider the sentence, "Joe took Father's shoe bench out." This silly sounding sentence is chosen because it is used in our laboratories for making tests on the efficiency of telephone systems. The sentence, together with its mate, "She was waiting at my lawn," contains all the fundamental sounds in the English language that contribute appreciably

toward the loudness of speech. As the sound wave produced by speaking this sentence travels along, each particle of air over which it passes executes a vibration through its original or undisturbed position. The successive positions occupied by the particle as it moves in the complicated series of vibrations corresponding to a spoken sound can be visualized in laboratory investigations from oscillographic records of the corresponding telephone currents.

Each successive particle of air along the line in which the sound is traveling executes a similar complicated series of vibrations but any particular oscillation is performed at a later instant by the particle which is farther away from the source of the sound. The disturbance in the air which represents a spoken sound may then be pictured

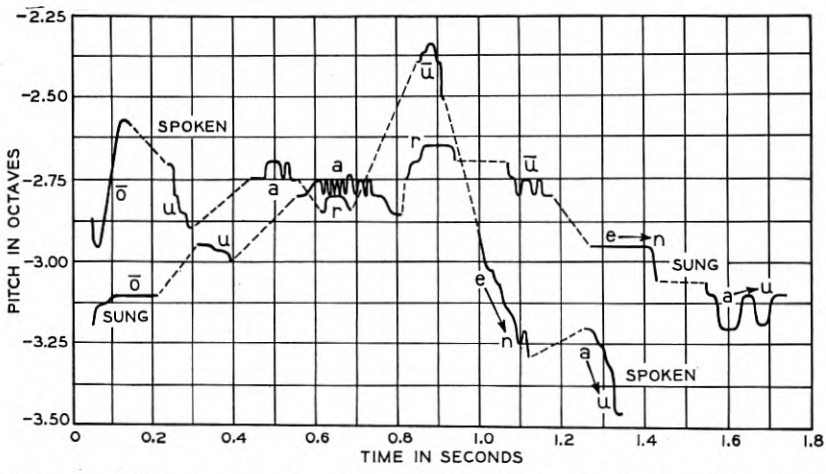


Fig. 1—Melodic curves showing the variation of pitch with time as the sentence "Joe took Father's shoe bench out" is spoken and sung.

either, as was first described, in terms of the successive positions of a single particle or in terms of the displacements at any instant of each of the particles along the line of travel of the sound wave. For example, for the sentence "Joe took Father's shoe bench out," the disturbance carrying the sound *j* in the word "Joe" is about fifteen hundred feet from the mouth by the time the sentence is finished. I have a record here which was taken in our laboratories which shows the intricate motion of each particle of air as this sentence is transmitted through the air.

If we analyze the wave when the sentence "Joe took Father's shoe bench out" is spoken, the variations in pitch of the speech sounds can be determined from the vibration rate. Such an analysis is shown in Fig. 1. The variations in pitch are represented on the

vertical axis. The duration of the sounds in fractions of a second is represented on the horizontal axis. It will be seen that the pitch rises and falls as the various sounds are spoken. This representation of the pitch variation is called the fundamental melodic stream. It is the melody in the same sense as this term is used in music, although it is evident that the pitch changes do not take place in musical intervals as would be the case if the sentence were sung.

To show the contrast, a graph was made when the sentence was intoned on the musical intervals *do, re, mi, fa, mi, re, do.* An analysis of the graph gave the result shown in Fig. 1. In the case of the sung sentence the pitch changes are in definite intervals on the musical scale, while for the spoken sentence the pitch varies irregularly, depending upon the emphasis given. The pitch of the fricative and stop consonants is ignored in the musical score, and since these consonants form no part of the music, they are generally slid over, making it difficult for a listener to understand the meaning of the words. Some of our friends in the musical profession may object to this statement of the situation, but I think it will be agreed that a singer's principal aim is to produce beautiful vowel quality and to manipulate the melodic stream so as to produce emotional effects. To do this, it is necessary in singing to lengthen the vowels and to shorten and give less emphasis to the stop and fricative consonants. It is for this reason that it is more difficult to understand song than speech.

There are two secondary melodic streams of speech represented by the second and third curves from the bottom of Fig. 2, which are due to the resonances imposed upon the speech sound by the throat and mouth cavities. The numbers on these curves give the number of the harmonic which is reenforced. These two secondary melodic streams are not sensed as changes in pitch, but rather as changes in the vowel quality. Then there is a fourth stream, or, it would probably be better to say, a fourth series of interrupted sounds which are very high in pitch and are the sounds which enable us to identify the fricative consonants. The secondary melodic streams produced while speaking the same sentence are approximately the same for different persons, even for a man and a woman, while the fundamental melodic stream is usually quite different. This latter stream is not used in identifying words, but it is used sometimes to give different meanings to the same words.

As one listens to this sentence he hears the variations in loudness as well as in pitch. Loudness is related to the amplitudes and frequencies of the components of the tone, but this relationship is very

complicated. It is dependent upon the action of the ear, including the nerve mechanism carrying the message to the brain. This relationship has been under study for a number of years so that we are now able to calculate from physical measurements the loudness for a typical ear and also to devise instruments for measuring approxi-

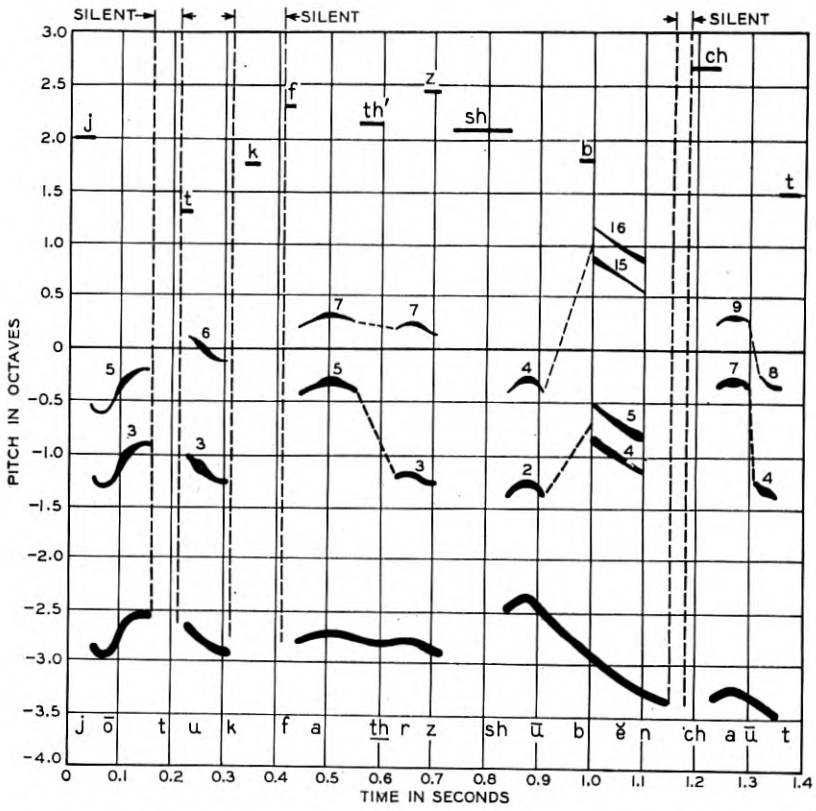


Fig. 2—Melodic curves showing the variation of pitch with time as the sentence "Joe took Father's shoe bench out" is intoned on the musical intervals do, re, mi, fa, mi, re, do. The pitch changes in regular intervals rather than in irregular intervals as shown in Fig. 1.

mately the loudness of any sound. The result of using such a device for recording the variations of loudness in the spoken sentence which we have been discussing is shown in Fig. 3. For comparison, the variations in pitch are also shown in this figure.

If the fifteen-hundred-foot wave carrying the sentence above mentioned could all be collected into an energy collector, the question

arises, "How much energy would be involved?" It is not possible here to describe the devices by which we were able to measure accurately the energies and frequencies involved in speech, but the results of this research work are interesting. When this sentence is spoken fairly rapidly, it will contain about two hundred ergs of energy. About 500,000,000 ergs of energy pass through the filament of an ordinary incandescent lamp each second. This shows that the acoustic

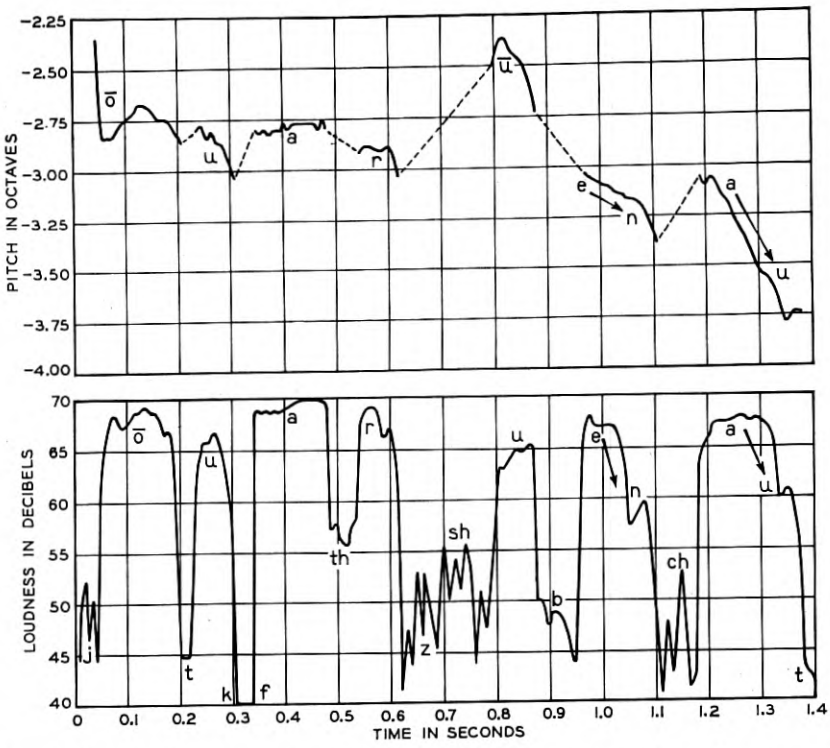


Fig. 3—Graph of the loudness of the various sound elements when the sentence "Joe took Father's shoe bench out" is spoken.

energy in this sentence is very small. Putting it in another way, it would require five hundred persons speaking this sentence continuously for a year to produce sufficient speech energy to heat a cup of tea.

An examination of the wave produced by this sentence shows that the vowels contain considerably more energy than the consonants. Exact measurements have indicated that in ordinary conversation the ratio of the intensity of the faintest speech sound, which is *th* as in "thin," to the loudest sound, which is *aw* as in "awl," is about one

to five hundred. The actual power used in producing the various sounds depends, of course, upon the speaker and the emphasis with which he pronounces the sound. The power in an accented syllable is three or four times that in a similar unaccented syllable. Measurements upon a number of voices during a conversation have indicated that the average power in the speech produced is ten microwatts (one one-hundred-thousandth watt). Some speak with more and others with less than this power. In Table I is shown how various

TABLE I

RELATIVE SPEECH POWERS USED BY INDIVIDUALS IN CONVERSATION

Ratio of power of individual speakers to average power	Below 1/16	1/16 to 1/8	1/8 to 1/4	1/4 to 1/2	1/2 to 1	1 to 2	2 to 4	4 to 8	Above 8
Per cent of speakers	7	9	14	18	22	17	9	4	0

voices in a sample group vary from the average. It is seen that seven per cent speak with less than one sixteenth the average power, eighteen per cent use powers lying between one quarter and one half the average, and four per cent between four and eight times the average. No speakers were found to use more than eight times the average for conversational purposes.

Now let us consider the variations for a typical speaker. As a conversation proceeds, the speech power varies from zero during the silent intervals to peak values which frequently are one hundred times the average power. Extensive measurements of these peak powers upon a number of speakers indicated a distribution about the average as shown in Table II. For example, if we should examine the speech

TABLE II

PEAK POWERS IN CONVERSATIONAL SPEECH

Power Boundaries In Terms of Average Power	Per Cent of Intervals
Below 1/8	12
1/8 to 1/4	4.0
1/4 to 1/2	4.5
1/2 to 1	5.5
1 to 2	8.3
2 to 4	12.7
4 to 8	18.6
8 to 16	17.0
16 to 32	10.5
32 to 64	5.1
64 to 128	1.7
Above 128	.1

during each one-eighth-second-interval throughout a typical conversation, we should find that for seventeen per cent of them the peak power would lie between eight to sixteen times the average over a long interval. It is seen that the most frequently occurring value of the peak power is about ten times the average.

Although a typical voice of a man and a typical voice of a woman are alike in that they use the same average power and variations of power from this average, they are different in other respects which we shall now consider. It is well known that the pitch of the voice of a woman is about one octave higher than that of a man. It was not known, however, until our experiments revealed it, that the intensity of the components having vibration rates above three thousand cycles per second was definitely greater for voices from women than from men. The following investigation shows the extent of this difference.

An apparatus has been devised in our laboratory which will receive the speech during a conversation and then sort out the components into groups depending upon their intensity and pitch. Those lying in each half-octave band on the pitch scale are automatically grouped together and the group power measured. Also, by means of another automatic device, a sorting process is accomplished within the group placing together all the components having powers between certain power boundaries so that they operate a particular recording meter. It was by means of an apparatus of this latter type that the results in Table II were obtained. It was found that the powers were distributed in each of these pitch bands in approximately the same manner as indicated in Table II for speech as a whole.

The relative values of the average speech power in each of the half-octave bands are shown in Fig. 4. The horizontal positions give the pitch in octaves above or below a tone having a vibration rate of one thousand cycles per second. The vertical positions give the fraction of the total power which comes into each half-octave band. For example, consider the half-octave from -2.25 to -1.75, which is the octave with its midpoint at middle "C" on the musical scale. The fraction of the power coming into this half-octave is about one quarter. It will be noted that for both types of voices the maximum power occurs in the second octave below one thousand cycles. This particular octave contains about one half of the total speech power. The octaves on either side of this one containing the maximum power contain slightly less than one quarter of the total power. No other octave contains more than about three per cent of the total power. It is seen that for the band of lowest pitch the voices from men contain

about eight times the power of those from women. Also, as stated above, for pitches above one—that is, for tones having vibration rates above two thousand cycles per second—the voice power for women is greater than for men. For the half-octave in the region of pitch three octaves above one thousand cycles, it is about ten times greater.

For some reason which is not very evident, women use higher pitch sounds for producing the fricative consonants, and this results in the greater power shown in the regions of higher pitch. Every one who is familiar with such transmission systems knows well that these high-frequency components are nearly always eliminated. While

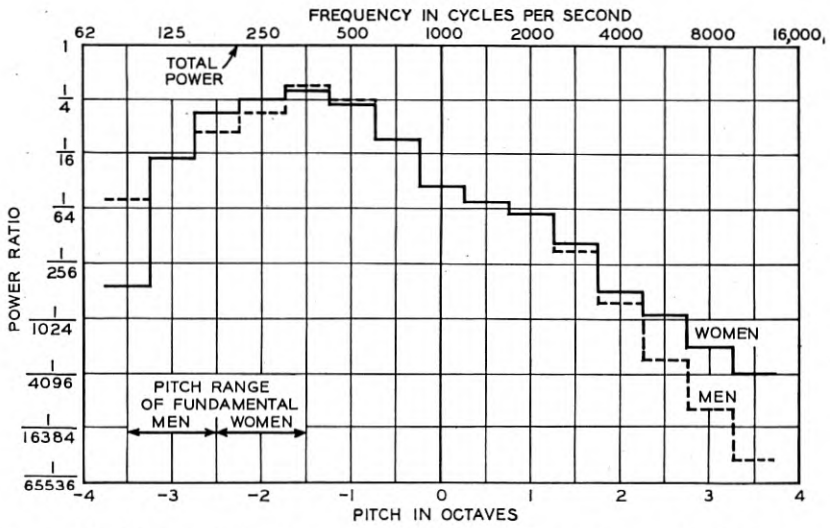


Fig. 4—Distribution of speech power in fractions of the total power for half-octave intervals above and below 1000 cycles.

these sounds are not of controlling importance in properly understanding speech, it is evident that the women's voices are somewhat handicapped as compared with men in systems which eliminate them.

Hearing

Paralleling our research on speech sounds, an investigation of hearing has been under way in Bell Telephone Laboratories. Broadly speaking, the aim has been to arrive at an accurate physical description and a measure of the mechanical operation of human ears in such terms that we may relate them directly to our electrical and acoustical instruments. We have measured the keenness of the sound-discriminating sense, and determined what is the smallest distortion which

the mind can perceive, and how it reacts to somewhat larger distortions. This information is utilized in determining a reasonable basis of design both for separate instruments and for transmission systems as a whole, to give a proper balance between cost and performance.

I can only indicate a few of the important results of our investigations. One of the first steps was to determine in a quantitative way the performance of our ears as machines. It was obviously important to know how faint a sound the ear can hear, and also how loud a

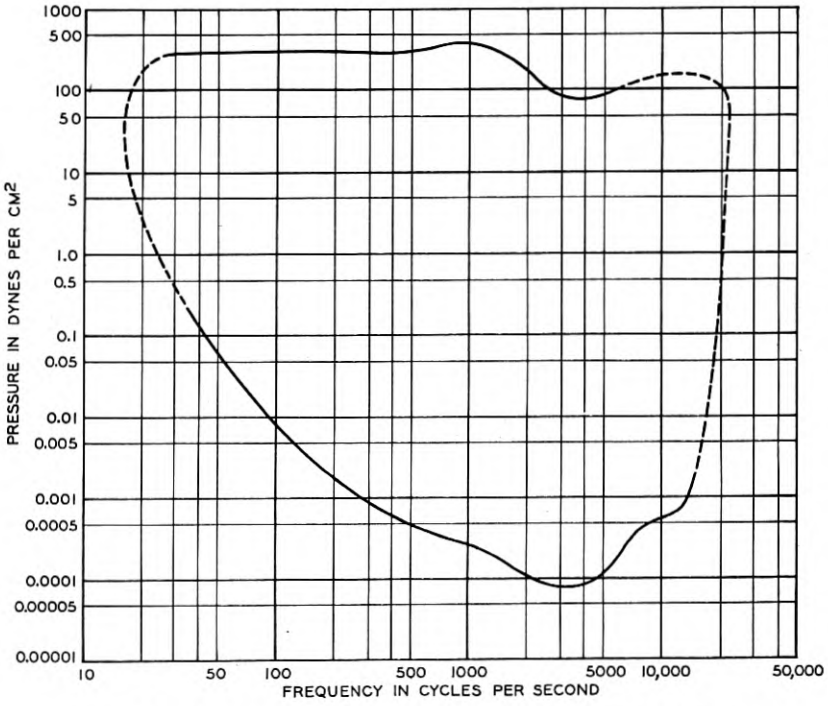


Fig. 5—Auditory sensation area for the typical ear of a young adult.

sound the ear can tolerate. With the advent of the vacuum tube, it was possible to develop methods of accurately measuring the intensity of faint sounds and of readily producing such sounds. Figure 5 gives the results of a large number of measurements made to determine the limits of hearing. This graph is called the auditory sensation area. The lower solid curve represents the minimum sound that an average young person can hear. The abscissa gives the frequency of the pure tone, and the ordinate the sound pressure in dynes per square centimeter. The top solid curve represents the

maximum intensity of sound that the ear is capable of handling. This curve was determined by noting that intensity which produced a feeling sensation. Intensities slightly higher than this result in pain and in some instances serious injury to the ear. The dotted lines on either side complete the enclosure and represent the upper and lower limits of pitch that can be heard. It is obvious from this figure that the upper or lower limit of pitch is greatly dependent upon the intensity at which the sound is produced. It will be seen that near the middle range of frequencies, the pressure range is one million to one. The pitch range of pure tones is from about 16 to 25,000 cycles per second.

These results are for young adults, and it may be of interest to note that as one becomes older the hearing acuity, at the higher frequencies particularly, becomes less. In the table below is shown some measurements to determine what the effect of age would be upon the hearing acuity:

TABLE III

DB LOSS IN HEARING WITH AGE

Frequency	60 to 1024 Cycles	2048 Cycles	4096 Cycles	8192 Cycles
Ages 20–29 (96 ears)	0	0	6	6
Ages 30–39 (162 ears)	0	0	16	11
Ages 40–49 (84 ears)	0	2	18	16
Ages 50–59 (28 ears)	0	5	30	32

These are average values obtained from measurements on a large number of persons.

Another important measurement of average hearing is that concerned with minimum perceptible differences in pitch and in intensity. Careful measurements on large groups of people have given us reliable data of this form. In Fig. 6 are shown the results of such measurements. They are plotted on the auditory sensation area. The ordinates are decibels above the reference pressure and the abscissas are centi-octaves above or below a pitch of 16.35 cycles per second. A frequency scale is also given for reference purposes. The numbers within the area indicate the minimum changes in the intensity level in db that the average ear is able to detect over that region of the auditory area. It will be seen that near the threshold fairly large changes are necessary to be perceptible, while at fairly high intensities about 1/4 decibel is all that is necessary for the change to be perceived.

In Fig. 7 are given similar data for minimum perceptible differences in pitch. The numbers in the figure in this case are given in centi-

octaves; that is, each unit corresponds to 1/100 of an octave. The results of this line of investigation have an important bearing on the physiological theory of hearing which I cannot enter into, and another important result has been the development of methods of determining the degree of impairment of hearing.

In telephony we are, of course, not directly concerned with simple sounds, but with the highly complex sounds of speech, and these are

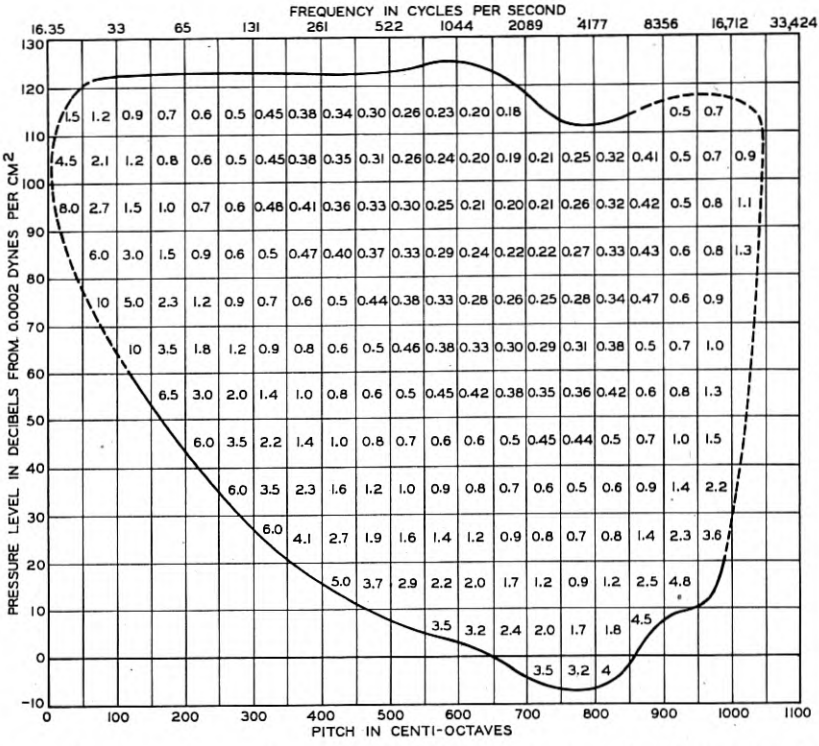


Fig. 6—Minimum changes in intensity level in db that the ear is able to appreciate at various positions in the auditory area.

on actual telephone circuits generally associated with extraneous sounds which we may group under the broad term of noise. Further, telephone instruments are not perfect, and could be made to approach perfection only at a great expense. In order to arrive at a quantitative understanding of the importance of departures from perfection in telephone transmission elements and conditions of use, we have in very general terms proceeded as follows: We set up transmission systems so nearly perfect that even the keenest ear could not find a

flaw in their transmission performance, and then introduce measured imperfections or variations.

By this general process, we were able to determine the effect of noise of chosen intensities either as noise present in the telephone receiver or as noise in the room. Similarly, the effect of a line or other characteristic such that voice frequencies above a certain value or below a certain value were not transmitted, was determined. The

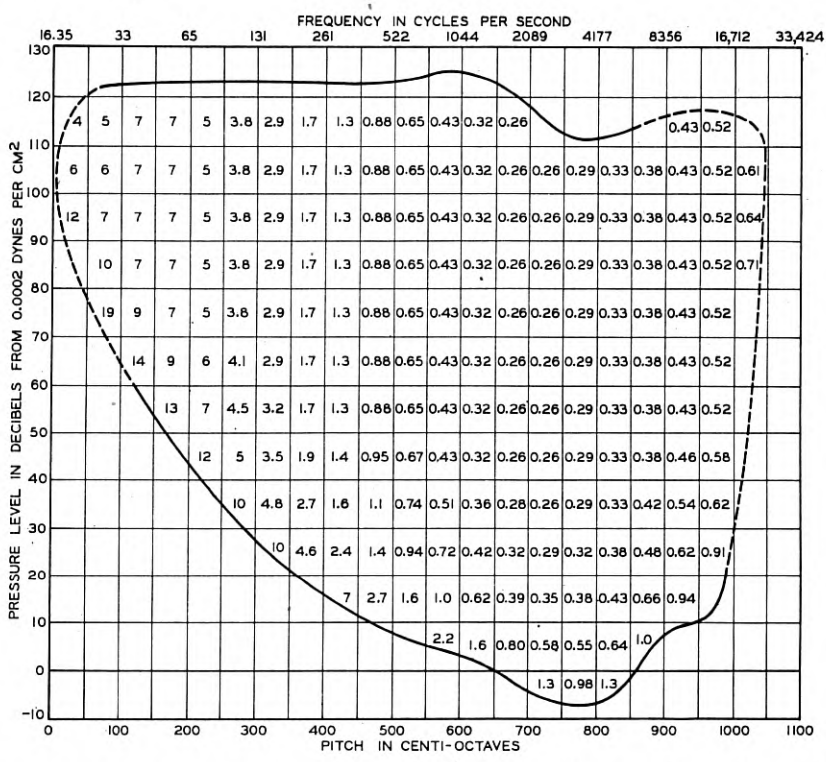


Fig. 7—Minimum perceptible differences in pitch in centi-octaves for various positions of the auditory area.

effect of introducing a highly resonant element or of a non-linear element was studied. The range in loudness of speech necessary for best reception was likewise measured.

As noise became recognized as a very real factor, a standard basis for noise measurements was established. Consequently we are now able to measure noise on a telephone circuit or in a room, and state the result in terms of a standard unit.

Materials

In the practical design of modern telephone instruments we owe a large debt to the chemist and the metallurgist. Modern molding materials and processes are utilized in order to secure forms of apparatus satisfactory from the standpoint of appearance and of mechanical strength. The newer types of permanent magnet steel, to the development of which your countrymen have contributed so largely, provide possibilities of light-weight and very efficient magnetic structures.

It is a most striking circumstance that commercial telephony is dependent upon the performance of a small mass of carbon granules in the transmitter. No single material entering into the construction of telephone apparatus has therefore greater importance. In America at least, transmitter carbons are largely derived from a certain specially selected anthracite coal. In its natural state, this coal exhibits none of the characteristics required for its use in a transmitter. These characteristics or properties are secured by heat treatment. These heat-treatment processes were for many years the result of empirical development and were not well understood or, as we now recognize, adequately controlled. This resulted in a product of uncertain quality. An important task of the Laboratories was therefore to study each step in the process of producing carbon and to develop a process definitely specified at each step, which would be capable of giving the desired uniform quality. The results so far obtained have had very important reactions upon transmitter performance. The Laboratories have also set themselves the more elementary task of understanding the fundamental properties of carbon contacts. One important element of this research is to determine the causes of resistance changes produced when the compressive force on a mass of carbon granules is changed. It is too early to report results from this research, but it seems clear that granular carbon will be an important element in the design of transmitters for many years to come, and we should seek to obtain complete fundamental knowledge of its operation.

Testing Methods

Broadly speaking, methods of testing have been developed, first, to enable the development and design engineer to determine quantitatively the various performance factors of the apparatus under development, and second, to determine how well the apparatus which has been developed performs under service conditions. In the Laboratories, we have over the last twenty years developed methods for measuring the physical constants of the apparatus involved so that

we can analyze this apparatus as electromechanical structures. Further, in the design of telephone transmission apparatus we are concerned with a power transmission system in which the design engineer has no control over the power source, the human voice, nor over the receiving agency, the human ear. His control is limited to the conveyance of power from speaker to listener. In the Laboratories, therefore, we have recognized that it is necessary, at least without present knowledge, to supplement physical measurements by measurements involving speech sounds and the human ear. Some years ago, these tests consisted of comparisons between different instruments or transmission elements made by the process of talking first over a circuit containing, for example, one instrument, and then over the same circuit containing a different instrument. Dependence was placed wholly upon the listener's skill to detect differences in volume, quality and intelligibility. It was recognized that this method of testing left much to be desired. Owing to the limitations of the human ear, small volume differences could not be detected, but even more important, this simple test furnished no very accurate measure of speech distortion affecting intelligibility, and obviously no definite information as to the relation between volume, various types of distortion, and overall effectiveness.

Dr. George A. Campbell, in 1910, proposed a method of testing which has been highly developed in our Laboratories. This method, termed "articulation testing," measures the relation between the reproduced and impressed sounds from the standpoint of effects on intelligibility of different kinds of distortion. This method has been described in a number of publications. Briefly, in this method, lists of syllables chosen at random and usually meaningless monosyllables are called over the circuits to be rated, and the percentage of syllables correctly understood gives a measure of the circuit performance. Further, the method has been extended to give quantitative measures in terms of the recognizability of reproduced speech sounds, of the effects of loudness of these sounds, and of the noise which may be present.

While various physical tests and the articulation test method are exceedingly useful tools in the hands of the research and development engineer, they do not give a direct measure of the transmission service performance of a circuit in terms of the ability of the user to carry on a conversation under actual commercial conditions. This ability of the user to carry on what may be termed a successful telephone conversation depends not only upon the performance of the telephone instruments and circuits but also, to a substantial extent, upon the

users' own performances—the subject material of conversation, how they talk into the transmitters, and how they hold the receivers—and upon the room noise conditions. In other words, there are a number of factors random in nature which, while beyond the control of those who design and engineer the telephone plant, must be taken account of in rating the service performance.

A large amount of thought and effort has been given to the problem of how best to determine transmission service performance. Very briefly stated, we have been led to the following steps: In order to take suitably weighted account of all the factors involved, service performance ratings should be based on service results, that is, transmission service performance should be measured by the success which users of the telephone circuit have in carrying on conversations over the circuit. With the various factors in mind, we have fixed upon what we have termed "effective transmission" ratings for transmission plant design. These ratings are based on a determination of the *repetition rate* in normal telephone conversations.

As the effect of a change in a circuit depends upon its initial characteristics, it is necessary in order to be able to compare numerical results to have a basic circuit for reference. By suitable choice of basic circuit, it is possible to express the effects of changes in any one transmission characteristic in terms of the attenuation of the trunk. For example, the effect of changes in sidetone level in the subscriber's set can be expressed as so many decibels change in trunk attenuation. Mr. W. H. Martin's paper, "Rating the Transmission Performance of Telephone Circuits," in the *Bell System Technical Journal*, January, 1931, discusses the method and general principles. It should be noted that the application of the method requires careful consideration of many factors and the accumulation and analysis of a very substantial amount of data. Based on these data, we have arrived at the following relationship:

$$\text{Relative effective loss in db} = 50 \log_{10} (r)$$

where r is the ratio of the repetition rates for the two conditions compared.

Association of Transmitter and Receiver

In order to furnish a convenient two-way talking circuit over a single pair of wires, the transmitter and the receiver at each end of the circuit must be continuously associated in the circuit. This has been accomplished by various circuit arrangements since the early days of the telephone, and as every user of the telephone knows, leads

to the condition that when speaking into the transmitter one hears his own voice in the receiver. Local speech so heard is designated as sidetone. The Laboratories have carried on research in order to determine the effect of sidetone on the overall efficiency of the circuit. We find that sidetone above a certain volume decreases the conversational efficiency of the circuit. Parallel with the study of the effects of sidetone, research has been carried on on methods which could be applied to limit sidetone in amount to more nearly its optimum value. This has led to the development of what are known as anti-sidetone circuits, which do not eliminate sidetone but reduce it to an amount which is more nearly that found to be desirable.

An important step in the association of the transmitter and the receiver is represented by the handset which provides a rigid mechanical connection between the two units. This rigid mechanical connection introduces mechanical coupling between the receiver and the transmitter, which had to be given very serious consideration in order to avoid speech distortion.

Trends in Instrument Development

I have broadly indicated to you fields of research which underlie the development and design of the telephone transmitter and receiver. It will now be of interest for us to note what application is likely to be made of the results of what has amounted to an enormous total of scientific effort. In this connection, it may be well again to emphasize that station apparatus is intimately associated with the user, and has therefore to be designed to fit him, his habits of using the telephone, and the conditions attending such use. The handset has to be designed to fit his head, the holes in the dial to fit the size of his finger, the bell to be loud enough, and so on. Our effective transmission rating system has been set up in an attempt to rate the performance of the telephone when employed by the customer in the way he wants to use it, under the conditions surrounding him. For this reason, this method of rating has been found particularly valuable in the development work on instruments.

Because of the wide range of customer usage and conditions, a number of factors have to be taken into account in the design of the apparatus. Also, because this apparatus is located on the customer's premises, where it is relatively inaccessible to the telephone personnel, it must be capable of standing up without undue trouble under this wide range of usage and conditions. To strike a proper balance in meeting all these factors requires an intimate knowledge of the field conditions as well as of the development and manufacturing possi-

bilities. A continuing close contact with field experience is employed to modify the designs towards securing the proper balance to meet these factors.

In order to indicate more clearly the present trends in design, I shall refer briefly to the earlier art. In the early development of transmitters and receivers, the matter of getting efficiency was of primary importance since this could be evaluated directly in terms of the amount of copper required in the connecting line. The early transmitters, which were of the same construction as the receiver, depended on the generator action of a diaphragm and coil and developed sufficient power to be heard over only a few miles of heavy-gauge wire. Some amplification was necessary before telephone communication could begin to assume the proportions of a widespread service. This amplification was obtained at a reasonable cost in the carbon contact transmitter. Transmitters of this type are in the order of 60 db more efficient as transducers of acoustic to electric energy than the earlier type.

Both the transmitter and the receiver operate by means of diaphragms which have natural periods of vibration. These resonances and the resonances of the air spaces on each side of the diaphragm were used to obtain as efficient a transfer of energy as possible. In the early design, a great deal of attention was also given to locating these resonances at the portion of the frequency range where they would tend to increase the intelligibility of the reproduced sound. As a result, both instruments were made very efficient in the region of 1000 cycles, which lies within the range where the ear and the sensation of loudness are most sensitive.

It was recognized that these resonances caused undesirable distortion, but under the conditions the resulting increase in efficiency more than compensated for this disadvantage. As time went on, the diaphragm resonances came to be looked upon as practically inherent in commercial transmitters and receivers, because no way was known of eliminating them without making a very material sacrifice in the efficiency of the instrument.

About twenty years ago, the development of the vacuum tube amplifier and the high quality condenser transmitter made it possible to demonstrate and measure quantitatively the advantages of reducing distortion. These high-quality instruments, the improvement in measuring technique and the development of improved methods of designing vibratory systems offered the promise of providing instruments in which the resonance effect could be reduced without unduly affecting efficiency.

The first commercial instrument for station use, which demonstrated the possibility of carrying out this promise, was the transmitter employed in the handset first supplied by the Bell System in 1927. This transmitter had to meet the requirement of giving the same transmission service as transmitters of the deskstand type, and at the same time meet the very exacting requirements imposed by the handset to make it free from howling and capable of performing over a wide range of positions. The diaphragm resonance was damped to a large extent by the use of paper rings and, by lightening the structure, the point of maximum response was moved up in frequency so that it no longer coincided with the peak of the receiver. The effect of this was not only to broaden the response characteristic and improve intelligibility, but also to reduce the gain in the local howling circuit which is, of course, a maximum when both transmitter and receiver have their greatest efficiency at the same frequency. The same separation of peaks resulted in the received speech being less loud, but in spite of this the overall performance was equivalent to that of the best deskstand type of instrument then available.

With this accomplishment, further work was directed toward maintaining the lower distortion and increasing the efficiency. The transmitter introduced in 1934 represented a marked improvement along this line. This instrument still further broadens the transmitted frequency range and is used with about the same efficiency in deskstands, handsets, wall sets, and coin-collect sets.

A new type of handset will be introduced in the Bell System in 1937 which, in addition to having a more pleasing and simplified design, will incorporate the new transmitter mounted in such a way as to make fullest use of its ability to transmit efficiently over a wide-frequency band.

During this evolution of the transmitter, the knowledge which had been gained as to the importance of transmitting different widths of frequency band over commercial telephone circuits led to the establishment of the range from 250 to 2750 cycles for designs of new circuits. It was not the intention in the establishment of this range that circuits should not do better than this where it is possible without materially increasing cost, but that all circuits should be at least as good as this. The establishment of this frequency range took into account a number of factors of which a very important one is that the overall utilization of this range from the sound entering the transmitter to the sound output of the receiver provides a grade of transmission which is highly satisfactory for the reproduction of conversational material.

The establishment of this frequency range played a part not only in the design of circuits, but also in guiding the evolution of the transmitter and receiver. The transmitter last referred to meets this requirement very well. In fact, its efficiency is fairly uniform for a frequency range extending beyond 4000 cycles.

The next step in the process was to improve the performance of the receiver. A pronounced resonance at 1000 cycles was no longer necessary since means had been found to improve the efficiency of instruments in other ways than by concentrating all the resonances at one frequency. The importance of the higher frequencies in transmitting and reproducing the transient sounds characteristic of the consonants in speech led to placing more emphasis on these frequencies and attempting to produce more uniformly the band of frequencies which was set as a limit for circuits. This has now been accomplished in a practical fashion in the receiver which is being introduced in 1937.

The effect of this evolution in the design of station instruments may be brought out by a comparison of the overall response characteristic —that is, the relation of the sound delivered to the ear to the sound available at the transmitter—for a typical telephone connection having, in one case, both terminal instruments of the 1920 type and, in the other case, the terminal instruments of the coming new 1937 type. In this typical circuit, the trunk has been taken as free from distortion so that its effect will not influence the indicated performance of the instruments, although the circuit does include two 22-gauge loops each three miles long.

At the resonance point of the old instruments, just over 1000 cycles, the overall response in going to the new instruments is reduced by almost 30 db while the response in the range from 2000 to 3000 cycles is increased by over 20 db. In the frequency range from 500 to 2000 cycles, the circuit employing the older instruments shows a variation of overall response of over 30 db. For the new type, the variation for this same frequency range is reduced to 15 db, and, furthermore, this variation of 15 db applies approximately for the range of 250 to 2750 cycles which was mentioned as the transmission range requirement for the design of new circuits. In regard to the variation of 15 db in this frequency range, there is good indication that this response is more desirable than one of no variation, from the standpoint of having the telephone performance approach that of direct air transmission.

In addition to these improvements in frequency response and efficiency, the intensive development program on these instruments has improved materially the stability of the carbon transmitter under

service conditions. This is an important factor in extending the useful life of these instruments and in reducing the cost of maintaining the desired transmission performance.

You will perhaps pardon me if, in concluding, I say a few words which I hope will not seem unduly laudatory of the work of my associates in the Bell Laboratories. The facts seem to be that twenty years ago or thereabouts, there was very little general scientific interest in sound and sound devices. As a result of work begun in

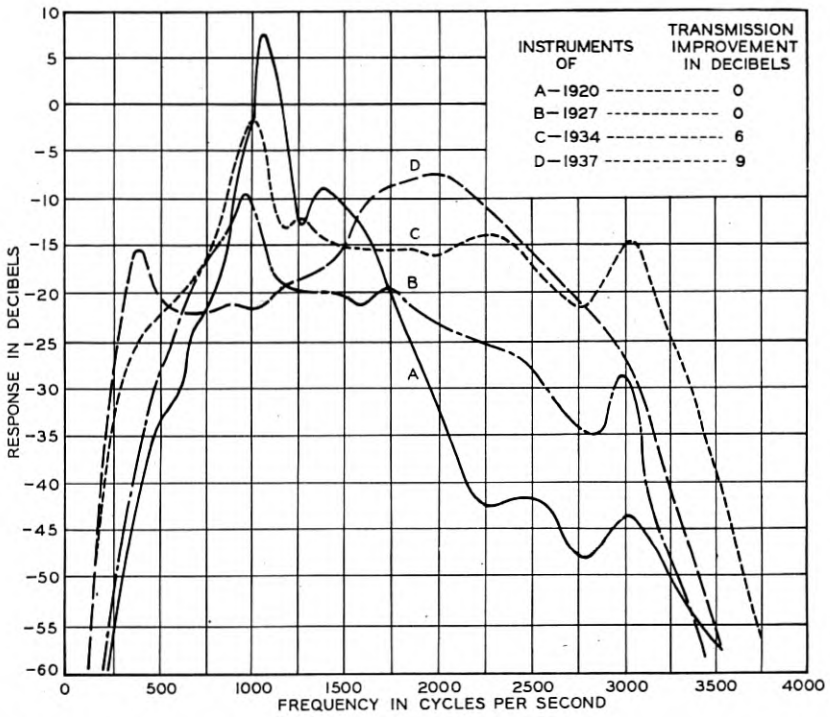


Fig. 8—Comparison of the response-frequency characteristics of telephone instruments since 1927.

these Laboratories, and as the possibilities of interesting and important applications became apparent, broad scientific interest was stimulated, and we have seen and welcomed increasing research activity in sound and acoustics in many of the university laboratories and in new industries based upon the results of scientific research in sound initiated by us. A number of my associates have attained world-wide recognition for their scientific and technical accomplishments. Our scientific investigations were undertaken to enable us to develop further the

telephone art, and the results of these investigations are serving to guide us not only in the development of telephone instruments but in all developments of telephone transmission. The Laboratories' scientific and design work has contributed in large measure to the improvement of methods of recording and reproducing sound in the phonograph and sound-picture arts. The art of radio broadcasting owes a large debt to the work of the Laboratories, not only for the fundamental scientific knowledge contributed but also for actual instrumentalities employed. To those with impaired hearing, the Laboratories' investigations have made possible improved means for determining the extent of their impairment, and improved hearing aids. Finally, at least in America, we are becoming what I may term as "noise conscious." In our cities, noise is being recognized as a factor affecting comfort, efficiency, and possibly even health. The development of accurate methods for the measurement of noise is contributing to studies looking towards the reduction of noise.

Lecturer's Note: The lecturer wishes to acknowledge assistance given in the preparation of this material, particularly by Dr. Harvey Fletcher and Mr. W. H. Martin of the Bell Telephone Laboratories' staff.

The Use of Coaxial and Balanced Transmission Lines in Filters and Wide-Band Transformers for High Radio Frequencies

By W. P. MASON and R. A. SYKES

At the high radio frequencies, filters and transformers become difficult to construct from conventional electrical coils and condensers, on account of the small sizes of the elements, the large effects of the interconnecting windings and the low ratios of reactance to resistance realizable in coils. It is shown in this paper that selective filters and wide-band transformers can be constructed using transmission lines and condensers as elements. The ratio of reactance to resistance in these elements can be made very high; consequently very selective filters and transformers with small losses, can be constructed. The effect of the distributed nature of the elements is taken account of in the design equations and methods are described for obtaining single-band filters and transformers. Experimental measurements of such filters and transformers are shown. The experimental loss curve is shown of a coaxial filter used in the Provincetown-Green Harbor short-wave radio circuit for the purpose of connecting a transmitter and receiver to the same antenna.

I. Introduction

At the higher radio frequencies, coil and condenser networks become difficult to construct on account of the small sizes of the elements and the large effects of the interconnecting windings. The Q realizable in high-frequency coils is about the same as can be obtained at the lower frequencies but due to the smaller percentage band widths, it is desirable to obtain a higher Q. There has been a tendency to replace coils by small lengths of transmission lines, and these have been used to some extent as tuned circuits, and as single-frequency transformers.[1,2,3]

It is the purpose of this paper to describe work which has been done in constructing selective filters and wide-band transformers from lengths of transmission lines and condensers. Due to the high ratio of reactance to resistance obtainable in both of these types of elements,

[1] "Transmission Lines for Short-Wave Radio Systems," E. J. Sterba and C. B. Feldman, *B. S. T. J.*, Vol. XI, No. 3, July 1932, page 411.

[2] "Resonant Lines for Radio Circuits," F. E. Terman, *Elec. Engg.*, Vol. 53, pp. 1046–1053, July 1934.

[3] "A Unicontrol Radio Receiver for Ultra-High Frequencies Using Concentric Lines as Interstage Couplers," F. W. Dunmore, *Proc. I. R. E.*, Vol. 24, No. 6, June 1936.

very selective networks can be obtained at the high radio frequencies. The effect of the distributed nature of the elements is considered and methods are described for obtaining single-band filters and transformers. Experimental measurements of such filters and transformers are shown, and these results indicate that such structures should be of use in short-wave radio circuits.

II. Characteristics of Transmission Lines

To facilitate an understanding of the following discussion, the equations of transmission lines as they apply to filter structures will be briefly reviewed first. The equations of propagation for any uniform transmission line can be expressed in the form of equations between the output voltage e_2, the output current i_2, the input voltage e_1, and the input current i_1 by the relations

$$\begin{aligned} e_2 &= e_1 \cosh Pl - i_1 Z_0 \sinh Pl, \\ i_2 &= i_1 \cosh Pl - \frac{e_1}{Z_0} \sinh Pl, \end{aligned} \tag{1}$$

where l is the length of the line, P the propagation constant, and Z_0 the characteristic impedance of the line. In terms of the distributed resistance R per unit length of line, L the distributed inductance, G the distributed conductance, and C the distributed capacitance, P and Z_0 can be expressed by the relations

$$P = \sqrt{(R + j\omega L)(G + j\omega C)}; \qquad Z_0 = \sqrt{\frac{R + j\omega L}{G + j\omega C}}, \tag{2}$$

where ω is 2π times the frequency.

The distributed conductance G is usually very low and can be neglected for coaxial or balanced transmission lines in dry atmospheres. For copper coaxial lines, the values of R, L, and C have been calculated[4] to be

$$\begin{aligned} R &= 41.6 \times 10^{-9} \sqrt{f} \left(\frac{1}{a} + \frac{1}{b} \right) \text{ ohms per centimeter}, \\ L &= 2 \log_e \frac{b}{a} \times 10^{-9} \text{ henries per centimeter}, \\ C &= \frac{1.11 \times 10^{-12}}{2 \log_e \frac{b}{a}} \text{ farads per centimeter}, \end{aligned} \tag{3}$$

where b is the inside radius of the outer conductor and a the outside radius of the inner conductor. If we define the Q of the conductor as

[4] See reference 1, page 415 and page 417.

the ratio of the series inductive reactance to the series resistance, the ratio will be

$$Q = .302b\sqrt{f}\,(\log_e k)/(k+1), \tag{4}$$

where $k = b/a$. It will be noted that this is the value of Q measured for a short-circuited conductor for low frequencies. As an example, a conductor 3 inches in diameter with the optimum ratio of $k = 3.6$ will have a Q of 3,200 at 100 megacycles. The value of the characteristic impedance Z_0 of a coaxial line is

$$Z_0 \doteq \sqrt{\frac{L}{C}} = 60 \log_e \frac{b}{a}. \tag{5}$$

For a balanced transmission line the values of R, L, and C are,[4] if D is much larger than a,

$$\begin{aligned} R &= \frac{83.2 \times 10^{-9}\sqrt{f}}{a} \text{ ohms per centimeter,} \\ L &= 4 \log_e \frac{D}{a} \times 10^{-9} \text{ henries per centimeter length,} \\ C &= \frac{1.111 \times 10^{-12}}{4 \log_e \frac{D}{a}} \text{ farads per centimeter,} \end{aligned} \tag{6}$$

where D is the spacing between wires, and a the radius of one of the pair of wires. With these values, the expressions for Q and Z_0 become

$$Q = .302\sqrt{f}a \log_e \frac{D}{a}; \qquad Z_0 = 120 \log_e \frac{D}{a}. \tag{7}$$

Another combination of some interest is obtained by using the inside conductors of two coaxial conductors adjacent to each other. Such a construction results in a balanced and shielded transmission line. All of the constants are double those given by equation (3), except for the capacitance which is halved.

III. Filters Employing Transmission Lines as Elements

One of the first uses of transmission lines as elements in wave filters is described in a patent of one of the writers.[5] In this patent are considered the characteristics obtainable by combining sections of trans-

[5] U. S. Patent 1,781,469 issued to W. P. Mason. Application filed June 25, 1927; Patent granted Nov. 11, 1930. Wave filters using transmission lines only are very similar to acoustic wave filters as pointed out in an article, "Acoustic Filters," *Bell Laboratories Record*, April 1928. Most of the equations and results of a former paper on acoustic filters, *B. S. T. J.*, April 1927, p. 258, are also applicable to transmission line filters.

mission lines in ladder filter structures. The results obtained are briefly reviewed here.

One of the simplest filters considered is shown on Fig. 1. The filter consists of a length $2l_1$ of transmission line shunted at its center by a short-circuited transmission line of length l_2. To determine the transmission bands of a filter, it is necessary to neglect the dissipation occurring in the elements and hence we assume that R and G are zero. In any case these values are small for transmission lines since they have a high Q. Neglecting R and G, equations (1) become

$$e_2 = e_1 \cos\frac{\omega l}{v} - j i_1 Z_0 \sin\frac{\omega l}{v},$$
$$i_2 = i_1 \cos\frac{\omega l}{v} - j\frac{e_1}{Z_0}\sin\frac{\omega l}{v}, \tag{8}$$

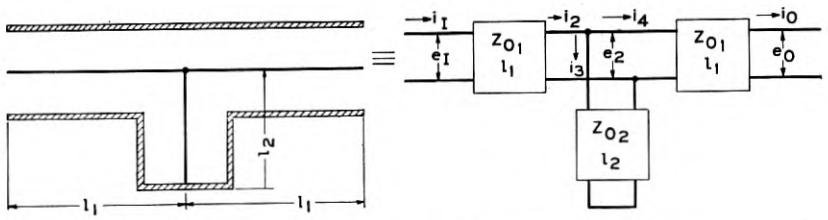


Fig. 1—Band-pass filter constructed from coaxial conductors.

where v, the velocity of propagation, and Z_0, the characteristic impedance, have the values

$$v = \frac{1}{\sqrt{LC}}; \qquad Z_0 = \sqrt{\frac{L}{C}}. \tag{9}$$

Using these equations the characteristics of the filter illustrated by Fig. 1 are easily calculated. With reference to Fig. 1 and equations (8) we can write the equations of the network as

$$e_2 = e_I \cos\frac{\omega l_1}{v} - j i_I Z_{0_1} \sin\frac{\omega l_1}{v},$$
$$i_2 = i_I \cos\frac{\omega l_1}{v} - j\frac{e_I}{Z_{0_1}}\sin\frac{\omega l_1}{v}, \qquad i_2 = i_3 + i_4,$$
$$e_0 = e_2 \cos\frac{\omega l_1}{v} - j i_4 Z_{0_1} \sin\frac{\omega l_1}{v}, \qquad i_3 = \frac{e_2}{j Z_{0_2} \tan\frac{\omega l_2}{v}}. \tag{10}$$
$$i_0 = i_4 \cos\frac{\omega l_1}{v} - j\frac{e_2}{Z_{0_1}}\sin\frac{\omega l_1}{v},$$

Combining these equations and eliminating all terms except the input

and output voltages and currents we have

$$e_0 = e_I\left[\cos\frac{2\omega l_1}{v}+\frac{Z_{0_1}}{2Z_{0_2}}\frac{\sin\frac{2\omega l_1}{v}}{\tan\frac{\omega l_2}{v}}\right] - ji_IZ_{0_1}\left[\sin\frac{2\omega l_1}{v}+\frac{Z_{0_1}}{Z_{0_2}}\frac{\sin^2\frac{\omega l_1}{v}}{\tan\frac{\omega l_2}{v}}\right];$$

$$i_0 = i_I\left[\cos\frac{2\omega l_1}{v}+\frac{Z_{0_1}}{2Z_{0_2}}\frac{\sin\frac{2\omega l_1}{v}}{\tan\frac{\omega l_2}{v}}\right] - j\frac{e_I}{Z_{0_1}}\left[\sin\frac{2\omega l_1}{v}-\frac{Z_{0_1}}{Z_{0_2}}\frac{\cos^2\frac{\omega l_1}{v}}{\tan\frac{\omega l_2}{v}}\right]. \tag{11}$$

The properties of symmetrical filters such as considered here are usually specified in terms of the propagation constant Γ and the iterative impedance K. These are similar to the line parameters used in equation (1) and it can be shown that the same relations exist between the input and output voltages and currents that exist in equation (1). Hence comparing equation (1) with equation (11) we find

$$\cosh\Gamma = \cos\frac{2\omega l_1}{v}+\frac{Z_{0_1}}{2Z_{0_2}}\frac{\sin\frac{2\omega l_1}{v}}{\tan\frac{\omega l_2}{v}},$$

$$K = Z_{0_1}\sqrt{\frac{1+\frac{Z_{0_1}}{2Z_{0_2}}\frac{\tan\frac{\omega l_1}{v}}{\tan\frac{\omega l_2}{v}}}{1-\frac{Z_{0_1}}{2Z_{0_2}}\frac{\cot\frac{\omega l_1}{v}}{\tan\frac{\omega l_2}{v}}}}. \tag{12}$$

The propagation constant Γ has the same significance as for a uniform line, namely that

$$e^{-\Gamma} = \frac{e_n}{e_{n-1}} = \frac{i_n}{i_{n-1}} = \text{ratio of currents or voltages between any two sections in an infinite sequence of sections,}$$

while K, the iterative impedance, is the impedance measured looking into an infinite sequence of such sections. The propagation constant Γ is in general a complex number $A + jB$. The real part is the attenuation constant and the complex part the phase constant. For a pass band, A must be zero, which will occur when cosh Γ is between $+1$ and -1. Hence to locate the pass band of the dissipationless filter, the value of the first equation of (12) must lie between $+1$ and -1.

In order to point out what types of filters result from a combination of transmission lines, two simplifying cases are considered. The first case is when $l_1 = l_2$. Then the equation for the propagation constant becomes

$$\cosh \Gamma = \left(1 + \frac{Z_{0_1}}{Z_{0_2}}\right)\left(\cos^2 \frac{\omega l_1}{v}\right) - \sin^2 \frac{\omega l_1}{v}$$

$$= \cos \frac{2\omega l_1}{v} \times \left[1 + \frac{Z_{0_1}}{2Z_{0_2}}\right] + \frac{Z_{0_1}}{2Z_{0_2}}. \quad (13)$$

The edges of the pass band occur when

$$\tan \frac{\omega l_1}{v} = \sqrt{\frac{Z_{0_1}}{2Z_0}} \quad (14)$$

and the centers of the bands occur when

$$\sin \frac{\omega l_1}{v} = 1; \qquad \frac{\omega l_1}{v} = \frac{(2n+1)\pi}{2} \qquad \text{and} \qquad f_m = \frac{(2n+1)v}{4l_1}, \quad (15)$$

where $n = 0, 1, 2$, etc. A plot of the propagation constant for several ratios of Z_{0_1}/Z_{0_2} is shown in Fig. 2. As is evident the filter is a multi-band filter with bands centered around odd harmonic frequencies. The ratio of Z_{0_1} to Z_{0_2} determines the band width of the filter. This ratio cannot be made very large because the characteristic impedance of a coaxial line or a balanced line cannot be widely varied from the mean value. For example when the ratio of b/a varies from 1.05 to 100 the characteristic impedance of a coaxial conductor changes from about 3 ohms to 275 ohms. This represents about as extreme a range as can be obtained. For a balanced conductor the impedance may range from 90 to 1100 ohms by taking extreme values of the ratio D/a. Taking 100 as the extreme range between Z_{0_1} and Z_{0_2} the band width for this case cannot be made less than 20 per cent.

We next consider the case given by taking $Z_{0_1} = 2Z_{0_2}$. For this

case the filter parameters take on the simple form

$$\cosh \Gamma = \frac{\sin \frac{\omega(2l_1 + l_2)}{v}}{\sin \frac{\omega l_2}{v}} \text{ and } K = Z_{0_1} \sqrt{- \tan \frac{\omega l_1}{v} \tan \frac{\omega(l_1 + l_2)}{v}}. \quad (16)$$

For this case the lower side of the pass band occurs when $\cosh \Gamma = +1$ and the upper side when $\cosh \Gamma = -1$. Hence the frequency limits

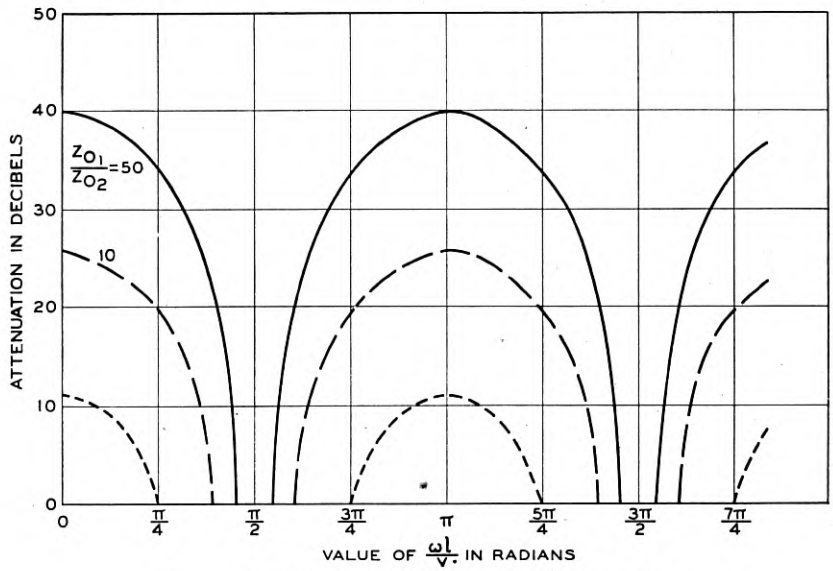


Fig. 2—Attenuation characteristic of band-pass filter.

of the first pass band can be obtained by solving the transcendental equation

$$\sin \frac{\omega(2l_1 + l_2)}{v} \mp \sin \frac{\omega l_2}{v} = 0,$$

which has the solutions

$$f_1 = \frac{v}{4(l_1 + l_2)}; \qquad f_2 = \frac{v}{4l_1}. \quad (17)$$

Hence by making the shunting line very short, it is possible to obtain a narrow band with this type of filter. The mid-frequency of the band occurs when

$$\sin \frac{\omega(2l_1 + l_2)}{v} = 0 \qquad \text{or when} \qquad f_m = \frac{v}{4l_1 + 2l_2}.$$

The characteristic impedance of this type of filter becomes very high for narrow-band filters as is shown by equation (16). At the mean

frequency f_m the impedance of the filter is

$$K = Z_{0_1}\sqrt{-\tan\left[\frac{\pi/2}{1+\frac{l_2}{2l_1}}\right]\tan\left[\frac{\pi}{2}\left(\frac{1+\frac{l_2}{l_1}}{1+\frac{l_2}{2l_1}}\right)\right]}$$

$$\doteq \frac{4l_1}{\pi l_2} Z_{0_1} \text{ for narrow bands.} \quad (18)$$

Such a filter would be useful for an interstage coupler to couple together the plate of one screen grid or pentode tube to the grid of another one. One such arrangement is shown in Fig. 3. This method of coupling together two stages of vacuum tubes has an advantage over using a coaxial conductor or a coil and condenser as a tuned circuit on several counts. In the first place the width of the band passed can be accurately controlled and a flatter gain characteristic is obtained. As will be shown later, distributed capacity in the plate and grid of the

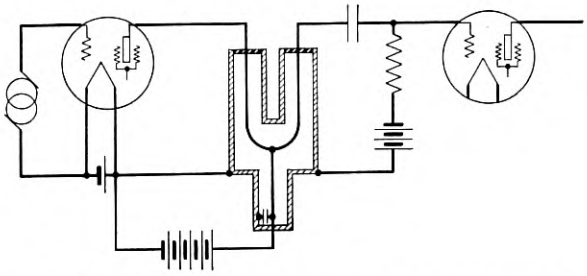


Fig. 3—Coaxial band-pass filter used to couple vacuum tubes.

vacuum tube can be absorbed in the filter by making the line length l_1 shorter. Since only half of the total distributed capacity has to be absorbed on each end of the filter, a higher impedance can be built up in the filter for the same band width, and hence more gain per section can be obtained than with a tuned circuit.

The filter will have other pass bands at $2f_m$, $3f_m$, etc., but these do not usually cause any trouble at the short-wave frequencies because the gain in the vacuum tube is falling off very rapidly and no appreciable signal is passed. If desired, however, another band-pass filter can be added which has the same fundamental bands but different overtone bands, and this will eliminate the effect of the additional pass bands. For very narrow bands, the line length l_2 can be made longer and hence more realizable by making the characteristic impedance Z_{0_2} lower. If anything is to be gained by making the impedance on the plate side different from that on the grid side this can be accomplished by making the filter an impedance transforming device, as discussed in the next section.

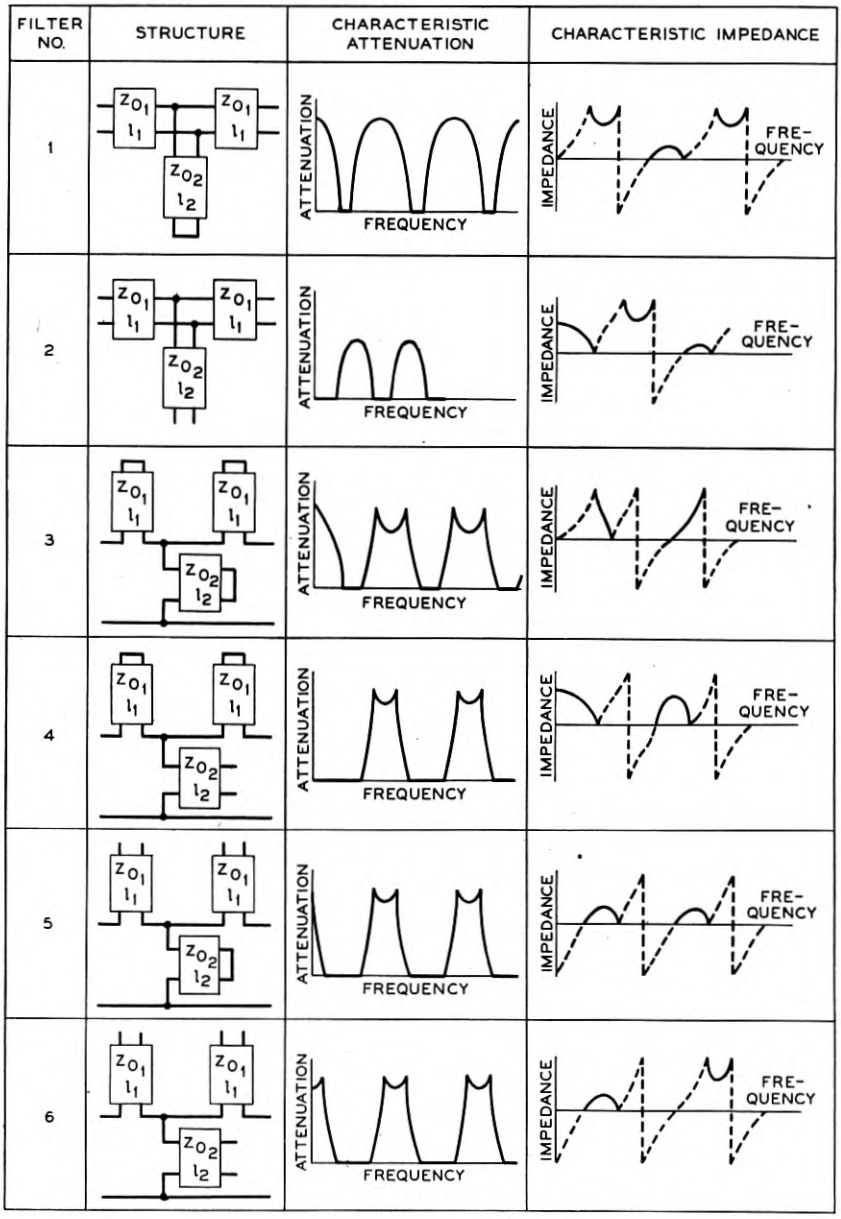


Fig. 4—A list of filters constructed from transmission lines.

The filter discussed above shows some of the possibilities and limitations of filters constructed from transmission lines. Many other types are also possible. Figure 4 lists a number of these and the types of filter characteristics they give. The design equations for a number of them are considered in detail in the patent referred to above and hence will not be worked out here.

IV. Impedance Transforming Band-Pass Filters Employing Transmission Lines as Elements

In a good many cases it is desirable to transform from one impedance to another over a wide range of frequencies. Examples of such uses are when antennas are connected to transmission lines, or when transmission lines are to be connected to vacuum tubes, etc. Transforming band-pass filters constructed out of transmission lines which will transform over wide frequency ranges are therefore of practical interest. Previously two types of single-frequency transformers, constructed

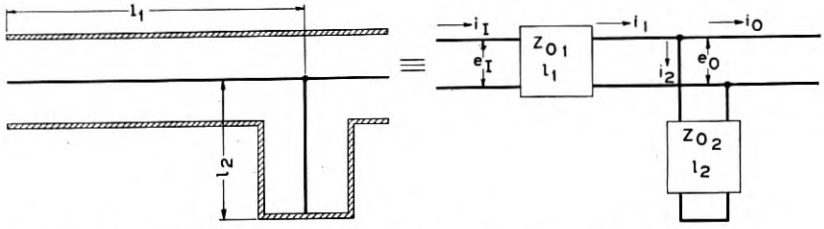


Fig. 5—A wide-band transformer—constructed from coaxial conductors.

from transmission lines, have been suggested [6] but these differ from the types investigated here in that they have a specified ratio of impedances for a single frequency only.

One of the simplest types of band transforming filters is shown by Fig. 5. It consists of a series transmission line of characteristic impedance Z_{0_1} and a shunt line having the characteristic impedance Z_{0_2}. Since the impedance of the short-circuited line is

$$jZ_{0_2} \tan \frac{\omega l_2}{v},$$

the equations for the structure are

$$\begin{aligned} e_1 &= e_I \cos\frac{\omega l_1}{v} - j i_I Z_{0_1} \sin\frac{\omega l_1}{v}; \qquad i_1 = i_2 + i_0; \qquad e_1 = e_0; \\ i_1 &= i_I \cos\frac{\omega l_1}{v} - j\frac{e_I}{Z_{0_1}} \sin\frac{\omega l_1}{v}; \qquad i_2 = -\frac{j e_0}{Z_{0_2} \tan\frac{\omega l_2}{v}}. \end{aligned} \tag{19}$$

[6] See reference 1, pages 430 and 431.

Combining these equations we have

$$e_0 = e_I \cos\frac{\omega l_1}{v} - j i_I Z_{0_1} \sin\frac{\omega l_1}{v},$$

$$i_0 = i_I \cos\frac{\omega l_1}{v}\left[1 + \frac{Z_{0_1}\tan\frac{\omega l_1}{v}}{Z_{0_2}\tan\frac{\omega l_2}{v}}\right] - j e_I \frac{\sin\frac{\omega l_1}{v}}{Z_{0_1}}\left[1 - \frac{Z_{0_1}\cot\frac{\omega l_1}{v}}{Z_{0_2}\tan\frac{\omega l_2}{v}}\right]. \quad (20)$$

In order to interpret this equation in terms of transformer theory, it can be shown that equations (20) are identical to the equations for a perfect transformer and a symmetrical filter. To show this, consider the circuit of Fig. 6, which consists of two half-sections of a symmetrical

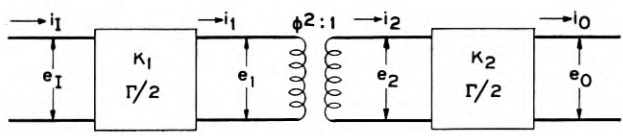


Fig. 6—Transformer and filter.

filter separated by a perfect transformer having an impedance step-down φ^2 to 1. The characteristic impedance of the first filter is φ^2 times that of the second filter. The equations for the first section, the transformer, and the last section are respectively

$$e_1 = e_1 \cosh\frac{\Gamma}{2} - i_I K_1 \sinh\frac{\Gamma}{2}; \quad i_1 = i_I \cosh\frac{\Gamma}{2} - \frac{e_I}{K_1}\sinh\frac{\Gamma}{2};$$

$$i_2 = \varphi i_1; \quad e_2 = e_1/\varphi; \quad (21)$$

$$e_0 = e_2 \cosh\frac{\Gamma}{2} - i_2 K_2 \sinh\frac{\Gamma}{2}; \quad i_0 = i_2 \cosh\frac{\Gamma}{2} - \frac{e_2}{K_2}\sinh\frac{\Gamma}{2}.$$

Combining these equations on the assumption that $K_1/K_2 = \varphi^2$, we have

$$e_0 = \frac{1}{\varphi}[e_I \cosh\Gamma - i_I K_1 \sinh\Gamma];$$

$$i_0 = \varphi\left[i_I \cosh\Gamma - \frac{e_I \sinh\Gamma}{K_1}\right]. \quad (22)$$

Similar results can be obtained by using the image impedance parameters of a dissymmetrical filter. For a dissymmetrical filter the most general relationship between the input and output voltages and currents can be written in the form

$$e_2 = e_1 A - i_1 B; \quad i_2 = i_1 C - e_1 D \quad \text{where} \quad AC - BD = 1. \tag{23}$$

In terms of these parameters the image transfer constant θ and the image transfer impedances are given by the relationships [7]

$$\cosh \theta = \sqrt{AC}; \quad K_1 = \sqrt{\frac{BC}{AD}}; \quad K_2 = \sqrt{\frac{AB}{CD}}. \tag{24}$$

The transformation ratio between the two ends of the network is then

$$\frac{K_1}{K_2} = \frac{C}{A} = \varphi^2 \tag{25}$$

in agreement with the results of equation (22).

Applying these results to equation (20) we find for the most general case

$$\cosh \theta = \cos \frac{\omega l_1}{v} \sqrt{1 + \frac{Z_{0_1} \tan \frac{\omega l_1}{v}}{Z_{0_2} \tan \frac{\omega l_2}{v}}};$$

$$K_1 = Z_{0_1} \sqrt{- \tan \frac{\omega l_1}{v} \left[\frac{\tan \frac{\omega l_1}{v} + \frac{Z_{0_2}}{Z_{0_1}} \tan \frac{\omega l_2}{v}}{1 - \frac{Z_{0_2}}{Z_{0_1}} \tan \frac{\omega l_1}{v} \tan \frac{\omega l_2}{v}} \right]};$$

$$\varphi^2 = 1 + \frac{Z_{0_1} \tan \frac{\omega l_1}{v}}{Z_{0_2} \tan \frac{\omega l_2}{v}}. \tag{26}$$

Two special cases are of interest here: the first when $Z_{0_1} = Z_{0_2}$, and the second when $l_1 = l_2$. The first case corresponds to the transformer disclosed by P. H. Smith,[8] which can be used to transform over a wide

[7] These relations were first proved for wave transmission networks in a paper of the writer's, *B. S. T. J.*, April 1927. See equations 67 and 68, page 291. They are proved by other methods in "Communication Networks," E. A. Guillemin, p. 139.

[8] See reference 1, p. 431.

range of impedance for a single frequency. For this case

$$\cosh\theta = \cos\frac{\omega l_1}{v}\sqrt{1+\frac{\tan\frac{\omega l_1}{v}}{\tan\frac{\omega l_2}{v}}};$$

$$K_1 = Z_{0_1}\sqrt{-\tan\frac{\omega l_1}{v}\tan\frac{\omega(l_1+l_2)}{v}}; \tag{27}$$

$$\varphi^2 = 1+\frac{\tan\frac{\omega l_1}{v}}{\tan\frac{\omega l_2}{v}}.$$

The pass band of the filter lies between the values

$$f_1 = \frac{v}{4(l_1+l_2)} \quad \text{and} \quad f_2 = \frac{v}{4l_1} \tag{28}$$

and hence between these frequencies the structure will act as a transformer. The ratio of the transformer, however, varies with frequency, and hence two given impedances can only be matched at one frequency. By adjusting the values of l_1 and l_2 it is possible to transform between any two resistances at a given frequency.

For a number of purposes it is desirable to transform a wide band of frequencies between two constant impedances. This requires a transformer with a constant transformation ratio over the whole band of frequencies. As can be seen from equation (26) this can be accomplished with the structure considered above if we let $l_1 = l_2 = l$. For this case

$$\cosh\theta = \sqrt{1+\frac{Z_{0_1}}{Z_{0_2}}}\cos\frac{\omega l}{v};$$

$$K_1 = Z_{0_1}\varphi\sqrt{\frac{1}{1-\frac{Z_{0_1}}{Z_{0_2}}\cot^2\frac{\omega l}{v}}}; \tag{29}$$

$$\varphi^2 = 1+\frac{Z_{0_1}}{Z_{0_2}}.$$

The mid-band frequency occurs when $\cosh\theta = \cos(\omega l/v) = 0$. Hence l is $\frac{1}{4}$ wave-length at the mid-band frequency. At the mid-band frequency the impedance of the transformer is

$$K_{1_0} = Z_{0_1}\varphi. \tag{30}$$

This is the impedance that the transformer should be connected to since

this is the characteristic impedance of the filter. The width of the pass band is determined by

$$1 \geqq \cosh \theta \geqq -1. \tag{31}$$

The cut-off frequencies of the filter transformer are given by the formula

$$\cos \frac{\omega l}{v} = \pm \frac{1}{\varphi}. \tag{32}$$

For relatively narrow bands, the ratio of the band width to the mean frequency is given by the simple formula

$$\frac{f_2 - f_1}{f_m} \doteq \frac{4}{\pi \varphi}, \tag{33}$$

as can readily be shown from equation (32) by using the approximation formula for the cosine in the neighborhood of the angle $\pi/2$.

Hence the structure shown in Fig. 5 is equivalent to a perfect transformer whose ratio is $\varphi^2 = 1 + (Z_{01}/Z_{02})$ to 1 and a filter whose band width is given by equation (33). Such a filter will have a flat attenuation loss over about 80 per cent of its theoretical band when it is terminated on each side by resistances equal to $Z_{01}\varphi$ and Z_{01}/φ on its high- and low-impedance sides respectively. Due to the high Q obtainable in the transmission lines, the loss in the band of the filter can be made very low and hence such a transformer will introduce a very small transmission loss. Furthermore, since it is constructed only of transmission lines, it can carry a large amount of power. The complete design equations for the transformer are

$$\varphi^2 = 1 + \frac{Z_{01}}{Z_{02}}; \qquad Z_{01} = \frac{R_I}{\varphi}; \qquad Z_{02} = \frac{R_I}{\varphi(\varphi^2 - 1)};$$
$$l = \frac{v}{4f_m}; \qquad R_0 = \frac{R_I}{\varphi^2}, \tag{34}$$

where R_I and R_0 are respectively the input and the output resistances that the transformer works between.

Many other types of transforming networks containing only transmission lines are also possible. Another simple network which is the inverse of the one considered above is shown in Fig. 7. It consists of a length of line l_1 and characteristic impedance Z_{01} in series with a balanced open-circuited line of length l_2 and characteristic impedance Z_{02}. It is easily shown by employing the equations for a line that

when $l_1=l_2=l$

$$e_0 = e_I\left(1+\frac{Z_{02}}{Z_{01}}\right)\cos\frac{\omega l}{v} - ji_I Z_{01}\sin\frac{\omega l}{v}\left[1-\frac{Z_{02}}{Z_{01}}\cot^2\frac{\omega l}{v}\right];$$
$$i_0 = i_I\cos\frac{\omega l}{v} - j\frac{e_I}{Z_{01}}\sin\frac{\omega l}{v}. \tag{35}$$

For this case

$$\varphi^2 = \frac{1}{1+\frac{Z_{02}}{Z_{01}}};\quad \cosh\theta = \frac{\cos\frac{\omega l}{v}}{\varphi};\quad K_1 = Z_{01}\varphi\sqrt{1-\frac{Z_{02}}{Z_{01}}\cot^2\frac{\omega l}{v}}. \tag{36}$$

The band width of the filter for narrow bands is given approximately by

$$\frac{f_2-f_1}{f_m} \doteq \frac{4\varphi}{\pi}.$$

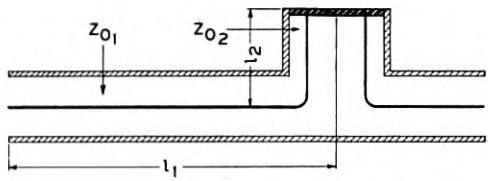


Fig. 7—A wide-band transformer constructed from coaxial conductors.

The design equations for the transformer are

$$\varphi^2 = \frac{1}{1+\frac{Z_{02}}{Z_{01}}};\qquad Z_{01} = R_I/\varphi;\qquad Z_{02} = R_I(1-\varphi^2)/\varphi^3;$$
$$l = v/4f_m;\qquad R_0 = R_I/\varphi^2. \tag{37}$$

The first transformer discussed is a step-down transformer while the one considered here has a step up from the input of the line to the output. The first filter had a mid-shunt impedance characteristic on each end, i.e., the impedance of the band is infinity at the two edges; whereas the transformer with the series open-circuited line has a mid-series impedance, since the impedance given by the last expression in equation (36) goes to zero at the two cut-off frequencies. The range of transformation is about the same for each and hence one type has no particular advantage over the other.

It is often desirable in filter work to be able to have the impedance of one end of the filter somewhat different from that of the other, i.e., to have the filter act as a transformer of a moderate ratio. An example of this occurs when using a structure composed of short lengths of

line to connect two high-frequency pentodes. As shown by Salzberg and Burnside,[9] the output impedance of a high-frequency pentode at 100 megacycles may be in the order of 30,000 ohms due to the high-frequency shunting loss of the tube. On the other hand, due to active grid loss, the impedance looking into the grid of the next tube may be in the order of 20,000 ohms. Hence in order to obtain the most gain from such tubes, the coupling circuit should be able to work from 30,000 ohms when connected with the output to 20,000 ohms for the input of the next tube. If we employ the simple coupling circuit shown in Section III, Fig. 1, this can be made impedance transforming by making the second-series conductor of a different characteristic impedance from the first-series conductor. Such a combination is shown in Fig. 8. The equations of the combination are easily solved

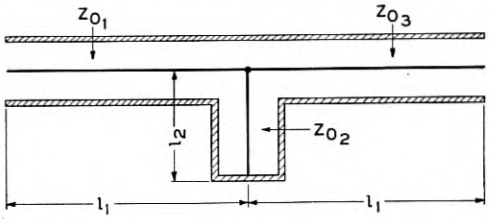


Fig. 8—A narrow-band transformer for coupling vacuum tubes.

with the result that

$$i_c = i_I\left[\cos^2\frac{\omega l_1}{v} - \frac{Z_{0_1}}{Z_{0_3}}\sin^2\frac{\omega l_1}{v} + \frac{Z_{0_1}}{Z_{0_2}}\frac{\sin\frac{\omega l_1}{v}\cos\frac{\omega l_1}{v}}{\tan\frac{\omega l_2}{v}}\right] - je_I\left[\frac{Z_{0_1}+Z_{0_3}}{Z_{0_1}Z_{0_3}}\sin\frac{\omega l_1}{v}\cos\frac{\omega l_1}{v} - \frac{\cos^2\frac{\omega l_1}{v}}{Z_{0_2}\tan\frac{\omega l_2}{v}}\right];$$

$$e_0 = e_I\left[\cos^2\frac{\omega l_1}{v} - \frac{Z_{0_3}}{Z_{0_1}}\sin^2\frac{\omega l_1}{v} + \frac{Z_{0_3}}{Z_{0_2}}\frac{\sin\frac{\omega l_1}{v}\cos\frac{\omega l_1}{v}}{\tan\frac{\omega l_2}{v}}\right] - ji_I\left[(Z_{0_1}+Z_{0_3})\sin\frac{\omega l_1}{v}\cos\frac{\omega l_1}{v} + \frac{Z_{0_1}Z_{0_3}}{Z_{0_2}}\frac{\sin^2\frac{\omega l_1}{v}}{\tan\frac{\omega l_2}{v}}\right]. \quad (38)$$

[9] "Recent Developments in Miniature Tubes," *Proc. I. R. E.*, Vol. 23, No. 10, p. 142, Oct. 1935.

For this case the image transfer constant and the impedance ratio of the transforming filter are given by the equations

$$\cosh\theta = \sqrt{\cos^2\frac{2\omega l_1}{v} + \left(\frac{Z_{01}+Z_{03}}{2Z_{02}}\right)\frac{\sin\frac{2\omega l_1}{v}\cos\frac{2\omega l_1}{v}}{\tan\frac{\omega l_2}{v}} + \sin^2\frac{2\omega l_1}{v}\left[\frac{Z_{01}Z_{03}}{4Z_{02}^2\tan^2\frac{\omega l_2}{v}} - \frac{(Z_{01}-Z_{03})^2}{4Z_{01}Z_{03}}\right]}; \tag{39}$$

$$\varphi^2 = \left[\frac{\cos^2\frac{\omega l_1}{v} - \frac{Z_{01}}{Z_{03}}\sin^2\frac{\omega l_1}{v} + \frac{Z_{01}}{Z_{02}}\frac{\sin\frac{\omega l_1}{v}\cos\frac{\omega l_1}{v}}{\tan\frac{\omega l_2}{v}}}{\cos^2\frac{\omega l_1}{v} - \frac{Z_{03}}{Z_{01}}\sin^2\frac{\omega l_1}{v} + \frac{Z_{03}}{Z_{02}}\frac{\sin\frac{\omega l_1}{v}\cos\frac{\omega l_1}{v}}{\tan\frac{\omega l_2}{v}}}\right].$$

When we solve the expression for $\cosh\theta$ for the cut-off frequencies, we find that one of them is given by the expression

$$f_2 = \frac{v}{4l_1} \tag{40}$$

and the other by

$$\tan\frac{2\omega l_1}{v} = \frac{-2\left(\frac{Z_{01}+Z_{03}}{Z_{01}Z_{03}}\right)Z_{02}\tan\frac{\omega l_2}{v}}{1-\frac{(Z_{01}+Z_{03})^2}{Z_{01}^2Z_{03}^2}Z_{02}^2\tan^2\frac{\omega l_2}{v}}. \tag{41}$$

If we consider the special case $Z_{02} = Z_{01}Z_{03}/(Z_{01}+Z_{03})$ equation (41) reduces to the simple form

$$\tan\frac{2\omega l_1}{v} = -\tan\frac{2\omega l_2}{v} \tag{42}$$

or for narrow bands

$$\frac{2\omega_1 l_1}{v} = \pi - \frac{2\omega_1 l_2}{v} \quad \text{and} \quad f_1 = \frac{v}{4(l_1+l_2)} \tag{43}$$

in agreement with equation (17).

At the two cut-off frequencies, it can be shown that

$$\varphi^2 = \frac{Z_{01}^2}{Z_{03}^2} \tag{44}$$

and throughout the band the value of φ does not differ much from this value for narrow-band filters.

Hence the structure of Fig. 8 acts as a narrow-band coupling unit which introduces a transformation from input to output. For narrow bands it is easily shown that the image impedance at the middle of the pass band is given by the expression

$$K_1 = \frac{4}{\pi}\frac{(f_m)}{(f_2 - f_1)}\sqrt{Z_{01}Z_{03}}\frac{Z_{01}}{Z_{03}}. \tag{45}$$

The design equations for this transforming filter are

$$l_1 = \frac{v}{4f_2}; \qquad l_2 = \frac{v}{4}\left[\frac{1}{f_1} - \frac{1}{f_2}\right]; \qquad Z_{01} = \frac{\pi}{4}\frac{R_I(f_2 - f_1)}{\sqrt{\varphi} f_m};$$

$$Z_{03} = \frac{\pi}{4}\frac{R_I(f_2 - f_1)}{f_m\varphi^{\frac{1}{2}}}; \qquad Z_{02} = \frac{\pi R_I(f_2 - f_1)}{4f_m\varphi^{\frac{1}{2}}(1 + \varphi)}; \qquad \frac{R_I}{R_0} = \varphi^2. \tag{46}$$

V. Filters and Transformers Employing Transmission Lines and Condensers

Condensers can be constructed for high radio frequencies which have little dissipation and hence they can be combined with short sections of transmission lines to produce filters and transformers. Combinations of lines with condensers have the advantages that much more isolated bands can be obtained, in general narrower pass bands can be obtained, and at the lower frequencies shorter sections of lines can be employed if they are resonated by capacities. Also, when such structures are used as interstage coupling units working between vacuum tubes, they usually will have to incorporate the grid to filament and plate to filament capacities as part of the coupling circuit. Hence it is desirable to consider combinations of transmission lines and condensers as filters and transformers.

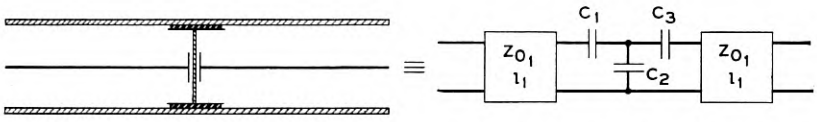


Fig. 9—A transformer or filter using condensers and coaxial conductors.

One of the simplest and most useful types of band-pass filter using transmission lines and condensers is shown in Fig. 9. This structure

has been used as a wide-band transformer and as a very narrow-band filter, and experimental curves are given in the next section. The equations connecting the output voltage and current with the input voltage and current can easily be calculated by the methods given above and are

$$e_0 = e_I\left[\frac{C_2+C_3}{C_3}\cos^2\frac{\omega l}{v} - \frac{Z_{02}}{Z_{01}}\left(\frac{C_1+C_2}{C_2}\right)\sin^2\frac{\omega l}{v} + \frac{\sin\frac{2\omega l}{v}}{2}\left(\frac{C_1+C_2+C_3}{\omega Z_{01}C_1C_3} - \omega C_2 Z_{02}\right)\right]$$
$$- ji_I Z_{01}\left[\frac{\sin\frac{2\omega l}{v}}{2}\left[\frac{C_2+C_3}{C_3} + \frac{Z_{02}}{Z_{01}}\left(\frac{C_1+C_2}{C_1}\right)\right] - \left[\frac{C_1+C_2+C_3}{\omega Z_{01}C_1C_3}\cos^2\frac{\omega l}{v} + Z_{02}\omega C_2\sin^2\frac{\omega l}{v}\right]\right];$$
$$(47)$$

$$i_0 = i_I\left[\frac{C_1+C_2}{C_1}\cos^2\frac{\omega l}{v} - \frac{Z_{01}}{Z_{02}}\left(\frac{C_2+C_3}{C_3}\right)\sin^2\frac{\omega l}{v} + \frac{\sin\frac{2\omega l}{v}}{2}\left[\frac{C_1+C_2+C_3}{\omega Z_{02}C_1C_3} - Z_{01}\omega C_2\right]\right]$$
$$- \frac{je_I}{Z_{01}}\left[\frac{\sin\frac{2\omega l}{v}}{2}\left[\frac{C_1+C_2}{C_1} + \frac{Z_{01}}{Z_{02}}\left(\frac{C_2+C_3}{C_3}\right)\right] + \sin^2\frac{\omega l}{v}\left[\frac{C_1+C_2+C_3}{\omega Z_{02}C_1C_3}\right] + \omega C_2 Z_{01}\cos^2\frac{\omega l}{v}\right].$$

In order that the structure shall transform uniformly over a band of frequencies we must have

$$\varphi^2 = \left(\frac{C_1+C_2}{C_1}\right)\left(\frac{C_3}{C_2+C_3}\right) = \frac{Z_{01}}{Z_{02}}$$
$$= \text{impedance transformation ratio.} \quad (48)$$

With this substitution, equations (47) simplify to

$$e_0 = e_I\left[\frac{C_2+C_3}{C_3}\cos\frac{2\omega l}{v}+\frac{\sin\frac{2\omega l}{v}}{2}\left[\frac{C_1+C_2+C_3}{\omega Z_{0_1}C_1C_3}-\omega C_2Z_{0_2}\right]\right]$$
$$-ji_IZ_{0_1}\left[\left(\frac{C_2+C_3}{C_3}\right)\sin\frac{2\omega l}{v}\right.$$
$$\left.-\left[\frac{C_1+C_2+C_3}{\omega Z_{0_1}C_1C_3}\cos^2\frac{\omega l}{v}+Z_{0_2}\omega C_2\sin^2\frac{\omega l}{v}\right]\right];$$

$$i_0 = \varphi^2\left[i_I\left[\frac{C_2+C_3}{C_3}\cos\frac{2\omega l}{v}\right.\right. \tag{49}$$
$$\left.+\frac{\sin\frac{2\omega l}{v}}{2}\left(\frac{C_1+C_2+C_3}{\omega Z_{0_1}C_1C_3}-Z_{0_2}\omega C_2\right)\right]$$
$$-\frac{je_I}{Z_{0_1}}\left[\left(\frac{C_2+C_3}{C_3}\right)\sin\frac{2\omega l}{v}+\frac{C_1+C_2+C_3}{\omega Z_{0_1}C_1C_3}\sin^2\frac{\omega l}{v}\right.$$
$$\left.\left.+\omega C_2Z_{0_2}\cos^2\frac{\omega l}{v}\right]\right].$$

Comparing these equations with equations (24) we have for the image parameters

$$\cosh\theta = \varphi\left[\frac{C_2+C_3}{C_3}\cos\frac{2\omega l}{v}\right.$$
$$\left.+\frac{\sin\frac{2\omega l}{v}}{2}\left[\frac{C_1+C_2+C_3}{\omega Z_{0_1}C_1C_3}-\omega C_2Z_{0_2}\right]\right];$$

$$K_1 = Z_{0_1}\sqrt{\frac{\frac{C_2+C_3}{C_3}\sin\frac{2\omega l}{v}-\left[\frac{C_1+C_2+C_3}{\omega Z_{0_1}C_1C_3}\cos^2\frac{\omega l}{v}+Z_{0_2}\omega C_2\sin^2\frac{\omega l}{v}\right]}{\frac{C_2+C_3}{C_3}\sin\frac{2\omega l}{v}+\frac{C_1+C_2+C_3}{\omega Z_{0_1}C_1C_3}\sin^2\frac{\omega l}{v}+Z_{0_2}\omega C_2\cos^2\frac{\omega l}{v}}}; \tag{50}$$

$$K_2 = \frac{K_1}{\varphi^2}.$$

These equations give the image parameters for a general transforming band-pass filter. The two uses to which such a structure will ordinarily be put are either to obtain a transformer with as wide a pass band as possible for a given impedance transformation or else to obtain a filter without transformation ratio. For the transformer case it can be shown that the widest pass-band occurs when $C_3 \rightarrow \infty$, or in other words the condenser C_3 is short-circuited. In order to obtain a simple design, it is assumed that each conductor is an eighth of a wave-length at the mid-band frequency or that

$$\frac{\omega_m l}{v} = \frac{\pi}{4}. \tag{51}$$

The mid-band frequency occurs when $\cosh \theta = 0$. Upon substituting the relation $Z_{0_1} = l/vC_0$ where C_0 is the total distributed capacity of the input line of the transformer, $\cosh \theta$ vanishes when

$$\frac{C_0}{C_1} = \frac{\pi^2}{16} \frac{C_2}{\varphi^2 C_0}. \tag{52}$$

Solving for the frequencies for which $\cosh \theta = \pm 1$, it is easily shown that the ratio of the band width to the mean frequency is given by the expression

$$\frac{f_2 - f_1}{f_m} \doteq \frac{\frac{4}{\pi\varphi}}{1 + \frac{C_2}{2\varphi^2 C_0}}. \tag{53}$$

The image impedance K_1 at the mid-frequency of the band is from equation (50)

$$K_{1_0} = Z_{0_1} \sqrt{\frac{1 - \frac{\pi}{4} \frac{C_2}{\varphi^2 C_0}}{1 + \frac{\pi}{4} \frac{C_2}{\varphi^2 C_0}}}. \tag{54}$$

From the above equations and noting that $\varphi^2 = 1 + C_2/C_1$, the design equations of the transformer become

$$Z_{0_1} = R_1 \sqrt{\frac{\varphi + \sqrt{\varphi^2 - 1}}{\varphi - \sqrt{\varphi^2 - 1}}}; \qquad Z_{0_2} = \frac{Z_{0_1}}{\varphi^2}; \qquad l = \frac{v}{8f_m};$$

$$C_0 = \frac{33.3l}{Z_{0_1}} \text{ in } \mu\mu\text{f}; \qquad C_1 = \frac{4C_0}{\pi} \sqrt{\frac{\varphi^2}{\varphi^2 - 1}}; \tag{55}$$

$$C_2 = \frac{4C_0}{\pi} \sqrt{(\varphi^2 - 1)\varphi^2},$$

where R_1 is the input impedance from which the transformer must work.

When the structure of Fig. 9 is used as a filter without transformation, we have $C_1 = C_3$; $Z_{0_1} = Z_{0_2} = Z_0$. For this case the image parameters become

$$\cosh\theta = \frac{C_1 + C_2}{C_1}\cos\frac{2\omega l}{v} + \frac{\sin\frac{2\omega l}{v}}{2}\left[\frac{2C_1 + C_2}{\omega Z_0 C_1^2} - \omega C_2 Z_0\right];$$

$$K_1 = K_2 = K = Z_0\sqrt{\frac{1 - \left[\frac{(2C_1 + C_2)\cot\frac{\omega l_1}{v}}{2\omega Z_0 C_1(C_1 + C_2)} + \frac{Z_0\omega C_1 C_2}{2(C_1 + C_2)}\tan\frac{\omega l_1}{v}\right]}{1 + \frac{(2C_1 + C_2)\tan\frac{\omega l_1}{v}}{2\omega Z_0 C_1(C_1 + C_2)} + \frac{Z_0\omega C_1 C_2}{2(C_1 + C_2)}\cot\frac{\omega l_1}{v}}}. \quad (56)$$

For narrow-band filters it is easily shown that

$$\frac{f_2 - f_1}{f_m} = \Delta = \frac{\frac{4}{\pi}}{1 + \frac{C_2}{C_1} + \frac{C_2}{2C_0}}; \qquad \frac{(2C_1 + C_2)C_0}{C_1^2} = \frac{\pi^2}{16}\frac{C_2}{C_0}. \quad (57)$$

At the mid-band of the filter, since the constants were worked out on the assumption that each conductor was an eighth of a wave-length at the mid-band frequency, the mid-band filter impedance can be obtained from the last part of (56) by setting $\omega l/v = \pi/4$, giving

$$K_0 = Z_0\sqrt{\frac{1 + \frac{C_2}{C_1} - \frac{\pi}{4}\frac{C_2}{C_0}}{1 + \frac{C_2}{C_1} + \frac{\pi}{4}\frac{C_2}{C_0}}}. \quad (58)$$

Hence solving for the constants of the filter on the assumption that Δ is a small quantity, we find

$$C_1 = \frac{4C_0}{\pi}; \qquad C_2 = \frac{16C_0}{\pi(2 + \pi)\Delta}; \qquad Z_0 = \frac{8R}{(\pi + 2)\Delta};$$

$$l = \frac{v}{8f_m}; \qquad C_0 = \frac{33.3l}{Z_0}\text{ in }\mu\mu\text{f}, \quad (59)$$

where R is a resistance equal to K at the mean-frequency of the filter. For narrow bands this gives a very large value for C_2, the shunt capacity. A more practical arrangement is to replace the two series condensers C_1 and the shunt condenser C_2 by a π network consisting of two shunt condensers C_A separated by a series condenser C_B. These have the values

$$C_A = \frac{C_1 C_2}{2C_1 + C_2}; \qquad C_B = \frac{C_1^2}{2C_1 + C_2}. \tag{60}$$

With this arrangement we find that for narrow bands $C_A \doteq C_1$ and C_B is a very small capacity. This can readily be obtained physically by inserting a partition with a small hole in it at the middle of the section. Then C_A will be the capacity of the inside conductors to the partition, and C_B will be the capacity of one inside conductor to the other looking through the small hole. By adjusting the size of this hole, this capacity can be made as small as desired.

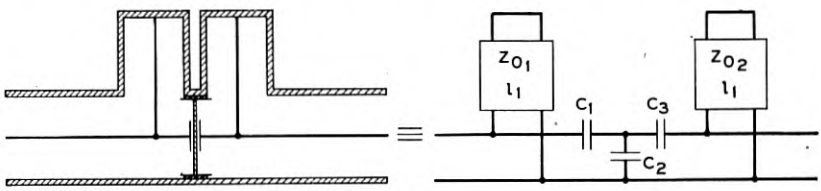


Fig. 10—A shunt terminated transformer.

The transformer discussed above is suitable for transforming from line impedances down to very low impedances, but cannot be used to transform from line impedances up to very high impedances such as the impedance of a vacuum tube. This generally requires a shunt type of termination rather than the series type discussed above. One such transformer is shown in Fig. 10. It consists of two shunt lines connected together by a T or π network of capacities. The constants of such a transformer are readily calculated, and for the condition of maximum transformation for a given band width—which occurs when $C_3 \to \infty$, and for eighth wave-length conductors on each end—these have been found to be

$$\varphi^2 = 1 + \frac{C_2}{C_1} = \frac{Z_{01}}{Z_{02}}; \qquad Z_{01} = \frac{K_{10}}{\varphi}; \qquad Z_{02} = \frac{K_{10}}{\varphi^3};$$
$$C_1 = \frac{4C_0}{\pi}; \qquad C_2 = \frac{4}{\pi} C_0(\varphi - 1); \qquad l = \frac{v}{8f_m}. \tag{61}$$

The theoretical band width for this type transformer is given approximately by the expression

$$\frac{f_2 - f_1}{f_m} \doteq \frac{4}{\varphi(\pi + 2)}. \tag{62}$$

Such a transformer is also suitable for connecting together vacuum tubes of high impedance.

Another type of transformer of some interest is one which will transform from very high impedances to very low impedances. Such a transformer is shown in Fig. 11. It has a shunt conductor on the high-impedance end and a series conductor on the low-impedance end. Such a transformer does not have a constant transformation ratio over the whole band, but for about 80 per cent of the theoretical band width

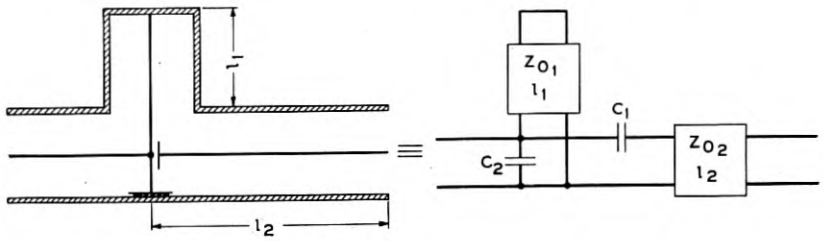


Fig. 11—A series shunt type transformer suitable for wide bands.

the transformation ratio is approximately constant. The design equations for such a transformer are

$$\begin{aligned}
&\frac{K_{10}}{K_{20}} = \varphi^2; \quad f_m = \sqrt{f_1 f_2}; \quad Z_{01} = \frac{K_{10}(f_2 - f_1)}{f_m}; \quad l_1 = \frac{v}{8f_m}; \\
&Z_{02} = \frac{K_{20} f_m}{(f_2 - f_1)\sqrt{1 - 1/\varphi}}; \quad l_2 = \frac{v\sqrt{1 - 1/\varphi}}{2\pi f_m}; \\
&C_1 = \frac{(f_2 - f_1)\varphi}{2\pi f_m^2 K_{10}}; \\
&C_2 = \frac{1}{2\pi K_{10}(f_2 - f_1)}\left[1 - \frac{(f_2 - f_1)^2(\varphi - 1)}{f_m^2}\right].
\end{aligned} \tag{63}$$

The transformer is especially useful since it will give the widest transmission band for a given transformation ratio of any of the transformers discussed.

Many other types of filters, transforming and nontransforming, are also possible using transmission lines and condensers. A partial list of such filters is shown in Fig. 12 together with their attenuation and

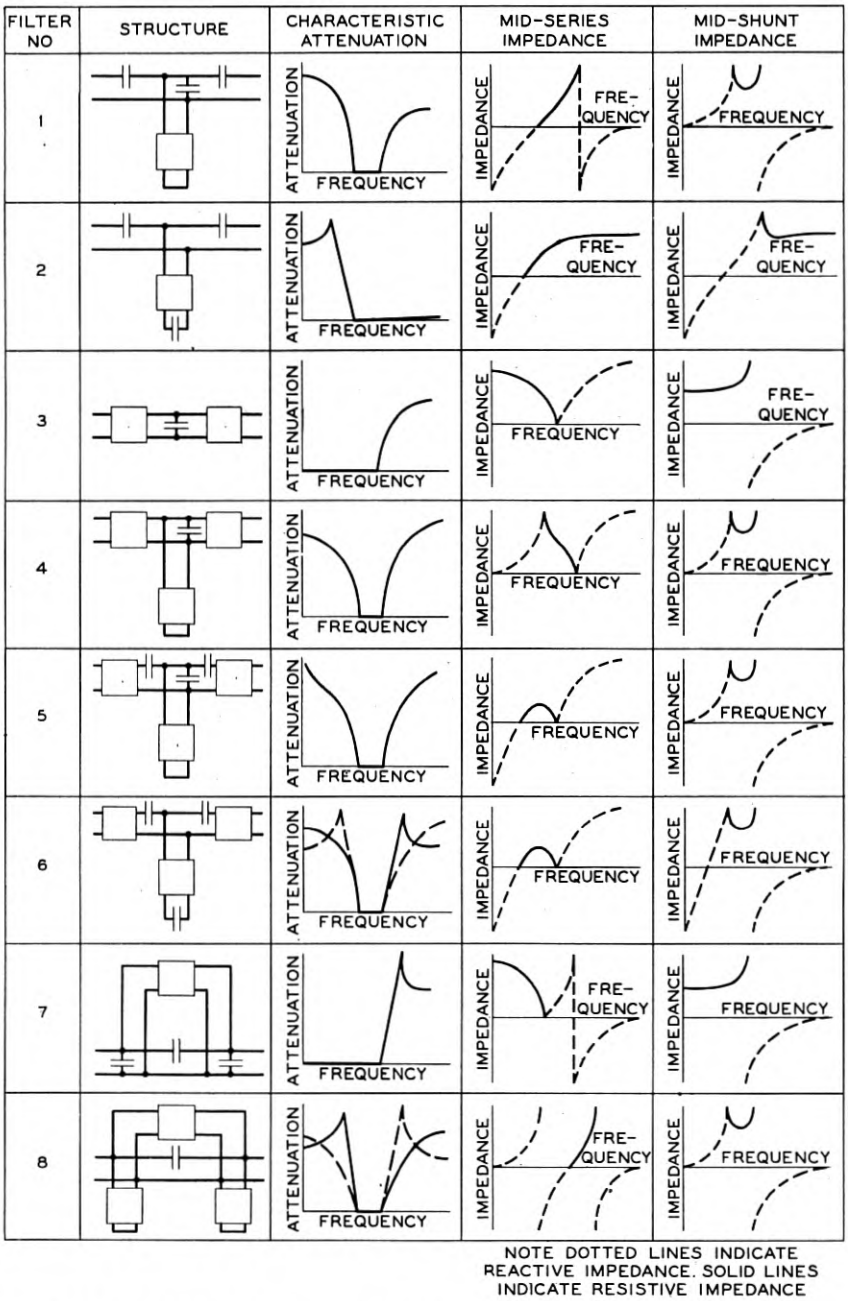


Fig. 12—A list of filter structures employing transmission lines and condensers.

iterative impedance characteristics. All of the band-pass filters can be made impedance-transforming by varying the ratios of the impedance elements for the two ends.

VI. Experimental Results

Several filters and transformers of the types discussed above have been tested experimentally and have been found to give operating characteristics in accordance with the calculated results. Figure 13

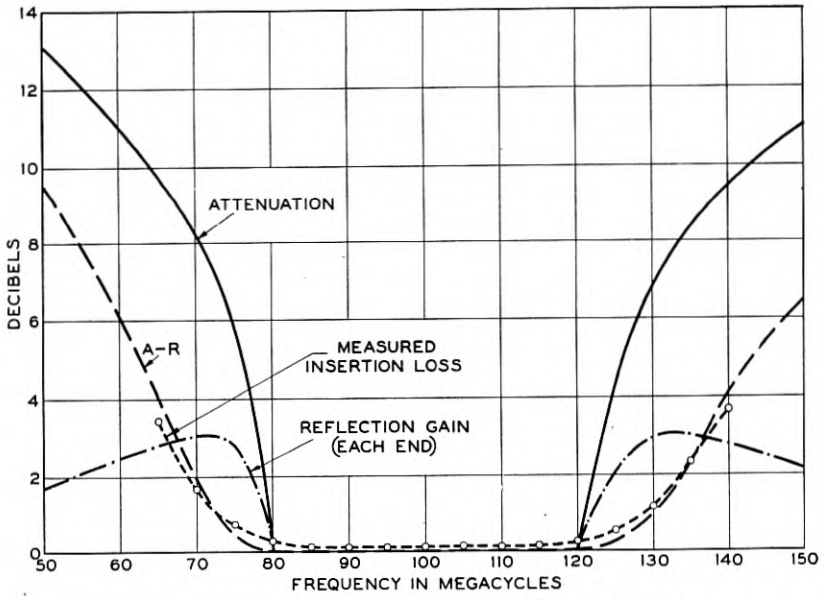


Fig. 13—Measured insertion loss of a wide-band transformer.

shows the measured insertion loss of a wide-band low-impedance transformer which transforms from 70 ohms down to an impedance of 17.5 ohms. The useful transformation band is from 80 megacycles to 120 megacycles. The measuring circuit is shown in Fig. 14. It consists

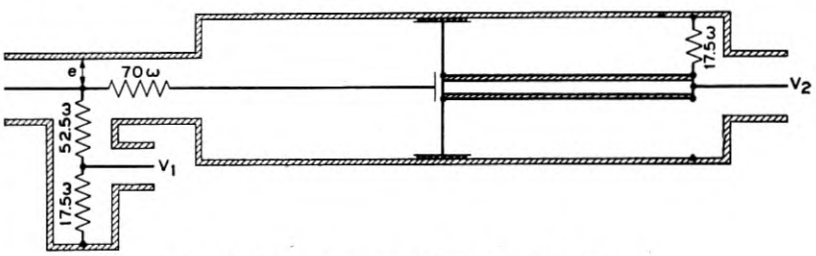


Fig. 14—Measuring circuit for a transformer.

of a source of high-frequency voltage impressed on a divided circuit. In one branch is a series resistance of 70 ohms connected to the 70-ohm side of the transformer. The output of the transformer is connected

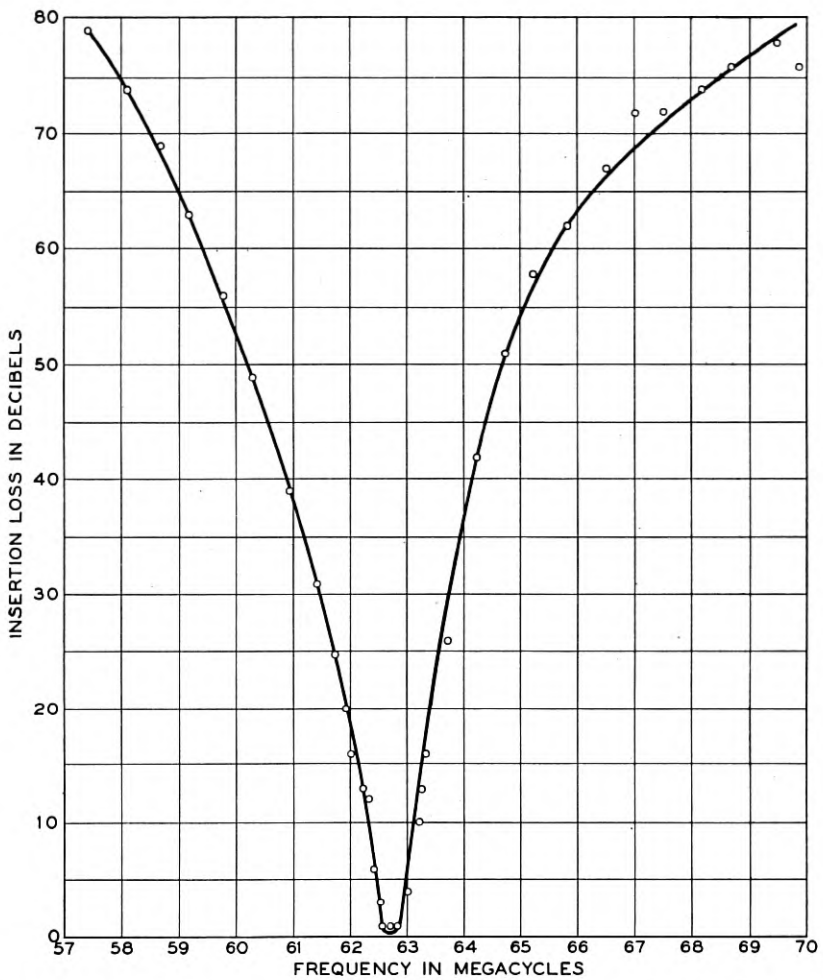


Fig. 15—Measured insertion loss of filter used on Green Harbor–Provincetown radio link.

to a 17.5 ohm resistance and a high-impedance voltmeter is connected across it. The other branch contains a 52.5-ohm and a 17.5-ohm resistance in series, with a high-impedance voltmeter shunted across the 17.5 ohms. If the transformer were a perfect transformer, the

current from the output of the transformer would be

$$i_0 = \frac{e}{140} \times \sqrt{\frac{70}{17.5}} = \frac{e}{70}, \tag{64}$$

where e is the voltage applied to ground at the common point. The voltage across the output should be

$$e_0 = \frac{e}{70} \times 17.5 = \frac{e}{4}. \tag{65}$$

But this is just the voltage that should occur across the voltmeter V_1. Hence the difference in reading between V_2 and V_1 will be a measure of the loss introduced by the transformer. From Fig. 13 we see that this is in the order of 0.1 db, which represents a small loss for a transformer.

Several of the narrow-band filters of the type shown in section V, Fig. 9, have also been constructed and tested. One of these has been used on an experimental radio system at Green Harbor, Massachusetts, since 1935, for the purpose of connecting a transmitter and receiver on the same antenna. This filter has been constructed and tested by Messrs. F. A. Polkinghorn and N. J. Pierce using the design data developed here. The filter used consisted of three sections of the type shown in Fig. 9 connected in tandem. The resulting insertion loss of the filter and associated transformers is shown in Fig. 15. The loss at mid-band is in the order of 1 db and an insertion loss of over 50 db is obtained 2 megacycles on either side of the center of the pass band.

A Ladder Network Theorem

By JOHN RIORDAN

The theorem of this paper gives four-terminal representation of ladder networks satisfying a prescribed condition on the side impedances, in terms of the three parameters specifying the network connected as a transducer, the driving-point impedance between short-circuited transducer terminal pairs, and an impedance ratio involving the side impedances only. This mode of representation has a special advantage in applications to electric railway networks in that the transducer parameters which alone involve the ladder shunt impedances (under the stated conditions) may be calculated in a relatively simple fashion, and extensive networks reduced to manageable form. The theorem is stated and proved, and its applications are sketched in some detail.

THE theorem of this paper gives a four-terminal representation of ladder networks satisfying a certain condition with respect to the side impedances. Ladder networks appearing in transmission and filter theory generally are connected as transducers, that is, such that the entry and exit terminals on the ladder sides are associated in pairs; the networks are two-terminal pairs. As is well known, passive transducers may be completely specified by three parameters (as is the case for three-terminal networks, with which transducers are similar in some, though not all, respects), the choice of which has been the occasion for much study and ingenuity.[1] The present theorem does not assume transducer connection and is thus quite distinct from earlier work; indeed it arose outside the communication field in the problem of the calculation of short-circuit currents and network current distribution of electric railway networks, where at present it seems to have chief application.

This paper gives a statement of the theorem, an indication of its applications, and finally its proof.

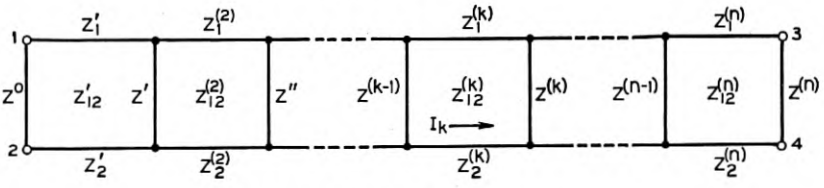
The Theorem

A ladder network, composed of any number of arbitrary shunt impedances forming sections whose side impedances $Z_1^{(k)}$, $Z_2^{(k)}$ *and* $Z_{12}^{(k)}$, $k = 1$,

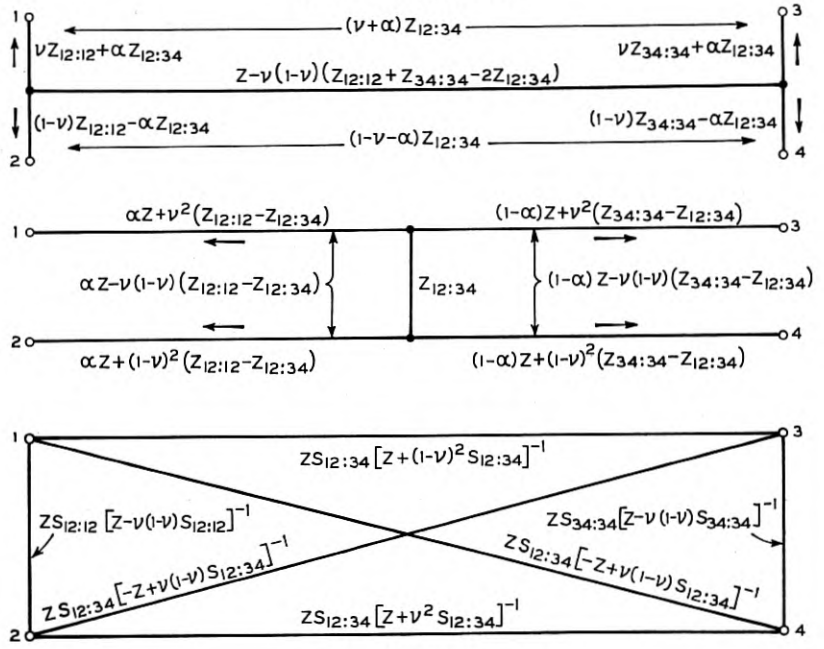
[1] Five types of equivalent networks by which a transducer may be replaced, including T, π, transformer and artificial line networks, and their interrelations are given on Table I of "Cisoidal Oscillations" by G. A. Campbell, *Trans. A. I. E. E.*, **30**, pp. 873–909 (1911). The most significant addition to the table would appear to be the image impedance representation due to O. J. Zobel.

$2 \cdots$, *are such that* $[Z_1^{(k)} - Z_{12}^{(k)}][Z_2^{(k)} - Z_{12}^{(k)}]^{-1}$ *is a constant, may be completely specified by its three transducer parameters (with transducer terminal pairs each made up of adjacent terminals on opposite ladder sides*[2]*), the driving-point impedance between short-circuited transducer*

A. Network Diagram



B. Network Equivalents



Notation

$$\nu = \nu_k = [Z_1^{(k)} - Z_{12}^{(k)}][Z_1^{(k)} + Z_2^{(k)} - 2Z_{12}^{(k)}]^{-1}$$

$$Z = \sum_1^n \frac{Z_1^{(k)}Z_2^{(k)} - (Z_{12}^{(k)})^2}{Z_1^{(k)} + Z_2^{(k)} - 2Z_{12}^{(k)}}$$

α = Arbitrary Constant

$$S_{12:12} = (Z_{12:12}Z_{34:34} - Z^2_{12:34})/Z_{34:34}$$

$$S_{34:34} = (Z_{12:12}Z_{34:34} - Z^2_{12:34})/Z_{12:12}$$

$$S_{12:34} = (Z_{12:12}Z_{34:34} - Z^2_{12:34})/Z_{12:34}$$

Fig. 1—A sample ladder network illustrating notation; and some network equivalents.

[2] The theorem also holds when the terminals of each pair are non-adjacent, that is, for terminal pairs 1, 4 and 3, 2 of Fig. 1*A*; this result is of no importance in the railway applications.

terminal pairs, and the constant

$$[Z_1^{(k)} - Z_{12}^{(k)}][Z_2^{(k)} - Z_{12}^{(k)}]^{-1}.$$

The current in any branch of the network for any condition of energization of the network terminals is a linear function of the currents in the same branch for energization at sending and receiving transducer terminals.

It should be observed that the result stated is independent of the number or values of the shunt impedances (except as they are included in the transducer parameters); hence in the diagram on Fig. 1*A* illustrating the ladder network in question, any of the shunt impedances may be allowed to vanish or become infinite, and their number $n + 1$ may be increased or decreased at pleasure provided that one shunt remains (this excludes the trivial case in which, the sides being completely insulated from each other, the network degenerates to a pair of single impedances).

When the impedances of the sides are linearly extended impedances, as is the case in electric railway applications, the section impedances may be written:

$$Z_1^{(k)} = s_k z_1,$$
$$Z_2^{(k)} = s_k z_2,$$
$$Z_{12}^{(k)} = s_k z_{12},$$

where z_1, z_2 and z_{12} are self and mutual impedances of the sides per unit length. The condition, $[Z_1^{(k)} - Z_{12}^{(k)}][Z_2^{(k)} - Z_{12}^{(k)}]^{-1} = \text{const.}$, is replaced by the condition that the shunt impedances connect corresponding points on the sides.

Since a four-terminal network requires six independent quantities for its specification, the conditions (a) that the network be of ladder type and (b) that the given section impedance ratio be constant may be regarded, at least intuitively, as replacing two (or more) of the measurable impedances at the terminals.[3]

With the sending and receiving transducer terminal pairs short-circuited, and with the side impedances satisfying the given condition, no current flows in any of the shunt impedances and the driving-point impedance required for the theorem is

$$Z = \sum_{k=1}^{n} \frac{Z_1^{(k)} Z_2^{(k)} - Z_{12}^{(k)2}}{Z_1^{(k)} + Z_2^{(k)} - 2Z_{12}^{(k)}} = \frac{\sum Z_1^{(k)} \sum Z_2^{(k)} - (\sum Z_{12}^{(k)})^2}{\sum Z_1^{(k)} + \sum Z_2^{(k)} - 2\sum Z_{12}^{(k)}}, \qquad (1)$$

[3] It is interesting to observe that, if short circuits between terminals are permitted, there are 64 measurable impedances for a four-terminal network. The network may be specified by any six of these which are independent; hence the number of ways of specifying the network is something less than the number of combinations of 64 things taken 6 at a time, which equals 74,974,368. The number of non-independent sets which make up this large total appears at the moment to be the smaller part, and possibly a very small part indeed. These remarks are inspired by Mr. R. M. Foster.

where the summations in the last expression extend over all the sections; this impedance then is simply the parallel impedance of the sides taken in their entirety.

The current in branch k of line 2, designated by I_k on Fig. 1A, for any condition of energization is expressed in terms of the currents in the same branch and in the same direction for unit current supplied between terminals 1 and 2, and 3 and 4 [terminals 3, 4 (1, 2) open, respectively], designated by $i_{k:12}$ and $i_{k:34}$, respectively, by the following equation:

$$I_k = \nu(I_3 + I_4) - i_{k:12}[\nu I_1 - (1 - \nu)I_2] - i_{k:34}[\nu I_3 - (1 - \nu)I_4], \quad (2)$$

where I_1, I_2, I_3 and I_4 are currents flowing out of the network from the respective terminals, and ν is the current in side 2 for unit current between short-circuited transducer terminals, as given on Fig. 1B. Thus I_k is a linear function of currents $i_{k:12}$ and $i_{k:34}$, as stated in the second half of the theorem.

Three types of networks completely equivalent to any ladder network satisfying the condition of the theorem are shown on Fig. 1B. The transducer impedances employed in the representation by these networks are the driving-point impedances between transducer terminals 1 and 2, and 3 and 4 [terminals 3, 4 (1, 2) open, respectively] and the corresponding transfer impedance between the ends of the transducer. These impedances are designated $Z_{12:12}$, $Z_{34:34}$ and $Z_{12:34}$, following a notation for Neumann integrals used by G. A. Campbell.[4] For present purposes the notation has the advantage of putting into evidence the terminals between which current is supplied and the terminals between which voltage is measured; thus $Z_{12:12}$ may be read as the voltage drop from 1 to 2 for unit current from 1 to 2 (terminals 3, 4 open), $Z_{12:34}$ the voltage drop from 3 to 4 for unit current from 1 to 2 under the same conditions.[5] By the reciprocity theorem $Z_{12:34} = Z_{34:12}$.

[4] "Mutual Impedance of Grounded Circuits," *Bell System Technical Journal*, 2, 1–30 (Oct. 1923).

[5] Further, the subscripts may be handled algebraically to give results following from the superposition theorem. For this purpose the numbers in each part of the two-part subscript are taken as separated by a minus sign and the colon is taken as a sign of multiplication; thus:

$$Z_{12:12} = Z_{(1-2)(1-2)} = Z_{11} + Z_{22} - 2Z_{12},$$

the last expression being formed by writing out the indicated product and separating the terms. The equation expresses the fact that the impedance of a circuit may be subdivided into the self-impedances of sides (real or fictional) associated with its terminals minus twice their mutual impedance. Moreover, any additional subscripts desired may be intercalated by adding and subtracting the same numeral; the expansion of bracketed terms then gives a relation between circuit impedances; thus:

$$\begin{aligned} Z_{(1-2)(1-2)} &= Z_{[(1-3)+(3-2)][(1-3)+(3-2)]} \\ &= Z_{(13+32)(13+32)} \\ &= Z_{13:13} + Z_{32:32} + 2Z_{13:32}. \end{aligned}$$

The first two equivalent networks are of the H type; [6] as seven impedances are shown on each, whereas only six are required for complete representation, an arbitrary constant α has been introduced so that the mutual impedance of the uprights may be varied at pleasure. Thus in the first H network, the condition $\alpha + \nu = 0$ puts all the mutual impedance between the uprights below the crossbar; the condition $1 - \nu - \alpha = 0$ puts it all above. The same type of shift may be made in the second H network.

The third equivalent is the network of direct impedances ("Cisoidal Oscillations," loc. cit. designation (b)); these are expressed in terms of the transducer parameters with opposite pairs of terminals short-circuited, which following Campbell are denoted by S's. Thus $S_{12:12}$ is the driving-point impedance between terminals 1 and 2 with terminals 3 and 4 short-circuited; $S_{12:34}$ is the ratio of current from 3 to 4 to voltage from 1 to 2 with 1, 2 energized and 3, 4 short-circuited, or its reciprocity theorem equivalent.

These three equivalents correspond respectively to transformer, T and π transducer equivalent networks. For the first H type the transducer condition that currents into terminals 1 and 2, and 3 and 4, shall be equal and opposite entails zero current in the H crossbar, which may be removed, leaving a transformer connection. For the second H type the transducer condition allows grouping the impedances of branches 1 and 2 and their mutual impedance, and of 3 and 4 and their mutual impedance, into single branches, say, branches 1 and 3, which gives the T equivalent network. The reduction of the direct impedance network is not so immediate.

Periodic Ladder Networks

When the network is periodic, the transducer impedances and current distribution may be expressed completely in terms of the

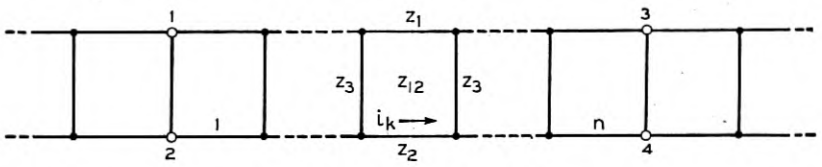
The justification of the operation lies in the fact that, as regards the current half of the subscript, a unit current from 1 to 2 is equivalent by the superposition theorem to unit currents 1 to 3 and 3 to 2 and similarly the voltage 1 to 2 for unit current 1 to 2 is the same as the sum of voltages 1 to 3 and 3 to 2. Thus the notation is a shorthand for application of the superposition theorem. Its use is illustrated further in the course of the proof of the theorem.

[6] This is a form of equivalent network falling under designation (c) of the list of equivalents for an arbitrary number of terminals given by G. A. Campbell ("Cisoidal Oscillations," p. 889, loc. cit.), which is described as branches radiating from a common concealed point, one to each of the terminals, with mutual impedances between pairs. This is not a unique representation since the number of elements is redundant, and the set of mutual impedances may be given values appropriate for particular purposes provided that the self-impedances are adjusted correspondingly. In the present application the mutual impedances of branches to terminals 1 and 4, and 2 and 3, have been set at zero and the mutual impedances of branches 1 and 2 and 3 and 4 in the first H diagram, and of branches 1 and 3, 2 and 4 in the second, have been eliminated in forming the cross bar of the H.

section impedances, giving a certain concreteness to the application of the theorem, which may be valuable. Partly on this account, and partly because periodic iterative impedances are in themselves useful in applications, a particular type of periodic network is considered in this section.

The network considered is infinite in extent, with section series impedances z_1, z_2 and z_{12} and shunt impedance z_3, as shown on Fig. 2*A*.

A. Network Diagram



B. Network Impedances $Z_{ab:cd}$

cd \ ab	12	13	14	24
12	K			
13	$\nu(K-T)$	$Z+2\nu^2(K-T)$		
14	$T+\nu(K-T)$	$Z-\nu(1-2\nu)(K-T)$	$Z+K-2\nu(1-\nu)(K-T)$	
23	$-K+\nu(K-T)$	$Z-\nu(1-2\nu)(K-T)$	$Z-T-2\nu(1-\nu)(K-T)$	$Z+(1-\nu)(1-2\nu)(K-T)$
24	$-(1-\nu)(K-T)$	$Z-2\nu(1-\nu)(K-T)$	$Z+(1-\nu)(1-2\nu)(K-T)$	$Z+2(1-\nu)^2(K-T)$
34	T	$-\nu(K-T)$	$(1-\nu)K+\nu T$	$(1-\nu)(K-T)$

C. Current Distribution

Unit Current Between:	i_k
1 and 2	$-\frac{K_1}{K_1+K_2}e^{-k\alpha}$
1 and 3	$\nu-\frac{K_1}{K_1+K_2}[\nu e^{-k\alpha}+\nu e^{-(n+1-k)\alpha}]$
1 and 4	$\nu-\frac{K_1}{K_1+K_2}[\nu e^{-k\alpha}-(1-\nu)e^{-(n+1-k)\alpha}]$
2 and 3	$\nu+\frac{K_1}{K_1+K_2}[(1-\nu)e^{-k\alpha}-\nu e^{-(n+1-k)\alpha}]$
2 and 4	$\nu+\frac{K_1}{K_1+K_2}[(1-\nu)e^{-k\alpha}+(1-\nu)e^{-(n+1-k)\alpha}]$
3 and 4	$\frac{K_1}{K_1+K_2}e^{-(n+1-k)\alpha}$

Fig. 2—Periodic ladder network of infinite extent; network diagram, impedances and current distribution.

The infinite network is the simplest to formulate since there are no points of reflection; it is of course symmetrical with respect to the terminal pairs 1, 2 and 3, 4.

The impedance across either of these terminal pairs is the parallel impedance of the full-series and full-shunt iterative impedances (or one-half the mid-shunt iterative impedance). The full-series and full-shunt iterative impedances are given by the following formulas:

$$\begin{aligned} \textit{Full-series}\ K_1 &= \tfrac{1}{2}[\sqrt{z(z+4z_3)} + z] = K_2 + z, \\ \textit{Full-shunt}\ K_2 &= \tfrac{1}{2}[\sqrt{z(z+4z_3)} - z] = \frac{K_1 z_3}{K_1 + z_3}, \end{aligned} \tag{3}$$

where, for brevity, $z = z_1 + z_2 - 2z_{12}$.

Then

$$Z_{12:12} = Z_{34:34} = K = \frac{K_1 K_2}{K_1 + K_2} = z_3 \left[1 + \frac{4z_3}{z} \right]^{-1/2}. \tag{4}$$

The voltage across lines is propagated as $\exp(-k\alpha)$ where α is the section propagation constant; hence

$$Z_{12:34} = T = Ke^{-n\alpha}. \tag{5}$$

The propagation factor $\exp(-\alpha)$ is defined in terms of the iterative impedances by

$$e^{-\alpha} = \frac{K_2}{K_1}. \tag{6}$$

The currents $i_{k:12}$ and $i_{k:34}$ are given by the following formulas:

$$i_{k:12} = -\frac{K_1}{K_1 + K_2} e^{-k\alpha}, \tag{7}$$

$$i_{k:34} = \frac{K_1}{K_1 + K_2} e^{-(n+1-k)\alpha}. \tag{8}$$

This completes the formulation, since the remaining quantities, ν and Z, are given immediately by

$$\nu = \frac{z_1 - z_{12}}{z_1 + z_2 - 2z_{12}},$$

$$Z = \sum_{k=1}^{n} \frac{z_1 z_2 - z_{12}^2}{z_1 + z_2 - 2z_{12}} = n \frac{z_1 z_2 - z_{12}^2}{z_1 + z_2 - 2z_{12}}.$$

Figure 2*B* shows driving-point and transfer impedances for energization between terminals, omitting certain impedances equal by symmetry. Figure 2*C* shows the corresponding k-section currents in side 2.

Applications to Electric Railroad Networks

A.-c. electric railroad networks in one-line diagram are predominantly of the ladder type. The series elements of sides 1 and 2 represent, for two-wire networks, impedances of sections of transmission lines and traction circuits, respectively; the shunt elements represent transformer impedances. For three-wire networks, the series elements may represent trolley-feeder (or feeder-rail) and trolley-rail impedance elements, the shunt elements autotransformer impedances.

The theorem may be used for representing portions of a network or a whole network of ladder form,[7] when the series impedances satisfy the condition of the theorem. As the circuits are linearly extended this is almost always the case except where the traction circuits change character, from two to four tracks, for example. For approximate purposes the H networks may be used even in these cases provided that the parameter ν is properly chosen. In many cases the transfer impedance $Z_{12:34}$ is negligible and a value of ν may be associated with each pair of terminals; the values for the sections immediately adjoining the terminal pairs 1–2 and 3–4 (sections 1 and n on Fig. 1A) are of dominant importance and serve for rough purposes. If the transfer impedance is not negligible a mean of these values may be sufficiently accurate.

In two-wire networks, generator circuits are connected directly to the transmission line (side 1), and the short circuits of chief interest (grounding points on the one-line diagram) are those on the traction circuits (side 2). Thus, for a single generating point the network is energized between points on sides 1 and 2, such as 1 and 4, for example; if the impedance in the generator connection is Z_g and the impedance of the short circuit is zero, the short-circuit driving-point impedance and the traction circuit currents are as follows:

$$\begin{aligned} Z_0 &= Z_g + Z_{14:14} \\ &= Z_g + Z + \nu^2 Z_{12:12} + (1-\nu)^2 Z_{34:34} + 2\nu(1-\nu)Z_{12:34}, \end{aligned} \tag{9}$$

$$I_k = [\nu + \nu i_{k:12} + (1-\nu) i_{k:34}] \frac{E}{Z_0}, \tag{10}$$

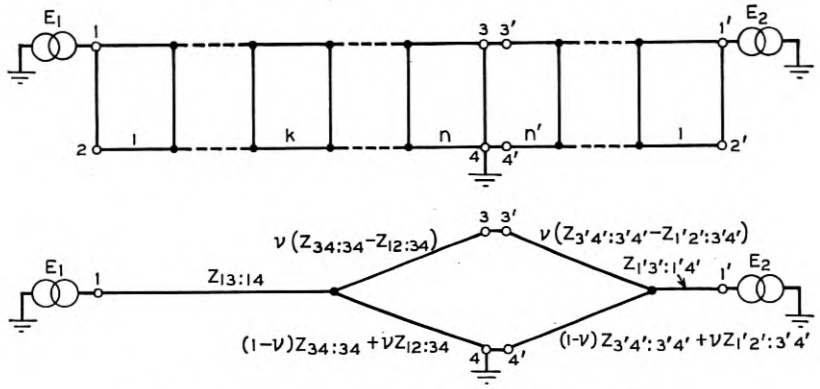
where E is the generator voltage. The impedance may be obtained immediately from either of the H networks—as the sum of the self-

[7] For multiple transmission-line two-wire networks the ladder form is obtained when all transmission lines are bussed at all generating stations and substations or when the generators, step-up transformers and substation transformers connected to each line are of similar impedances and are similarly connected. When these conditions are not met the network is of multiple-side ladder form, for the representation of which an extension of the theorem would be required. Similar remarks apply to three-wire networks.

impedances of legs 1 and 4 and of the crossbar. The current expression follows from equation (2) with $I_2 = I_3 = 0$; $-I_1 = I_4 = E/Z_0$.

For multiple generating points (or for multiple points of short

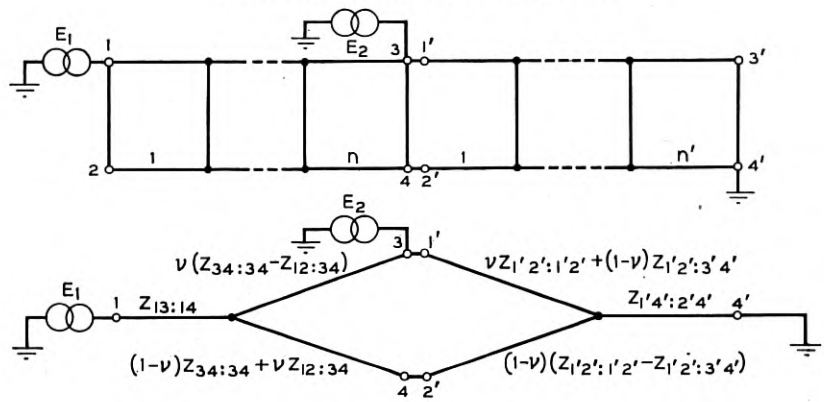
A. Short Circuit Between Generator Points



$$Z_{13:14} = Z + \nu^2 (Z_{12:12} - Z_{12:34}) - \nu\,(1-\nu)(Z_{34:34} - Z_{12:34})$$

$$Z_{1'3':1'4'} = Z' + \nu^2 (Z_{1'2':1'2'} - Z_{1'2':3'4'}) - \nu\,(1-\nu)\,(Z_{3'4':3'4'} - Z_{1'2':3'4'})$$

B. Short Circuit Beyond Generator Points



$$Z_{1'4':2'4'} = Z' - \nu\,(1-\nu)\,(Z_{1'2':1'2'} - Z_{1'2':3'4'}) + (1-\nu)^2\,(Z_{3'4':3'4'} - Z_{1'2':3'4'})$$

Fig. 3—Equivalent networks for electrified railways; two-wire system, two sources.

circuit) the theorem may be used to represent portions of the network. Examples for the case of two generators are shown on Fig. 3; on Fig. 3*A* the short circuit is located between generator points, on Fig. 3*B* beyond them. On Fig. 3*A* the network is supplied with duplicate pairs of terminals at the short-circuit point, separated by an infinitesimal

difference; the parts of the original network thus formed are represented by Y-connected impedances which may be derived from the first H network. On Fig. 3B the network is broken at the intermediate generator point and similarly represented. A similar process may be followed for any number of generator points but in some cases it may be expedient to superpose additional generators; the network impedances required for superposition may be formulated in the manner followed in the proof of the theorem.

The solution of these reduced networks supplies the currents I_1, I_3, I_4; I_1', I_3', I_4', etc., from which the current in branch k of side 2 of any of the ladder sections may be found from equation (2). Thus, for example the current I_k in the kth section of the ladder network with terminals 1, 2, 3 and 4 on Figs. 3A and 3B is formulated as follows:

$$I_k = [\nu + \nu i_{k:12} - \nu i_{k:34}]I_3 + [\nu + \nu i_{k:12} + (1 - \nu)i_{k:34}]I_4, \quad (11)$$

which follows from equation (2) with $I_2 = 0$; $-I_1 = I_3 + I_4$. Similar formulas apply to the other two ladder networks on Fig. 3.

In three-wire networks, generators are usually connected to the traction network by three-winding transformers which may be represented on the network diagram by three impedances connected in star. The traction network may be represented on a trolley-feeder, trolley-rail or a feeder-rail, trolley-rail base and it is well known that the three-winding transformer equivalent impedances for the two bases are related. Using the notation shown on Figs. 4A and 4A', with primes distinguishing the feeder-rail, trolley-rail base, the relations are as follows:

$$\begin{aligned} V_{tf}Z_a &= V_{tf}Z_a' + V_{tr}Z_b', \\ V_{tf}Z_b &= V_{fr}Z_b', \\ V_{tf}^2Z_c &= V_{fr}^2Z_c' - V_{tr}V_{fr}Z_b', \end{aligned} \quad (12)$$

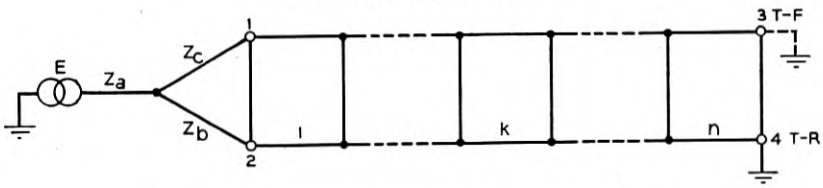
where V_{tr}, V_{tf} and V_{fr} are the trolley-rail, trolley-feeder and feeder-rail circuit voltages, respectively.

From Figs. 4B and 4B' showing the reduced networks for trolley-rail short-circuits on the two bases for a single source feed, it is apparent that the impedances involved in the equivalent networks must be similarly related. The relations are found to be as follows:

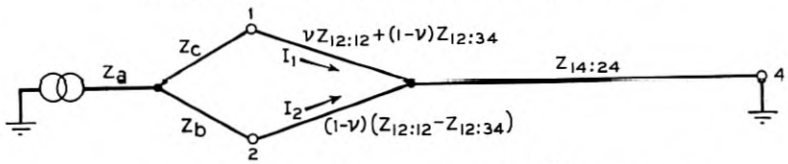
$$\begin{aligned} V_{tf}^2Z_{12:12} &= V_{fr}^2Z'_{12:12}, \\ V_{tf}^2Z_{34:34} &= V_{fr}^2Z'_{34:34}, \\ V_{tf}^2Z_{12:34} &= V_{fr}^2Z'_{12:34}, \\ V_{fr}(1 - \nu) &= V_{tf}(1 - \nu'), \\ Z &= Z', \\ i_{k:12} &= i'_{k:12}, \\ i_{k:34} &= i'_{k:34}. \end{aligned} \quad (13)$$

TROLLEY-FEEDER, TROLLEY-RAIL BASE

A. Actual Network Connections

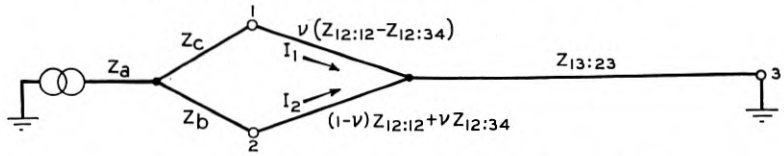


B. Equivalent Network for Trolley-Rail Short Circuits



$$Z_{14:24} = Z - \nu(1-\nu)(Z_{12:12} - Z_{12:34}) + (1-\nu)^2 (Z_{34:34} - Z_{12:34})$$

C. Equivalent Network for Trolley-Feeder Short Circuits



$$Z_{13:23} = Z - \nu(1-\nu)(Z_{12:12} - Z_{12:34}) + \nu^2 (Z_{34:34} - Z_{12:34})$$

Fig. 4—Equivalent networks for electrified railways; three-wire system, single source, trolley-feeder, trolley-rail and feeder-rail, trolley-rail bases.

Thus the complete set of short-circuit currents (trolley-rail, trolley-feeder and feeder-rail short circuits) may be made from a single determination of the transducer impedances and current distributions on either of the two bases, whenever the theorem is applicable.

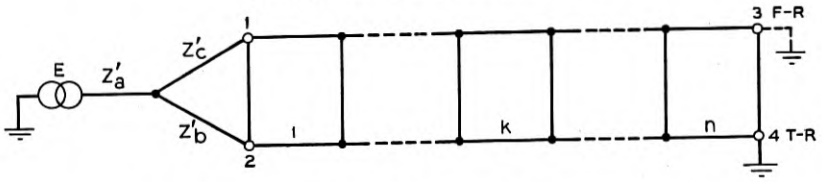
For multiple generator three-wire systems, and for three-wire systems with auxiliary transmission lines, the theorem may be used to represent portions of the network, possibly broken as in the two-wire cases illustrated above, four-terminal representation being necessary in general. The application follows the lines indicated above.

PROOF OF THEOREM

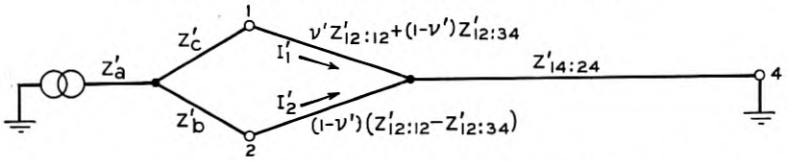
For energization between terminals 1 and 2, the sum of the currents in sides 1 and 2 at any point on the ladder is zero; the current $i_{k:12}$ may be taken as flowing in a mesh made up of the k-section sides and its terminating shunt impedances. For unit current supplied, the

FEEDER-RAIL, TROLLEY-RAIL BASE

A'. Actual Network Connections

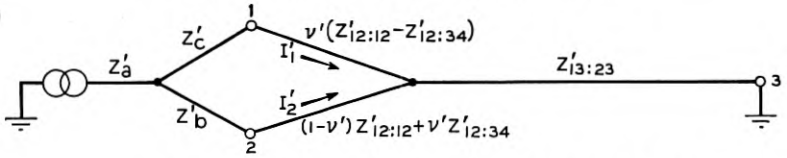


B'. Equivalent Network for Trolley-Rail Short Circuits



$$Z'_{14:24} = Z'-\nu'(1-\nu')(Z'_{12:12}-Z'_{12:34})+(1-\nu')^2(Z'_{34:34}-Z'_{12:34})$$

C'. Equivalent Network for Feeder-Rail Short Circuits



$$Z'_{13:23} = Z'-\nu'(1-\nu')(Z'_{12:12}-Z'_{12:34})+(\nu')^2(Z'_{34:34}-Z'_{12:34})$$

Fig. 4—*Continued from page 346.*

driving-point impedance between terminals 1, 2 and the transfer impedance to terminals 3, 4 for the network shown on Fig. 1*A* may be formulated immediately as:

$$\begin{aligned} Z_{12:12} &= (1 + i_{1:12})Z^0, \\ Z_{12:34} &= -\, i_{n:12}Z^{(n)}. \end{aligned} \tag{14}$$

The positive sense for currents $i_{1:12}$ and $i_{n:12}$ is taken as indicated on Fig. 1*A*, namely, in the direction from terminal 2 to terminal 4 on side 2.

From the voltage equation around the loop formed from sides 1 and 2 in their entirety and the terminal shunt impedances, the difference of these impedances may be expressed by:

$$Z_{12:12} - Z_{12:34} = -\sum_{k=1}^{n} (Z_1^{(k)} + Z_2^{(k)} - 2Z_{12}^{(k)})i_{k:12}. \tag{15}$$

The transfer impedances with respect to the side terminals 1, 3 and 2, 4 are formulated as:

$$Z_{12:13} = -\sum_{k=1}^{n} (Z_1^{(k)} - Z_{12}^{(k)}) i_{k:12},$$
$$Z_{12:24} = \sum_{k=1}^{n} (Z_2^{(k)} - Z_{12}^{(k)}) i_{k:12}. \tag{16}$$

From the condition $[Z_1^{(k)} - Z_{12}^{(k)}][Z_2^{(k)} - Z_{12}^{(k)}]^{-1} = \text{const.}$, a constant ν may be defined such that:

$$\nu = \nu_k = [Z_1^{(k)} - Z_{12}^{(k)}][Z_1^{(k)} + Z_2^{(k)} - 2Z_{12}^{(k)}]^{-1},$$
$$1 - \nu = [Z_2^{(k)} - Z_{12}^{(k)}][Z_1^{(k)} + Z_2^{(k)} - 2Z_{12}^{(k)}]^{-1},$$

and equations (16) may be combined with (15) to give:

$$Z_{12:13} = \nu Z_{12:12} - \nu Z_{12:34},$$
$$Z_{12:24} = -(1 - \nu) Z_{12:12} + (1 - \nu) Z_{12:34}. \tag{17}$$

The remaining transfer impedances follow by superposition; thus

$$Z_{12:14} = Z_{12:13} + Z_{12:34} = \nu Z_{12:12} + (1 - \nu) Z_{12:34},$$
$$Z_{12:23} = Z_{12:13} - Z_{12:12} = -(1 - \nu) Z_{12:12} - \nu Z_{12:34}. \tag{18}$$

It may be observed that

$$Z_{12:24} = Z_{12:13} + Z_{12:34} - Z_{12:12}. \tag{19}$$

Similarly

$$Z_{34:13} = -\nu Z_{34:34} + \nu Z_{12:34},$$
$$Z_{34:14} = (1 - \nu) Z_{34:34} + \nu Z_{12:34},$$
$$Z_{34:23} = -\nu Z_{34:34} - (1 - \nu) Z_{12:34},$$
$$Z_{34:24} = (1 - \nu) Z_{34:34} - (1 - \nu) Z_{12:34}. \tag{20}$$

These impedances, together with $Z_{34:34}$ and $Z_{34:12}$, form a set of 12 impedances of which only five are independent. There are three independent impedances determined by energization at each pair of terminals, including $Z_{12:34}$ and $Z_{34:12}$, which are equal by the reciprocity theorem; one independent set, for example, is $Z_{12:12}$, $Z_{12:13}$, $Z_{12:34}$, $Z_{34:34}$, $Z_{34:13}$. Consequently the network may be completely specified by the addition of a single impedance; for the set illustrated, the impedance required is $Z_{13:13}$.

This impedance may be formulated as:

$$\begin{aligned} Z_{13:13} &= \sum Z_1^{(k)} - \sum (Z_1^{(k)} - Z_{12}^{(k)}) i_{k:13} \\ &= \sum Z_1^{(k)} - \nu \sum (Z_1^{(k)} + Z_2^{(k)} - 2Z_{12}^{(k)}) i_{k:13}, \end{aligned} \tag{21}$$

where the summations extend as above from 1 to n.

Writing the equation around the loop used in deriving equation (15) it is found that:

$$\begin{aligned} &\sum (Z_1^{(k)} + Z_2^{(k)} - 2Z_{12}^{(k)}) i_{k:13} \\ &\qquad = \sum (Z_1^{(k)} - Z_{12}^{(k)}) - Z_{13:12} + Z_{13:34} \\ &\qquad = \sum (Z_1^{(k)} - Z_{12}^{(k)}) - \nu (Z_{12:12} + Z_{34:34} - 2Z_{12:34}), \end{aligned} \tag{22}$$

the last step being made by use of the reciprocity theorem and the formulas already developed. Thus, finally:

$$Z_{13:13} = Z + \nu^2 (Z_{12:12} + Z_{34:34} - 2Z_{12:34}), \tag{23}$$

where

$$Z = \sum_{k=1}^{n} \frac{Z_1^{(k)} Z_2^{(k)} - (Z_{12}^{(k)})^2}{Z_1^{(k)} + Z_2^{(k)} - 2Z_{12}^{(k)}}.$$

As already mentioned, Z is the impedance between short-circuited terminals 1, 2 and 3, 4; this may be verified in a number of ways.

The remaining impedances follow by superposition, which can be carried out formally through the impedance notation in the manner suggested. There are 21 driving-point and transfer impedances between terminals which can be displayed in a triangular array similar to that shown on Fig. 2*B*. The additional measurable impedances at the terminals arise as follows: 36 from short-circuiting two terminals, 4 from short-circuiting three terminals and 3 from short-circuiting terminals in pairs.

Equation (22) may also be written in terms of currents $i_{k:12}$ and $i_{k:34}$, since $Z_{13:12}$ and $Z_{13:34}$ may be expressed in terms of the latter; this suggests the following relation:

$$i_{k:13} = \nu + \nu i_{k:12} - \nu i_{k:34}. \tag{24}$$

The relation is verified by substituting into the mesh equations for currents $i_{k:13}$; the typical equation is as follows:

$$\begin{aligned} - Z^{(k-1)} i_{k-1:13} + [Z_1^{(k)} + Z_2^{(k)} - 2Z_{12}^{(k)} + Z^{(k-1)} + Z^{(k)}] i_{k:13} \\ - Z^{(k)} i_{k+1:13} = Z_1^{(k)} - Z_{12}^{(k)}. \end{aligned}$$

The remaining current relations then follow by superposition, as follows:

$$\begin{aligned} i_{k:14} &= i_{k:13} + i_{k:34} = \nu + \nu i_{k:12} + (1 - \nu) i_{k:34}, \\ i_{k:23} &= i_{k:13} - i_{k:12} = \nu - (1 - \nu) i_{k:12} - \nu i_{k:34}, \\ i_{k:24} &= i_{k:14} - i_{k:12} = \nu - (1 - \nu) i_{k:12} + (1 - \nu) i_{k:34}. \end{aligned} \tag{25}$$

It will be observed that only three of the six currents $i_{k:12}$, $i_{k:13}$, $i_{k:14}$, $i_{k:23}$, $i_{k:24}$ and $i_{k:34}$ are independent; one independent set is $i_{k:12}$, $i_{k:34}$ and $i_{k:13}$. Hence any arbitrary set of currents I_1, I_2, I_3 and I_4 flowing out of the network at the terminals may be resolved into three flows, such as those illustrated in the independent set above, which leads to equation (2).

The first H network may be obtained in the following manner. The value of $Z_{12:13}$, namely, $\nu Z_{12:12} - \nu Z_{12:34}$, in conjunction with the condition $Z_1 + Z_2 = Z_{12:12}$, Z_1 and Z_2 being the impedances of branches to terminals 1 and 2, respectively, suggests the following values of branch self and mutual impedances:

$$\begin{aligned} Z_1 &= \nu Z_{12:12}, \\ Z_2 &= (1 - \nu) Z_{12:12}, \\ Z_{13} &= \nu Z_{12:34}. \end{aligned}$$

The value of Z_{13} is verified by inspection of $Z_{13:34}$ if

$$Z_3 = \nu Z_{34:34}.$$

Similarly, by inspection of $Z_{12:14}$ and $Z_{34:32}$, the values of Z_{24} and Z_4 may be tentatively set at

$$\begin{aligned} Z_{24} &= (1 - \nu) Z_{12:34}, \\ Z_4 &= (1 - \nu) Z_{34:34}. \end{aligned}$$

The impedance of the crossbar, say Z_0, may be found from any of the impedances $Z_{13:13}$, $Z_{14:14}$, $Z_{23:23}$, $Z_{24:24}$; e.g.,

$$\begin{aligned} Z_0 &= Z_{13:13} - (Z_1 + Z_3 - 2Z_{13}) \\ &= Z - \nu(1 - \nu)(Z_{12:12} + Z_{34:34} - 2Z_{12:34}). \end{aligned}$$

But the presence of seven elements, as already mentioned, suggests an arbitrariness which may be put into evidence by adding $\alpha Z_{12:34}$ to Z_1, which entails a similar addition to Z_3 and Z_{13}, and a similar subtraction from Z_2, Z_4 and Z_{24}.

Similar considerations apply to the derivation of the second H network.

The direct impedances may be found, in a well-known manner, by energizing the network between one terminal and the other three

short-circuited terminals or by applying a formula due to G. A. Campbell.[8]

Acknowledgments

I am indebted to Mr. E. D. Sunde for a suggestion which brought this theorem into focus for me; Dr. H. M. Trueblood has aided in eliminating certain weak points in the proof; to Mr. R. M. Foster I owe a number of improvements, aside from that mentioned in the footnote, in the general spirit of the paper. My chief acknowledgment, however, as indicated by the footnotes referring to his work, is to Dr. G. A. Campbell; I should like the paper to be taken as an instance of the fertility and generality of the methods and results of network analysis he has introduced.

[8] "Direct Capacity Measurement," *Bell System Technical Journal*, I, 1, pp. 18–38 (July, 1922). The formula, given on page 34, requires modification only to the extent of substituting impedances for capacities; for four terminals the direct impedance, D_{ij}, between terminals i and j is given by:

$$D_{ij} = \frac{\Delta}{2\Delta_{ij}},$$

where Δ_{ij} is the co-factor of the element in row i, column j of the determinant:

$$\Delta = \begin{vmatrix} 0 & Z_{12} & Z_{13} & Z_{14} & 1 \\ Z_{12} & 0 & Z_{23} & Z_{24} & 1 \\ Z_{13} & Z_{23} & 0 & Z_{34} & 1 \\ Z_{14} & Z_{24} & Z_{34} & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{vmatrix}.$$

The elements of the determinant are the driving-point impedances between terminals indicated by the subscripts (all other terminals open), namely, $Z_{12:12}$, $Z_{13:13}$, etc., written for brevity Z_{12}, Z_{13}, etc.

Contemporary Advances in Physics, XXXI—Spinning Atoms and Spinning Electrons [1]

By KARL K. DARROW

NO doubt you are all accustomed to thinking of atoms as objects—very small objects, of course—which are endowed with *weight.* I can say that with perfect safety to an audience of engineers and physicists; but indeed it can be said with safety to any audience—I mean, of course, any audience literate enough to attach any meaning at all to such a word as "atom." It may be that philosophers of the past have imagined weightless atoms—I am not historian enough to deny that, nor to affirm it; but if such have ever been invented, they have remained quite outside the currents of modern thought. For us, weight is a property which we attribute to the atom. Since this is, after all, a professional audience, I will now change over to that other word which many people have such difficulty in distinguishing from "weight": I will say that *mass is a property which we attribute to the atom.* In a way, that is a negative statement. It means that we do not hope to explain mass in terms of something more fundamental; it means that we accept mass as being itself so fundamental that even the elementary particles have it. When I say "elementary particles," I am still referring in part to the atoms, though it is a somewhat careless usage to do so; but I am referring also to electrons both positive and negative, to protons, to alpha-particles, to nuclei—to all the particles, in effect, of which the atoms are built up. Also I ought to include the corpuscles of light, but this lecture will be quite long enough if I leave them almost unmentioned. All of these particles, then, are endowed with mass; each of them has a characteristic mass of its own, which we do not attempt to explain, but which we do try to measure as closely as we can.

There is another property, familiar to you though not to everyone, which we accept as equally fundamental and equally unexplainable with mass: it is *electric charge.* We attribute it also to the elementary particles, though not, it is true, to all of them. We assign it to the electrons, of course, and to protons and alpha-particles and all of the hundreds of nuclei which distinguish the elements and the isotopes

[1] A lecture delivered before the American Physical Society at Chicago on November 27, 1936, and before the American Institute of Electrical Engineers at New York on May 6, 1937.

thereof from each other. When I say that we "assign" it, I do not for a moment mean that we are doing something arbitrary or untestable. We *know* that these elementary particles are charged, and indeed we have measured their charges. Atoms do normally appear to us uncharged, but we know that that is because the elementary particles of which they in their turn are made up are some of them positive, some of them negative, and the balance normally happening to be perfect. Some particles, the neutrons and the corpuscles of light, seem permanently chargeless; but perhaps some day we shall find it expedient to regard them as groups of smaller particles having charges which balance one another. Apart from these two cases, we may say with assurance that whenever we penetrate as far into the fine structure of substance as we are able to go, we find the elementary particles invested with mass and with charge.

And now I arrive at the subject of this talk, the *third* property of the elementary particles: the property which is called "angular momentum," or "spin" for short. Now of course I am speaking as to an audience of physicists, for if this were an audience of laymen it would certainly be frightened by such a term as "angular momentum." This is a misfortune, and perhaps a defect of general education; for angular momentum is about as important as mass or charge, not only on the scale of the elementary particles but also on the scale of the visible world. Think what it would mean if there were no such thing as the conservation of angular momentum! the earth might cease from turning, it might cease from providing us with the regular alternation of day and night, and with our standard of the flow of time; it might even cease from traversing its regular orbit, and fly off into space or fall into the sun. Well, I do not wish to scare you with any such dire imaginings—I only want to remark on the fact that the human race has been acquainted for a very long time indeed with angular momentum as something which is unvarying, imperturbable, incessant; for of all the unvarying and imperturbable and incessant things in the world, the rotation of the earth is the most obvious and the most striking. So striking it is, that you might reasonably expect that all the philosophers and all the physicists of the past would have conferred the property of spin on all the atoms which they have invented. Well, they did not; the notions of the spinning atom, the spinning electron, the spinning nucleus are among the newest in physics. I think that some of the reasons for the delay will be evident later on in this talk, but it remains partly mysterious, at least to me. Looking back on the situation with the well-known advantages of hindsight, I do feel a good deal of surprise that the

spinning atom did not make an earlier entry upon the scientific stage. Perhaps some of you will remember hearing the words "vortex atom" and "vortex theory" which used to be so prominent in physics, and will take the spinning atom of today for a lineal descendant of those vortices of old. If this were correct, we could trace back the ancestry of the spinning atom for about three hundred years; but I think that it is not correct. The vortices of which Descartes and Malebranche were dreaming three centuries ago were more like whirlpools of streaming particles, and the vortices which were imagined by Helmholtz and Lord Kelvin some fifty years ago were also whirlpools, but they were whirls in an idealized continuous frictionless fluid. Let us pause for a moment to notice how the attitude of physicists has altered in these fifty years! Kelvin and Helmholtz began with the idea of an aethereal fluid pervading the whole of space, and valiantly tried to represent the atoms as whirlpools in that fluid; but we have long since discarded that aether, and our spinning atoms and other elementary particles are small delimited rotating bodies voyaging in a void.

It is nor therefore the vortex which I will introduce to you as the ancestor of the spinning atom, but rather the "Amperian whirl" as it still is sometimes called. You remember, of course, how Ampère in 1820 made a very great achievement which for the purposes of this talk I will divide into three. First, he discovered the fact that an electric current flowing in a circuit is equivalent to a magnet. Next, he worked out the mathematical laws whereby, given a current and the circuit in which that current is flowing, we may calculate the strength or the moment of the equivalent magnet. I will write down the formula for the case of a current i, flowing in a plane loop of area A: the magnetic moment of the equivalent magnet, μ, is given thus:

$$\mu = iA/c.$$

Here c is a factor of transformation which we are now obliged to employ because we habitually use, in atomic physics especially, a unit-system different from Ampère's. The third part of the great achievement was this: Ampère founded what remains to this day the theory of magnetism, by presuming that *the individual atoms of any magnetizable substance are themselves little magnets*, and that *the atoms are magnets because they have little whirls of current in them.*

This notion—the notion that atoms are magnets, and that they are magnets because they have internal circulating currents—is the true forerunner of our present conception of the spinning atom. It is, however, only a very primitive form of the modern conception, and there is much to be added to it. First of all and above all, there is

the question of angular momentum. Is angular momentum an attribute of these whirling intra-atomic currents, or is it not? You may think that the answer to this question is self-evidently "yes!" but remember that for many generations of our forefathers electricity was an *imponderable* fluid. Weber, however, did consider the affirmative answer, and Maxwell even attempted to ask the question of Nature by experiment—vainly, as it turned out. Not till the electron was discovered did the mass of electricity become a prominent part of experience. A moment ago I divided Ampère's achievement into three parts; similarly I wish now to divide the discovery of the electron into three. Those who isolated and identified and measured the electron were proving three things: first, that negative electricity consists of corpuscles of a definite charge, e; second, that these corpuscles have a mass, m; and following from these two, the principle which I have called the third part of the discovery, *viz.* that an electron revolving in an orbit has an angular momentum.

I will designate angular momentum in general by the letter p, and now I will show you a formula for the ratio of μ to p in an atom in which an electron is running around in an orbit and constituting an Amperian whirl. The formula, like this other one, for μ, is valid for an orbit of any shape, but to get it quickly I will simplify by postulating a circular orbit. The radius of the circle being r, the area A is πr^2; the current around it is equal to the electron-charge e, multiplied by the number of times per second that the electron runs around the orbit; if I denote the velocity of the electron by v, this number is $v/2\pi r$; hence the product iA/c is equal to $evr/2c$. Now the angular momentum p of the electron, as you all know, is mvr; and hence for the ratio I derive:

$$\mu/p = e/2mc,$$

which is one of the most important formulae in the whole of atomic physics. You notice that it does not involve in any way the size or shape of the orbit or the frequency with which the electron travels around it. It is the same for any or all of the revolving electrons of any atom of any kind.

Now let us see how this formula may be tested. Imagine a rod of some highly magnetizable metal, iron for instance, and imagine it to be unmagnetized at the start. This means, that at the start the little atomic magnets are pointing at random in all directions; that is to say, the vectors which represent their magnetic moments are pointing every way, and so are the vectors which represent their angular momenta, the latter being parallel to the former. Since these atomic vectors of angular momentum are pointing every way at

random, they add up to zero, and the rod as a whole possesses no resultant angular momentum; it is just standing still. Now let the rod be surrounded with a solenoid, and by means of a current in the solenoid let it be magnetized to saturation. Now all the arrows representing magnetic moments are pointing parallel to the axis of the rod. But so are all the arrows representing atomic angular momenta! their resultant is no longer zero—suddenly there has arisen a *resultant angular momentum*, belonging to the totality of all the atomic magnets, and quite large enough to be detected, instead of being tiny like the angular momentum of an individual atom. Unless our theory is fundamentally wrong somewhere, we should be able to observe this resultant angular momentum.

The experiment is done by hanging the rod vertically from a fine suspension, and sending the magnetizing current through the solenoid. At the instant of the magnetization, the rod turns sharply on its axis, twisting the suspending fibre. Thus it manifests the angular momentum of which I have just been speaking—though I ought to say that what we observe is of the nature of a recoil, or back-kick: when the totality of the little atomic magnets suddenly acquires its resultant angular momentum, the substance of the rod as a whole acquires an equal and opposite amount (so as to keep constant the total amount of angular momentum in the universe) and it is the latter which we detect. The experiment is quite a delicate one, but its technique has been developed to a remarkable degree since it was first attempted twenty years ago by Einstein and de Haas. What we measure is the ratio of the magnetization of the rod-as-a-whole to the angular momentum of the rod-as-a-whole; and this is just the same as the ratio of μ to p for the elementary atomic magnets. There are not many properties of matter of which we can say that the value measured on a large piece of matter is the same as the value for the individual atom; but there are a few, and this is one of them.

Now in giving you the result, let me first emphasize the general principle that here we have evidence of the spinning of elementary particles, and of the interrelation between spinning and magnetism. Next, I give you the numerical result itself. For iron and nearly all of the other ferromagnetic materials, we find:

$$\mu/p = e/mc$$

or *twice* the theoretical value which I quoted a moment ago.

This cannot be explained by assuming any peculiarity of size or shape or frequency of the electron-orbits in the atoms, for as I just said the theoretical formula is independent of all these things. We

are obliged to make some more drastic assumption. If I had unlimited time before me, I might sketch the history of our assumption; but as I don't, I will come straight to the present situation. We assume first, that in the iron atoms in the rod the electron-orbits are so oriented with respect to each other that their magnetic moments kill one another off completely. We then assume that every electron has a magnetic moment and an angular momentum of its own, intrinsic to it and inherent in it, and altogether independent of whether or not the electron is revolving in an orbit. Just as the earth has a rotation of its own in addition to its elliptical course around the sun, so we imagine that the electron has a rotation of its own; this rotation has an angular momentum, and with it there is connected a magnetic moment. (You will remember doubtless that the earth also has a magnetic moment, but this is one of the analogies which it is better not to force too far.) When we magnetize the iron rod, it is the *electrons* which we are turning; the vectors which we cause to point all in the same direction are the magnetic moments and the angular momenta which are inherent in the electrons, and the value of their ratio is the value which is characteristic of the "spinning electron," as we call it. Therefore, amplifying the notation a little, I write:

$$\mu/p = g(e/2mc)\begin{cases} g = 1 \text{ for electron-orbits,} \\ g = 2 \text{ for spinning electron,} \end{cases}$$

and now I leave the spinning electron for a few minutes, in order to turn again to the theory of electrons revolving in their orbits.

You all realized, of course, that when I converted the Amperian whirl of current into an electron running around an orbit, I was adopting the atom-model known by the names of Rutherford and Bohr; for these were the original thinkers who impelled all the rest of us, following in their footsteps, to think of the atom as a positively-charged nucleus around which electrons are revolving like planets around the sun. This is an atom-model in which magnetism is inherent—*a Rutherford-Bohr atom is intrinsically a magnet.* Anyone who did not know the history of the model might well assume that it was designed expressly to account for magnetism, and any such person might also quite reasonably assume that all the physicists of the early nineteenth century thought of it simultaneously as soon as the electron was discovered. Well, it was *not* designed expressly to account for magnetism, and most of the physicists of the early nineteenth did *not* think of it—or if they did, they thought of it only to reject it. At that time, the atom-model with the orbital electrons seemed to be disqualified by a very potent reason; for according to the classical

electromagnetic theory, an electron revolving in an orbit ought to radiate all of its energy in a very short time and fall into the nucleus. Bohr was the man who overrode this objection. He overrode it, not in order to construct a theory of magnetism in defiance of it, but in order to construct a theory of spectra in defiance of it. This theory has been extraordinarily successful. Our theory of magnetism is hardly more than a by-product of that theory of spectra; and this, in an odd sort of way, enhances its credit. A theory devised expressly for a certain purpose is always less impressive than one which follows incidentally from a successful theory devised for quite another purpose; and the contemporary theory of magnetism is a wonderful example of this latter and more impressive type.

The main element of Bohr's theory of spectra—if one can speak of one element as the main one, which is really not quite proper—is an assumption about the angular momentum of the electron in its orbit or, let me say, the angular momentum p of the electron-orbit. It was assumed that the electron may revolve, without radiating its energy, in any orbit of which the angular momentum is an integer multiple of $h/2\pi$, $-$ h now standing, of course, for the famous quantum-constant of Planck which is the emblem of modern physics. I write this down as follows:

$$p = (1, 2, 3, 4, \cdots)(h/2\pi).$$

Bohr was thinking at first about the hydrogen atom; but hydrogen is an inconvenient example to use in talking about magnetism, and iron is a very complicated case indeed, so I will talk entirely about the sodium atom.

The sodium atom has a nucleus with a charge of $+ 11e$, and eleven electrons circulating in orbits around it. This certainly sounds formidably complex, but it happens—and I shall later remind you of this fact—that the orbits and also the spins of ten of the electrons are so oriented with respect to one another that their angular momenta and their magnetic moments completely neutralize each other. I shall therefore ask you to imagine these ten inner electrons as a sort of cloud. The eleventh electron of the sodium atom—known technically as the "valence" electron—cruises around this system; sometimes it is traveling in an orbit completely outside the cloud, sometimes in an orbit which cuts across the cloud, but never in an orbit which is entirely or even mainly inside the cloud. The ten inner electrons which constitute the cloud neutralize a part of the force with which the nucleus acts upon the valence-electron; but they make—I repeat —not the slightest contribution to the angular momentum or to the magnetic moment of the atom.

I have given above, the permitted values of the angular momentum of this orbit of the valence-electron. Now I point out that to each of these permitted values of p corresponds a permitted value of the magnetic moment μ, which I obtain by multiplying the former with $e/2mc$:

$$\mu = (1, 2, 3, 4, \cdots)(eh/4\pi mc).$$

However, only one (at most) of these values can be appropriate to the normal state of the sodium atom; all the rest must correspond to abnormal, unusual, or "excited" states. We are going to be interested primarily in the normal state, so we must identify the right one. In the early days of the Bohr theory, the right one was supposed to be the first which I have written down. However, the theory has been greatly remodelled and bettered since those days, with the aid of what is known as "quantum mechanics"; and it now seems quite certain that these lists of the permitted values of p and μ for electron-orbits are both incomplete. I must add to each of them the value *zero*, so that the two lists become

$$p = (0, 1, 2, 3, 4, \cdots)h/2\pi,$$
$$\mu = (0, 1, 2, 3, 4, \cdots)eh/4\pi mc.$$

Moreover, it is precisely this new value *zero* which belongs to the normal state of the sodium atom. So the theory, in this stage, quite definitely prescribes that the sodium atom in its normal state should have no magnetic moment and no angular momentum. But now let us look at the data.

It would do no good in this connection to make measurements on solid or on liquid sodium, for in those "condensed phases" the atoms are crowded so closely together as to be badly distorted. We can, however, experiment on free atoms of sodium, in their normal state, in the way illustrated by Fig. 1. In the upper portion, A represents an "oven," consisting of a small box heated electrically and containing some liquid sodium which is steadily being vaporized. There is a hole in the wall of the box through which free sodium atoms are steadily shooting in all directions, with the distribution-in-speed which we know from the theory of thermal agitation; and beyond, there is a sequence of diaphragms with slits in them which delimit a straight and narrow beam of these fast-moving atoms. Disregarding what the theory has just said, let us suppose that each of these atoms is a magnet—a bar magnet, with a north pole and a south pole. As they emerge from the oven, these atoms must surely be oriented at random in all directions.

Continuing to look at the upper part of Fig. 1, we have two large magnet-poles, and the beam travels between them, having no trouble with molecules of air as it shoots along, for the whole of this apparatus is enclosed in a highly-evacuated tube. It may seem natural to visualize these magnet-poles as the two broad flat extremities of a horseshoe-magnet, with a uniform magnetic field pervading all the space between them. Such an arrangement, however, would make the experiment futile.

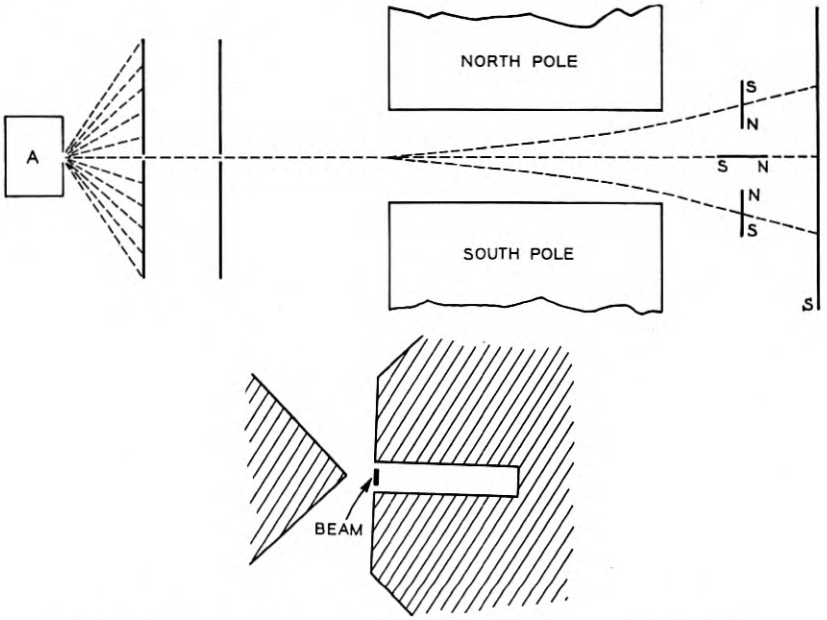


Fig. 1—Longitudinal section and cross-section of apparatus for the Gerlach-Stern experiment.

Nothing would happen to any of the atoms, for in the uniform field the north pole of each atomic magnet would be pressed downward just as hard as and no harder than the south pole is drawn upward, and the net force would be zero. The beam would go on unbroadened and undeflected, and make a small spot on the screen S, the spot being just opposite the slits in the diaphragms. Something else must be tried; and what we do—or rather, what Gerlach and Stern did in Hamburg some fourteen years ago—is, to shape one of the magnet-poles in the form of a wedge and hollow out the other, so that the field between the two shall no longer be uniform. The lower part of Fig. 1 represents the cross-section of such a pair of pole-pieces. The field-strength is now much greater near the wedge than near the opposite

pole-piece, and there is a vertical gradient of field-strength which may be made fairly constant over most of the interspace. Rabi at Columbia achieves the same result more efficiently by peculiar arrangements of current-carrying wires instead of iron magnets.

Now consider (thinking classically!) a few of the atomic magnets as they shoot across this non-uniform field. Think of one which originally is oriented vertically, with north pole down and south pole up—the south pole will be in a stronger part of the field than its mate; it will be drawn upward harder than the north pole is pushed downward; the atom will sweep in a parabolic arc upward. Think of another which originally is oriented vertically with north pole up and south pole down—it will be swept in a parabolic arc downward. Think of another which originally has its axis pointing horizontally—it will shoot in a straight line across the field, as though the pole-pieces were not there.[2] Now think of all those which are oblique to the vertical; they will describe parabolic arcs of intermediate curvatures, upward or downward as the case may be. One infers that the beam must be spread out into a continuous fan, making a continuous vertical band upon the photographic plate.

Moreover, from the upper edge or from the lower edge of this continuous band, it should be possible to determine the magnetic moment μ of the atoms. For let us consider one of the vertically-oriented atoms, and call its pole-strength M and the length of its magnet r. The upward force on the north pole is MH,—H standing for the field-strength at the point where the north pole is. The downward force on the south pole is $M(H + dH/dz \cdot r)$. The net force is $Mr \cdot dH/dz$, which is $\mu(dH/dz)$, because Mr is the magnetic moment μ of the magnet by definition. The acceleration of the little atom is equal to this force divided by m_a, the mass of the sodium atom. The deflection is equal to half the acceleration by the square of the time during which the atom is exposed to the force. This time is equal to the distance D which the atoms traverse across the field, divided by v the speed which they have when they come out of the furnace. So finally, we have:

$$\text{Deflection} = \frac{1}{2}\frac{\mu(dH/dz)}{m_a}(D/v)^2.$$

We ascertain the deflection by looking at the end of the band on the photographic plate, and we can ascertain all the other things in the

[2] When thinking classically, we must not expect the atomic magnets to turn their axes toward the field-direction as they shoot across the field; the gyroscopic quality of these magnets, due to their angular momentum, inhibits this.

equation excepting μ,—v is the hardest to estimate accurately—and so we can solve the equation for μ.

When the experiment is done there appears, however, a very remarkable thing. Instead of there being a long band upon the plate, there are just two spots. Instead of the beam having been broadened out into a continuous fan, it has evidently been split into two separate pencils. It looks as though the field had acted first of all upon the magnets, by setting them all vertical,—half of them with north pole up, and half with north pole down. Not this phenomenon alone, but many others in Nature show us that this is just what happens. You may perhaps feel for the moment that it is intelligible, after all; the compass-needle turns to the north—why should not the little atomic magnets, as soon as they enter the field, turn their south poles toward the north pole of the magnet which attracts them? Well, this would not account for the magnets which constitute the beam which bends away from the wedge-shaped north pole, instead of toward it; and indeed it does not even account for the beam which bends toward the wedge-shaped pole. Classically the field should have no orienting effect whatsoever upon the atoms, and yet it evidently does.[2] This is one of the phenomena of the atomic world which we cannot properly visualize in terms of the behavior of objects large enough to be tangible and visible. All that I can do is to assert it, and to say that it justifies us in using this formula to calculate μ. When we use it, the value which we find for the magnetic moment of the sodium atom in its normal state is

$$eh/4\pi mc,$$

which happens to be one of the values in the sequence which I just wrote down.

I repeat that according to the theory in its present stage, the electron-orbits in the normal sodium atom have a net magnetic moment of zero. This value $eh/4\pi mc$ is, therefore, the magnetic moment due to the spin of the valence-electron—it is the magnetic moment of the spinning electron. I write it in the appropriate place, and then with the aid of the g-value derived from the gyromagnetic effect I write down the value of angular momentum which we assign to the spinning electron:

$$p = \tfrac{1}{2}(h/2\pi).$$

To this roster of three statements about the spinning electron I now make a final addition. The Gerlach-Stern experiment on sodium shows that a beam of sodium atoms—which for this purpose is the equivalent of a beam of spinning electrons—is divided into two by a

magnetic field. I write down "2" to indicate this number of separated beams; but I will call it by preference the "number of orientations in the field," because that is the fundamental point. The spinning electron always sets itself in one or the other of two orientations, with respect to whatever field it happens to be traversing. We call them the "parallel" and the "anti-parallel" orientations, though according to quantum mechanics these terms are a little too strong. Here then is the list of the properties of the electron-spin: g equal to 2—angular momentum equal to $\frac{1}{2}(h/2\pi)$—magnetic moment equal to $eh/4\pi mc$—two permitted orientations in any field.

It has doubtless struck you as rather odd that I began by talking about the angular momenta and the magnetic moments of electron-orbits, and then carefully picked out a couple of special cases in which these neutralized each other altogether and there was nothing left over except what I ascribed to the electron-spin. Is there no point at all, then, in talking about the electron-orbits? Oh, very much so! Indeed there are cases in which the electron-spins neutralize each other altogether, and we have nothing left over except what is attributed to the orbits. To do this I may choose an atom like magnesium, which has a nuclear charge of $+12e$, a cloud of ten inner electrons which neutralize one another completely as to angular momentum and magnetic moment (just as in sodium), and *two* valence electrons instead of one. In some of the states of the magnesium atom—not in all of its states, but in *some* of them—the spins of the two valence electrons are oriented opposite to each other in the atom, and cancel each other out. When the atom is in a state of this kind, then nothing is left over except the angular momenta and the magnetic moments of the orbits of the two valence-electrons; and then, all the statements of the orbital theory (page 326) are applicable—g is equal to unity, the angular momentum takes one of the values $nh/2\pi$ and the magnetic moment takes one of the values $n(eh/4\pi mc)$. Moreover, there is another theorem derived from quantum mechanics which turns out to be valid: the number of orientations of such an atom in a field, the number of separated beams which appear in the Gerlach-Stern experiments, is chosen from among the members of this sequence: 1, 3, 5, 7. . . . (It is a most interesting historical fact, that Gerlach and Stern were moved to undertake their difficult experiment by the wish to test this remarkable assertion of quantal theory.) You notice that the number 2 does not appear in the sequence; were it not for the electron-spin, we never could obtain it; it is distinctive of the spinning electron.

I must, however, admit that all these cases of which I have been speaking are special, and comparatively rare. Both the cases in which the spins neutralize each other perfectly, and the cases in which the orbital moments neutralize each other perfectly—both types are unusual. Still more unusual, and yet occurring here and there, is the most special of all cases—that in which *all* of the moments and *all* the momenta, both those of the spins and those of the orbits, balance one another perfectly so that the sums are zero. An atom in such a state is completely devoid both of magnetism and of spin; such atoms are those of helium, of neon, of argon and the other noble gases, when in their normal states. Usually, however, we find ourselves confronted with some example of the general case, in which neither the spins nor the orbits are completely neutralized. The atom has an angular momentum which is a sort of composite or resultant of the angular momenta of the spins and the orbits, and it has a magnetic moment which also is a sort of composite or resultant.

If I were to embark on the description of the general case this lecture might go on interminably, and at its end you would probably not remember anything except what you had already known at its beginning. The laws of the composition of spins and orbits are so foreign to our customary ways of thinking, and the formulæ which express them are so curiously built, that to work once only through them is not sufficient: one has to memorize the derivations and the results alike, and go over them incessantly until they are imprinted on the brain. I think you will agree to this readily enough, when I remind you that this theory is none other than the general theory of spectra; for even quite outside the ranks of physicists, the theory of spectra is beginning to be notorious for its complexity. I shall not venture even to give the formulæ, much less their derivations; it must suffice to fill out the two lists of p-values and μ-values on which I have already begun, and the list of n-values or numbers-of-permitted-orientations.

The spinning atom is a congeries of electrons, all of them always possessing spin, most of them usually possessing orbital motions; and these motions are compounded with each other in such ways, that:

First, the angular momentum of the spinning atom has one of the values,

$$p = (0, \tfrac{1}{2}, 1, \tfrac{3}{2}, 2, \tfrac{5}{2}, 3 \cdots)h/2\pi.$$

Second, the number of beams in the Gerlach-Stern experiment, or the number of permitted orientations of the atom in the field, has one of the values,

$$n = 1, 2, 3, 4, \cdots.$$

Third—and now comes a surprise, for you will probably expect me to say of the magnetic moment that it is $eh/4\pi mc$ multiplied either by an integer or a half-integer; but this is not so. The actual state of affairs is described by a formula which is called the g-formula, because it gives g in terms of the moments, both spin and orbit, of the individual electrons. It was discovered by Landé and interpreted in terms of the spinning electron by Goudsmit. The g-formula gives unity, as of course it must, in the special cases where the electron-spins cancel each other and only the orbital moments are left over, and it gives 2 in the special cases where the orbital moments are neutralized with only the spins left over. In the other cases it may give any one of a large variety of values: mostly one gets simple-looking fractions such as 9/8 and 4/3 and 5/6. The magnetic moments are then computed by multiplying the appropriate g-values times $e/2mc$, into the existing values of angular momentum.[3]

I now turn to that component of the atom of which the spin remains to be discussed—to the *nucleus*.

As I have already intimated, the values of p and n and μ for the electron-family of any atom are mostly ascertained by analyzing their spectra and utilizing the great general theory of spectra. Such magnetic experiments as I used for my examples are relatively few, and feasible for relatively few substances. The reason for making this remark at this late moment is, that by analyzing spectra we may also learn something about the spins of nuclei; for nuclei also are invested with these properties of angular momentum and of magnetic moment. I take in particular the case of the *proton*—that lightest of all nuclei, the nucleus of the lightest known kind of atom which is ordinary hydrogen, so called to distinguish it from "heavy" hydrogen. Analysis of the spectrum of hydrogen shows us that the proton is capable of taking two permitted orientations in a field; thus, our first piece of information about the proton-spin is conveyed by writing $n = 2$.

Now that we have this piece of information, we deduce that as the spinning electron has an n-value of 2 and an angular momentum of $\frac{1}{2}(h/2\pi)$, so the proton with its n-value of 2 must have an angular momentum of $\frac{1}{2}(h/2\pi)$. Continuing along this line of thought, we are further tempted to infer that the proton should have a g-value of 2 and a magnetic moment of $(eh/4\pi mc)$. But what shall we suppose

[3] Should any reader intend to proceed from this article to a thorough study of atomic theory, he should be warned in advance that according to the latest form of quantal theory, the p-values and the μ-values here given are the values of the projections of these vectors upon the field-direction, the magnitudes of the vectors themselves being somewhat greater: I have given details as to this in "Contemporary Advances in Physics, XXIX, . . ." This *Journal*, **14**, pp. 293 ff.(1935).

about m? Formerly it represented the mass of the electron; now we are dealing with a different sort of particle, having a mass which (as many other kinds of experiments show us) is about 1835 times as great as the electron-mass. I denote this mass by M. It seems natural, then, to expect for the magnetic moment of the proton the value $eh/4\pi Mc$, or about 1/1835 of that of the spinning electron.

This is a formidably small magnetic moment to hope to measure, nay even to detect! yet Stern and his pupils undertook to measure it, and they succeeded. Of course, modifications had to be made in the technique which worked so well for sodium. Hydrogen being a gas at room-temperature, no heated oven was required; nevertheless they used an "oven," but it was refrigerated instead of being heated—a sort of super-ice-box; this was in order to obtain slow-moving atoms, for the slower the atoms traverse the field, the more accurately the experiment can be made. I just said "atoms"; but as most people know, the particles of gaseous hydrogen are not atoms, but diatomic molecules—systems composed of two protons and two electrons apiece. This is a circumstance which in many desirable tests of modern theoretical physics is a great inconvenience, for usually our simplest theoretical affirmations refer to hydrogen atoms and we should like to be able to experiment on them directly. Here, however, it turns out to be a great convenience, indeed perhaps the only thing that makes the experiment possible. For if we had an isolated hydrogen atom, the magnetic moment of its electron would so far exceed that of its proton that the latter would be indetectable. (Perhaps it is not superfluous to mention that bare protons could not be used in the experiment either, as the magnetic field would exert so large a force upon their moving charges that the forces upon their magnetic poles would be insignificant by comparison.)

But if in a single hydrogen atom the magnetic moment of the electron swamps that of the proton, how shall this fate be avoided for a system composed of two electrons and two protons? Here enters in, and in a very important and significant way, that law of the permitted orientations. Just as a spinning particle of angular momentum $\frac{1}{2}(h/2\pi)$ can take only two permitted orientations in a field, so it can take only two with respect to another particle of its kind—the parallel and the anti-parallel. It chances—or rather it does not chance, it follows from the underlying laws of Nature—that in the hydrogen molecule the two electrons are oriented anti-parallel to each other. Their magnetic moments cancel each other, and do not trouble the experimenter.

May not, however, the same thing happen in respect to the two protons, so rendering the experiment hopeless? It turns out that for these the spins are anti-parallel in some molecules but parallel in others. Molecules of the former type, which is called para-hydrogen, are indeed useless for the experiment; but molecules of the latter type, which is called ortho-hydrogen, are available, and in them the magnetic moments of the two protons collaborate so that the magnetic moment of the molecule-as-a-whole is twice as great as that of the single proton, a welcome assistance! In ordinary gaseous hydrogen at room temperature, about three-quarters of the molecules are ortho-hydrogen.

When the experiment was at last achieved by the school of Stern, it was found that the foregoing inference as to the μ-value of the proton is roughly but not exactly correct! The latest information is, that the magnetic moment of the proton is close to $2\frac{1}{2}$ times $eh/4\pi Mc$. Measurements with another method by Rabi and his school have confirmed these results; and we are definitively debarred from believing that for the proton and the electron, the magnetic moments stand in the inverse ratio of the masses. Perhaps this signifies that the proton is itself a composite particle, a notion for which there is some support from other sources.

I mention briefly the characteristics of a few other nuclei. After the proton, the next simplest is the deuteron or nucleus of the heavy-hydrogen atom. It is composed of a proton and a neutron, the latter being a neutral particle of about the same mass as the proton. We can observe free neutrons wandering about in space, but we cannot determine their spins nor their magnetic moments. The deuteron, however, has three permitted orientations (this we discern from the spectrum of heavy hydrogen) and consequently an angular momentum of $(h/2\pi)$. It is inferred that the neutron has $\frac{1}{2}(h/2\pi)$ for its angular momentum, and that in the deuteron these two constituent particles —proton and neutron—are oriented with their equal spins parallel to one another. The magnetic moment of the deuteron is less than that of the proton, and accordingly the neutron must have its magnetic moment oppositely directed to that of its companion in the system, even though their angular momenta be similarly directed—a strange complication!

After the deuteron, the next simplest among the nuclei (except two which are much too rare for investigation) is the alpha-particle or helium nucleus. It is composed of two neutrons and two protons. We find that its angular momentum and its magnetic moment are zero, a clear indication that the four spins of its components are cancelling each other two by two, and the four magnetic moments

likewise. Next comes a nucleus composed of three protons and three neutrons, belonging to the element lithium. It is found that in angular momentum as in magnetic moment it is practically a duplicate of the deuteron, its six components having disposed themselves into a nearly normal deuteron attached to a nearly normal alpha-particle. I might proceed some distance farther along the list of the known nuclei after this fashion, were there space; but it is best to close this section with a general rule: *nuclei with an even number of constituent particles* (protons and neutrons) *have even spins, nuclei with an odd number of particles have odd spins.* "Even" and "odd" in this formulation mean that the angular momentum is an even or an odd integer multiple of $\frac{1}{2}(h/2\pi)$, respectively. One sees that if any two spins of magnitude $\frac{1}{2}(h/2\pi)$ are allowed to choose only between parallel and anti-parallel orientations, the rule follows inevitably; reversely, from the rule (which is based on experience with some fifty or sixty kinds of atom), we derive extra strength for that theorem about orientations.

Now to come to the conclusion and the climax. Although this property of angular momentum, of being allowed to take only a limited number of permitted orientations—although this strange and wonderful property of angular momentum was introduced in this lecture as though it pertained only to atoms subjected to applied external fields, yet it manifests itself far more broadly. Indeed, it manifests itself universally, and the stability and the character of the world are due to it. I have already mentioned in half-a-dozen places how it manifests itself within the molecule and within the atom: how in the atom, it is responsible for those laws of composition which determine the angular momentum and the magnetic moment of the electron-family—how in the hydrogen molecule, it establishes a difference between ortho-hydrogen and para-hydrogen—how in the nucleus it fixes the angular momenta and the magnetic moments of composite nuclei as sharply as those of their constituents the proton and the neutron. Perhaps these seem to be remote and unimportant qualities; but what has just been said of them may be said with equal truth and equal force of *all* the chemical and physical properties of all the elements, mass alone excepted (and even mass not fully excepted). If this feature of angular momentum did not prevail, there could not be the fixity of properties which characterizes each element by itself and the variety of properties which characterizes the totality of the elements. Gold would not be gold, lead would not be lead, oxygen would not be oxygen, helium would not be helium; for though it is commonly said that each element is distinguished by its nuclear charge and the number of electrons in its electron-family, this is not

adequate. It is the law governing angular momentum which imprints upon these electron-families the characters which we recognize as the properties distinctive of the elements. It seems strange indeed that character should depend upon motion, and fixity upon the laws of whirling things; but however strange it may seem, there is no doubt about it. In the construction of houses the builder requires raw material in the form of brick and stone and wood and steel; but he requires also principles of architecture, whereby the raw materials may be parceled off and integrated into the general design. In the construction of the physical world, mass and charge fulfil the role of raw materials, and the laws of angular momentum furnish the principles of the architecture thereof.

A Multiple Unit Steerable Antenna for Short-Wave Reception *

By H. T. FRIIS and C. B. FELDMAN

This paper discusses a receiving system employing sharp vertical-plane directivity, capable of being steered to meet the varying angles at which short radio waves arrive at a receiving location. The system is the culmination of some four years effort to determine the degree to which receiving antenna directivity may be carried to increase the reliability of short-wave transatlantic telephone circuits. The system consists of an end-on array of antennas, of fixed directivity, whose outputs are combined in phase for the desired angle. The antenna outputs are conducted over coaxial transmission lines to the receiving building where the phasing is accomplished by means of rotatable phase shifters operating at intermediate frequency. These phase shifters, one for each antenna, are geared together, and the favored direction in the vertical plane may be steered by rotating the assembly. Several sets of these phase shifters are paralleled, each set constituting a separately steerable branch. One of these branches serves as an exploring or monitoring circuit for determining the angles at which waves are arriving. The remaining branches may then be set to receive at these angles. The several receiving branches have common automatic gain control and thus provide a diversity on an angle basis. To obtain the full benefit of the angular resolution afforded by the sharp directivity, the different transmission times, corresponding to the different angles, are equalized by audio delay networks, before combining in the final output.

The experimental system, located at the Bell Telephone Laboratories' field laboratory near Holmdel, New Jersey, is described. This system comprises six rhombic antennas extending three quarters of a mile along the direction to England. Two receiving branches, in addition to a monitoring branch, are provided. Experience obtained with this system since the spring of 1935 is discussed. The benefits ascribable to it are (1) a signal-to-noise improvement of seven to eight decibels, referred to one of the six antennas alone, and (2) a substantial quality improvement due jointly to the diversity action and the reduction of selective fading.

While a three-quarter-mile short-wave antenna system is an unusually long one, the steerability feature permits the employment of considerably more directivity, afforded by further increasing the length. A system two miles long is believed to be practicable and desirable. It could be expected to perform more consistently better than the three-quarter-mile trial installation, and should yield a signal-to-noise improvement of twelve to thirteen decibels

* Presented before Silver Anniversary Convention of the Institute of Radio Engineers, New York City, May 10, 1937. Published in *Proc. I. R. E.*, July, 1937.

referred to one rhombic antenna. With the object of predicting the performance of larger systems, the performance of the experimental system is examined in great detail and compared with theory.

I. Introduction

FOR more than a decade, point-to-point short-wave radio services have employed directional antennas both in transmitting and receiving. Transmitting antenna directivity results in increased field intensity at the receiving location and receiving antenna directivity discriminates against noise. Both directivities improve the signal-to-noise ratio of a given circuit and permit operation under more adverse transmission conditions. Arrays of simple antennas as well as extensive configurations of long wires have been used to produce these directivities in both the vertical and the horizontal planes.

Antennas in present use on the longer circuits, such as the New York-London telephone facilities, represent about the limit of fixed directivity. Further increase or "sharpening" of the directivity would seriously encroach upon the angular range of directions which are effective in the propagation of waves from transmitter to receiver. The vertical angle range useful in transmitting and receiving short waves is considerable. The horizontal range is appreciable although considerably less than the vertical range. To confine the principal antenna response to only a portion of these ranges penalizes the circuit when that portion is ineffective.

Much experience and considerable statistical data have been obtained which determine this useful range of directions for the New York-London circuits, and antennas have been designed in conformity with these results. However, too much weight must not be given to statistical results which indicate, for instance, that ninety per cent of the time the effective angles are, say, in the range from ten to twenty degrees. For, if the remaining ten per cent includes much of the time that has been lost with existing facilities, an antenna designed for a ten- to twenty-degree response may really be of no value, or even detrimental as a means of extending the usefulness of the circuit. Owing to the great variability in conditions on the north Atlantic path and to the relatively small amount of significant data which has been accumulated during times when gain is most needed it might be detrimental to carry fixed directivity further than present practice has adopted.[1]

[1] One way of attacking the problem of obtaining increased antenna gain has been proposed by John Stone Stone in U. S. Patent 1,954,898. This patent relates to fixed antennas but has certain features, such as delay equalization, in common with the system to be described in this paper.

If, however, the directivity can be varied or "steered" to meet the various conditions imposed by nature, a new field is opened in which a new order of antenna sharpness and gain is possible. In addition to the gain in signal-to-noise ratio afforded by directivity, a reduction in selective fading is possible if the sharpness is increased to the point where a separation of differently delayed waves is achieved. As early as 1927, Edmond Bruce [2, 3] found remarkable reductions in short-wave fading by using a receiving antenna having an extremely sharp directional pattern. The successful employment of sharp directivity is, of course, predicated upon considerable stability of wave directions. The experiments reported by R. K. Potter [4] in 1930 suggested that short waves are propagated in a more or less orderly manner and that stable wave directions might exist. Later experiments,[5] made in co-operation with the British Post Office, using pulse transmission to resolve angles in time, gave confirming data and demonstrated clearly the physical facts upon which is based the system to be described in the present paper. These fundamental facts, outlined in the paper describing the experiments just mentioned, are recapitulated here because a clear understanding of their nature and significance is an essential introduction to the subject in hand. In the pulse tests it was found that:

"1. To the extent that we have been able to resolve the propagation into separate (vertical) angles, the separate angles are found not to be erratic; they vary slowly.

"2. There appears to be at least a qualitative relation between angle and delay; the greater the delay the greater the angle above the horizontal.

"The existence of the many waves of different delay, which is known to make fading selective with respect to frequency, greatly impairs the quality of a short-wave radio telephone circuit. . . . The experimental facts, tentatively established, that individual wave angles are fairly stable and that waves of different delay invariably possess different vertical angles, make this problem hold considerable promise.

"The simple antennas described . . . are suitable for angle determination because of their ability to reject a single wave but they are not

[2] E. Bruce, "Developments in Short-Wave Directive Antennas," *Proc. I. R. E.*, vol. 19, pp. 1406–1433, August, 1931; *Bell Sys. Tech. Jour.*, vol. 10, pp. 656–683, October, 1931.

[3] E. Bruce and A. C. Beck, "Experiments with Directivity Steering for Fading Reduction," *Proc. I. R. E.*, vol. 23, pp. 357–371, April, 1935; *Bell Sys. Tech. Jour.*, vol. 14, pp. 195–210, April, 1935.

[4] R. K. Potter, "Transmission Characteristics of a Short-Wave Telephone Circuit," *Proc. I. R. E.*, vol. 18, pp. 581–648, April, 1930.

[5] Friis, Feldman, and Sharpless, "The Determination of the Direction of Arrival of Short Radio Waves," *Proc. I. R. E.*, vol. 22, pp. 47–78, January, 1934.

in general suitable for quality improvement. For such studies it would be preferable to construct a more elaborate antenna whose directional pattern has a single major lobe which is steerable in the vertical plane. Such an antenna would aim to select a narrow range of angles in which occur waves of substantially the same delay."

The present paper describes a steerable antenna receiving system of the general character suggested by the above quotation, and which has been in experimental operation at the Holmdel, New Jersey, field laboratory of the Bell Telephone Laboratories for the past two years. Certain other important features are incorporated in the system, notably an arrangement whereby individual wave groups arriving at different vertical angles are received separately and, after separate delay equalization, combined, thereby incorporating a unique form of diversity. Another important feature possessed by the system is its frequency range which permits operation on all of the frequencies used in short-wave transatlantic services.

II. Principles of Steering Antenna Directivity

An old and elemental type of steering of receiving antenna directivity is found in direction finders. The steering of a directional lobe as distinguished from the steering of a null has been accomplished in recent years. Schelleng [6] reported a moderate degree of horizontal plane steering, accomplished by means of phase shifters. Jansky [7] has obtained horizontal steering by bodily rotating an entire broadside array. Bruce and Beck [3] obtained vertical steering by varying the shape of a rhombic antenna by means of ropes, and demonstrated the value of steering in the reduction of selective fading. The present authors [5] have employed rotatable phase shifters to steer the nulls in the directional patterns of two spaced antennas. In that work the value of the rapid adjustments possible with phase shifters was very apparent. In the linear end-on MUSA [8] system to be described rotatable phase shifters are again employed to steer the vertical response.[9]

In Fig. 1 is shown a schematic representation of a linear end-on array of N equally spaced unit antennas in free space. The antennas are indicated by the numbered points. For simplicity it is assumed, in the following preliminary analysis, that the antennas are spaced far

[6] J. C. Schelleng, "Some Problems in Short-Wave Telephone Transmission," *Proc. I. R. E.*, vol. 18, pp. 913–938, June, 1930.

[7] K. G. Jansky, "Directional Studies of Atmospherics at High Frequencies," *Proc. I. R. E.*, vol. 20, pp. 1920–1932, December, 1932.

[8] The word MUSA is coined from the initial letters of "multiple unit steerable antenna."

[9] U. S. Patent No. 2,041,600.

enough to be substantially isolated from each other. Choosing antenna No. 1 for reference and considering a plane wave arriving at an angle δ with the axis of the array, it is clear that the output of No. 2 will add in phase with that of No. 1 if the phase advance ϕ is made equal to $2\pi ac/v - 2\pi a \cos \delta$, where c = velocity of light and v = the phase velocity of the transmission lines. Similarly, the output of No. 3 will add to that of No. 1 and No. 2 if its phase is advanced 2ϕ, etc. If the spacing, $a\lambda$, is sufficient there will be other angles for which the N outputs add in phase; at intermediate angles the outputs interfere with the result that zeros and minor maxima occur. By properly designing the unit antenna the undesired maxima may be suppressed.

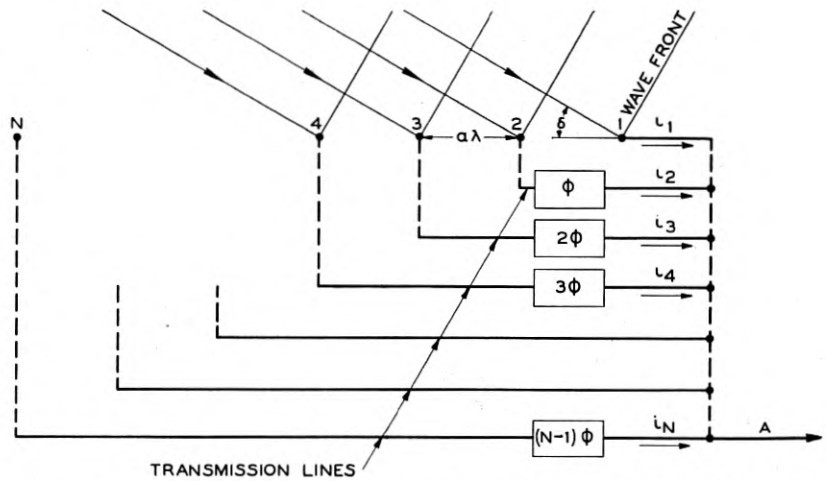


Fig. 1—A steerable antenna array using variable phase shifts ϕ, 2ϕ, 3ϕ, etc. The transmission lines indicated by broken lines are assumed to be of zero length; a is the spacing in free space wave-lengths.

In the Holmdel experimental system the unit antennas are of the rhombic type. An aerial view of the six antennas, which are located on the great circle through England, is shown in Fig. 2. These six antennas, combined as in Fig. 1, yield polar directional patterns such as those shown at the top of Fig. 3. The solid line pattern and the dashed line pattern correspond to different values of the phase shift ϕ. The multiple phase shifts of Fig. 1 are obtained by gearing the phase shifters to a common shaft which enables the directional pattern to be steered simply by rotating the shaft.

Thus far we have discussed the problem of sharp steerable directivity from the point of view of a single plane wave, whereas it is well

known that multiple ionosphere reflections usually produce several more or less discrete waves, or bundles of waves, having different vertical angles and different transmission delays. To obtain the maximum advantage, however, requires that all of the several wave bundles be separately received and suitably combined after the transmission delays have been equalized. The achievement of this objective

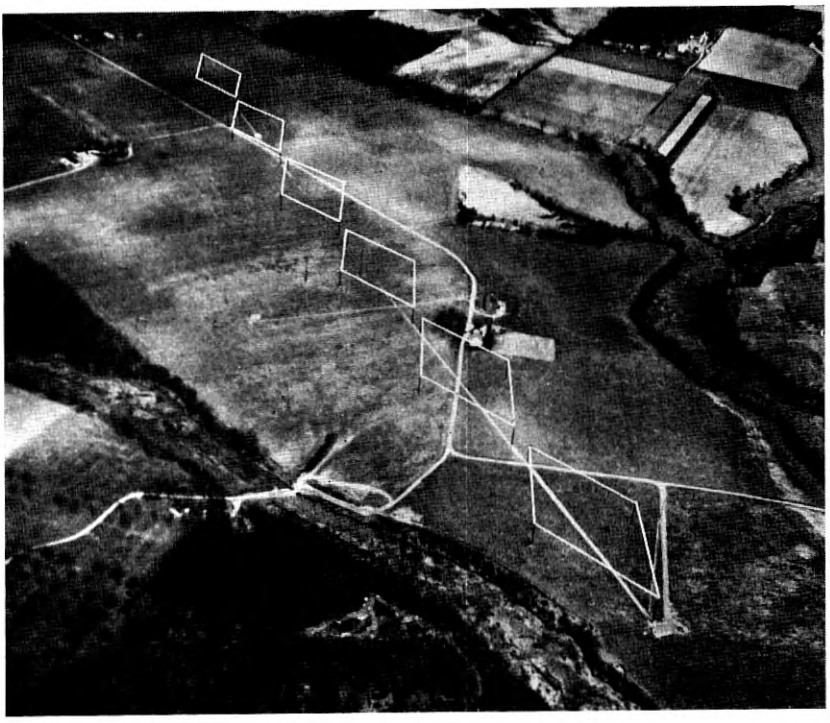


Fig. 2—Airplane view of the three-quarter-mile experimental MUSA on the receiving laboratory site located near Holmdel, New Jersey. The white line beneath the antennas is the newly filled trench in which coaxial transmission lines are buried. The building appearing in the right-hand foreground houses the receiving apparatus. The ground is flat to within ± 4 feet.

not only yields the ultimate gain in signal-to-noise ratio but at the same time *reduces the distortion associated with selective fading.*

The method of obtaining sharp steerable directivity by combining the output of fixed antennas through phase shifters makes it possible to use the same antennas and transmission lines to provide several separately steerable lobes each of which is in effect an independent MUSA.[10] In the experimental system, shown schematically in Fig. 3,

[10] R. K. Potter, U. S. Patent No. 2,030,181.

the antenna outputs are combined at intermediate frequency, and the separately steerable lobes are obtained by branching each of the six first detectors into three phase shifters and combining the outputs of

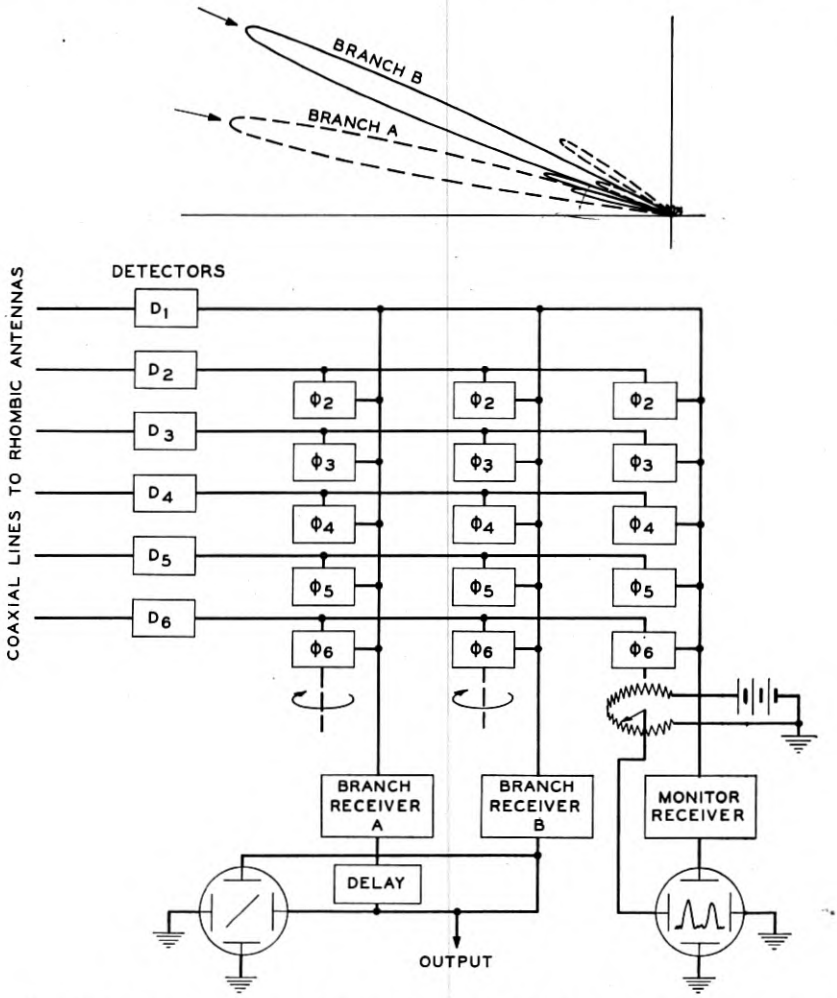


Fig. 3—Schematic diagram of the experimental MUSA receiver. The five phase shifters ϕ_2, ϕ_3, etc., of each branch, are geared to a shaft to provide the phase shifts ϕ, 2ϕ, 3ϕ, etc., of Fig. 1. The inset at the top shows the directional patterns of the two branches when steered at angles of 12 and 23 degrees, at a wave length of 25 meters.

the phase shifters to form three steerable branches. One branch is used continuously to explore the angle range to determine at which angles the waves are arriving. The other two branches are set accord-

ingly and their outputs are "received" by conventional receivers, with common automatic gain control. The demodulated audio outputs are equalized for difference in transmission time and then combined. A cathode-ray oscilloscope displays the output of the exploring or monitoring branch. It plots amplitude (provided by a linear rectifier) as the ordinate, against phase shift ϕ_2 (corresponding to ϕ in Fig. 1). The screen of the oscilloscope is of the retentive type and thus displays several consecutive sweeps at once. A pattern corresponding to two waves is illustrated. The other cathode-ray oscilloscope is used in the adjustment which equalizes the delay of the two waves. Delay is added to the low angle branch until the oscilloscope shows a line (or compact elongated figure) which oscillates between the two axes as the two waves fade differently. This means that all of the audio frequencies of one branch are combining in phase with those of the other.

The above brief description was introduced to acquaint the reader with the essentially simple features of the MUSA system. Before describing the details and the results obtained with the experimental system, a more comprehensive analysis of steering principles will be given.

Returning to Fig. 1, it is assumed, of course, that the transmission lines are terminated in their characteristic impedance at the receiving terminal [11] (the phase shifters of Fig. 1) so that the phase is distributed linearly along the lines. Neglecting line loss (or equalizing it), the N currents, equal in magnitude and different in phase, are

$$\left.\begin{aligned} i_1 &= I\epsilon^{j\omega t} \\ i_2 &= I\epsilon^{j[\omega t+\phi-2\pi a(v-\cos\delta)]} \\ i_3 &= I\epsilon^{j[\omega t+2[\phi-2\pi a(v-\cos\delta)]]} \\ &\cdots\cdots\cdots\cdots \\ i_N &= I\epsilon^{j[\omega t+(N-1)[\phi-2\pi a(v-\cos\delta)]]} \end{aligned}\right\} \quad (1)$$

where i = instantaneous current in exponential notation
ω = angular frequency
N = total number of unit antennas
a = spacing in free space wave-lengths
$v = c/v$ = the ratio of the velocity of light to that of the transmission line.

The sum of the N currents is

$$A = I\epsilon^{j\omega t}\{1 + \epsilon^{j[\phi-2\pi a(v-\cos\delta]} + \cdots + \epsilon^{j(N-1)[\phi-2\pi a(v-\cos\delta)]}\}. \quad (2)$$

[11] Non-characteristic terminations at the receiving ends of the lines are permissible if all terminations are identical and if the antennas are matched to the characteristic line impedance. Conversely, characteristic terminations at the receiving ends suffice if the antenna impedances are merely identical.

This exponential series may be evaluated with the aid of the identity [12]

$$1 + \epsilon^{j\theta} + \epsilon^{j2\theta} + \cdots + \epsilon^{j(n-1)\theta} \equiv \frac{\sin\frac{n\theta}{2}}{\sin\frac{\theta}{2}} \epsilon^{j\frac{(n-1)\theta}{2}}.$$

Using this summation we have

$$A = I \frac{\sin\frac{N}{2}[\phi - 2\pi a(v - \cos\delta)]}{\sin\frac{1}{2}[\phi - 2\pi a(v - \cos\delta)]} \epsilon^{j[\omega t + (N-1)/2[\phi - 2\pi a(v - \cos\delta)]]}. \quad (3)$$

The amplitude of A in (3) is the array directional pattern or array factor. It is zero when the numerator alone is zero, i.e., when

$$\frac{1}{2}[\phi - 2\pi a(v - \cos\delta)] \neq 0, \pm\pi, \pm 2\pi \cdots$$

and simultaneously

$$\frac{N}{2}[\phi - 2\pi a(v - \cos\delta)] = 0, \pm\pi, \pm 2\pi \cdots.$$

It attains its maximum value of NI when the denominator and numerator are zero simultaneously, i.e., when

$$\frac{1}{2}[\phi - 2\pi a(v - \cos\delta)] = 0, \pm\pi, \pm 2\pi \cdots$$

and

$$\frac{N}{2}[\phi - 2\pi a(v - \cos\delta)] = 0, \pm\pi, \pm 2\pi \cdots.$$

Plots of (3) for ten unit antennas ($N = 10$) spaced five wave-lengths ($a = 5$) are shown in Fig. 4 for two arbitrary values of ϕ. The same array used at twice the frequency ($N = 10$, $a = 10$) has the directional patterns shown in Fig. 5. The abscissas are labeled earth angle although nothing has been said thus far concerning the disposition of the N antennas with respect to the earth. In order that the simple multiple phase shifts of Fig. 1 shall suffice to steer the array, reflection from the ground must affect the phase of all antenna outputs identically. This is assured by constructing the array over, and parallel to, a flat expanse of ground. Since the angle δ measures the direction of

[12] This identity may be deduced by substituting $\epsilon^{j\theta}$ for r in the well-known formula for the sum of a geometrical progression

$$1 + r + r^2 + r^3 + \cdots + r^{n-1} \equiv \frac{r^n - 1}{r - 1}.$$

the wave referred to the direction of the array axis the array factor represents a surface of revolution. Figures 4 and 5 show merely axial cross sections which, for a horizontal array, may be considered vertical plane patterns.

Equation (3), as well as Figs. 4 and 5, shows that the sharpness of the principal lobe depends upon the total length of the array in wave-

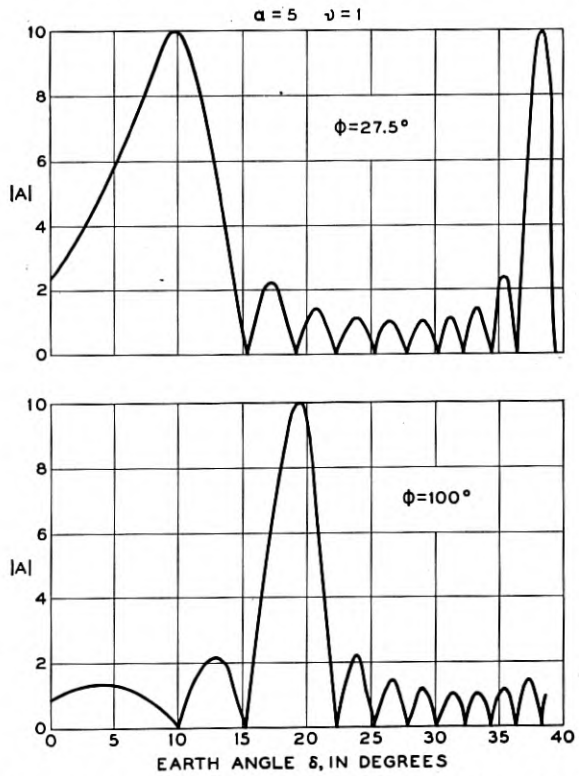


Fig. 4—Plots of the array factor for a 45-wave-length horizontal end-on array.

lengths, i.e., upon Na, while the angular spacing of adjacent principal lobes depends inversely upon the spacing "a." Thus, a single lobed pattern results if the array consists of a large number of closely spaced units.

A single lobed pattern is desirable, but to obtain it by using a large number of unit antennas [13] with separate transmission lines and phase

[13] The reader may observe that the reduction of the spacing would, if carried so far as to make "a" a fraction of a wave-length, violate the assumption that there is negligible reaction or coupling between unit antennas. As stated, this assumption is made in the interest of simplicity. It is theoretically possible to compensate for coupling between antennas so that (1), (2), and (3) still hold.

shifters would be a rather extensive undertaking. Provided a restricted range of steering is permissible, a simpler solution is to employ comparatively few large unit antennas and to let their directional pattern suppress the undesired principal lobes of the array pattern. Useful angles for transatlantic circuits are confined to the range from zero, or some low undetermined limit, to some higher limit. In what follows let δ_m represent an angle a little above the useful range so that a

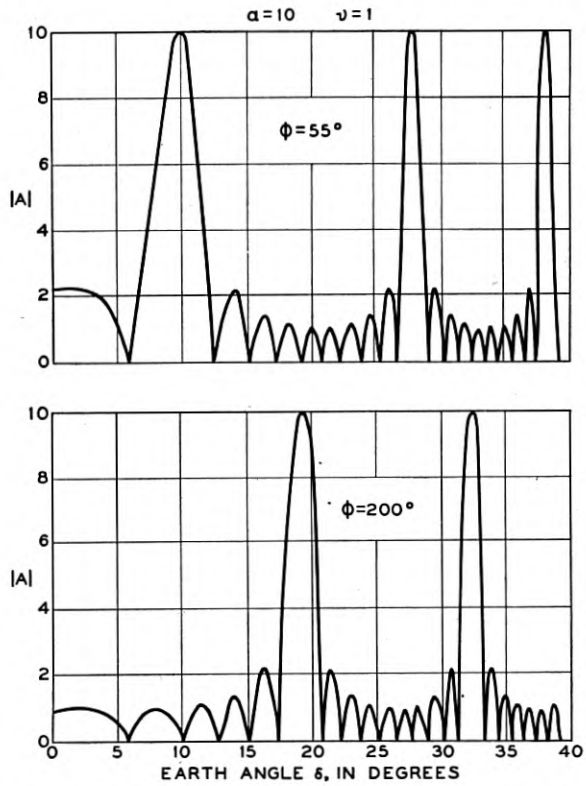


Fig. 5—Plots of the array factor for a 90-wave-length array; that of Fig. 4 used at twice the frequency.

null may be located at δ_m without imposing an excessive loss. The array may then be designed so that when the first principal lobe is steered at zero angle the second falls at δ_m or beyond. The question of whether the array design permits the construction of a suitable unit antenna in the length $a\lambda$ allotted to it is considered in the following paragraph. As a matter of fact, this analysis closely follows the actual steps in the development of the MUSA system.

Turning back to the ideal system comprising a very large number of closely spaced unit antennas, which yields the single lobed pattern, let us divide the antennas into N groups with n antennas in each group. Calling the group spacing "a" and the phase shift between adjacent antennas ϕ the application of (3) gives, dropping the exponential factor,

$$A = \frac{\sin \frac{nN}{2}\left[\phi - 2\pi \frac{a}{n}(v - \cos \delta)\right]}{\sin \frac{1}{2}\left[\phi - 2\pi \frac{a}{n}(v - \cos \delta)\right]}. \tag{4}$$

Multiplying numerator and denominator by

$$\sin \frac{n}{2}\left[\phi - 2\pi \frac{a}{n}(v - \cos \delta)\right]$$

results in

$$A' = \frac{\sin \frac{n}{2}\left[\phi - 2\pi \frac{a}{n}(v - \cos \delta)\right]}{\sin \frac{1}{2}\left[\phi - 2\pi \frac{a}{n}(v - \cos \delta)\right]} \times \frac{\sin \frac{N}{2}[n\phi - 2\pi a(v - \cos \delta)]}{\sin \frac{1}{2}[n\phi - 2\pi a(v - \cos \delta)]}. \tag{5}$$

Equation (5), which appears as the product of two array factors, is merely another way of writing the array factor for the large number (Nn) of unit antennas. The first factor represents an array of n "sub-unit" antennas of spacing a/n and phase shift ϕ. The second represents the array of these arrays with a spacing of $a\lambda$ and a phase shift $n\phi$. We now proceed to treat these two factors independently and assign the values ϕ_f and ϕ_v to replace ϕ and $n\phi$, respectively. Figure 6 depicts such an array of arrays. If now we regard the array of n sub-units as constituting a fixed unit antenna and adjust it to receive at zero angle (by putting $\phi_f = 2\pi a/n(v - 1)$) in accordance with the lower limit of the useful range, we obtain

$$A'' = \frac{\sin \frac{n}{2}\left[2\pi \frac{a}{n}(1 - \cos \delta)\right]}{\sin \frac{1}{2}\left[2\pi \frac{a}{n}(1 - \cos \delta)\right]} \times \frac{\sin \frac{N}{2}[\phi_v - 2\pi a(v - \cos \delta)]}{\sin \frac{1}{2}[\phi_v - 2\pi a(v - \cos \delta)]}. \tag{6}$$

The first factor in (6) represents the pattern of the unit antenna. It is a relatively broad single lobed pattern with maximum response at

$\delta = 0$. It drops to zero at $\delta = \cos^{-1}(1 - 1/a)$. For higher angles nothing but minor maxima occur since "a" is small and n is large. The second factor represents the steerable array pattern of the N unit antennas. With ϕ_v adjusted for maximum response at zero angle (this makes $\phi_v = n\phi_f$) this system is identical with the array of Nn subunits. In this case, all principal lobes of the array factor for the N units, excepting the first, coincide exactly with nulls of the array factor for the n subunits, and the familiar tapered distribution of minor maxima associated with the array of Nn subunits results. As ϕ_v is varied to steer for other angles than $\delta = 0$, the coincidence of nulls and undesired principal lobes no longer occurs. Since, however, the

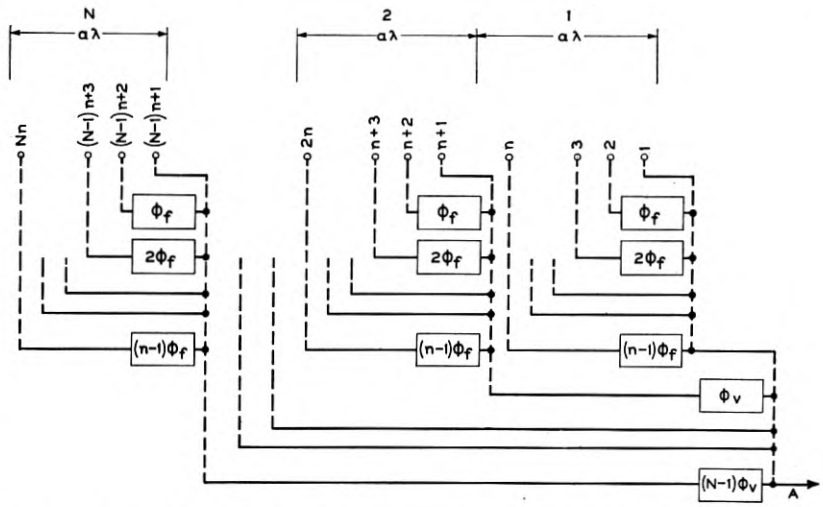


Fig. 6—A steerable array formed by dividing the antennas of Fig. 1 into N groups of n each. The subscripts "f" and "v" refer to fixed and variable phase shifts.

fixed unit antenna has only minor response beyond its first null, those undesired principal lobes are adequately suppressed, and the array may therefore be steered anywhere within the range from $\delta = 0$ to $\delta = \cos^{-1}(1 - 1/a)$, with single lobed response. As the principal lobe is steered away from $\delta = 0$ the maximum amplitude falls off in comparison with that of the array of Nn subunits. This represents a loss of signal-to-noise ratio and is to be regarded as a penalty for compromising to the extent of using fixed arrays as unit antennas. The loss is appreciable, however, only if the array is steered near the upper cutoff angle of the unit antenna. It remains but to select "a" so that $\cos^{-1}(1 - 1/a)$ represents the upper limit of the range, δ_m.

For a fixed physical spacing, "a" varies inversely with the wave-length, which results in an increasing steering range with increasing wave-length. Since the critical angle of reflection from the ionosphere increases with wave-length the upper limit of the range of useful angles can be expected to increase also. By selecting the proper spacing, the steering range and the critical angle can be made to agree satisfactorily. Figure 7 shows a plot of $\delta = \cos^{-1}(1 - 1/a)$ against wave-length for the unit antenna spacing of 200 meters which was

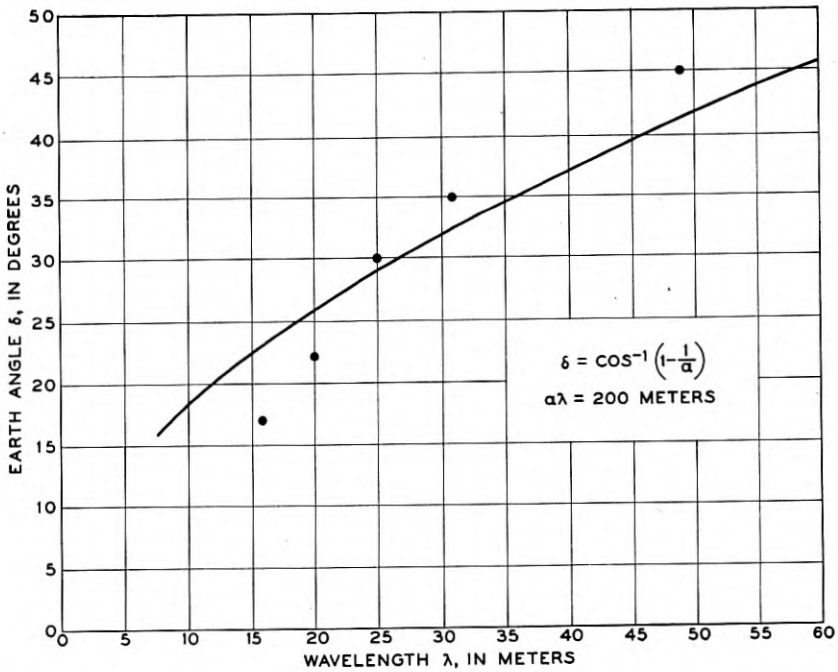


Fig. 7—Measured upper limits of vertical angles as a function of wave-length, compared with the upper limit of the array depicted in Fig. 6. The measured values represent the highest angles observed; usually stronger waves of lower angles predominate.

adopted for the experimental MUSA system to be described. The points denote upper limits of earth angles obtained from measurements made during the years 1933–1936 on signals from Rugby [5] and Daventry, England.

The foregoing analysis shows that

(1) A MUSA system may be so proportioned that the upper limit of its steering range follows, with fair accuracy, the upper limit of the range of useful angles, as the wave-length is varied.

(2) It is theoretically possible to construct a suitable unit antenna in the space provided for it when (1) is satisfied.

III. Description of the Experimental MUSA System

Antennas and Transmission Lines

Any type of unit antenna whose directional pattern suppresses the undesired principal lobes over the required wave-length range is basically suitable for use in a MUSA system. The rhombic antenna [14] does not fulfill this requirement as well as the linear array of subunits discussed in the preceding section. It was, however, selected on account of its advanced state of development. The manner in which it fits into the MUSA array factor will be discussed later.

The coupling or "cross talk" between antennas need not be of negligible magnitude in a MUSA system. For, to a first approximation,

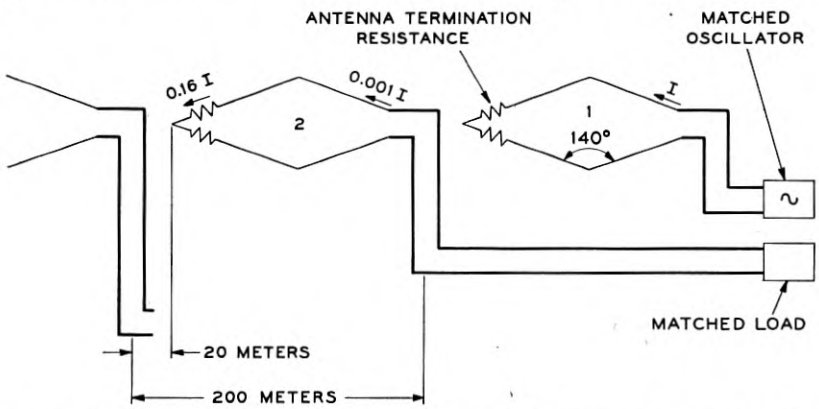


Fig. 8—Measurements of cross talk between adjacent antennas in the MUSA as made from the transmitting point of view.

the coupling is confined to adjacent antennas and is similar for all pairs so that only the *end* antennas could be expected to fail to combine properly with the others. At the ends, "dummy" antennas, not connected with the receiver but terminated like the others, could be erected to supply the coupling necessary to make all antennas alike. Measurements made on the experimental MUSA (Fig. 2) indicated that the cross talk is small enough to be neglected, however, so that dummy antennas ahead of or behind the six regular ones were considered unnecessary. The performance of the system in subsequent tests corroborates this conclusion.

The crosstalk measurements yielded the results indicated on Fig. 8. The small amount of crosstalk current ($0.001I$) measured at the trans-

[14] Bruce, Beck, and Lowry, "Horizontal Rhombic Antennas," *Proc. I. R. E.*, vol. 23, pp. 24–46, January, 1935; *Bell Sys. Tech. Jour.*, January, 1935.

mission line end of the forward antenna (No. 2) and the larger current ($0.16I$) at the other end reflect the fact that the rhombic antenna is "unidirectional." To a first approximation the current in such an aperiodic antenna accumulates progressively towards the output end. Therefore, the "effective" cross talk current is probably less than $(0.16I + 0.001I)/2 = 0.08I$; i.e., the effect upon the field radiated in the principal lobe will be altered by less than ten per cent due to the parasitic excitation of the antenna ahead. Antennas farther ahead as well as those behind contribute relatively nothing.

Since by the reciprocal theorem the directional pattern of any antenna is the same for transmitting and receiving, the crosstalk should likewise result in less than 10 per cent effect in the receiving case.

The measurements of Fig. 8 were made at 18 megacycles. At this frequency the rhombic antennas are proportioned to give maximum radiation approximately end-on. At lower frequencies the crosstalk is probably less.

The coaxial transmission lines are constructed of 60-foot lengths of one-inch copper plumbing pipe spliced with screw type plumbing unions. The inner conductor is one-fourth inch in diameter and is supported by isolantite insulators. The characteristic impedance of the lines is 78 ohms. The lines extend up the poles where they are connected to the antennas through balanced-to-unbalanced matching transformers.[14] At the receiving building the lines terminate on a special jack strip. Nitrogen pressure is maintained in all lines to exclude moisture.

In order to operate the MUSA system it is not essential that the velocity of the transmission lines be known. The velocity must be known accurately, however, in order to determine the angle of the waves as they are selected by the steerable lobe. Accordingly, the velocity was calculated (taking the insulators into account) and also measured. The calculated ratio of the line velocity to the velocity of light is 0.941; measurements yielded 0.933 ± 0.004. Using the value of 0.933, angles less than zero have occasionally been measured. A value of 0.937 would have made the lowest indicated angle just zero.

The longest line is about 1000 meters in length. Its impedance measured at one end when the other end is terminated by a resistance of 78 ohms shows some variation as the frequency is varied. In Fig. 9 are shown the results of impedance measurements made by substituting for the line an equivalent parallel combination of resistance and reactance. The two notable variations occurring at approximately 7.7 and 15.4 megacycles are believed to be caused by a slight irregularity at each joint, which adds a shunt capacitance of the order of 1.8 micro-

microfarads. When spaced regularly at 60-foot intervals these capacitances have a somewhat cumulative effect at frequencies for which 60 feet (18.3 meters) is a multiple of the half wave-length. Sixty feet, when increased by the line velocity ratio, corresponds to 7.7 and 15.4 megacycles. Clearly, line sections which are not short compared with the shortest wave-length should be made unequal so that joint irregularities will not be harmful. The smaller variations of the order of $\pm$ 10 ohms may be due to random eccentricities produced by slight buckling of the inner conductor between insulators. With the possible

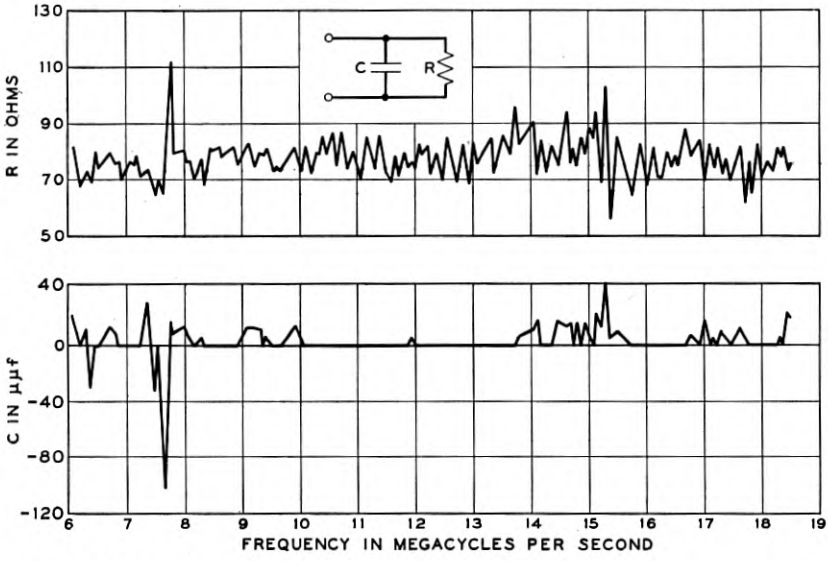


Fig. 9—Impedance measurements made upon the 1000-meter line terminated in a resistance of 78 ohms. The reactance is expressed as shunt capacitance, negative values meaning an inductive reactance numerically equal to the corresponding capacitive reactance.

exception of the two large variations this line is sufficiently smooth for use in a MUSA, as both theory and subsequent experience indicate.

Input Circuit and First Detectors

The MUSA system imposes requirements upon the input circuits and detectors which do not apply to conventional receivers. These requirements are as follows:

(1) The circuits must suppress standing waves on the transmission lines.[15]

[15] This requirement was more easily met than the alternate requirement mentioned in footnote (11).

(2) The phase shift from the transmission line to the phase shifter stage must be alike in all six circuits, independent of wave-length.

In order to simplify the experimental job it was decided to dispense with the selectivity afforded by high-frequency amplifiers and to use the simple circuits shown in Fig. 10. The capacitive coupling to the transmission line is a convenient means of matching the low-impedance lines to the high-impedance circuits. Plug-in coils (L) are used to cover the range from 4.5 to 22 megacycles.

The first detectors are of the two-tube balanced type which suppresses interference from two signals differing by the intermediate

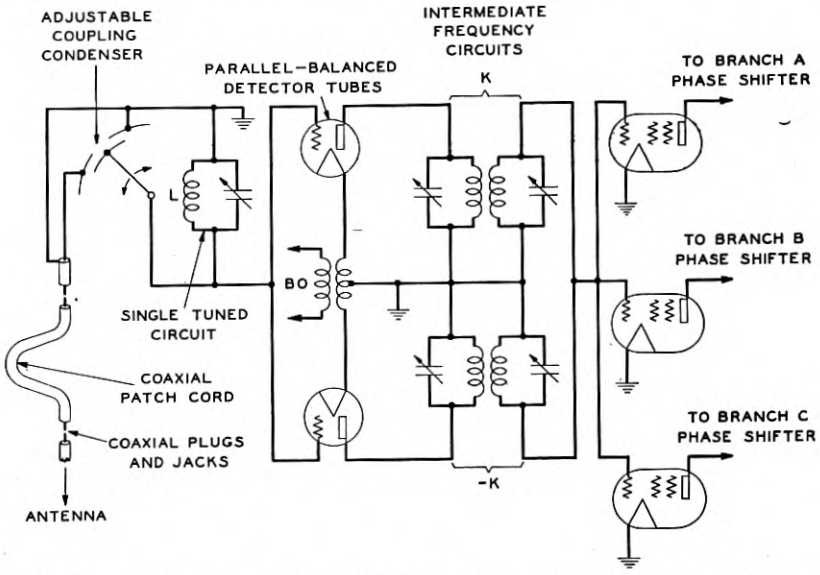


Fig. 10—Input circuit, first detector, and first intermediate-frequency tubes.

frequency and isolates the beating oscillator supply from the input circuits. The latter prevents crosstalk between the six inputs, and assures independence in the tuning of the input circuits. The beating oscillator voltage is introduced, at low impedance, between cathodes [16] by means of the distributing system of equal length coaxial lines shown in Fig. 11. This distributing system gives equiphase beating oscillator inputs to all detectors and makes requirement (2) attainable by having nominal similarity in the remaining parts of the six circuits.

Requirement (1) is met by feeding a test oscillator of 78 ohms impedance into the first circuit jack and adjusting the tuning condenser

[16] W. A. Harris, "Superheterodyne Frequency Conversion Systems," *Proc. I. R. E.*, vol. 22, pp. 279–294, April, 1935.

and the coupling condenser (Fig. 10) alternately until the maximum signal voltage appears on an indicating meter in one of the three intermediate-frequency branches. The three-terminal coupling condenser is an aid in this procedure since varying the coupling imposes only a slight variation in the capacitance across the coil. When the indicating instrument is a square-law vacuum tube voltmeter with the main

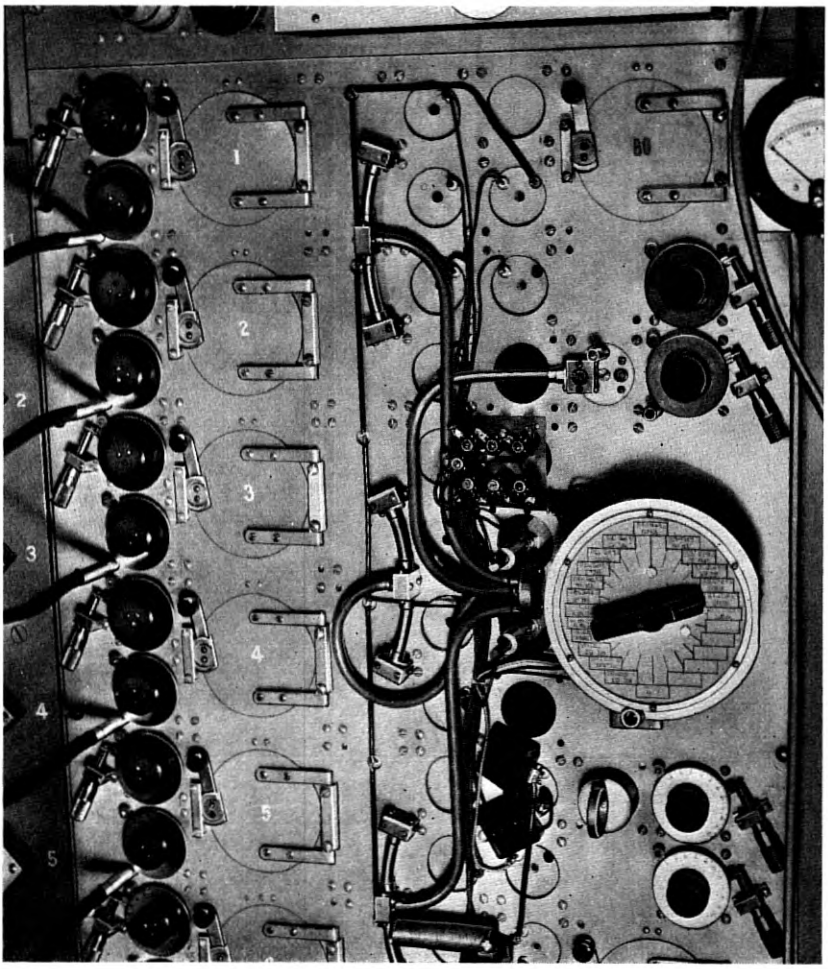


Fig. 11—Close-up view of high-frequency panel with cover removed. The beating oscillator supply line originates in the upper right-hand corner. It supplies the six detectors with equiphase and equiamplitude voltages. Plug-in coils fit into the compartments covered by the six circular doors. Micrometer heads which are used to adjust the six tuning condensers appear. The coaxial patch cords appear at the extreme left.

current balanced out and the remainder indicated by a 30-microampere meter, the sensitivity is more than sufficient to tune the circuits correctly.

The criterion of correct tune is the degree of suppression of standing waves on the transmission lines. To determine whether or not the maximizing adjustment insures an adequate standing wave suppression, a standing wave detector was incorporated in the experimental design. This is shown in Fig. 12. It consists of about 16 meters of 78-ohm coaxial line arranged in a coil and terminated by the

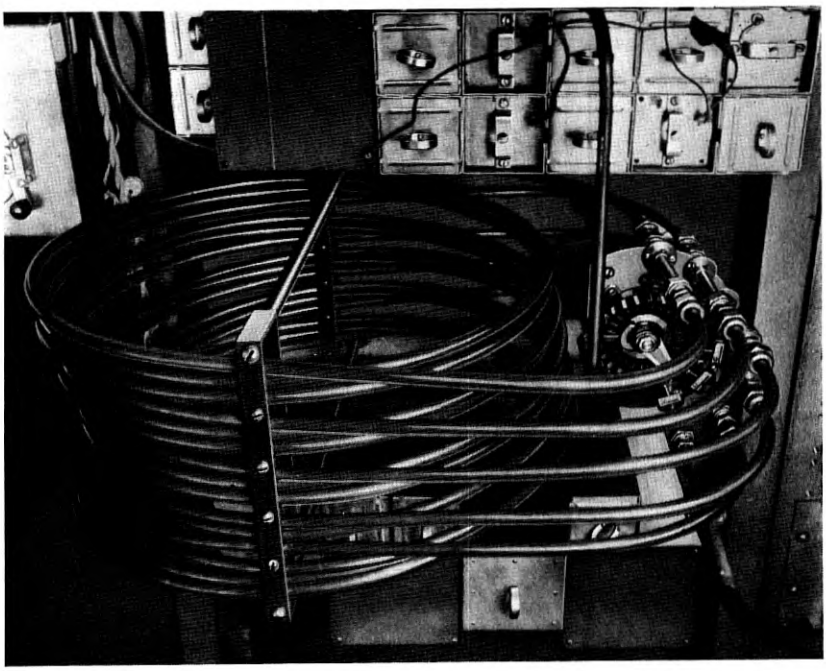


Fig. 12—The standing wave detector comprising 50 feet of 3/8-inch coaxial line, which may be used to test the correctness of the input circuit adjustment.

first circuit to be tested. It is fed at the other end by a test oscillator. Six capacitively coupled taps are brought to the low-capacitance switch shown in the photograph. The selector arm connects the taps to an auxiliary receiver with a high-input impedance. The absence of standing waves is shown by equal readings at the six positions. It was found that the maximizing adjustment results in a standing wave with less than ten per cent total variation, which represents nearly as much suppression as the smoothness of the line allows.

With nominally correct resistance termination standing waves of five per cent usually occur. For standing waves not exceeding ten per cent the accompanying phase distribution along the line does not depart more than a few degrees from the desired linear distribution. The use of the standing wave detector in routine operation was therefore not required.

Phase Shifters

Of the numerous methods of shifting phase the method [17] illustrated in Fig. 13 is the one chosen for the 18 circuits (3 branches, 6 antennas) of the experimental MUSA. Here points a, b, d, and c have voltages

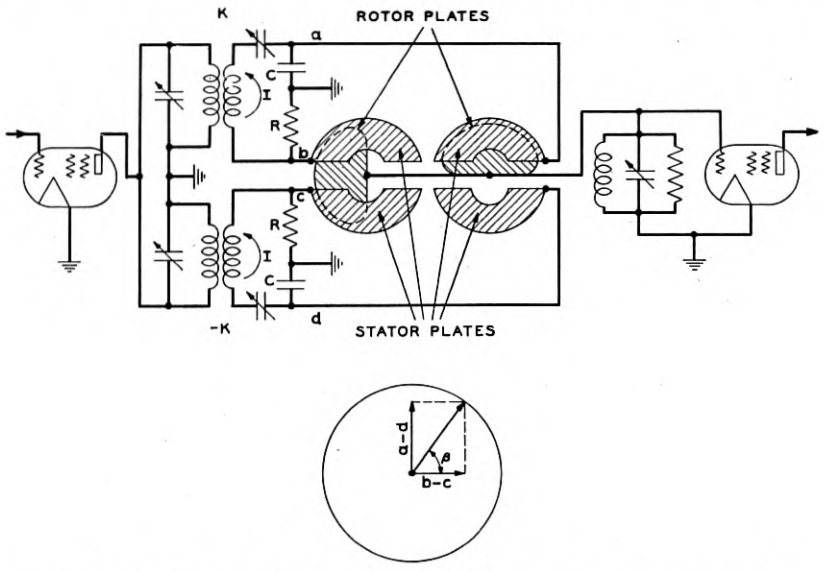


Fig. 13—Circuit diagram and vector diagram of the phase shifter. The rotor plates are especially designed to give a phase shift proportional to shaft angle.

to ground 90 degrees apart. The potential of point b is IR; that of c is $-IR$; that of a is $jI/\omega C$; that of d is $-jI/\omega C$. The resistance R and reactance $1/\omega C$ are made equal at the mid-band frequency so that four equal voltages, distributed equally over 360 degrees of phase, appear on the four stators of the special condenser. A photograph of this condenser appears in Fig. 14. Two specially shaped eccentric rotors mounted in quadrature to each other on the same shaft comprise the output terminal. It will be noted that voltages of opposite phase are connected to adjacent stators. Thus, with the rotors in

[17] L. A. Meacham, U. S. Patent No. 2,004,613.

the position shown dotted in Fig. 13 the output comes from point a since d is not coupled and b and c cancel each other. By shaping the two rotors so that the difference in exposure to opposite stator plates is proportional, respectively, to the sine and cosine of the angle of shaft rotation, the total current flowing from the two rotors will be constant and of phase proportional to the shaft angle. This is illustrated by the vector diagram in Fig. 13 in which β is the shaft angle and vectors $a - d$ and $b - c$ are the quadrature rotor outputs proportional to $\sin \beta$ and $\cos \beta$.

These phase shifters vary in output by less than $\pm$ 5 per cent as the shaft is rotated. The departure from linearity of phase shift is correspondingly small; i.e., less than $\pm$ 5 degrees.

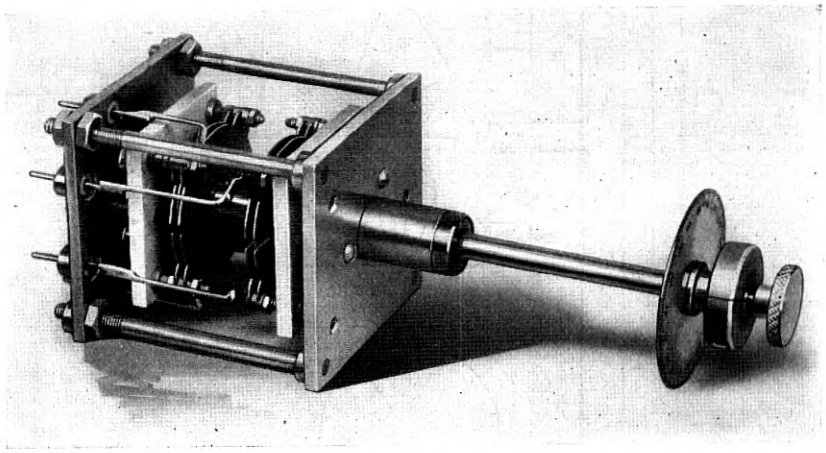


Fig. 14—The phase shifting condenser.

The useful band width of this type of phase shifter is fundamentally limited by the fact that $1/\omega C$ varies with frequency while R does not. However, this limitation does not appear in the Holmdel MUSA in which the percentage band width is small because the phase shifters operate at the intermediate frequency of 396 kilocycles.

The phase shifters are connected to the steering shaft with helical gears of multiple ratios as shown in Fig. 15. The phase shifter shafts may be slipped with respect to the main shaft. After they have been aligned so that locally supplied equiphase inputs to all detectors add in phase at the point where the phase shifter outputs are combined they are locked. This adjustment is independent of signal frequency. Provision is made for adjusting the gain of each of the six phase shifter circuits so that the differences in transmission-line loss may be com-

pensated and any other desired amplitude adjustments made. The photograph of Fig. 15 shows the monitoring or exploring branch whose steering shaft is motor driven at one revolution per second.

Before leaving the subject of phase shifting it may be well to distinguish between phase shift and delay as here used. All electrical net-

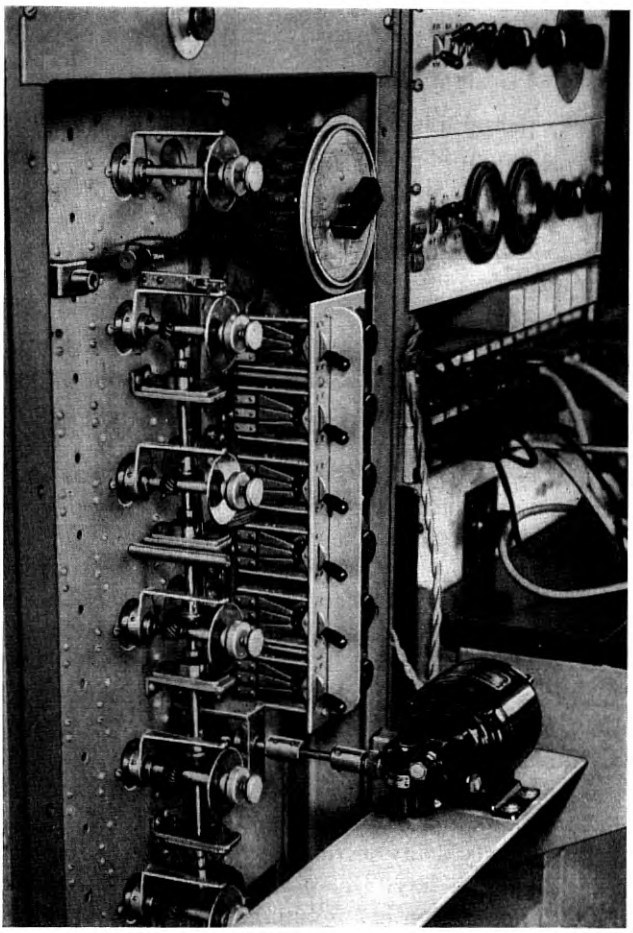


Fig. 15—Phase shifting panel of the monitoring branch. Only five of the six phase shifters are rotated for steering purposes. They are geared to the steering shaft in ratios of 1 : 1, 1 : 2, 1 : 3, 1 : 4, and 1 : 5.

works, except for certain highly distortive ones, possess a phase-frequency characteristic which is such that higher frequencies have their phases retarded with respect to lower frequencies. The ratio of the ncrement of phase retardation to the increment of frequency, i.e., the

slope of the phase characteristic, is the delay. It is sometimes called the group delay or group transmission time as distinguished from the "phase time." [18] The delay is the only time which can be measured. It does not determine the phase shift of a particular frequency nor is it determined by the phase shift. A phase shifter applied to the network merely moves the phase curve intact up or down on the phase axis.

General Description of the System

The preceding paragraphs have described features which distinguish the MUSA system from conventional receiving systems. There remain to describe several auxiliary features and to present a unified picture of the whole.

The experimental system was designed for double side-band reception and all of the results reported in this paper refer to double side band. There has recently been completed equipment which may be substituted for the double side-band equipment for the reception of reduced carrier single side-band signals. The new equipment may also be used to select, with crystal filters, one side band of double side-band signals.

The delay to be inserted in the low-angle branch as indicated in Fig. 3 is obtained electrically from an audio-frequency delay network. The delay could theoretically be provided at the intermediate frequency but no advantage would result. The audio-frequency delay network is a special artificial line composed of forty sections and terminated by its characteristic impedance. Each section has a delay of 68 microseconds. A special switch is arranged to tap a high impedance output circuit across any desired section, thus providing a delay of 2.7 milliseconds variable in 0.068-millisecond steps. A special equalizing network [19] which makes the transmission loss the same for all steps and which also equalizes the frequency-loss characteristic so that the response is flat to 5000 cycles for all steps is automatically controlled by this switch. The forty delay sections appear in Fig. 16 just under the shelf on the right-hand bay. The maximum delay which has been required in actual operation is 2.5 milliseconds.

Both linear rectifiers and square-law detectors are provided for final demodulation and either may be switched into service as desired. The

[18] This distinction is brought out by J. C. Schelleng in a "Note on the Determination of the Ionization of the Upper Atmosphere," *Proc. I. R. E.*, vol. 16, pp. 1471–1476, November, 1928.

A general discussion of delay distortion (phase distortion) is to be found in three papers appearing in the *Bell Sys. Tech. Jour.*, vol. 9, July, 1930.

[19] This network and the delay sections were designed by P. H. Richardson of Bell Telephone Laboratories, Inc.

automatic gain control for use with either demodulator is obtained from linear rectifiers but a different diversity connection is made for each type of demodulator, in the interest of output volume constancy. A choice of time constants of 0.06, 0.5, and 4 seconds is provided.

Keys are provided, the ganged manipulation of which makes it possible, among other things, to compare (1) the MUSA output versus any one of the six antennas connected to one branch receiver, and (2) any pair of antennas in ordinary diversity using both branch receivers, versus one antenna using one receiver.

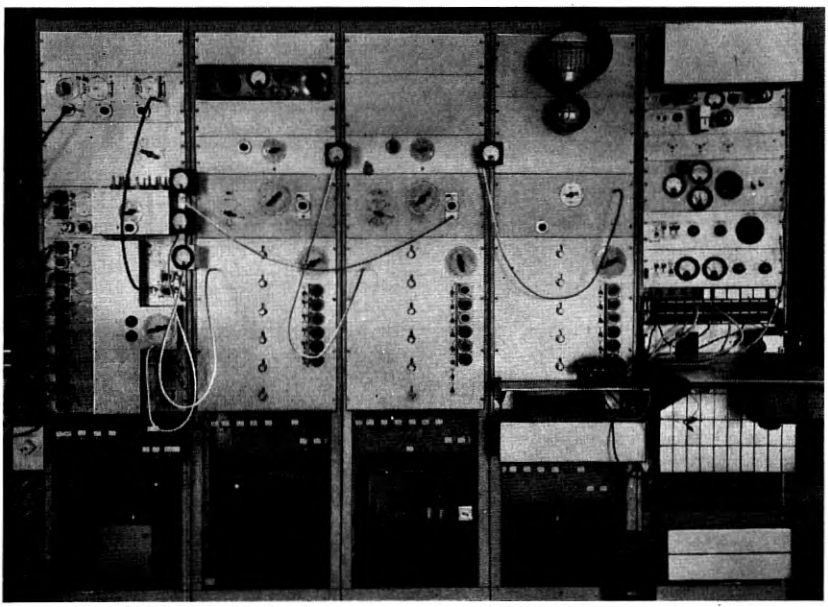


Fig. 16—Front view of the MUSA receiving equipment. The high-frequency bay is at the left and the audio-frequency bay at the right. The branch receivers are the panels directly above the phase shifting panels. The pulse receivers appear above these. At the top of the bay containing the monitoring branch equipment are the two oscilloscopes referred to in Fig. 3. The large tube with the ruled face is the monitoring oscilloscope.

In addition to the regular branch receivers with a 12-kilocycle band width and the monitoring branch receiver with a 2.5-kilocycle band width, two other receivers are provided in the experimental system. These receivers have a 30-kilocycle band width and are used for pulse reception. They are bridged across the inputs of the two regular branch receivers and are connected to a cathode-ray oscilloscope through a commutator.[5]

Various photographs of the MUSA receiver appear with explanatory captions in Figs. 16, 17, and 18.

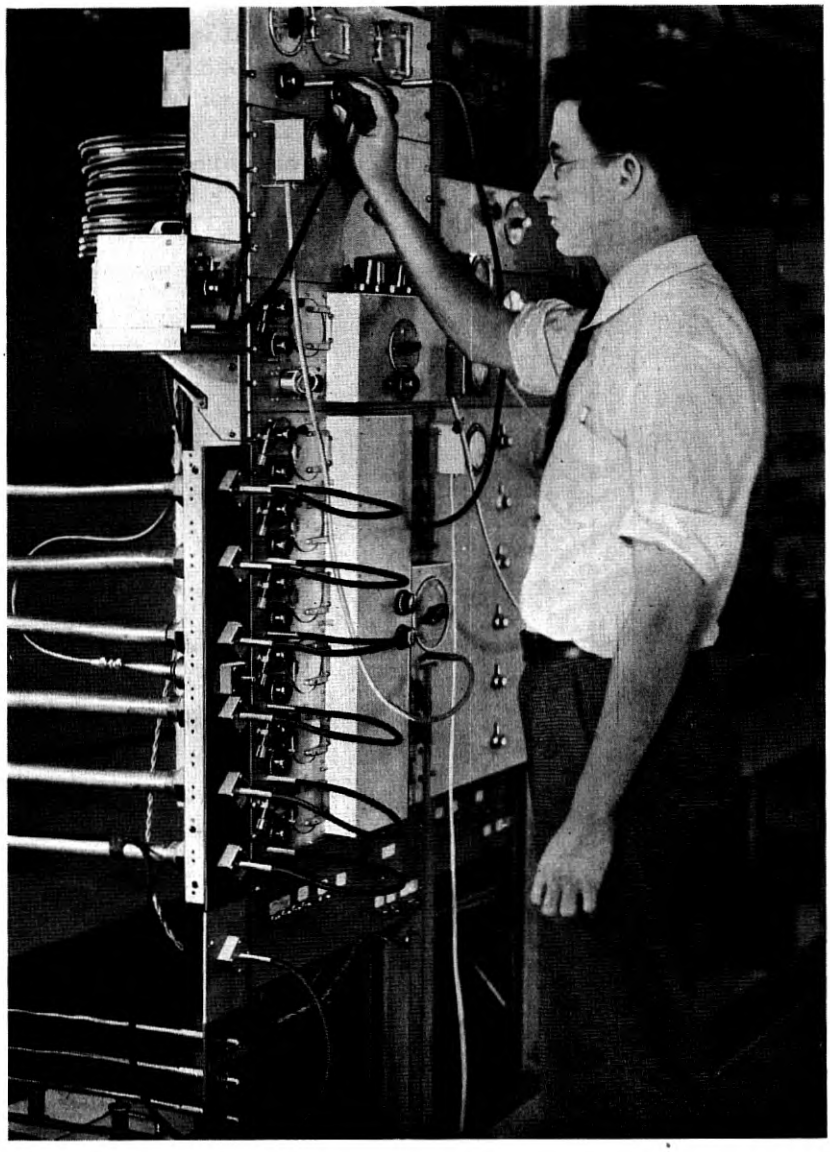


Fig. 17—View showing the six transmission lines and coaxial patch cords. The beating oscillator is mounted upon the shelf and is connected to the power amplifier (which is being adjusted by Mr. Edwards) at the top of the bay.

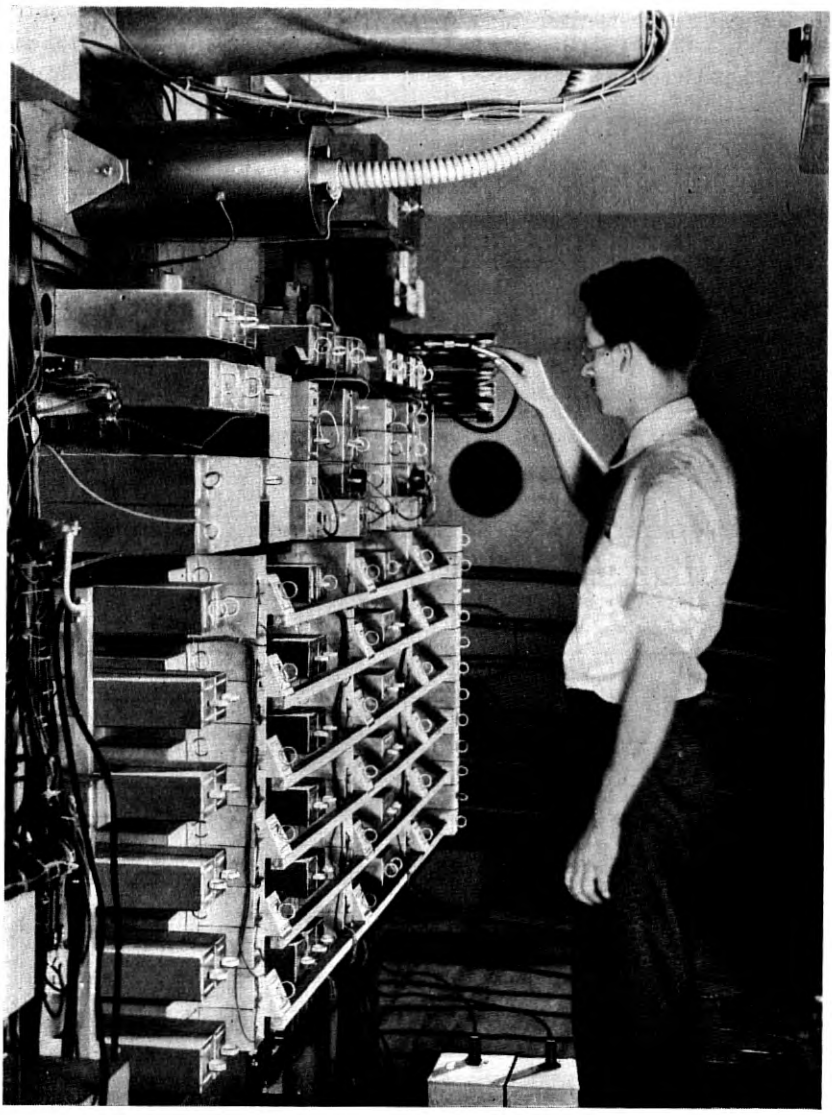


Fig. 18—Rear view of the receiving equipment. The six detector outputs feed the three branches via the square transmission lines.

A family of calculated directional patterns of the experimental MUSA is shown in Figs. 19 and 20. At the top of each column is shown the principal lobe of the vertical directional pattern of the unit rhombic antenna, calculated in the median plane. Beneath are shown six vertical patterns of the MUSA, which are obtained by multiplying the

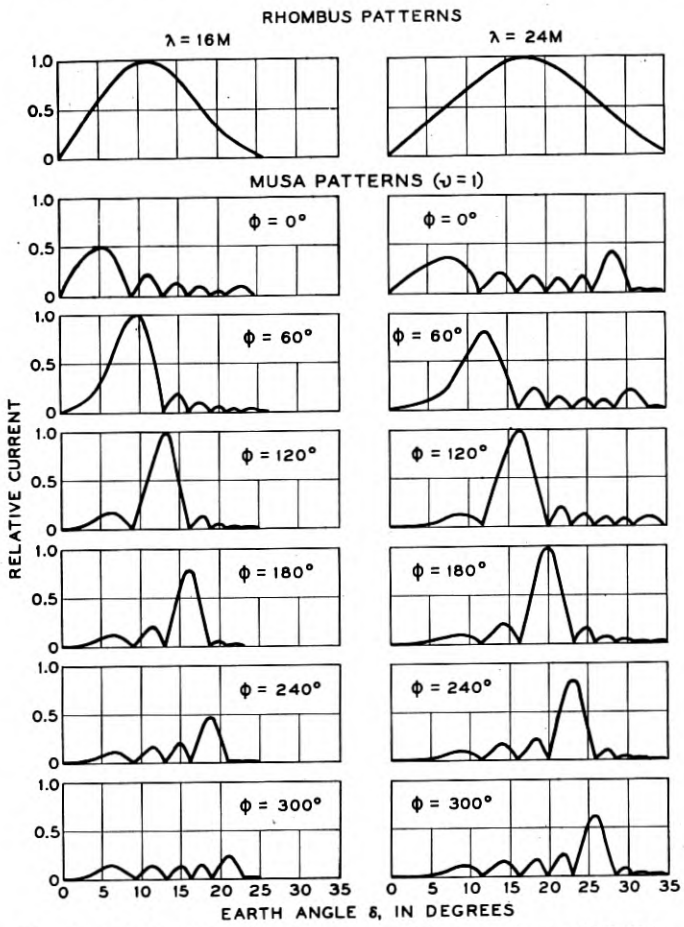


Fig. 19—Vertical directional patterns of the experimental MUSA.

array factor [20] by the unit antenna pattern. The upper pattern corresponds to phasing for zero angle. The remaining ones are plotted for increments of 60 degrees of phase.

These patterns fall short of the "ideal," which the reader may have visualized while reading Section II, in two ways. First, the unit an-

[20] Calculated from (3) putting $v = 1$.

tenna does not suppress the second lobe of the array factor as well as could be desired. By design, it does so for the short waves but inherently fails to do so for the longer waves. Second, the principal lobe of the unit antenna shifts bodily towards higher angles with increasing wave-length, whereas it is desirable to have only the upper cutoff move

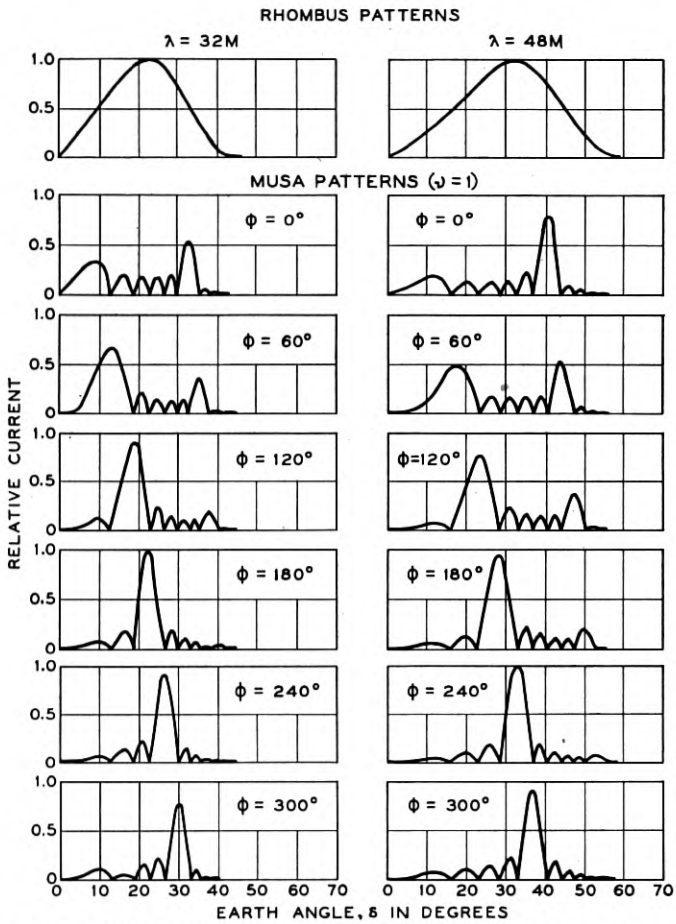


Fig. 20—Vertical directional patterns of the experimental MUSA. Note that the angle scale is half of that in Fig. 19.

upward while retaining low-angle response. Both of these shortcomings restrict the steering range through which the MUSA is essentially single lobed. The upward shift with wave-length of the unit antenna response is in fair agreement with the way in which the mean angle of arrival has been found to vary with wave-length and is,

on that account, not altogether objectionable. The Holmdel MUSA employing such unit antennas represents, however, a considerable departure from present antennas of fixed directivity designed from statistical data, and approaches the ideal MUSA steerable over the entire useful angle range.

The curves as plotted assume that the differences in transmission line loss for the various line-lengths have been equalized in the intermediate-frequency circuits. By slightly tapering the amplitudes so that the antennas in the middle of the array contribute more than those near the ends a reduction of the minor lobes has been obtained at the cost of slightly widening the principal lobe. As a result of this, the directional discrimination of the experimental MUSA has been improved. All data and photographic records reported in this paper, however, were obtained before this improvement was introduced.

IV. Tests and General Evaluation [21]

Tests and Experience

Numerous experiments and tests had been carried out on the various parts of the MUSA system before it was first tuned to a transatlantic signal. Despite the fact that all tests concurred in predicting that the system would perform as designed, it was with considerable gratification that a pattern was observed on the monitoring oscilloscope, during one of the early trials, which was almost exactly as calculated for a single wave. Patterns corresponding to two or more waves in various degrees of resolution were observed from time to time. To increase the angle resolution, for test purposes, pulses were transmitted by the British Post Office on several occasions. Turning the steering shaft during these tests clearly showed the principal lobe sweeping through the angle range. When fairly discrete pulses were received the minor lobes could be readily identified. In Fig. 21 is shown a sample of motion picture oscillograms of pulse reception. Two principal waves or, more accurately, wave bundles occurred and were separated by the two MUSA branches as shown. For details of the pulse technique employed in these tests the reader is referred to a previous publication.[5]

Before exhibiting sample motion pictures of typical patterns displayed by the angle monitoring oscilloscope and the delay indicator oscilloscope, further discussion of the former is desirable. The photographs of Fig. 22 show the monitoring oscilloscope pattern with a locally produced equiphase, equiamplitude input supplied to each

[21] The theory and test results of the signal-to-noise advantage are considered together in Part V.

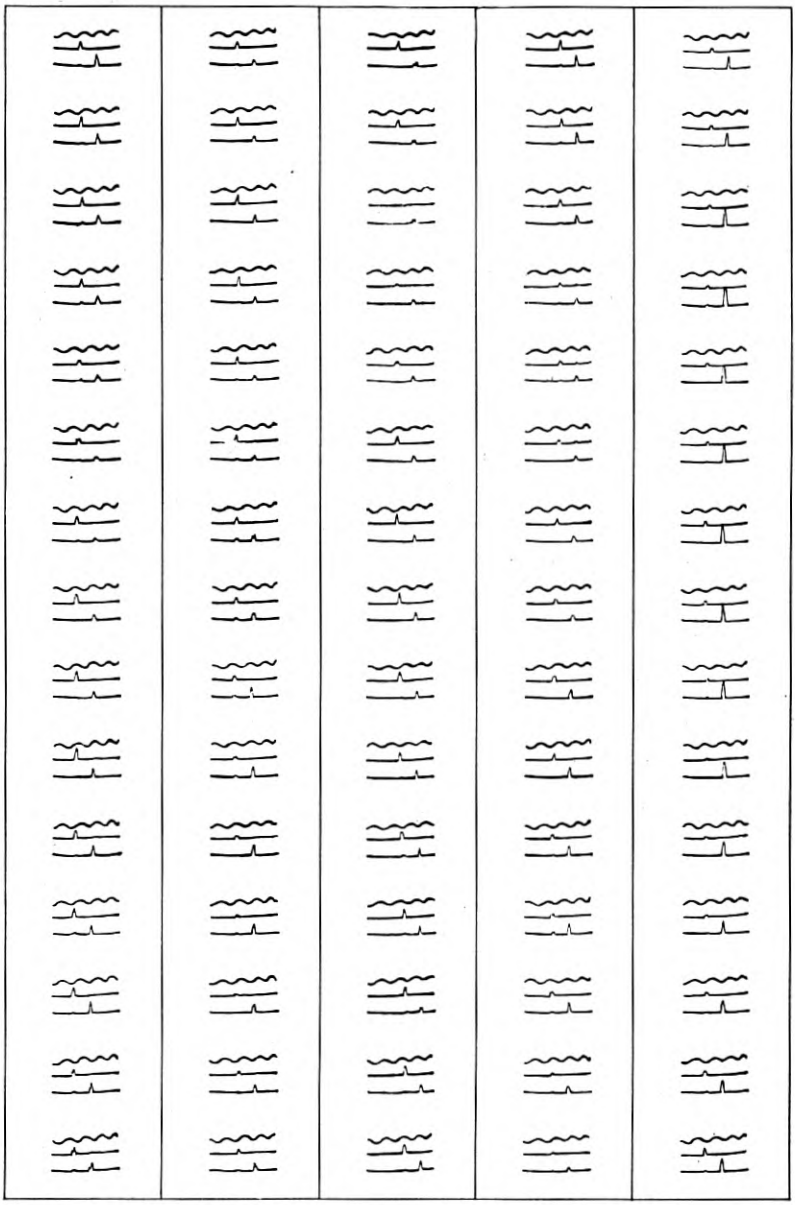


Fig. 21—This retouched plate shows pulses received with MUSA. On each frame time advances from left to right and is measured by the thousand-cycle timing wave. The center trace shows the output of one MUSA branch steered at 25 degrees. The bottom trace shows the output of the second MUSA branch steered at 32 degrees. Wide band amplifiers for pulse reception are bridged in parallel with the speech band intermediate-frequency amplifiers. The transmitted pulses are about 200 microseconds long. GCS (9020 kilocycles) Rugby, February 25, 1936, at 4:11 P.M., E.S.T.

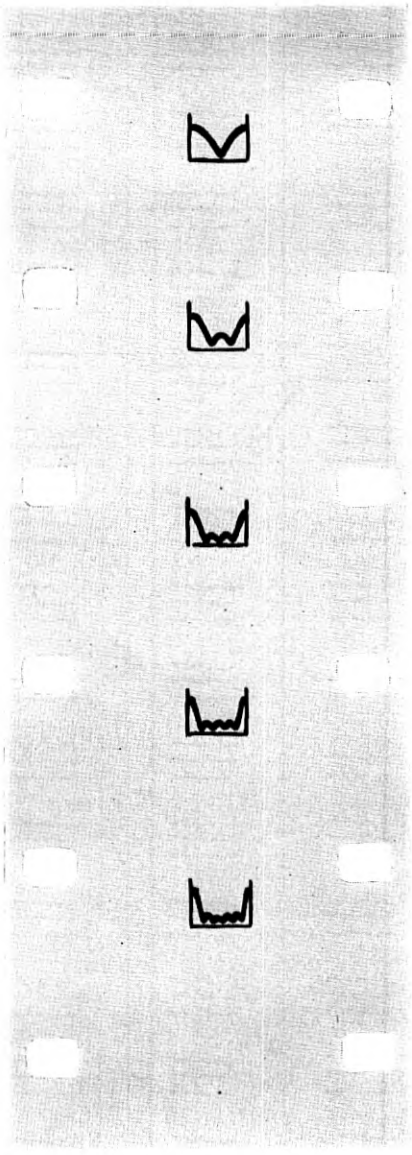


Fig. 22—These five frames show the angle monitoring pattern when a local signal is used to simulate a wave. The bottom frame shows the ideal MUSA pattern for one wave. The remaining frames show the effect of reducing the number of antennas from six (1-2-3-4-5-6) to five (2-3-4-5-6) to four (3-4-5-6) to three (4-5-6) to two (5-6). These films were taken before the amplitude tapering was introduced. Tapered amplitudes reduce the minor lobes to about half of the amplitudes shown.

detector. This figure illustrates the manner in which the pattern is built from the six components after the manner of a Fourier synthesis. The vertical and horizontal axes visible on the monitoring oscilloscope in Fig. 16, but which do not appear naturally on the photographs, were drawn in Fig. 22. As mentioned previously, the oscilloscope sweep axis represents one revolution of the "fundamental" phase shifter so that the beginning and end of the sweep represent the same condition. The ends of the sweep are arbitrarily fixed to represent zero (or 360) degrees of phase shift referred to the output of the first antenna, whose phase is not varied. Consequently equiphase inputs result in a principal lobe half of which appears at each end. This would correspond to a wave of zero angle if the velocity of the transmission lines was equal to that of light. For a lesser velocity, zero angle may occur at any point on the phase axis, depending upon the wave-length. (See Fig. 28 for a sample angle calibration curve.) The principal lobe as well as the four minor lobes of the monitoring oscilloscope represents the output from one wave as the MUSA is steered through its entire range. The oscilloscope pattern, unlike the directional pattern, does not appear sharper for short wave-lengths than for long wave-lengths; the principal lobe is always 120 degrees wide and the minor ones 60 degrees wide on the phase axis. One degree of phase difference, however, represents a difference in steering angle which depends upon the wave-length and the earth angle.

The samples of motion picture film shown in Fig. 23 represent fairly typical "two-path" patterns. The camera was focused to include both oscilloscopes and was manipulated by means of a special step-by-step crank. The operator endeavored to expose each frame during one sweep of the monitoring tube. The delay indicator tube shows a continuous pattern produced by the audio frequencies. A correct delay setting is indicated by a straight line. Here, with the two branches steered at the indicated angles of 8.5 and 20.5 degrees, a delay of 950 microseconds was required to produce the straight line. The diversity action is apparent in the tilting of this line. When the low angle wave, which corresponds to the left-hand peak on the monitoring tube, is predominant the delay indicator line becomes horizontal and, conversely, when the high angle wave is predominant the line approaches the vertical axis. Automatic gain control is used on the branch receivers supplying the speech outputs but is not used on the monitoring branch.

Figure 24 shows, in samples 1 and 2, reception of two waves which are just separable by the directivity present in the Holmdel MUSA. The angles are 15 and 22 degrees and the wave-length is 31.6 meters. The

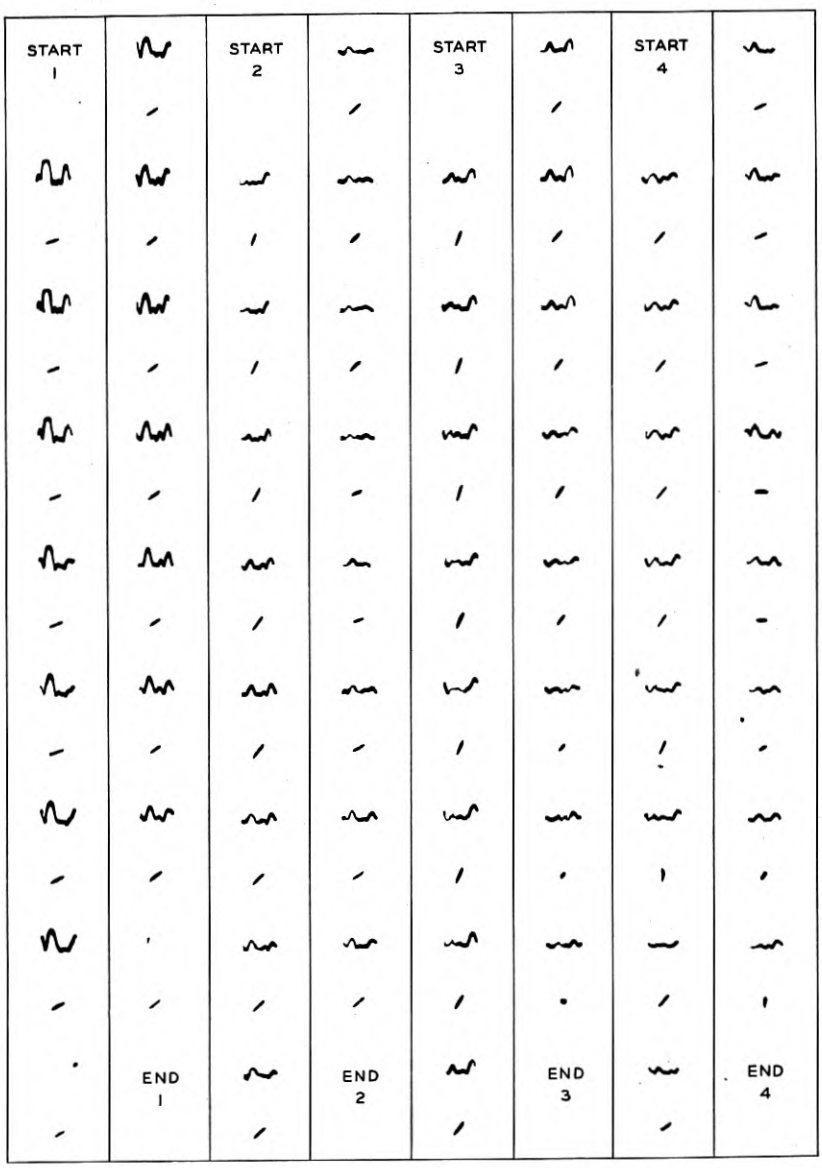


Fig. 23—Pictures of the angle monitoring oscilloscope and the delay indicator tube. The right-hand end of the sweep of the monitoring tube represents zero phase. The indicated angles are 8.5 and 20.5 degrees. The MUSA branches were set at these angles ($\phi_A = 240°$, $\phi_B = 30°$) using 950 microseconds delay. GSE (11,860 kilocycles) Daventry, February 21, 1936, 11:05 A.M., E.S.T. Musical program.

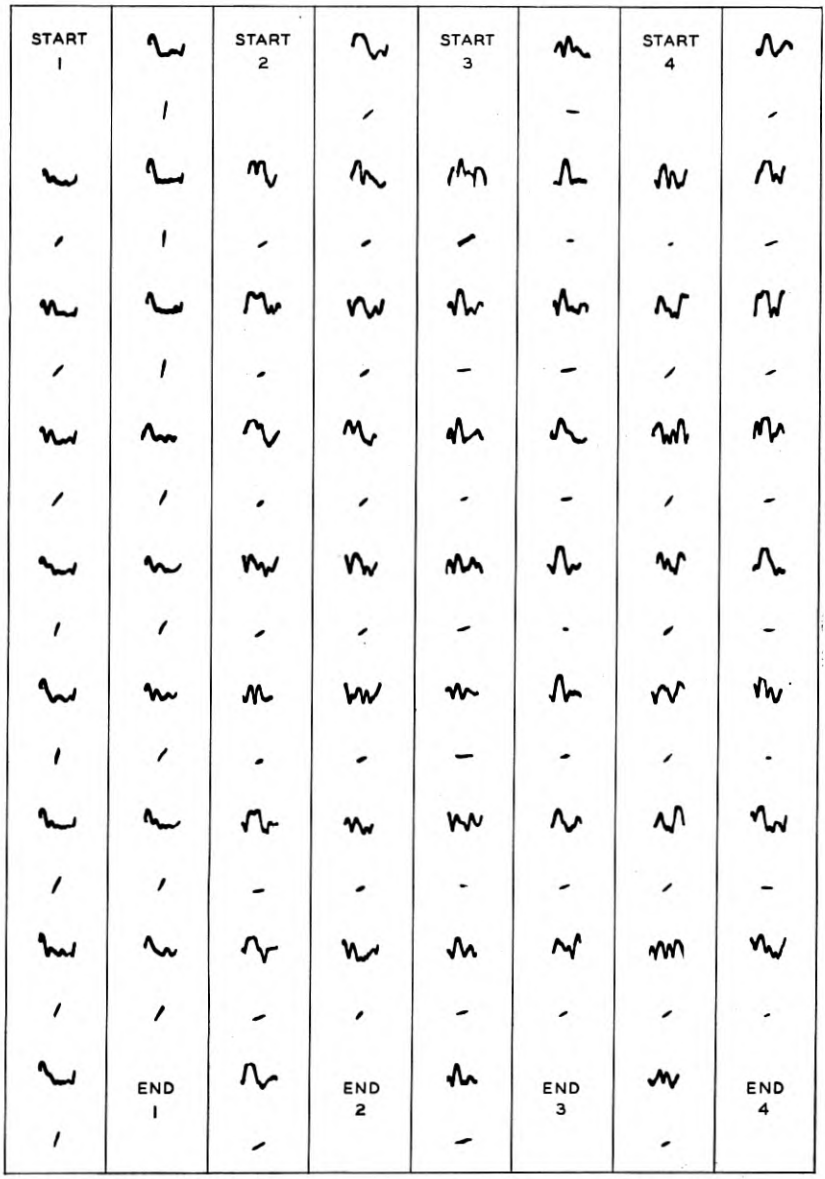


Fig. 24—Pictures of the angle monitoring oscilloscope and the delay indicator tube. In films 1 and 2 the indicated angles are 15 and 22 degrees. The MUSA branches were set at these angles ($\phi_A = 240°$, $\phi_B = 340°$) and a delay of 400 microseconds was used. The time was 3:20 P.M., E.S.T. Films 3 and 4 were taken at 3:15 when a third wave of 26 degrees was present. One branch was steered at this wave ($\phi_B = 50°$); the other at 16 degrees ($\phi_A = 250°$). A delay of 1000 microseconds was required. GSB (9510 kilocycles) Daventry, March 10, 1936.

delay used is 400 microseconds. In samples 3 and 4 a third wave of 26 degrees is present. One branch was steered at this wave; the other was steered at the 15-degree wave and a delay of 1000 microseconds was used.

It is of interest to compare these samples showing the manner in which the MUSA branch outputs combine, with the samples in Fig. 25 which were obtained with a two-antenna space diversity setup. Six antennas were retained in the monitoring branch but five were cut out of each receiving branch, leaving one antenna to supply each branch. In samples 1 and 2, antennas 1 and 6 (1000 meters apart) were retained. In samples 3 and 4, adjacent antennas (Nos. 1 and 2) 200 meters apart were used. These records were obtained about 15 minutes later than those of Fig. 23 and show the same two waves at 8.5 and 20.5 degrees. No delay was used. Note that the outputs combine in phase only when one wave predominates. Inserting delay in either branch is, of course, not effective in improving the audio combination. To do so would impair the addition when one wave is predominant and would not be beneficial when both waves are comparable.

Figure 26 shows, in samples 1, 2, 3, and 4, how the delay indicator tube pattern is affected by the delay adjustment. The two branches were steered at the same angle, thus making both branch outputs identical so that perfect delay adjustment occurs with zero delay. This is the condition depicted in sample 1. In samples 2, 3, and 4 the delays are 340, 680, and 2700 microseconds, respectively.

A number of tests were carried out with the cooperation of the British Post Office in which twelve tones were transmitted. These tones were nonharmonically related. They were separated at the output of the receiver by means of filters, and commutated to appear successively on an oscilloscope. The reader is referred to a paper [4] by R. K. Potter describing this technique. Figure 27 shows a sample of motion pictures made of the oscilloscope patterns. Two receiving systems are compared; the right-hand pattern shows the output of the MUSA while the left shows the output of a conventional receiver connected to a horizontal half-wave antenna. The tones trace the horizontal lines in sequence from top (425 cycles) to bottom (2125 cycles). After one pattern is executed the commutator switches from one receiver to the other. The twelfth tone is omitted to provide time for the switching. The complete double pattern is traced in about one-sixth of a second and the camera is operated at a speed which exposes each frame a little longer than one-sixth of a second.

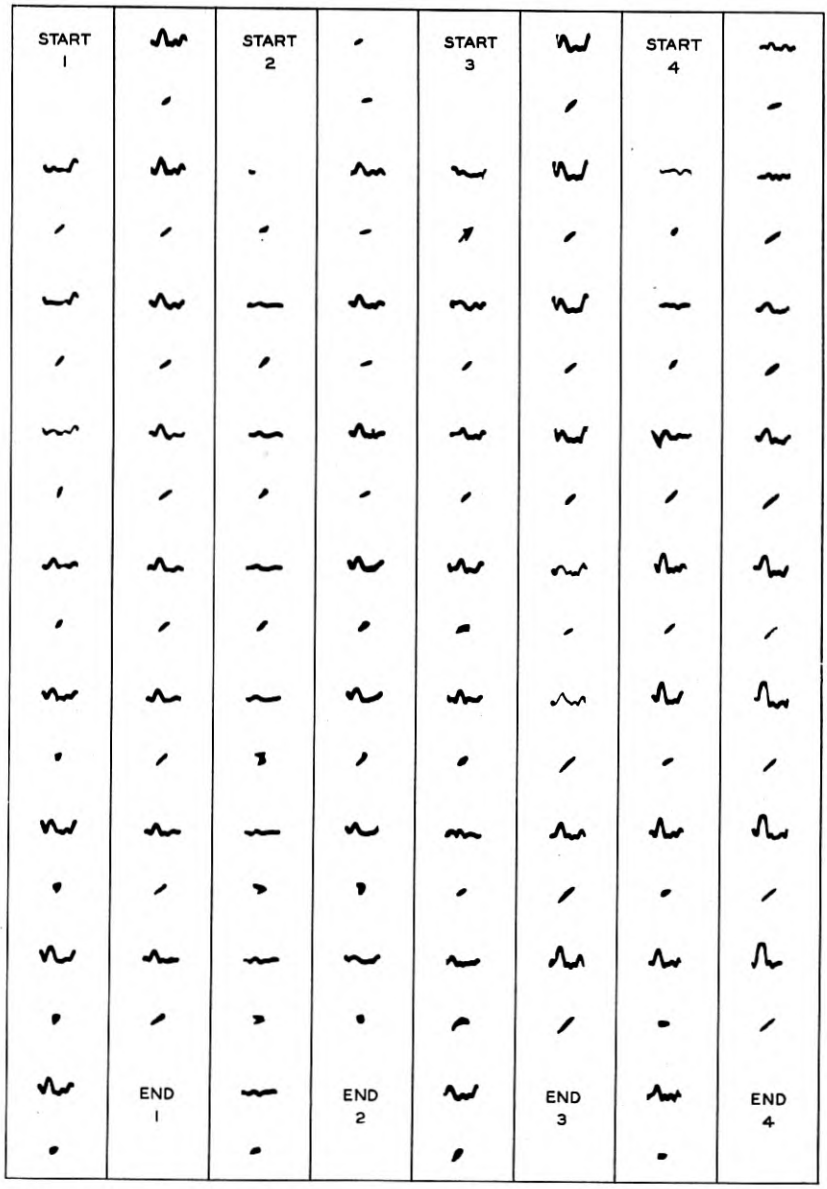


Fig. 25—This plate, made immediately following that of Fig. 23, shows for comparison the manner in which the audio outputs add in two-station space diversity. The angle monitor shows the 8.5- and 20.5-degree waves as before. Films 1 and 2 taken at 11:15 A.M., E.S.T., were obtained with rhombic antennas 1 and 6 (40 wavelengths apart). Films 3 and 4 taken at 11:20 were obtained with antennas 1 and 2 (8 wave-lengths apart). Note the second harmonic in film 2 particularly. GSE (11,860 kilocycles) Daventry, February 21, 1936. Musical program. Zero delay.

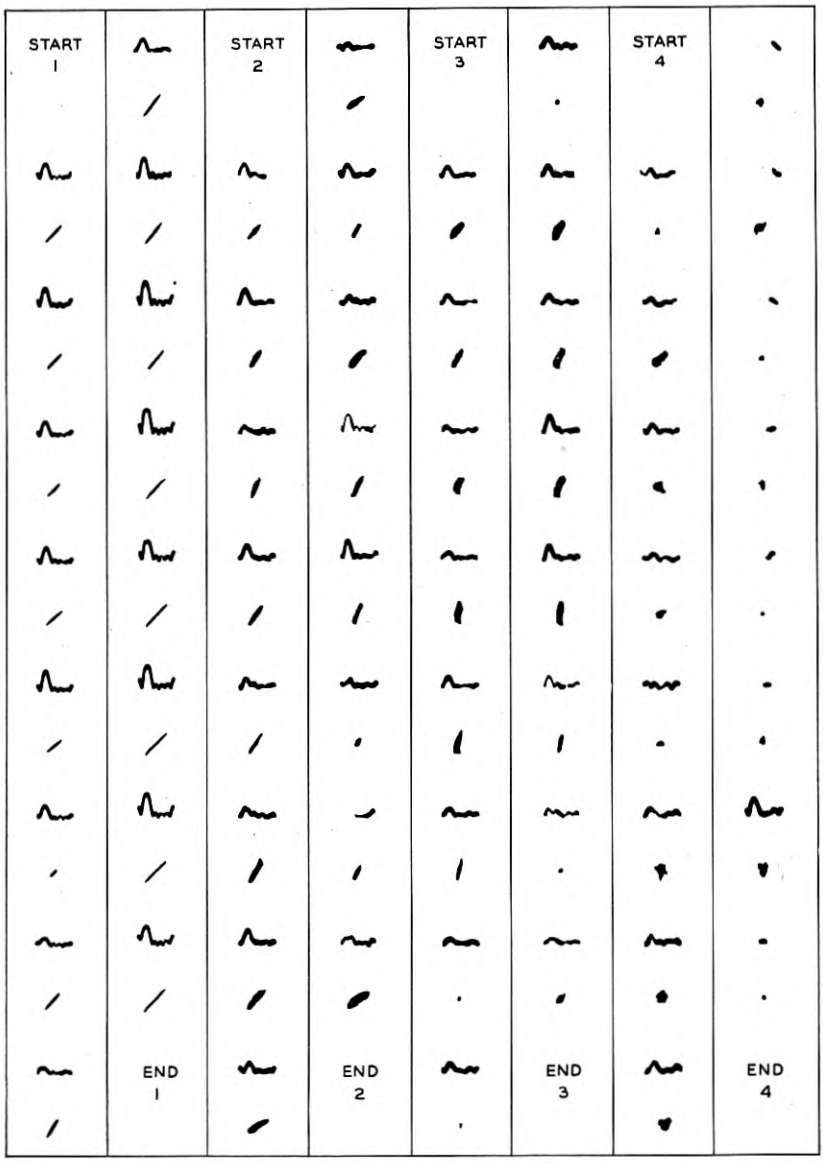


Fig. 26—Film showing the effect upon the audio addition of unequalized delay. Both MUSA branches were steered at the same angle,—that of the major wave shown. Film 1 shows no delay added and since each branch receives the same wave the audio outputs add perfectly. Films 2, 3, and 4 show the effect of adding 340, 680, and 2700 microseconds delay, respectively.

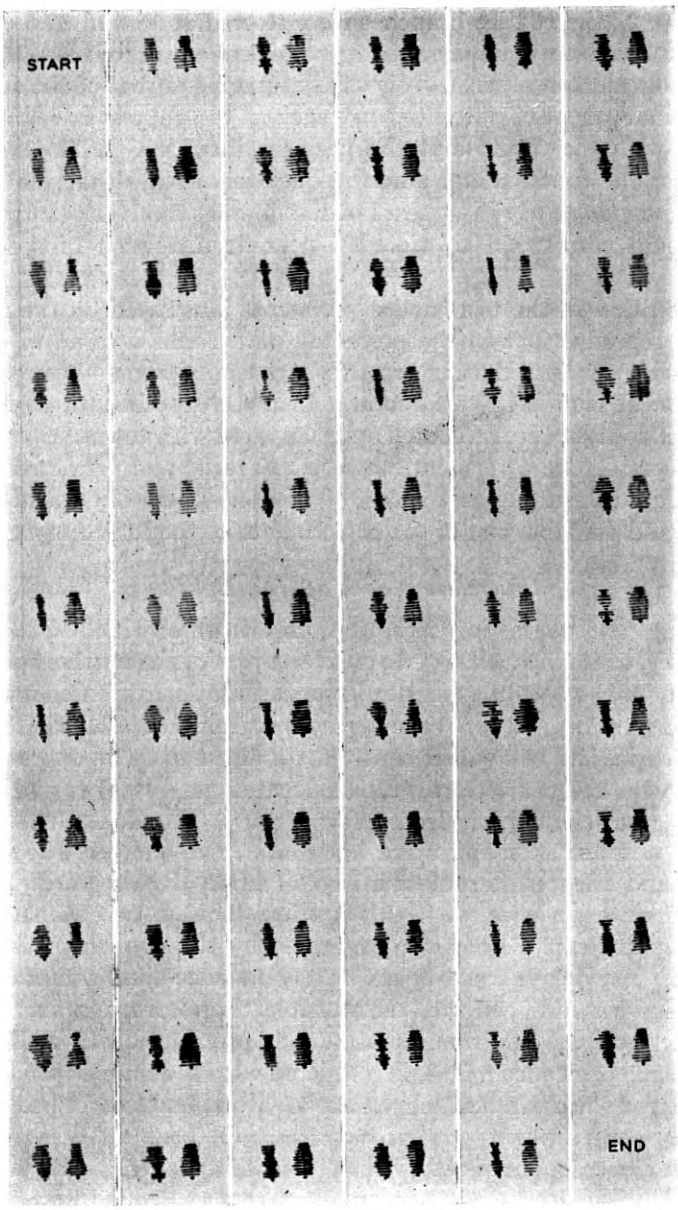


Fig. 27—This is a cathode-ray "multitone" record comparing the MUSA output with that of a horizontal half-wave antenna. The tones on the right-hand side of each frame are the MUSA output; those on the left are the output of the horizontal half-wave antenna. The two MUSA branches were steered at 15 and 22.5 degrees. A delay of 470 microseconds was used to equalize the transmission time. GCS (9020 kilocycles) Rugby, February 24, 1936, at 3:54 P.M., E.S.T.

In Fig. 27 the MUSA branches were steered at 15 and 22.5 degrees and employed an equalizing delay of 470 microseconds. While the MUSA output is not perfect it is vastly superior to that of the doublet. The tone frequencies and filters are such as to suppress harmonic distortion with the result that the patterns show mainly the selective fading of the fundamental audio frequencies. Note that the fundamental output nearly disappears in the doublet receiver. In practice this would correspond to violent harmonic distortion of speech or music.

In addition to the tests and experiments illustrated by the motion picture reproductions in the preceding paragraphs a series of experiments were conducted using broadcast transmission on 49 meters from a station at Halifax, Nova Scotia. In these experiments angles and delay differences were measured and compared with the multiple reflection theory. The agreement between measured and predicted values is not only interesting as a study of the ionosphere but constitutes a unique and valuable test of the performance of the MUSA system.

Observations on VE9HX, Halifax

During the course of reception experiments with GSL (BBC, Daventry, 6110 kilocycles) performed as a part of the routine operating program for the MUSA system, a broadcast station appeared on GSL's frequency. This station carried the programs of CHNS, Halifax, Nova Scotia, and was subsequently determined to be an experimental station with the call letters VE9HX located hear Halifax and nearly on the great-circle path from New York to London. The transmitting antenna is a half-wave horizontal, one-quarter wave above ground and oriented to radiate in the direction of New York.

The first experience with this station showed two stable transmission paths capable of being separated by the two branches of the MUSA. The delays could be accurately equalized and rather definite correlation was obtained with the multiple "hop" propagation picture. This fact and the additional reason that propagation from England on the same frequency might be compared with the simpler phenomena encountered with Halifax led to the measurements described in the following paragraphs.

About eleven hours of observation, distributed over fifteen days, are included. The log aimed to record all changes which occurred during an observation period. The procedure was as follows: The two branches of the receiver were steered at the angles indicated by the monitoring oscilloscope. Delay was added to the lower angle branch until the two audio outputs added. The delay setting was usually

critical to one section of the network (67.5 microseconds) and always to two sections. The angles were determined from the calibration curve reproduced in Fig. 28. The phase readings observed on the monitoring oscilloscope were recorded to within ± 10 degrees and the earth angles determined by them are liable to be in error by one degree (possibly 1.5 degrees) apart from the ambiguity due to the multiple lobe characteristics of the MUSA. At this wave-length, the major lobe of the unit rhombic antennas is broad, the first null occurring at 58 degrees, so that two angles had to be considered possible.

The multiple hop picture is illustrated in Fig. 29. Here the delay

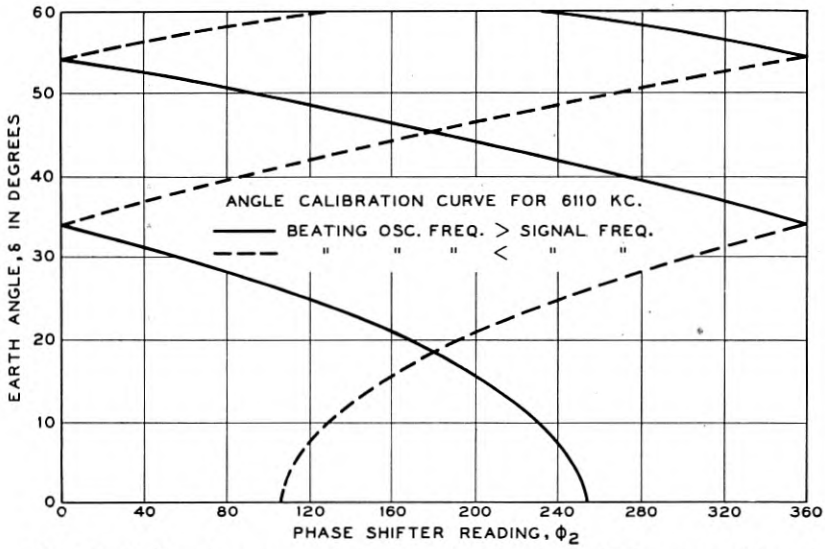
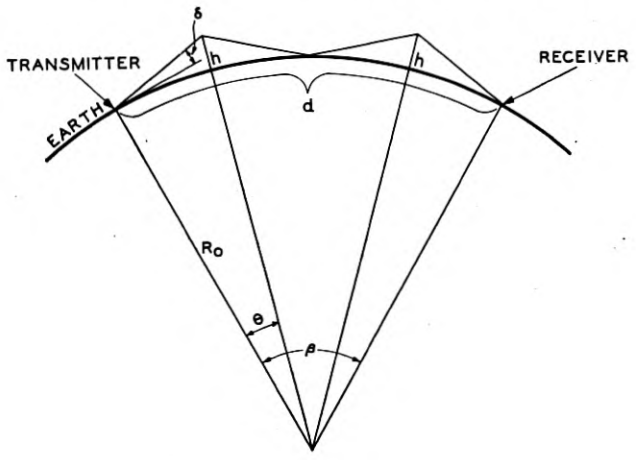


Fig. 28—Calibration curves of the Holmdel MUSA for 49.1 meters, giving the angle of the principal lobe as a function of phase advance ϕ_2 (Fig. 3). Note that the sense of the phase shift depends upon the beating oscillator frequency. The curves are calculated for a velocity ratio $v = 1/0.933$.

referred to the ground wave is expressed in terms of earth angles δ and n, the number of hops or ionosphere reflections. The height h and angle δ are also related through n as shown in Fig. 29. Using the first relation, the curves of Fig. 30 were drawn; using the second relation, points corresponding to various heights were located on the curves. For the Holmdel-Halifax circuit d is 643 miles (1030 kilometers) making $\beta = 9° 21'$. Corresponding to each measured angle there is a delay (referred arbitrarily to the ground wave which, of course, was not received) and a layer height, for each of the modes or orders. Both angles together yield a delay difference which is to be compared with the measured value.



$$\text{DELAY} = \frac{2nR_0 \sin\theta}{c\cos(\delta+\theta)} - \frac{d}{c}\ ;\ 1+\frac{h}{R_0} = \frac{\cos\delta}{\cos(\delta+\theta)}\ ;\ 2n\theta = \beta$$

Fig. 29—Delay and angle relations for multiple reflection from a uniform reflecting surface. The number of ionosphere reflections is designated by n.

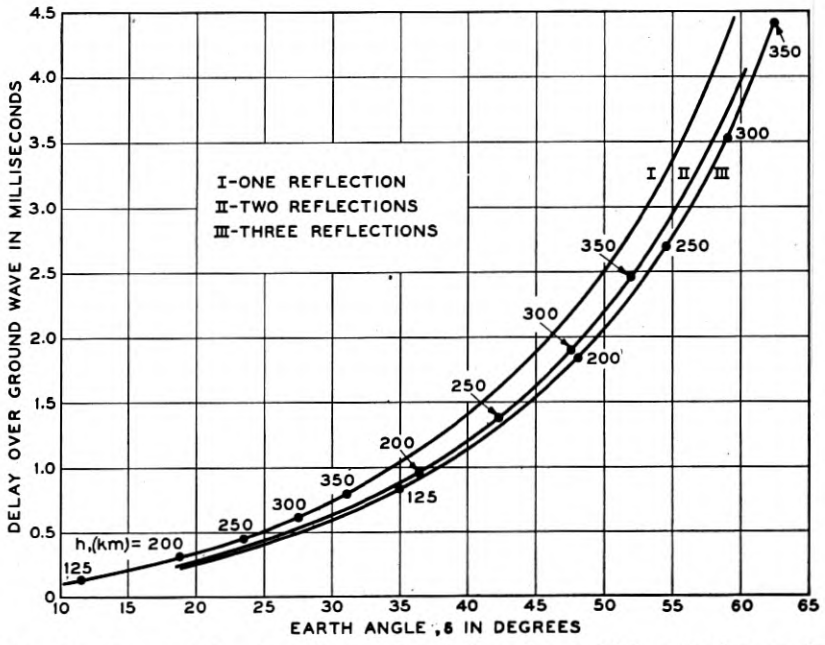


Fig. 30—Curves giving the delay-angle relations for multiple reflection on the Halifax-to-Holmdel path.

TABLE I

OBSERVATIONS ON VE9HX, HALIFAX, NOVA SCOTIA
6110 kilocycles 49.1 meters

	E.S.T.	Date	$\delta_1°$	$\delta_2°$	Relative Delay milliseconds		Virtual Height kilometers			Estimated		Field decibels above 1μv/m.
					meas.	calc.	1st	2nd	3rd	1st	2nd	
		1935										
	P.M. 5:00	11–25	18.2	38	1.01	0.77	195	215				
	P.M. 4:45	11–26	24.5	42	1.01	0.85	260	250				
	P.M. 4:40	12–17	26.4	44	0.95	0.97	290	270				27
	P.M. 4:41	12–17	24.5	43	0.95	0.95	265	255				
	P.M. 5:01	12–17	24.5	42	0.95	0.85	265	250				
	P.M. 4:20	12–18	25.5	44	0.95	1.05	275	270				27
	P.M. 4:25	12–18	24.0	42.5	0.95	0.90	250	250				
A	P.M. 4:40	12–18	23.0	42	—	0.90	245	250				
	P.M. 4:30	12–23	24.5	44	0.95	1.05	265	270				25
	A.M. 10:29	12–26	28.0	44	0.95	0.90	305	270				−14
	A.M. 10:46	12–26	28.0	44	0.81	0.90	305	270				
	A.M. 10:52	12–26	29.5	44	0.78	0.85	330	270				
	P.M. 4:33	12–26	24.5	44	0.88	1.05	265	270		242	245	18
	P.M. 4:49	12–26	24.5	42	0.95	0.85	265	250				
	P.M. 4:58	12–26	24.0	42.5	0.88	0.90	250	250				
	A.M. 10:07	12–27	25.5	43	0.95	0.95	275	255				−14
	A.M. 10:57	12–27	29.5	42	0.68	0.65	330	250		130	243	
	A.M. 11:03	12–27	31.0	42	0.88	0.70		250	100			
	A.M. 11:20	12–27	<8.0	45.5	1.28	1.6+	<120	280				
B						1.5+	<120		185			
	A.M. 11:48	12–27	8.0	47.5	1.28	1.8+	<120	300				
						1.7	<120		195			
	A.M. 11:56	12–27	8.0	45.5	1.28	1.6+	<120	280		130	247	
						1.5+	<120		185			
C	P.M. 4:32	12–27	25.5	44	0.88	1.05	275	270				14
	P.M. 4:59	12–27	27.0	43	0.88	0.85	295	255				
D	A.M. 10:45	12–31	31.0	42	0.50	0.55	350	250				0
	A.M. 11:45	12–31	8.0	35.5	0.71	0.8+	<120	195				
		1936										
	A.M. 10:30	1–2	24.5	42	0.88	0.85	265	250				− 2
	P.M. 6:05	1–14	20.5	42	1.01	1.00	215	250				8
	P.M. 6:35	1–14	18.4	38	1.01	0.77	200	215				
E	P.M. 6:15	1–15	23.0	42	1.08	0.90	245	250				22
	P.M. 6:20	1–15	24.0	42	1.11	0.85	250	250				
	P.M. 6:40	1–16	25.5	42	1.08	0.85	275	250				14
	P.M. 7:21	1–16	24.5	43	1.18	0.95	265	255				
F	P.M. 8:39	1–16	31.5	37	0.27	0.20	355	205				
	P.M. 9:35	1–16	26.4	37	0.47	0.42	290	205				27
	P.M. 5:50	1–21	24.5	42	0.95	0.85	265	250		267	247	22
G	P.M. 6:10	1–21	26.4	44	1.01	0.97	290	270				
	P.M. 6:16	1–21	22.0	40	0.95	0.80	235	230		232	245	
	A.M. 10:40	1–22	24.5	43	0.95	0.95	265	255				− 2
	A.M. 11:05	1–22	24.5	34	0.41	0.35	265	185				
						0.30	265		120			
H	A.M. 11:09	1–22	34.0	43	0.60	0.60	185	255				
						0.65		255	120			
	A.M. 11:30	1–22	24.5	43	0.95	0.95	265	255				
	A.M. 11:35	1–22	24.5	34	0.41	0.35	265	185				
						0.30	265		120			
I	P.M. 6:45	1–24	18.4	38	0.74	0.77	200	215				2

In Table I the virtual heights are deduced from the curves for the assumed hop orders. The calculated relative delay is the delay difference corresponding to these heights. All angles below 60 degrees were considered and all combinations of hop orders were considered for each angle, subject to the experimental knowledge of the sense of the delay. The values shown in the table are the ones which give the best agreement with the measured delay. In most instances there was no question concerning the interpretation; in a few doubtful cases two possibilities are presented (December 27 and January 22).

Examination of the table shows that except near noon, the propagation comprises the first and second reflections from the F region of the ionosphere. Groups A, C, E and G illustrate this. In the majority of instances the agreement is excellent; these cases constitute strong evidence that the MUSA performs correctly.

The discrepancies in the table between layer heights for the first and second hops and between measured and calculated delay are not entirely experimental error. Assuming errors in measured angles sufficient to make the delays agree will, in some cases, increase the discrepancy in heights. An interpretation one might make of this is that the ionosphere is not uniform over the circuit and the regular reflection basis of calculating is not strictly in accord with facts. However, there are other theoretical explanations for discrepancies in height. Under usual conditions, the second reflection height should be slightly greater than the first but for certain ionizations in the E region, the first F reflection may be retarded more than the second F reflection in passing through the E region. Thus the heights may differ in either direction without demanding horizontal non-uniformity. The discrepancies between measured and calculated delay may be explained by horizontal non-uniformity in the ionosphere. For an essentially non-dissipative atmosphere of ions having any vertical distribution but no horizontal gradient, and neglecting the earth's magnetic field, the group delay is identical with that calculated from triangular paths coinciding with the initial earth angles. Breit and Tuve showed this in their 1926 paper. With horizontal variations in the ionosphere such as tilting layers, no kind of agreement could be expected; the waves might even travel via other than great circle routes.

During three days of our observations W. M. Goodall made measurements of virtual height and of critical frequency which enabled him to predict the results we might be expected to observe. His estimates are shown in the next to the last column of the table.

The data for December 27 (B) are interesting in that after 11 o'clock the first F reflection apparently disappeared. Instead, a first reflection

from the E layer is indicated. This was predicted by Mr. Goodall on the basis that the E region ionization at noon became so great that 24-degree waves should be reflected. For completeness the table shows an alternative interpretation of a first E reflection and a third reflection from a 185- to 195-kilometer height. The first E reflection and second F reflection are perhaps more likely. The 11:03 record is not explained.

Something similar appeared to happen on December 31 (D). On January 22 (H) normal first and second F reflections occurred with angles of 24.5 and 43 degrees. In addition a third wave of 34 degrees appeared. Two interpretations of this are shown but neither seems very plausible.

As a general rule propagation from Halifax is simpler than from Daventry on the same wave-length. In particular GSL waves received by the two MUSA branches are definitely less discrete and include sufficient delay differences in themselves to prevent the nicety of equalization possible with VE9HX. If multiple reflection takes place, which we have no reason to doubt, it is generally so distorted by non-uniformity over the path or by other factors as to be unrecognizable. In view of the occasional complexity of the Halifax circuit, only one-sixth as long, this is perhaps to be expected.

The absence from these observations on Halifax of any third reflections from the F layer is likely due to the fact that they would fall in the neighborhood of the first null of the rhombic antenna and would have to be much stronger in space in order to appear comparable with the second or first. There have been momentary appearances of waves which might have been third reflections but they did not persist long enough to work with.

When single waves were present, which was not unusual in the later evening hours, the angle more often corresponded with the first F reflection rather than the second.

Additional Numerical Data on Reception with the MUSA

The data shown in Fig. 31 are submitted to supplement the rather meager numerical data on transatlantic reception thus far presented. Here, relative delays and angles taken from the MUSA operating log are shown in plots A, B, and C. Only the end points of the lines are significant; they denote by their abscissas the angles at which the two receiving branches were set. The ordinates of the upper end points denote the equalizing delay. The lines merely connect coexistent points. The data shown were selected from the rather extensive log to present a fair cross section of conditions, omitting, however, all cases

in which both branches were steered at the same wave bundle. They cover winter and summer and were obtained with frequencies appropriate to the time and season. Most of the observations were made on transmission from Daventry, the remainder on transmission from Rugby. In D are shown the results of pulse measurements made before the MUSA was in use. Here the angles were measured by the two antenna null method and the delays were observed directly on the oscilloscope time axis.[5] Although as many as five points, each denoting a wave bundle, are shown, generally not more than three were

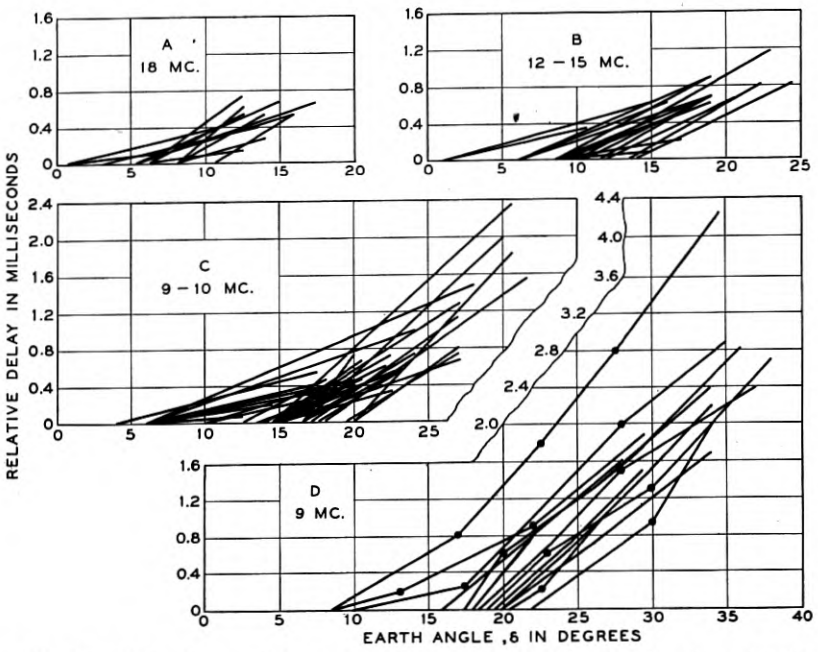


Fig. 31—Pairs of measured angles and relative delay denoted by the end points of the line segments. The data in *A*, *B*, and *C* were obtained with the MUSA; that of *D* was obtained by the use of pulses.

important at once. These measurements were made on transmission from Rugby.

It will be noticed that all four groups of data show that the relative delay per degree of angle difference is small at low angles and increases with the angle, roughly as the multiple reflection theory indicates. (This characteristic is distinctly favorable to the performance of the MUSA since its angle resolving power falls off at very low angles.) The scattering of the data indicates that an equalizing delay determined by the angle settings would not be successful; i.e., the delay

must be capable of adjustment to meet the various transmission conditions.

Quality Improvement with the MUSA

The distortion of speech and musical quality which characterizes short-wave circuits is due entirely to the interference of differently delayed waves each of which individually is fundamentally free from all kinds of distortion except non-selective fading. This conclusion is almost self-evident and is corroborated by the results of several years of pulse investigation [5] made in cooperation with the British Post Office.

A MUSA system can be expected to select one out of several multiple reflections. However, these reflections suffer more or less scattering with the result that they appear as bundles of waves of various degrees of compactness. These bundles possess a small spread of both angle and delay. The delay interval included in a bundle of waves is rarely less than 100 microseconds. Double refraction or "magnetoionic splitting" occurring in the ionosphere doubtless accounts for the existence of a small minimum delay. A delay interval of fifty microseconds or so may be detected even in the unusually compact bundles represented by the pulses of Fig. 21. Transmission from Halifax appears to include a delay interval of this order, also. With transatlantic propagation it is not uncommon to have a bundle containing numerous weaker components extending over several hundred microseconds. On rare occasions these have extended over two milliseconds, masking any multiple reflections which may have been present.

The quality associated with one MUSA branch which selects one out of several bundles of waves is thus not perfect. The effect of a delay interval of a few hundred microseconds is scarcely noticeable, however, except during deep carrier fades. Therefore, if diversity action between two branches steered at the low and high angle parts of the same bundle is employed, deep fades are avoided to a large extent, and the quality is almost perfect. When more than one wave bundle is present diversity action between branches steered at the principal bundles accomplishes this escape from deep fades. It is desirable to utilize all of the principal bundles in diversity in order to preserve the discrimination of the MUSA. For, one of the bundles, if not provided with a branch to receive it, would cross talk into the other branches when it momentarily became strong and those provided with branch receivers became weak. Signal-to-noise ratio considerations discussed in Section V constitute an equally important (and related) reason for utilizing all principal bundles.

As distinguished from selective fading, which is greatly reduced by the rejection of all but one wave bundle, general fading is by no means eliminated. The reader may expect, however, that when the MUSA selects one wave bundle from several it restricts the waves accepted to those which have traveled more nearly a common path, and for a given degree of turbulence in the ionosphere, the fading should be slower, since only relative changes among the several waves result in interference fading. Such a tendency no doubt exists and has been noticed occasionally in the operation of the MUSA but rarely has there been a marked effect (excepting certain cases of flutter fading to be described later). This will be understood when it is recalled that even a fifty-microsecond delay interval means that a difference of 500 wave-lengths is involved for a wave-length of thirty meters. In order that the fading rate be sharply reduced it is required that the ionosphere shall preserve this difference, to within a half wave-length, more effectively than it does if larger differences are involved. Since a half wave-length is only 0.1 per cent of 500 wave-lengths a rather high degree of balance is thus required. Evidently, the turbulence of the ionosphere usually prevents such a balance.

Using broadcast signals (double side band) from Daventry a thousand or more comparisons were made of the MUSA versus a single antenna and receiver, using the switching arrangement mentioned in Section III. Remarkable improvements were sometimes observed and some improvement was almost always noted. The exceptions were the instances when distortion was not detectable using one antenna, and the rare occasions when particularly violent flutter fading occurred.

Space diversity reception using two antennas showed a substantial improvement, usually, but failed ever to show the order of improvement demonstrated by the two-branch MUSA when two or more wave bundles of comparable amplitude occurred. Figures 23 and 25 suggest, by the way in which the audio outputs are seen to combine, that the distortion with MUSA reception is slight compared to that with diversity reception.

The increased naturalness which results from reducing the distortion is, of course, pleasing to the ear and has some value in telephone circuits on account of the subscribers' satisfaction. In addition, it increases the intelligibility particularly when considerable noise is present. It is impossible to evaluate the increased intelligibility definitely but, in certain cases at least, it permits the signal-to-noise ratio to be two or three decibels lower. From the point of view of picking up short-wave broadcasts for rebroadcasting, a more substantial value can be attached to the MUSA quality improvement.

To a considerable extent, the magnitude of the quality improvement ascribed to the MUSA in the preceding paragraphs depends upon the fact that double side-band signals were employed. For, with double side-band signals the selective fading caused by the interference of the differently delayed waves results not only in selective fading of the audio output, but also produces non-linear distortion when the carrier fades selectively. This non-linear distortion sounds much like over-modulation, and when it occurs in its more violent forms it completely ruins the quality and intelligibility. With single side-band transmission it is possible to demodulate with such a strong carrier that non-linear distortion is virtually eliminated. The fading of the audio output is sometimes more selective than with double side-band but the resulting quality is substantially better.

Single side-band transatlantic signals were not available during the trial of the MUSA system. However, as mentioned in Section III, receiving equipment was available which rejects one side band and reduces the percentage modulation by a factor of ten or more. It was found that this equipment, applied to the one-antenna system, resulted in substantially reduced non-linear distortion and that the quality could be still further improved by the reduction of selective fading afforded by the MUSA. With MUSA reception there was apparently no quality improvement in going from double to single side band.

Summarizing Discussion

In this section the general performance of the experimental MUSA has been described in a necessarily qualitative manner. Motion picture oscillograms were shown to illustrate the performance under fairly typical transatlantic conditions. An investigation of propagation from Halifax in which the MUSA was employed to identify ionosphere reflections was included to supplement the rather fragmentary evidence available in motion picture oscillograms. The improved quality obtained with MUSA reception was discussed from several points of view. The evaluation of the MUSA has been general; it serves partly to introduce the following section which deals specifically with the signal-to-noise ratio evaluation.

Before closing this section it is appropriate to discuss conditions with which the experimental MUSA could not adequately cope.

On numerous occasions the fact that only two branches are provided has definitely handicapped the performance. More often, however, the need for greater angular discrimination or resolving power has been apparent. Except on infrequent occasions a MUSA two to three times the length of the experimental one and equipped with three

branch receivers could be expected to perform as well as the experimental one now performs at its best. The occasions when it might not are the infrequent times when violent flutter fading occurs.

At least one type of flutter fading appears to be associated with a pronounced scattering which results in a kind of shower of erratic waves arriving over a wide range of directions. Receiving antenna directivity has been found definitely helpful in all except the most violent cases. Apparently when improvements due to directivity occur they occur principally by selecting a more or less normally propagated wave bundle and rejecting the shower of erratic scattered waves. When, in the most violent cases, no reduction of the flutter can be achieved the reason may be that the unit antenna accepts too wide a horizontal range to permit the MUSA to discriminate sufficiently against the shower. (It will be remembered that the MUSA array factor is of the form of a semiconical shell and thus the MUSA will, in general, accept as wide a horizontal range as the unit antenna permits.)

V. The Signal-to-Noise Improvement of the MUSA Receiving System

Because of the complicated nature of short-wave transmission and also because of the uncertain state of noise measuring technique, it is not a simple matter to give a satisfactory answer to the question: "What is the signal-to-noise improvement of a MUSA system?" In this section an attempt has been made to simplify the problem by separating the various factors involved. The section begins with an analysis of the problem assuming simple types of wave transmission. This is followed by experimental studies and discussions.

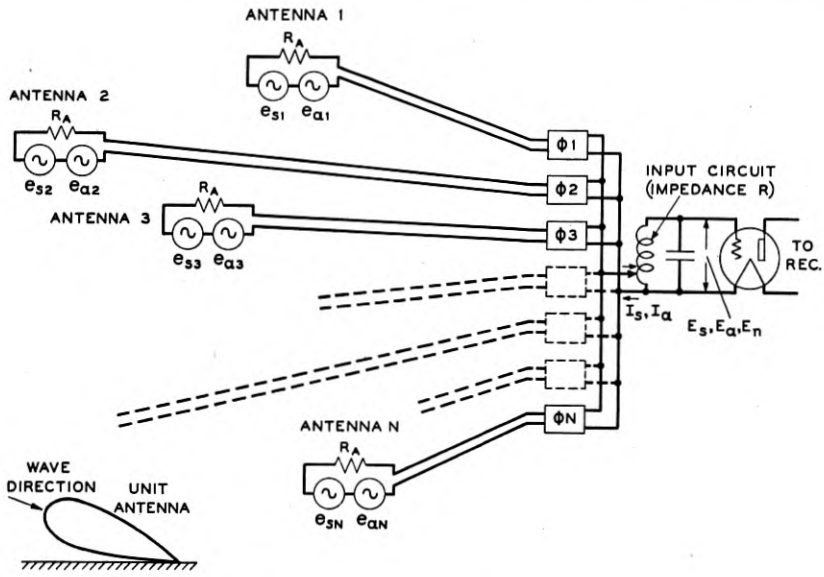
In discussing the signal-to-noise advantage of a MUSA it is understood that a reference receiving system must be adopted, and for this purpose one of the unit antennas connected to an automatic gain controlled receiver was chosen. Other types of antennas as, for instance, a simple vertical or horizontal doublet might have been used but other factors not significant to the MUSA would then have been involved.

Simple Analysis of the Signal-to-Noise Ratio Improvement

The MUSA differs from other directional antennas in that it is an array of antennas between which there is negligible electromagnetic coupling. This allows (but does not require) a different point of view, not explicitly involving directivity, in considering the signal-to-noise advantage of the array. The following analysis is made from this point of view. In Figs. 32 to 34 antennas are represented by signal

generators, e_s, static generators, e_a, and resistances R_A. The input circuits of the receivers are matched to the antennas.

In Fig. 32, N spaced antennas are shown connected in parallel. The root-mean-square noise voltage, E_n, at the input to the receivers represents the thermal noise originating in the receiver input circuits.



FIRST ASSUMPTION:
ALL LINES MATCHED
(LINE IMPEDANCES = R_A
PHASE SHIFTER IMP. = R_A
INPUT CIRCUIT MATCHED
TO IMP. R_A/N)

$$I_s = \sum_{K=1}^{K=N} \frac{e_{sK}}{2R_A}, \quad E_s = I_s\sqrt{\frac{RR_A}{N}},$$

$$I_a = \sum_{K=1}^{K=N} \frac{e_{aK}}{2R_A}, \quad E_a = I_a\sqrt{\frac{RR_A}{N}},$$

$$E_n = \text{CONST} \times \sqrt{R}$$

SECOND ASSUMPTION:
SINGLE WAVE. SIMILAR ANT.
($e_{s1} = e_{s2} \cdots e_{sN} = e_s$
$e_{a1} = e_{a2} \cdots e_{aN} = e_a$)
SIGNAL CURR. PHASED
STATIC CURR. RANDOM

$$I_s = \frac{N}{2R_a} e_s, \quad E_s = \frac{1}{2} e_s \sqrt{N}\sqrt{\frac{R}{R_a}}$$

$$I_a = \frac{\sqrt{N}}{2R_a} e_a, \quad E_a = \frac{1}{2} e_s \sqrt{\frac{R}{R_A}}$$

$$E_n = \text{CONST.} \times \sqrt{R}$$

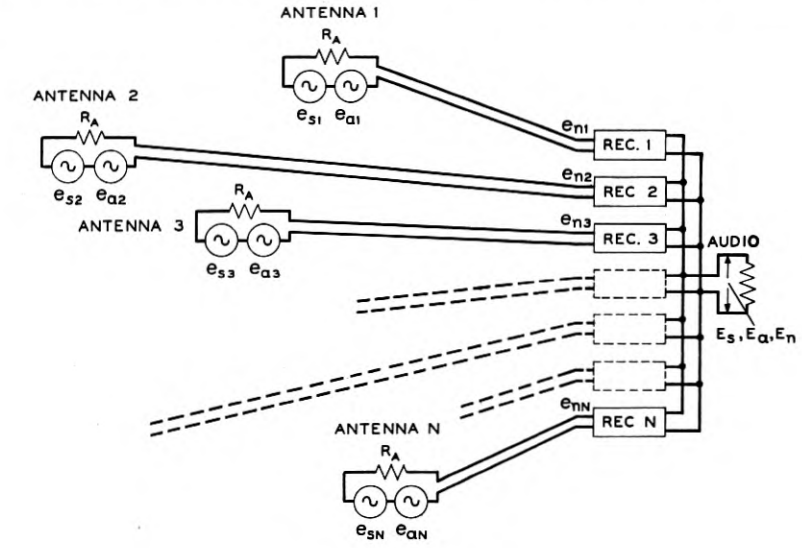
$$\frac{E_s}{E_a} = \sqrt{N}\frac{e_s}{e_a}$$

$$\frac{E_s}{E_n} = \sqrt{N}\frac{e_s}{\sqrt{R_A}} \text{ CONST.}$$

Fig. 32—Simple signal-to-noise analysis of a system of N spaced antennas. Signal currents are phased and combined at the incoming frequency. The summation signs include addition on the power basis.

For the matched condition this noise is constant and independent of the number of antennas. A single wave is assumed and the signal outputs of the antennas are phased by means of the phase shifters ϕ. The maximum signal power obtainable from N antennas obviously is N times that obtainable from one antenna. In terms of receiver noise,

e_n, the improvement in signal-to-noise ratio is 10 log N decibels referred to one antenna. If, instead of receiver noise, static is the predominating noise, the signal power received is not significant but the same improvement is realized for the general case in which the static is distributed randomly among the N antennas.[22] In that case the N signals are phased to add on a current basis while the N noise sources



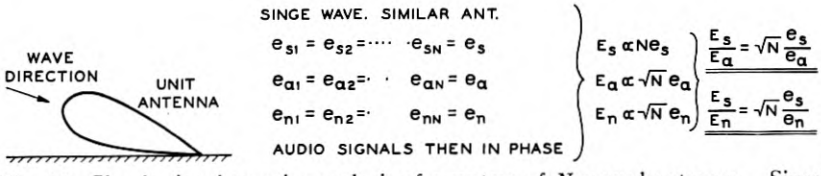


Fig. 33—Simple signal-to-noise analysis of a system of N spaced antennas. Signal currents are combined at audio frequency.

add on a power basis. Analogous arguments apply to a series connection of N antennas and result in the same improvement of 10 log N decibels.

[22] If static comes from all directions simultaneously, its distribution is random among the ideal unit antennas discussed in Section II. This is deduced from calculations which show that gain (signal-to-noise ratio) is proportional to the length of the system; i.e., to the number of unit antennas. The assumption of randomness requires that the spacing of unit antennas having a certain angular discrimination must be equal to or greater than the antenna length required to produce that discrimination in the simple linear end-on type of unit antenna.

That static is, on the average, distributed randomly among the rhombic antennas of the experimental MUSA is shown by measurements described later in this section.

The system described above has been shown mainly to introduce the system shown in Fig. 33. This diagram shows the audio addition of the outputs of N receivers fed by N antennas. Note that this system has no high-frequency phase shifters in the transmission lines. It is in fact similar to the diversity receiving system described by H. H. Beverage and H. O. Peterson.[23] For a *single wave* this is seen to be equivalent to the phased addition at carrier frequency shown in Fig. 32.

The signal-to-noise improvements shown on Figs. 32 and 33 were easily calculated because a single non-fading wave was assumed. In actual practice several fading waves are involved and it is then difficult, if not impossible, to make significant calculations. Later in this section, however, some of the general features of the system shown in Fig. 33 will be discussed from the point of view of several waves.

The MUSA system is characterized by the ability to separate waves and it is therefore possible to analyze it in a simple manner for cases of more than one wave. The arrangement in Fig. 34 corresponds to the Holmdel MUSA. The signals from the equally spaced antennas are here phased at the intermediate frequencies. Since random static and first circuit noise give identical results the analysis is given for static only.

As shown in Case I, if only one wave is present and both branches are phased for it the system functions as in Fig. 33 and it yields the same improvement of 10 log N decibels. If as shown in Case II the second branch is not phased for it (i.e., if the wave falls upon a minor lobe or a minimum of the MUSA directional pattern) less than the full improvement occurs. On the basis of linear audio detectors the reduction of improvement is 20 log x where x lies between 2 and $\sqrt{2}$. This quantity refers to the manner in which the noise from sources 1, 2, $\cdots$ N in Branch A adds with the noise *from the same sources* after having been phased differently and perhaps delayed differently in Branch B.[24] This involves the audio-frequency band width and method of noise measurement. As will be shown later x is usually not much different from $\sqrt{2}$. Taking $x = \sqrt{2}$ the loss in Case II is three decibels. If an audio detector is used which does not demodulate noise when the signal is absent (a square-law detector accomplishes this for practical purposes) this loss disappears, and branches may be phased for temporarily non-existent waves without incurring a penalty. Case III is the important one. It assumes two equal waves. Branch A is

[23] "Diversity Receiving System of RCA Communications, Inc.," *Proc. I. R. E.*, vol. 19, pp. 531–561, April, 1931.

[24] The case of $x = 2$ (in-phase addition) arises only when the phasing and delay of the two branches are alike.

phased for one; Branch B for the other. Again taking $x = \sqrt{2}$, the improvement referred to e_s/e_a is $10 \log N + 3$ decibels. Here e_s/e_a denotes the signal-to-noise ratio in each antenna due to one wave. Referring the improvement to the signal-to-noise ratio of one antenna

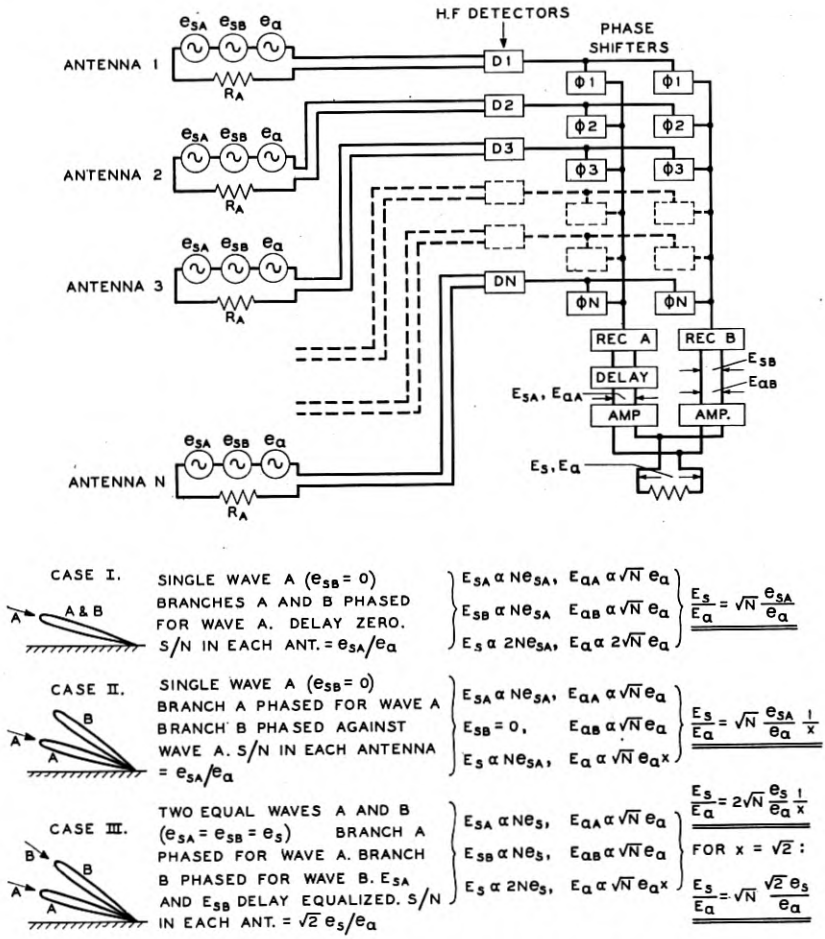


Fig. 34—Simple signal-to-noise analysis of the MUSA system.

receiving both waves, assumed to add randomly, increases the reference by three decibels and reduces the improvement to 10 log N decibels.

This analysis may be expanded to include K waves in which case K *branches* would be required to obtain the gain of *10 log N decibels* referred to one unit antenna.

It has been tacitly assumed in the foregoing analysis of Fig. 34 that the audio outputs of the several branches are delay equalized to add and that there is no diversity action (all of the waves are assumed to remain equal). The influence of fading is difficult to predict and will be discussed later in connection with experimental results.

Some readers, not concerned with details, may omit reading the following subsections and find it sufficient to read only the Summarizing Discussion of this section.

Test Method

From a practical point of view the best way of testing a MUSA system would seem to be to operate it on transatlantic telephone signals and compare its output with that of the reference system. Speech volume and noise could then be measured in the conventional manner. So far as the signal-to-noise improvement is concerned it would be a laborious and lengthy task to get satisfactory data because so often, during the test period,[25] static and receiver noise is masked by transmitted noise, interfering signals, and other man-made noise. To test the experimental MUSA, therefore, a different method was selected which gave significant data in a shorter time.

Since the success of the MUSA is related so fundamentally to the nature of the arriving signal the important thing to be determined by the measurements is how well the MUSA is able to cope with the various conditions of wave arrival. For instance, in the case of a single bundle of arriving waves how close does the actual signal-to-noise improvement come to the 10 log N decibel calculated for Case I (Fig. 34) in which a single non-fading wave was assumed? Likewise, for the case of two-wave bundles do the calculations of Case III agree with measurements?

For these purposes the signal-to-noise measurements would have to be free from directional static, interference, and transmitted noise; otherwise the measured improvements would be distorted. To insure uniform and desirable noise conditions it was decided to use thermal noise originating in the receiver input circuits [26] instead of whatever noise might be present on the radio channel. This was accomplished by inserting resistance pads in the antenna transmission lines to reduce the signal (and external noise) to a level where thermal noise greatly exceeded other noise. Signal-to-noise ratios in the range between fifteen and forty decibels were obtained in this manner, free of interference and directional static, and of transmitted noise.

[25] Transmission conditions during 1935 were comparatively undisturbed.

[26] A portion of the noise originates in the plate circuit of the first detector. For the present purposes this is equivalent to first circuit noise.

Substituting thermal noise for external static may at first seem far-fetched. Except for the fact that static is sometimes sufficiently directional to be received with different intensity as the MUSA is steered differently, the substitution is sound. In general, the static output does not vary with steering, as the measurements described later indicate but to avoid the distortion of results which would occur when this is not so, it was desired specifically to substitute non-directional noise. Studies of the characteristics of static and thermal noise have shown that both are alike so far as the effect of band width upon average and effective values is concerned, and have indicated that both consist of extremely short, randomly distributed pulses which overlap when received and detected by receivers of ordinary band widths. In a given band width, the envelope of the currents produced by static sources is highly irregular in comparison with that produced by thermal agitation. It appears, however, that the *character* of either envelope is not sensibly affected by the number of antennas combined nor by the manner in which the branch outputs are combined, so that both give the same improvement figures using any arbitrary noise measuring method.

There were several possibilities with respect to the signal to be employed in these tests. A single tone, a large number of tones distributed throughout the audio band and other special signals were considered. A simple method requiring no modulation was finally adopted. It consisted in alternately connecting the output of the antenna to be tested and that of the reference antenna to the same receiver. Assuming that the automatic gain control of this receiver would maintain a constant audio output level the signal-to-noise advantage is the ratio of the noise levels. The automatic gain control of the MUSA receiver did not, of course, hold the output level absolutely constant but a correction was easily made for the small variations in level.

The circuits of the measuring equipment are shown in Fig. 35. The rectified carrier appearing in the linear speech rectifier is taken to be proportional to signal and is measured simultaneously with the noise demodulated in the rectifier. When the keys are thrown to position 2 (by a gang arrangement) the signal meter shows the sum of the two rectified carriers and the noise meter reads the combined noise in the output of the diversity mixing amplifiers. Using the sum of the two rectifier currents to represent the signal implies that actual audio outputs from the two branches could be delay equalized to add arithmetically. As applied to a MUSA system this assumption is justified, in general. When the keys are switched to position 1, the rectified

carrier of branch B alone appears on the signal meter and noise from branch B appears alone on the noise meter. At the same time the diversity connection is broken and all except one of the six-phase shifter amplifiers in branch B are biased to cutoff; i.e., only one unit antenna is used. The pad "L" is adjusted to give the same audio gain from rectifier B to the noise meter for connection 1 as for connection 2.

By manipulating the keys which control the cutoff biases on the phase shifter tubes the "1" to "2" switchover may also be used to compare one antenna (one receiver) with two antennas in ordinary space diversity or one antenna with all six in a single branch.

The use of receiver noise as a noise source depends upon (1) having the noise equal in all six circuits and (2) upon having it originate ahead of the point where the gain is varied. In well-designed receivers the noise should approach the thermal noise limit of the first circuit. It was found possible to have the signal-to-noise ratio, for a given signal level, of all six high-frequency input circuits equal to within $\pm$ 0.5 decibel and within a few decibels of the thermal limit.

The first tests were made with a local oscillator supplying the signal. They really constituted tests of the measuring set up. All six input circuits were fed simultaneously through 80-ohm pads giving equiphase and equiamplitude signals on each detector grid. This corresponds to receiving a single steady wave, and one branch was "steered" as if to receive such a wave. When the multiple switch was manipulated as to compare one antenna with the steered branch the indicated signal-to-noise improvement was usually between seven and eight decibels, compared with the theoretical value of 7.8 decibels (10 log 6).

Such a local test using the switchover with all associated equipment was made before and after every transatlantic test. Corrections based upon 7.8 decibels were made to the data in cases where the local tests showed a slightly different improvement factor. In all of the work the gains of the phase shifters were adjusted to equalize the difference in line loss. The effect of this is, however, trivial.

In measuring on transatlantic waves with automatic gain control the noise variation, corresponding nearly to the reciprocal of fading, rendered visual noise readings too rough to be suitable. A Weston high-speed db meter (copper-oxide bridge type) having a calibrated range of 16 decibels was used as a noise meter. To this instrument was added a fluxmeter (Fig. 35) of low restoring torque which automatically averaged the variations of the meter pointer over the 15-second periods of observation used in these tests. The fact that the noise meter rectifier is linear means that the noise current is averaged arithmetically

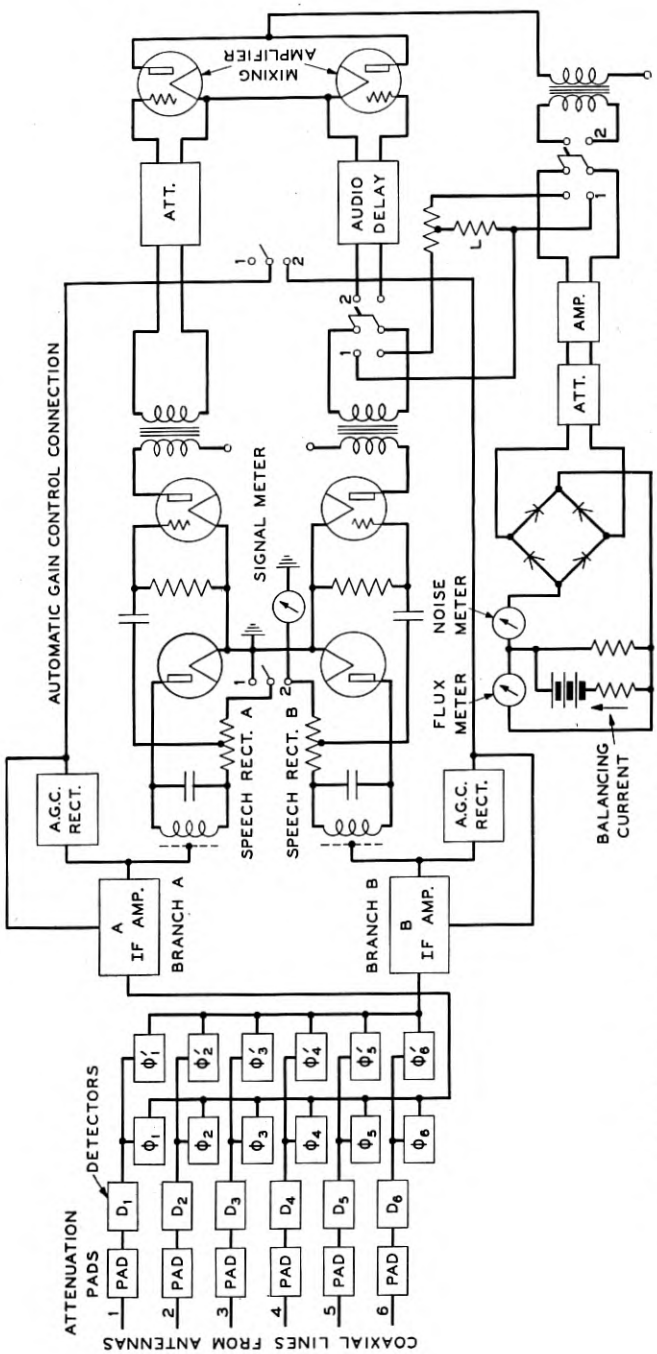


Fig. 35—Circuit used in measuring the signal-to-noise improvement of the experimental MUSA system.

along the time axis. (A discussion of other ways of averaging will be given later.) To permit the maximum time interval during which the restoring torque had negligible effect a balancing current was applied. A telechron motor marked time intervals of 15 seconds with a bell. The switchover between "1" and "2" (Fig. 35) was made and the fluxmeter restored to zero reading at the ring of the bell. The signal meter reading (calibrated in decibels) was maintained fairly steady with the automatic gain control, and could be averaged accurately by eye.

Test Results

The first measurements on transatlantic signals were made on 16 meters in September, 1935. Tests were made on the British Post Office Station GAU (18,620 kilocycles) between noon and 2 P.M., E.D.S.T., on September 16, 17, 18, and 19. At our request the transmitter operated on reduced power, presumably fifteen decibels less than normal. The propagation during these tests was characterized by low angles and slow fading. The angular discrimination of the MUSA for these low angles is so slight that it was decided to use only one branch and defer the question of diversity in these first tests.

Owing to the power reduction and to the fact that the period from September 15 to 18 inclusive was disturbed some of the data had to be obtained without antenna pads. Table II summarizes the results of the measurements.

When thermal noise is a contributing or predominating factor, the signal-to-noise ratio of the six-antenna branch must be compared with both No. 1 and No. 6 antennas in order that the line loss may be accounted for. With thermal noise predominating, the difference between the improvements referred to No. 1 and No. 6 should of course equal the line loss of No. 6. It may be shown that the arithmetic means of the two improvement ratios (voltage) should give very closely the improvement corresponding to random static (10 log 6 = 7.8 decibels). In Table II the arithmetic means of the improvement ratios are, therefore, called the equivalent improvement.

During these tests the indicated angle of arrival was from one to three degrees (such low angle determinations are not trustworthy within perhaps two degrees) and the receiving branch was set correspondingly and not altered during a test. The fading on No. 1 and No. 6 antennas was usually but not always unlike. Adjacent antennas always showed substantially synchronous fading.

A sample of the plots from which the figures in the table were obtained is shown in Fig. 36. This shows the noise readings reduced to

a constant signal meter reading for Test 13. Each horizontal line segment represents one 15-second fluxmeter period (the actual period was 14 seconds, one second being required for switching). The arithmetic averages of the segments are shown by dashed lines. It is to be emphasized that the scattering of improvements taken from adjacent readings is not experimental error but is due to the fact that pairs of adjacent readings were obtained at different stages of the fading cycle. Two separate systems (receivers and antennas and measuring equipment) permitting simultaneous measurements would not help this

TABLE II

ONE BRANCH

GAU 18,620 kilocycles

Date	Test No.	Pads in Ant. db	Reference Antenna	Number of Readings	S/N Improvement db	Group Average db
1935						
9–16	1	20	1	19	3.7	
9–17	6	20	1	18	3.0	
9–17	10	20	1	16	3.0	
						3.3
9–16	2	20	6	20	8.8	
9–17	7	20	6	20	11.0	
						10.0
9–16	4	0	1	17	6.2	
9–17	9	0	1	11	6.5	
9–19	11	0	1	18	6.0	
9–19	13	0	1	28	5.6	
9–19	15	0	1	20	8.0	
						6.4
9–16	5	0	6	30	6.4	
9–17	8	0	6	17	7.9	
9–19	12	0	6	20	5.0	
9–19	14	0	6	29	6.1	
						6.4

The apparent line loss is 6.7 db (4.3 calculated).
The equivalent improvement obtained from the data with pads is 7.3 db; without pads, 6.4 db. Average 6.9 or 7 db.

situation, since fading would not be synchronous on two systems. Evidently a considerable number of readings must be taken to insure that all stages of fading are equally represented in both of the systems being compared.

Returning to Table II the discrepancy between the apparent line loss and the calculated value (the loss could really be five decibels perhaps) suggests that insufficient data were obtained with the pads. There is a possibility that the measurements without pads involve directional static. However, taking the data as they stand yields an average of seven decibels which is less than one decibel below the value 10 log N

= 7.8 decibels calculated for the non-fading wave of Case I, Fig. 34. (The fact that only one branch was used in the measurements instead of the two branches of Case I does not affect the situation since the idle branch was disconnected and so contributed no signal and no noise.) A reduction of the order of one decibel from the calculated improvement could be expected since all of the waves of one bundle

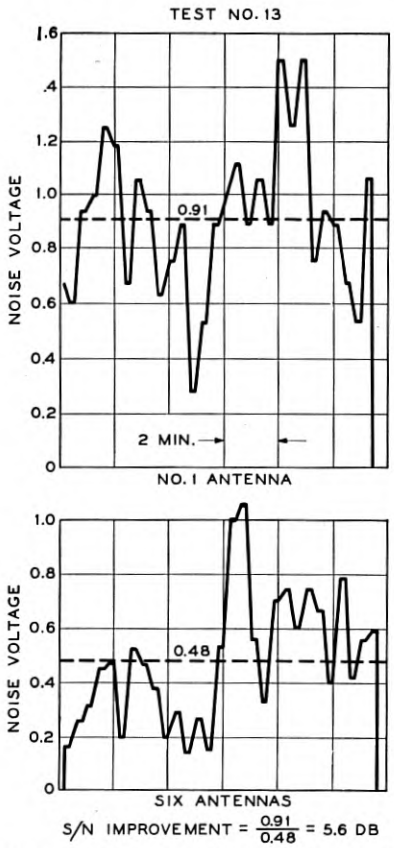


Fig. 36—Sample of measured noise plotted against time. GAU (18,620 kilocycles) Rugby, September 19, 1935.

cannot fall on the apex of the directional pattern. It was encouraging to find that no more loss occurred; i.e., to find that the waves in a single bundle may be phased so effectively.

Before leaving these tests, the results for September 18 should be mentioned. On this day the signal-to-noise ratio was so low, even without antenna pads, that measurements could not be made. The noise on this day was first taken to be thermal noise but was found

during the course of experimentation to be external noise [27] some ten decibels higher than thermal noise, as received on a single rhombus. At the end of the test the operator at Rugby keyed the transmitter with tone, advising us that the schedule was completed and wishing us "good night." With one antenna the signal was hopelessly lost in noise; with the six antennas the code was readable.

TABLE III

TWO BRANCHES

GBW 14,440 kilocycles

Date	Test No.	Pads in Ant. db	Reference Antenna	Number of Readings	S/N Improvement db	Group Average db	Number of Wave Bundles
1935							
10–2	21	40	1	20	2.7		1
10–2	23	40	1	16	5.3		1
10–2	24	40	1	16	4.4		1
10–8	29	40	1	20	6.3		1
10–10	37	40	1	20	4.4		1
						4.7	
10–2	20	40	6	19	6.9		1
10–2	22	40	6	20	8.4		1
10–2	25	40	6	16	9.0		1
10–8	28	40	6	17	10.9		1
10–9	30	40	6	38	8.0		
						8.5	
9–30	17	40	1	22	3.4		2
9–30	19	40	1	14	4.1		2
10–2	26	40	1	16	4.6		2
10–10	33	40	1	19	2.0		2
						3.5	
9–30	16	40	6	17	8.0		2
9–30	18	40	6	17	7.9		2
10–2	27	40	6	20	6.7		2
10–10	34	40	6	20	7.3		2
						7.4	

The apparent line loss is 3.8 and 3.9 db for the one-wave and two-wave groups, respectively. The calculated loss is 3.8 db.

The equivalent improvement for one-wave group is 6.8 db to which may be added the later determined correction of 1.2 db for the effect of delay, giving *8.0 db.*

The equivalent improvement for two-wave groups is 5.7 db to which may be added 0.8 db for the effect of delay, giving *6.5 db.*

More comprehensive measurements were made on GBW (14,440 kilocycles) and a few on GCW (9790 kilocycles), using the two branches. Since an unmodulated carrier was used, rectified carrier being taken to represent signal, there was no criterion for setting the audio delay. Accordingly, it was kept at zero and a correction introduced later. The results are shown in Tables III and IV.

[27] This noise, which was directive to the extent that four-decibel variation occurred with steering the MUSA, was doubtless a sample of the "star static." It was encountered also on 31 meters in October. See footnote (32).

TABLE IV

TWO BRANCHES

GCW 9790 kilocycles

Date	Test No.	Pads in Ant. db	Reference Antenna	Number of Readings	S/N Improvement db	Group Average db	Number of Wave Bundles
1935							
12–13 . . .	39	40	1	15	4.8		2
12–13 . . .	40	40	1	14	4.7		2
						4.7	
12–13 . . .	38	40	6	14	7.9		2
12–13 . . .	41	40	6	13	6.7		2
						7.3	

The apparent line loss is 2.6 db. The calculated loss is 3.1 db.
The equivalent improvement is 6.1 db to which may be added 0.9 db for the effect of delay, giving *7.0 db.*

The data in the tables are classified roughly according to whether two bundles or one bundle of waves was present. In the latter case the two branches were steered, one on each side of the bundle, a few degrees apart. During these tests slight adjustments in steering were made when indicated by the angle monitoring tube, as in normal operation of the system. The large amount of data taken with GBW makes the results in Table III particularly reliable. This is reflected in the close agreement between measured and calculated line loss. Before discussing the results further the effect of delay needs to be analyzed.

Correction Due to Delay

The effect upon the noise, of delaying one audio output, is shown in Fig. 37. The curves were obtained with the circuit shown in Fig. 35

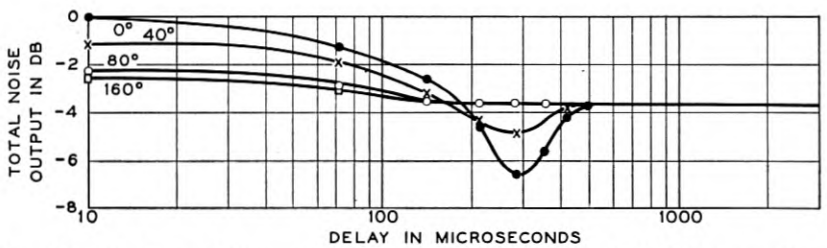


Fig. 37—The effect of delay and phasing upon noise output of the MUSA receiver.

using thermal noise. The equiphase, equiamplitude signal source mentioned previously was used to supply the inputs. Branch A was kept phased so that the six signals added, and branch B was varied. The

curves show the effect of delay upon noise in a 250- to 2750-cycle audio band, as measured with the Weston db meter, for various differences in steering. The curves are labeled in terms of the difference in the phase settings of the two branches. The 40-degree or 80-degree curves correspond in practice to steering on each side of a single bundle of waves. The 160-degree curve typifies steering at two separate wave bundles. The use of 100 microseconds (or more) delay is generally advantageous for the audio addition when steering at one bundle. Since this amount of delay makes the audio noise addition nearly random and for widely different steering the addition is also random the assumption that the noise from the two branches adds on a power basis, made in reference to Fig. 34 ($x = \sqrt{2}$), is justified.

The effect of delay is to produce an interference pattern in the audio noise spectrum. This accounts for the dip in the noise curve for a delay of 300 microseconds which locates the first interference minimum at about the center of the audio band. The asymptotic approach to 3.5-decibel reduction corresponds more nearly to a reduction ratio of $2/\pi$ than to $1/\sqrt{2}$ due to the fact that the Weston db meter is nearer linear than square law in response.

In obtaining these curves it was desired to simulate the reception of two waves for which the corresponding branches were phased to add. It was not convenient to set up locally such a two-wave case but the single wave input should give identically the same results provided phases were avoided which resulted in a signal at the second detector too low to demodulate the noise. A signal level so high that further increase did not affect the noise output was used for all points. The real purpose of the signal was to insure that the demodulated noise was not dependent upon the intermediate-frequency bands and that the results would be unaffected by possible differences in intermediate-frequency bands.[28]

As mentioned, noise has been measured with an unweighted 250- to 2750-cycle frequency band. Had a weighting network [29] which emphasizes frequencies in the vicinity of 1000 cycles been used the dips in the curves marked 0° and 40° would have been deeper and would have occurred in the vicinity of 500 microseconds delay.

Returning now to Tables III and IV the measured improvements were corrected to correspond to the effect of the delay which would probably have been used to obtain the best audio addition for the signal. The 1.2-decibel correction for the one-bundle case represents

[28] This precaution was subsequently found to be unnecessary; i.e., similar results were obtained with no input signal.

[29] Barstow, Blye, and Kent, "Measurement of Telephone Noise and Power Wave Shape," *Elec. Engg.*, vol. 54, pp. 1307–1315, December, 1935. Technical Digest published in *Bell Sys. Tech. Jour.*, January, 1936.

the reduction of noise obtained with 60-degree phase difference with the addition of 100 or 150 microseconds delay. The 0.8- or 0.9-decibel correction for the two-bundle case corresponds to any large delay and a large phase difference, say 160 degrees. These corrections would not have been much different if a weighting network had been used.

The measured difference of 1.5 decibels, shown in Table III, between the one-wave bundle and the two-wave bundle measurements is probably real and due to the fact that when the branches are steered at two separated bundles some loss is incurred when one wave disappears for a few minutes. This loss could have been at least partly recovered by using square-law detectors. Tests showing the advantage of square-law detectors over linear detectors are described later in this section. Employing square-law detectors would justify a correction of about one decibel to be added to the two-bundle improvement measurements in Tables III and IV. Applying this correction we summarize the results in Table V.

TABLE V

SUMMARY OF SIGNAL-TO-NOISE MEASUREMENTS

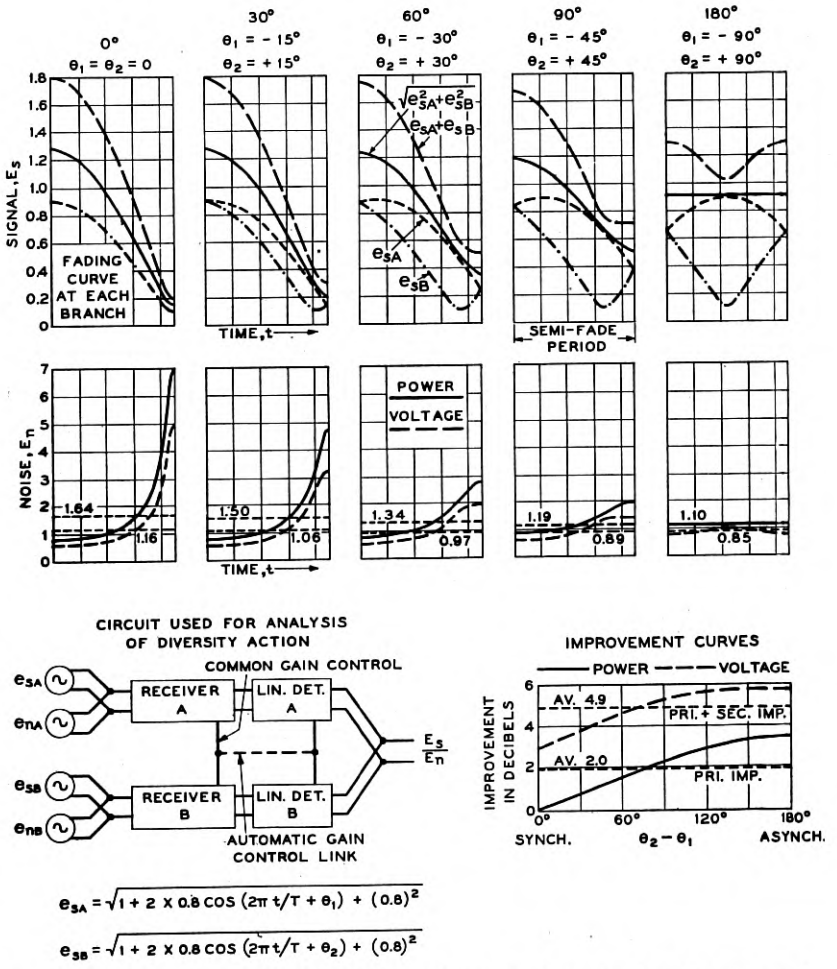
One Bundle of Waves		*Two Bundles of Waves*
One Branch only	*Two Branches*	*Two Branches*
7 db (Table II)	8 db (Table III)	7.5–8 db (Tables III and IV)

These improvement figures for two branches as they stand are approximately equal to 10 log 6 = 7.8 decibels as calculated for non-fading waves, and leave nothing to be ascribed to diversity action. Since a loss of perhaps one decibel occurs in the case of one branch (Table II), the recovery of that one decibel with two branches is to be ascribed to diversity action. Originally, considerably more was expected of the angle diversity. It appears however from theoretical and experimental evidence that one decibel is about what should be expected for the case in hand. It seems appropriate to include this study of diversity here.

Diversity Action

The first attempt to analyze diversity action was made with a graphical approach to the problem. On Fig. 38 is shown a schematic diagram of the system to be analyzed. Two receivers A and B with linear audio detectors may be regarded as fed from two angle branches of a MUSA. The noise generators e_{nA} and e_{nB} are assumed to be of

equal power but of random phase. The signal generators e_{sA} and e_{sB} are assumed to fade according to the equations shown beneath the diagram. They represent the carrier amplitude but since fading is assumed to be essentially non-selective in each branch they also repre-



$$e_{sA} = \sqrt{1 + 2 \times 0.8 \cos(2\pi t/T + \theta_1) + (0.8)^2}$$

$$e_{sB} = \sqrt{1 + 2 \times 0.8 \cos(2\pi t/T + \theta_2) + (0.8)^2}$$

Fig. 38—Graphical analysis of diversity action as it relates to signal-to-noise ratio.

sent the side bands. This type of fading might result from interference between two waves of small relative delay whose amplitude ratio is 0.8, such a pair being received by each branch. The automatic gain control is assumed to be perfect; i.e., the audio output, E_S, is maintained constant.

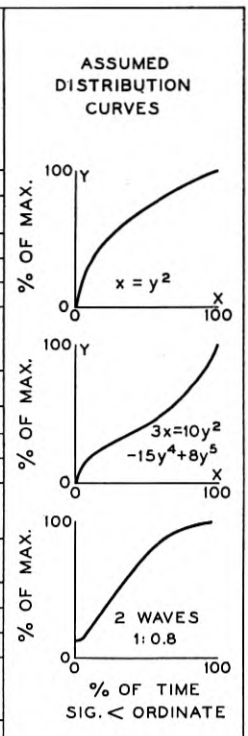
By definition, diversity in fading occurs when the fading of the two branches is not synchronous. If all degrees of asynchronism are equally probable the diversity is random. This is the case considered. In Fig. 38 five stages in the cycle of variation from synchronous to asynchronous fading are used for calculation. In each of these, fading curves corresponding to the two assumed waves are shown displaced from each other by 0, 30, 60, 90, and 180 degrees. With "ideal" automatic gain control, the two receiver gains, always equal, will be proportional to the reciprocal of the resultant of e_{SA} and e_{SB}. For "ideal" linear detectors the noise output of the receivers will be proportional only to the gain. The noise curves plot the noise variation on this basis. Two cases of signal addition are considered—voltage and power addition. The corresponding noise curves differ only as the reciprocals of the resultant signal curves differ. These noise curves are averaged (with a planimeter) and the resulting average signal-to-noise ratios are used to plot the improvement curves shown in the figure. The improvement curve for power addition of signal is located on the improvement axis so that zero improvement is shown for synchronous fading. The curve for voltage addition of signal is located three decibels higher at synchronous fading. These curves are again averaged over the cycle from synchronism to asynchronism (by averaging noise voltage). The improvements are 2.0 decibels and 4.9 decibels.

Power addition of the two signals corresponds in practice to the case in which the delay is unequalized and sufficient to cause the audio outputs of each branch to combine on a power basis like noise. The two-decibel improvement might appropriately be called the primary improvement since it is due solely to the diversified fading. The additional improvement of 2.9 decibels found with voltage addition of the signals is due to favorable discrimination in the addition of signal and of noise and might be called the secondary improvement. The secondary improvement occurs in reception with the MUSA; it has already been included in the 10 log N decibel improvement calculation.

In practice, it would be undesirable to use the "ideal" automatic gain control assumed in the above analysis; the action must be smoothed out with, for instance, a capacitance-resistance network. The effect of this is to reduce the primary gain since the noise peaks, whose avoidance by diversity action results in the primary improvement, are reduced. An analysis of diversity action without automatic gain control was made. In this case the signal was averaged while the noise remained constant. The results are included in the table shown in Fig. 39 which is introduced later.

This treatment of diversity action has been made from the point of view of MUSA reception but is applicable to ordinary space diversity using two stations or antennas. In the case of a single bundle of waves no modification of the analysis need be made; the signal generators then represent the spaced antenna outputs which are fading randomly. In this case voltage addition of the audio outputs may be

NO. OF STATIONS OR BRANCHES	$E_s = \sqrt{e_{sA}^2 + \cdots e_{sK}^2}$ $E_n = \sqrt{e_{nA}^2 + \cdots e_{nK}^2}$ PRIMARY IMPROVEMENT		$E_s = e_{sA} + \cdots e_{sK}$ $E_n = \sqrt{e_{nA}^2 + \cdots e_{nK}^2}$ PRI. + SEC. IMPROVEMENT	
K	NO AGC	AGC*	NO AGC	AGC
QUADRATIC DISTRIBUTION				
1	2.5	0	2.5	0
2	2.8	2.2	5.5	4.8
3	2.9	2.5	7.3	6.8
FOUR WAVE DISTRIBUTION				
1	3.4	0	3.4	0
2	—	—	6.4	5.2
3	—	—	8.2	7.5
TWO WAVE NEAR-SINUSOIDAL DISTRIBUTION 19 DECIBEL FADE				
1	2.8	0	2.8	0
2	3.2	2 0	5.8	4.9
3	—	—	—	—
AVERAGED RESULTS				
1	3	0	3	0
2	3	2	6	5
3	3	2.5	8	7



* AUTOMATIC GAIN CONTROL

Fig. 39—Summary of results of diversity analysis.

expected to occur since a single bundle will typically include only a small delay interval. In the general case of two or more wave bundles the signal generators must be interpreted to represent not the carrier but the side-band average, for fading will then be essentially selective and the audio output will not be proportional to carrier as was assumed in the analysis. If this interpretation is made the signal-to-noise ratio in each receiver becomes proportional to the generator amplitudes e_{sA}

and e_{sB}, and the analysis of Fig. 38 is applicable. In this case voltage addition does not occur since the audio outputs are essentially different owing to the selective fading.[30] They add, in general, to a value intermediate between the power and voltage sum, although for the more complicated conditions they combine on a power basis.

The above analysis has been based upon simple two-wave interference and the results might not be applicable to the more complicated and changing conditions of actual transmission. Accordingly, R. L. Dietzold has made a statistical analysis for other types of fading and for three stations as well as for two. The results appear in Fig. 39 together with the results of the above graphical analysis for two-wave interference fading. The time sequence of amplitude in the more complicated types of fading encountered in practice is not significant; the percentage distribution determines the results. The "four-wave" distribution curve corresponds to four equal waves of random phase. The quadratic distribution curve was deduced experimentally by R. S. Ohl. Except that these different distributions were assumed, the assumptions were the same as those of Fig. 38. The improvements are expressed in decibels referred to the signal-to-noise ratio for one station or branch with ideal automatic gain control.

The small effect upon the results of assuming different time distributions lends significance to these calculations. The averaged round numbers are probably about right.

With no automatic gain control (or with one which acts slowly compared with the fading) there is little or no primary gain. With infinitely fast and stiff gain control action there is a 2- and 2.5-decibel primary gain for two and three stations, respectively.

A few measurements were made at Holmdel on two-station diversity (antennas 1 and 6 of the MUSA). The thermal noise and rectified carrier technique was used. The results appear in Table VI.

The measuring technique was exactly the same as used in obtaining the data for Tables III and IV in which voltage addition of the signal was assumed. A 3-decibel improvement is therefore included in the 3.6-decibel figure. This leaves only 0.6 decibel (possibly one decibel or even 1.5 decibels since the measurements are too meager to be reliable to better than one decibel) for primary gain compared with a possible 2.0 decibels. We are inclined to use about one decibel for primary gain. The time constant on the automatic gain control was of the order of 0.06 second in this and all signal-to-noise comparisons. That this time constant was not fast enough to produce the high noise

[30] This refers to speech signals; in the case of telegraph signals the frequency band is so narrow that fading is always essentially non-selective, and voltage addition occurs.

peaks corresponding to the inverse of fading is shown by the transcribed motion picture record of the signal and noise meter variations, shown in Fig. 40. Note the signal fades.

A secondary improvement of 3 decibels is too high for two-station space diversity; i.e., the signals do not add on a voltage basis.[30]

TABLE VI

ANTENNAS 1 AND 6 IN DIVERSITY
GBW 14,440 kilocycles

Date	Test No.	Pads in Antennas db	Reference Antenna	No. of Readings	S/N Improvement db	Average
1935						
10–9 ...	32	40	1	19	1.7	
10–10 ...	35	40	1	20	2.2	2.0
10–9 ...	31	40	6	14	4.9	
10–10 ...	36	40	6	20	5.2	5.1

The apparent line loss is 3.1 db. The calculated loss is 3.8 db.
The equivalent improvement is 3.6 db.

Oscilloscope observations of the diversity combinations of the audio outputs of two spaced antennas (No. 1 and No. 6 of the Holmdel MUSA) indicate, however, that on the average the secondary improvement is appreciable and probably about 2 decibels for two antennas and 3 decibels for three antennas. This improvement depends upon

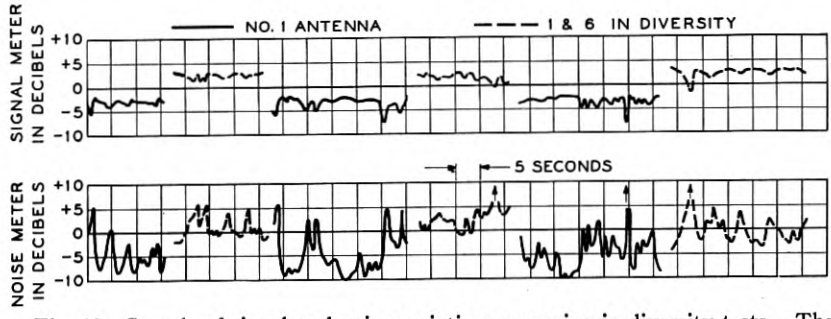


Fig. 40—Sample of signal and noise variations occurring in diversity tests. The arrows indicate noise levels beyond the scale of the meter. GBW (14,440 kilocycles) Rugby, October 10, 1935. 1920 G.M.T.

the number of wave bundles, their angular separation, their relative delays, the spacing of the antennas and the frequency band occupied by the signal. For a single compact wave bundle the secondary improvements will be nearly 3 decibels and 4.8 decibels for two- and

three-antenna systems, respectively, but for several bundles of large relative delay the secondary improvement may disappear.

The results of some recent tests of a three-antenna diversity system on trial at Netcong, N. J., carried out under the direction of F. A. Polkinghorn, showed a signal-to-noise improvement of 3 to 3.5 decibels. Assuming a 3-decibel secondary improvement there remains something of the order of 0.5 decibel for the primary improvement. This is plausible in view of the time constant of one second used on the automatic gain controls. Linear audio detectors were used in these tests. As will be discussed later the employment of square-law detectors could be expected to add 0.5 decibel to this figure. Linear detectors are to be preferred, however, on the basis of quality distortion. Table VII is based upon the theoretical and experimental study of diversity action and gives typical results for space diversity systems.

TABLE VII

SUMMARY OF SPACE DIVERSITY IMPROVEMENTS

Number of Antennas	Primary Improvement in db Automatic Gain Control		Secondary Improvements in db Number of Wave Bundles		
	0.06 Sec.	1 Sec.	1	2	3–5
2	1	0.5	3	2	1
3	1	0.5	4.5	3	1.5

Add 0.5 db to primary improvement when square-law detectors are used.

Table VII shows that on the average the secondary improvement is larger than the primary improvement. In other words, the advantage which accrues from the similarity of the antenna outputs exceeds that which accrues from their diversification. This result had not been expected.

It should be emphasized here that the improvements summarized in Table VII for space diversity systems and in Table V for a MUSA system refer to signal-to-noise ratios only; i.e., quality improvement is not included.

An important advantage of a MUSA system over a space diversity system is its ability to maintain its improvement when more than one wave bundle occur, and since two or more bundles are common, the advantage is distinctly real. A further advantage not discussed thus far relates to interfering signals as distinguished from static. Unless the interfering signals fall upon the principal lobe of the MUSA array pattern when it is steered to receive the desired signal, important directional discrimination against the interference occurs. Little or no

discrimination against interference can occur in a space diversity system since it lacks the phasing which produces directional discrimination.

The Time Constant of the Automatic Gain Control

Thus far no comments have been made on the improvement figures relating to "no automatic gain control" shown in the table of Fig. 39. This table shows that the signal-to-noise ratio for one antenna ($K = 1$) is from 2.5 to 3.5 decibels higher when no automatic gain control is used; i.e., perfect automatic gain control penalizes the signal-to-noise ratio to that extent.[31] The advantage of automatic gain control is a constant output volume. In practice, a compromise is effected by retarding the action of the control. A time constant of 0.5 or one second is usually used. (This compromise is influenced by quality considerations as well as noise considerations.)

In the MUSA system signal-to-noise ratio measurements the time constant of the automatic gain control circuit (0.06 second) was not changed during the switchover from the MUSA to the single antenna. If a time constant of 0.5 second had been used with the MUSA and a one-second time constant with the reference receiver, the measured improvement would probably have been reduced by a little less than one decibel.

Method of Averaging Noise

In all of the signal-to-noise measurements and in the diversity analysis noise voltage has been averaged arithmetically along the time axis. Owing to a rather marked reduction of noise peaks with the MUSA compared with a unit antenna different improvements would result if different ways of measuring it had been adopted. To investigate this, motion pictures were made of the signal meter and noise meter variations for the MUSA and for the single antenna. The transcribed records appear in Fig. 41. Some calculations have been carried out for the noise distributions marked A, B, and C in Fig. 41. If the noise ratio of B/A measured by arithmetically averaging noise voltage is called 0 decibels, it becomes + 2.4 decibels by averaging power arithmetically. The corresponding figures for B/C are 0 and + 2.7 decibels. Thus, if noise power is averaged instead of noise voltage the measured primary diversity improvement is substantially increased.

[31] The action of the automatic gain control does not change the instantaneous signal-to-noise ratio. Interpreting signal-to-noise ratio as *average signal* divided by *average noise* rather than the average of the *signal divided by the noise* results in this difference.

From the point of view of the interfering effect upon speech it is not clear which method of averaging is more significant. This matter is probably related too closely to the distortion which incidentally accompanies the noise peaks to be considered alone.

In the light of the discussion presented in the preceding pages it appears that the signal-to-noise improvement of the experimental MUSA can be expressed as 8 ± 1 decibels.

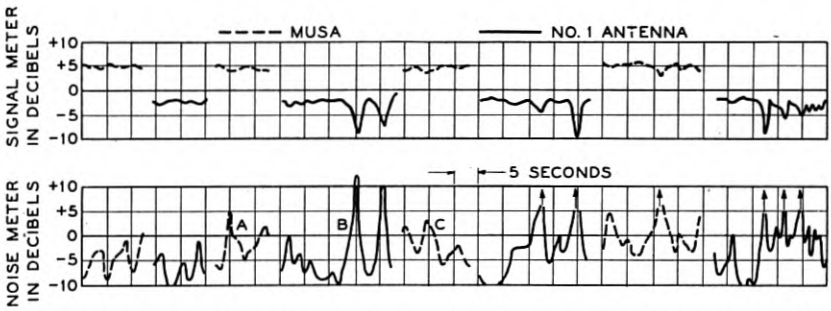


Fig. 41—Sample of signal and noise variations occurring in MUSA tests. The arrows indicate noise levels beyond the scale of the meter. This record was obtained five minutes before that in Fig. 40.

Square-Law Detectors

In the discussion of the signal-to-noise measurements it was stated that the measured improvements would have been higher had square-law audio detectors been used instead of linear rectifiers. For the two-bundle MUSA measurements one decibel was allowed for this and for the three-antenna diversity measurements at Netcong 0.5 decibel was allowed. These figures are based upon tests to be described in the following paragraphs. First, however, an analysis will be made of the effect upon the signal-to-noise ratio of various types of detectors in a MUSA system. Figure 42 is a schematic representation of the system to be analyzed, comprising K branches. The K signal generators e_{sA}, e_{sB}, $\cdots$ e_{sK} represent the various wave bundles as received by the steerable branches. The noise generators e_{nA}, e_{nB}, $\cdots$ e_{nK} are equal in amplitude but random in phase. The detectors are generalized to the extent that the audio output is proportional to the u power of the input.

Assuming that $e_s \gg e_n$ the audio outputs are proportional to ${e^u}_{sA}$, ${e^u}_{sB}$, $\cdots$ ${e^u}_{sK}$. The noise outputs are then proportional to $e_{nA}{e_{sA}}^{u-1}$, $e_{nB}{e_{sB}}^{u-1}$, $\cdots$ $e_{nK}{e_{sK}}^{u-1}$ since the signal-to-noise ratio in each branch must be independent of u. Assuming the signals to be delay equalized,

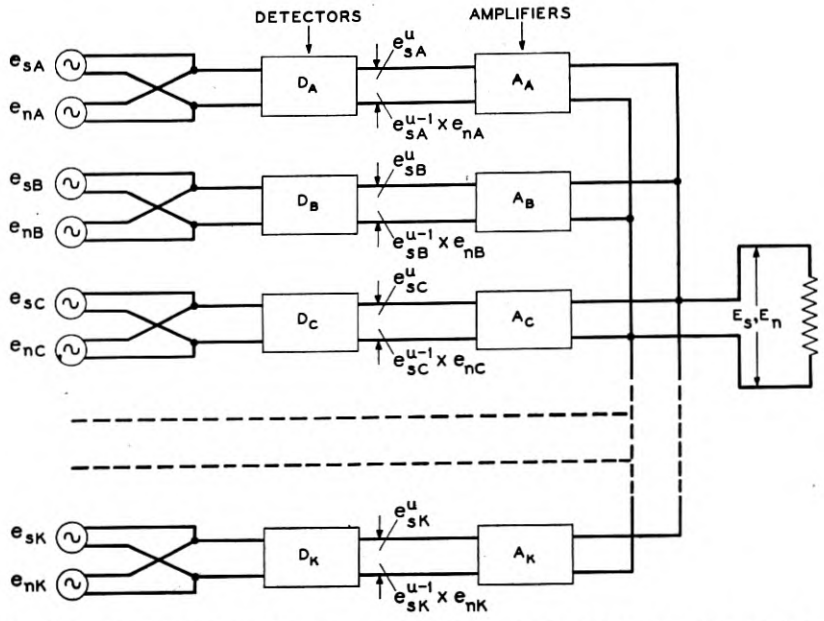


Fig. 42—Circuit employed in the analysis of the effect of detector characteristics upon signal-to-noise ratio.

the signal-to-noise ratio of the final output is

$$\frac{E_s}{E_n} = \frac{e^u{}_{sA} + e^u{}_{sB} + \cdots + e^u{}_{sK}}{e_n\sqrt{e_{sA}{}^{2u-2} + e_{sB}{}^{2u-2} + \cdots + e_{sK}{}^{2u-2}}} . \tag{7}$$

Maximizing this expression with respect to u shows that the maximum occurs for $u = 2$ (square law) and is

$$\frac{E_s}{E_n} = \frac{\sqrt{e^2{}_{sA} + e^2{}_{sB} + \cdots + e^2{}_{sK}}}{e_n} . \tag{8}$$
$$(u = 2)$$

That this expression represents the maximum signal-to-noise ratio may also be concluded by observing that it is proportional to the square root of the total energy.

For linear detectors $u = 1$ and the signal-to-noise ratio becomes

$$\frac{E_s}{E_n} = \frac{e_{sA} + e_{sB} + \cdots + e_{sK}}{e_n\sqrt{K}} . \tag{9}$$
$$(u = 1)$$

If the branch signals are all equal, i.e., if $e_{sA} = e_{sB} = \cdots = e_{sK}$, (8)

and (9) give the same result, but for unequal amplitudes there is an advantage in using square-law detectors.

This analysis shows that square-law detection introduces just the correct amount of emphasis upon the stronger waves and that any additional expansion or contraction of the differences among the waves is detrimental. This means that the gains in all branches should be equal. It also indicates that any arrangement in which the stronger of the several waves is automatically switched in and the remaining ones switched out is inferior.

The experimental MUSA receiver is equipped with both linear and square-law detectors, and some signal-to-noise ratio comparisons were made using locally generated signals. Figure 43 shows schematically the essential parts of the test circuit. The noise generators represent the thermal noise originating in the receiver input circuits. The input signal e_s was modulated with a tone. The calculated curves shown in the figure are obtained from (8) and (9) which reduce to

$$\frac{E_s}{E_n} \propto \sqrt{1 + \left(\frac{e_{sB}}{e_{sA}}\right)^2} \qquad (10)$$
$$u = 2$$

and to

$$\frac{E_s}{E_n} \propto \left(1 + \frac{e_{sB}}{e_{sA}}\right). \qquad (11)$$
$$u = 1$$

The equation for the square-law detector is sound and was verified by the measurements. The equation for the linear detector should apply only over a certain range of signal and noise levels. The measurements indicate this.

Automatic gain control was not used in these tests since the two gain controls could not be relied upon to "track" sufficiently well. To make the measurements significant manual gain control was used to maintain the receiver gains equal and the output normal. It may be pointed out here that in receiving actual radio signals with linear detectors accurate equality of gains is not required. Moderate differences in gain (of a few decibels) can be depended upon to be beneficial as often as detrimental. With square-law detectors no departure from equality can be beneficial.

The curves of Fig. 43 show that 10- and 20-decibel differences in signal level give the square-law detector an advantage of one and two decibels, respectively. In receiving two bundles of waves the branch outputs commonly fade in and out with the result that their average ratio is of the order of 10 decibels. Such were the conditions, as well

as could be estimated, during the two-bundle tests in Tables III and IV. The one-decibel correction applied to those results in Table V is therefore justified. In the one-wave bundle measurements, the percentage of time during which the branch outputs were substantially different was so small that no correction was applied in Table V. In the case of space diversity the correction is also small; about 0.5 decibel seems reasonable.

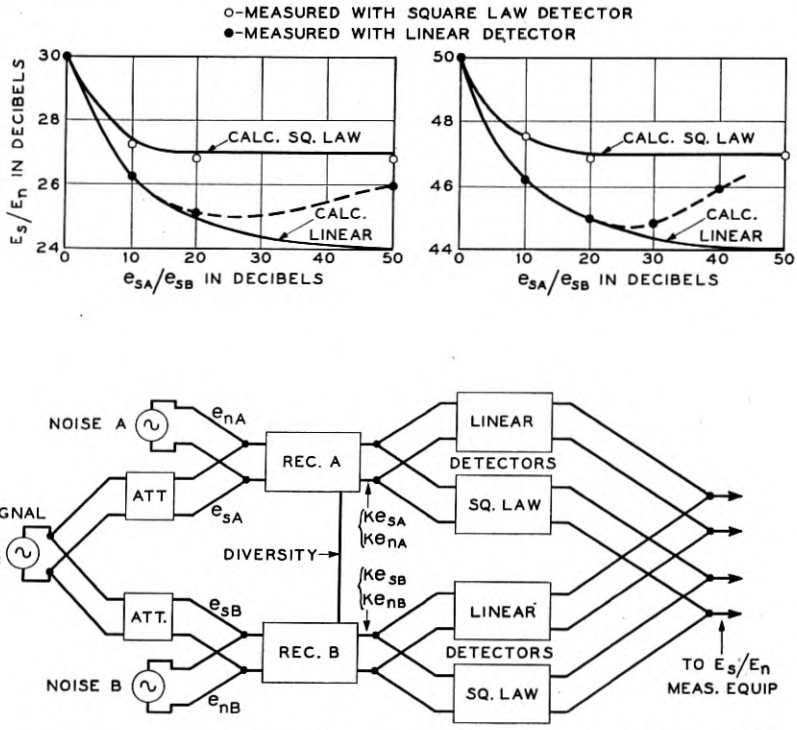


Fig. 43—Test circuit and experimental results of the study of detector characteristics.

From the point of view of distortion in receiving double side-band signals linear detectors are superior to square-law detectors and this compensates their inferiority in signal-to-noise ratio. The principal reason for using linear detection in the signal-to-noise ratio tests of the MUSA was, however, their experimental advantage in simplicity and accuracy (Fig. 35).

Random Addition of Static

In analyzing the spaced antenna systems at the beginning of this section it was assumed that the static outputs of the antennas add on a power basis. An experimental study of this was made by measuring

the static output of one unit antenna and comparing it with the static output of the six antennas combined as one MUSA branch. The circuit shown in Fig. 35 was used for these experiments. The results are tabulated in Table VIII.

TABLE VIII

Date	GMT	f_{mc}	Type of Static	Addition Max. db	Addition Min. db	Thermal Noise db	QRM	Method
1935								
9–19	1530	18.6	star	8.5		−12	light	Rdg. DB Meter
10–15	1500	9.51	star	8.0				Rdg. DB Meter
10–16		9.51	distant	7.5		− 6		Rdg. DB Meter
10–22	1500	9.51	crash	8		−20	none	Vary I-F gain
10–23	1500	9.51	distant	8		− 9	light	Vary I-F gain
	1820	11.86	distant	8.5			light	Vary I-F gain
10–24	1500	9.51	star	11.4	5.4	−12	light	Vary I-F gain
	1510	9.51	star	11.0	6.0	−12	light	Vary I-F gain
	2045	9.51	crash	7.5		−30	light	Vary I-F gain
11–1	1450	9.51	distant	9.0		− 8	none	Vary I-F gain
	1830	9.51	distant	8.0		− 7	—	Vary I-F gain
1936								
1–7	1500	9.51	distant	7.5		− 8	light	Vary I-F gain
	1505	9.51	distant	8.0		− 8	light	Vary I-F gain
1–14	0300	4.82	crash	8.2	7.3	−20	none	Fluxmeter
1–15	0300	4.82	crash	6.8	3.0	−30	none	Fluxmeter
		Average		8.0				

The column headed f_{mc} indicates the frequency to which the receiver was tuned. The MUSA was tuned to a desired station which possessed a comparatively clear channel, and, following the sign-off of the station, static was measured without a demodulating carrier. The receiver was, of course, operated with manual instead of automatic gain control. Care was taken to insure that overloading did not occur. The column headed "addition" gives the ratio in decibels of the static output of the six-antenna branch to that of one antenna. Where figures are entered in the middle of the column no effect of steering was noticed. (Effects of the order of one decibel could have been overlooked, however.) The column headed thermal noise gives the ratio in decibels of receiver noise originating in one of the first circuits (measured with a resistance replacing the antenna) to the total noise measured with the unit antenna connected. It shows that thermal noise was negligible.

The star static [32] was steady and therefore accurately measurable. The crash static on 4.82 megacycles was so intermittent that it required the use of the fluxmeter to obtain a satisfactory measurement.

[32] K. G. Jansky, "Electrical Disturbances Apparently of Extraterrestrial Origin," *Proc. I. R. E.*, vol. 21, pp. 1387–1398, October, 1935.

The average of all those measurements showing no effect of steering is 8.0 decibels compared with the theoretical figure of 10 log 6 = 7.8 decibels, for random addition. The random assumption employed in the analysis is thus justified on the average.

Summarizing Discussion

The aim of the signal-to-noise study described in this section has been not so much to evaluate the intrinsic merit of the experimental MUSA system as to compare its behavior with a simple theory. The element of research has been to find out how well the transatlantic waves fit into the background of the simple theory. To this end somewhat artificial devices, thermal noise and rectified carrier, were substituted for static and speech signals.

The study included an analysis of diversity action in which the effect upon the signal-to-noise ratio of (1) delay equalization, (2) detector characteristics, and (3) automatic gain control action was displayed prominently. Although those effects, taken together, are important, they are individually small and could be separately evaluated only by locally controlled test methods.

We propose now to review the results of the tests and studies. The slight decrease in improvement which would have occurred had the speed of the automatic gain control been reduced, and the increase in improvement which would have resulted had noise *power* been averaged do not affect the fundamental considerations and will be neglected here.

One wave bundle received with one branch (Table V) yielded an improvement of 7 decibels which is about one decibel less than 10 log N = 7.8 decibels calculated by simple theory (Fig. 34, Case I). Comments relating to this have been given.

One wave bundle received with two branches steered on each side of the center of the bundle yielded eight decibels improvement (Table V). Of this, one decibel is due to primary diversity action and three decibels are due to the secondary diversity gain. This three-decibel gain was assumed in the simple theory although it was not then designated as secondary diversity gain. It accrues by virtue of voltage addition of the signals and power addition of the noise. Both of these conditions are satisfied in Table V. There remain four decibels which represent the signal-to-noise improvement in each branch, referred to one antenna. This indicates a loss of three decibels as compared with a single branch steered at the center of the bundle, which gives seven decibels; this is reasonable when it is remembered that the branches were steered apart by a phase difference of about 60 degrees ($\phi_A - \phi_B$

$= 60°$). A loss of three decibels is about what one would estimate upon inspecting the directional pattern for $\phi = 60$ and 120, say, on Fig. 20. The procedure in which two branches are steered at one bundle as in the above is frequently employed and is an important factor in the operation of a MUSA.

The case of two wave bundles tabulated in Table V also yields an improvement of 7.5 to 8 decibels. Of this, one decibel and three decibels are due to primary and secondary diversity action, respectively, as in the case of one bundle. This leaves 3.5 to 4 decibels for the signal-to-noise improvement in each branch referred to a unit antenna. But the unit antenna has the advantage of two bundles, whereas the MUSA branch excludes one of them, a three-decibel difference. In comparison with a unit antenna receiving only one bundle, the improvement to be ascribed to one branch thus is increased to 6.5 or 7 decibels. This result compares favorably with the seven decibels yielded by one branch steered at one bundle. It is in this case of two bundles that square-law detectors are most important. Their advantage, amounting to an estimated one decibel, has already been included in Table V, it will be remembered.

The measurements which have permitted the above analysis of the MUSA signal-to-noise improvement were of course supplemented by aural observations made over the course of a year and a half. The listening tests corroborate the analytical results as well as can be expected of such observations. Not infrequently they showed somewhat less than the full eight-decibel improvement. The indications are, however, that a larger MUSA with three (or possibly four) branches would have yielded more nearly its full gain of 10 log N decibels. For, a MUSA receiving system does not perform its functions properly unless it is sharp enough to separate the waves sufficiently to permit effective delay equalization; also, to obtain the full gain, enough branches must be provided to utilize all of the important wave bundles. The Holmdel experimental MUSA is really a conservative approach to the field of steerable directivity. There is, of course, an upper limit to the size of a MUSA, beyond which (1) technical difficulties in phasing, etc., will occur, (2) the cost of the improvement may be less if introduced at the transmitter, and (3) the directional sharpness becomes too great to permit practical operation with waves of the stability encountered in transatlantic transmission. At present, a system about three times the length of the experimental MUSA comprising eighteen antennas and equipped with three branches seems practical. It should yield an improvement of $10 \log 18 = 12.5$

decibels more consistently than the present MUSA yields eight decibels.

It may be worth while here to point out that as the number of antennas in a MUSA system is increased there is no tendency for static to become subordinate to thermal noise (set noise) or vice versa when static, like thermal noise, adds on a power basis. Only to the extent that transmission-line loss increases with the number of antennas will the ratio of thermal noise to static increase.

A type of transmission sometimes occurs for which the experimental MUSA gives only small signal-to-noise improvement. We refer to the highly scattered propagation associated with flutter fading, discussed at the close of Section IV. In such cases signal-to-noise improvement is not highly significant, however, since at least in the worst cases, the distortion renders the circuit worthless. Thus, increasing the transmitting power is likewise ineffective. On the other hand the experimental MUSA can accomplish something by rejecting some of the scattered waves which appear to be responsible for the flutter fading. This is accomplished without a corresponding *loss* of signal-to-noise ratio since, of course, noise is rejected, too. Fortunately, flutter fading does not seem to be associated prominently with greatly depressed field intensity so the failure to secure signal-to-noise improvement with flutter fading does not appreciably penalize the MUSA as a means of extending operation through periods of depressed field conditions.

VI. Recapitulation

The MUSA receiving system described in this paper is the culmination of some four years effort to determine the extent to which receiving antenna directivity may be carried to increase the reliability of short-wave transatlantic telephone circuits.[33] Fundamental experimental studies of wave propagation were made with particular emphasis upon how the waves arrive. Based upon the results of these studies a system was evolved in which a new technique of phasings was required. The result is a steerable antenna whose signal-to-noise advantage is seven to eight decibels compared with the largest fixed antenna that can be employed effectively. By analyzing this improvement and comparing the various contributing factors with theory, it is possible to estimate that a system three times larger than the experimental one will yield an additional four to five decibels, and will perform better consistently. In addition to the signal-to-noise improvement a

[33] Potter and Peterson, "The Reliability of Short-Wave Radio Telephone Circuits," *Bell Sys. Tech. Jour.*, vol. 15, pp. 181–196, April, 1936.

substantial improvement in quality is obtained by reducing the distortion associated with selective fading. It is both interesting and important to note that whereas so often one advantage is gained only at the expense of another, in the MUSA system the best quality improvement and the greatest signal-to-noise advantage are obtained together, without compromising.

The system developed is expensive and might be thought to illustrate the law of diminishing returns. As a part of a point-to-point radiotelephone system, however, it has certain compensating features not mentioned thus far. One of these is the broad frequency band feature.

With essentially aperiodic unit antennas the MUSA possesses a broad frequency range; i.e., the directional pattern, despite its sharpness, is substantially the same over a band of a hundred or more kilocycles provided the terminal equipment is made sufficiently broad. (See Appendix I.) The broad-band feature is important for its possibilities in multiplexed operation of telephone circuits; i.e., it makes possible, insofar as the antenna system is concerned, the adaptation of some of the carrier telephone methods to radio circuits. It is to be expected that, excepting certain critical cases, fairly large percentage frequency bands will follow virtually the same paths. This assumption was verified by a few experiments in which pulses were received simultaneously from GBS (Rugby, 12,150 kilocycles) and GBU (Rugby, 12,290 kilocycles) 140 kilocycles apart. These tests showed that, although the pulse fading was, of course, not synchronous, the angles involved were alike.

Another compensating feature of the MUSA receiving system is that, with suitable terminal equipment, reception may be carried on from several points at once provided they lie within the horizontal angular range of the unit antenna. Some sacrifice in vertical angular selectivity occurs but this is confined to low angles where it is least important.

Certain features of the system make for economies in plant cost. The fact that a great many components are identical permits manufacturing economies. Also, spare units need be provided only for a few vital functions, since the failure of one of the many similar parts does not disrupt service.

The development of steerable directivity has thus far been concerned with receiving antennas. In receiving, one has the obvious advantage of having, in the monitoring branch, a criterion to dictate the steering adjustments. The lack of such a direct criterion for adjusting transmitting directivity does not, however, rule out the possibility, at

some future time, of a MUSA transmitting system. That horizontal steering of transmitting directivity may be decidedly important is strongly suggested by observations made on transmissions from Daventry in which significant effects upon flutter fading have been found to be associated with the orientation of the directional transmitting antennas.

Acknowledgment

The experiments described in this paper necessarily involved the coordinated effort of many individuals, in both the British Post Office and the Bell System, and their help has been appreciated. Mr. E. Bruce had charge of the design of the rhombic antennas and transmission lines, and Messrs. L. R. Lowry and W. M. Sharpless had important parts in the various phases of the work.

The authors are particularly indebted to Mr. R. K. Potter who contributed much through his keen interest throughout the entire work.

Appendix I

Broad-Band Characteristic of the MUSA

The frequency characteristic of the MUSA may be calculated from (3). Frequency and angle appear only in the form $2\pi a(\upsilon - \cos \delta)$ where a is inversely proportional to frequency. By writing the equation

$$\frac{2\pi Df}{c}[\upsilon - \cos \delta] = \frac{2\pi D(f + \Delta f)}{c}[\upsilon - \cos (\delta + \Delta\delta)]$$

we express the angular shift, from δ to $(\delta + \Delta\delta)$, of a given point on the directional pattern as the frequency is varied from f to $(f + \Delta f)$. This equation may be rewritten as

$$1 + \frac{\Delta f}{f} = \frac{\upsilon - \cos \delta}{\upsilon - \cos (\delta + \Delta\delta)}.$$

As an example consider $\Delta f = 200$ kilocycles, $f = 10$ megacycles, $\delta = 30$ degrees, and $\upsilon = 1.05$. Then $\Delta\delta = -0.4$ degree. For lower values of δ, $\Delta\delta$ becomes still smaller.

The frequency characteristic expressed in terms of percentage band and angular shift given by the above equation is independent of the size of the MUSA. It relates to the over-all length of the system, however, by the fact that for greater lengths a given angular shift has more effect.

The broad band of the MUSA reflects the fact that, with the terminal at the "leeward" end as assumed heretofore, the delay of the space paths is nearly the same as that of the transmission-line paths so that if the antenna outputs are phased to add at one frequency they will nearly add at other frequencies. If the terminal is located at the center of the MUSA to economize on transmission line, the frequency range is greatly reduced. The broad band may be regained, however, by delays introduced in the receiving equipment. With a center location, the antennas in the forward and rearward sections of the MUSA must have their phases shifted oppositely, and, unless certain other compensating networks are provided, the two phase shifts must be coupled in different phase relations for different wave-lengths.

Abstracts of Technical Articles from Bell System Sources

Modern Theater Loud Speakers and their Development.[1] C. FLANNAGAN, R. WOLF, and W. C. JONES. Although many of the basic ideas involved in the operation of present-day loud speakers were conceived during the early stages of the development of the telephone, it was not until the advent of the vacuum tube amplifier that these principles were applied to the design of structures capable of delivering sufficient acoustical power to be audible throughout a room or auditorium. Having reached this stage, however, the developments that culminated in the sound reproducing systems employed with present-day sound pictures came in rapid succession. These developments have embraced all phases of loud speaker design, with the result that systems are now available that convert from 25 to 50 per cent of the electrical input into acoustical output, and maintain conversion efficiencies of this order of magnitude over a frequency range of 50 to 10,000 cps. These systems are so designed as to be capable of reproducing the recorded sound at intensities that not only greatly enhance the dramatic effect of the presentation in the theater, but also open entirely new fields in recording. All these improvements have been attained with a reduction in distortion and improved fidelity of the reproduced sound. The directional properties of the loud speakers also have been markedly improved, with the result that the better quality of reproduction achieved is available throughout the entire seating area and the undesirable beam effects previously experienced have been eliminated.

Power System Faults to Ground—Part I: Characteristics.[2] C. L. GILKESON, P. A. JEANNE and J. C. DAVENPORT, JR. The results of an extensive oscillographic study of power-system faults to ground are presented herewith. While this study was made primarily to obtain data useful in inductive coordination problems, the results are believed to be of general interest as well. They provide data on such items as frequency of occurrence of ground-current disturbances, their monthly distribution, duration, cause, method of clearance, and wave-trace characteristics. Data on fault resistance are given in part II, a companion paper.

[1] *Jour. S. M. P. E.*, March 1937.
[2] *Elec. Engg.*, April 1937.

Direct Recording and Reproducing Materials for Disk Recording.[3] A. C. KELLER. Recently materials for direct recording and reproducing work have been improved so that they are now suitable for many uses. These materials, as they are available on the market, are classified chemically into five groups and measurements are given of frequency characteristic, surface noise, life, distortion, etc. These data have been taken with both lateral and vertical recording.

[3] *Jour. Acous. Soc. Amer.*, April 1937.

Contributors to this Issue

EDWIN H. COLPITTS was introduced to readers of the *Journal*—if it can be said that he needed an introduction—in the April issue.

KARL K. DARROW, B.S., University of Chicago, 1911; University of Paris, 1911–12; University of Berlin, 1912; Ph.D., University of Chicago, 1917. Western Electric Company, 1917–25; Bell Telephone Laboratories, 1925–. Dr. Darrow has been engaged largely in writing on various fields of physics and the allied sciences.

C. B. FELDMAN, B.Sc., University of Minnesota, 1926; Teaching Fellow, University of Minnesota, 1926–28; M.Sc., University of Minnesota, 1928. Bell Telephone Laboratories, 1928–. Mr. Feldman has been engaged in short-wave radio receiving. His work has been mainly on transmission lines, antennas, and wave propagation problems.

H. T. FRIIS, E.E., Royal Technical College in Copenhagen, 1916; Columbia University, 1919–20. Research Department, Western Electric Company, 1920–24; Bell Telephone Laboratories, 1925–. Mr. Friis' work has been largely in connection with radio reception methods and measurements. He has published papers on vacuum tubes as generators, radio transmission measurements and static interference. As Radio Research Engineer he now directs studies of new methods of short-wave reception.

W. P. MASON, B.S. in Electrical Engineering, University of Kansas, 1921; M.A., Columbia University, 1924; Ph.D., 1928. Bell Telephone Laboratories, 1921–. Dr. Mason has been engaged in investigations on carrier transmission systems and more recently in work on wave transmission networks, both electrical and mechanical.

JOHN RIORDAN, B.S., Sheffield Scientific School, Yale University, 1923. American Telephone and Telegraph Company, Department of Development and Research, 1926–34; Bell Telephone Laboratories, 1934–. Mr. Riordan's work has been mainly on problems associated with inductive effects of electrified railways.

R. A. SYKES, Massachusetts Institute of Technology, B.S. 1929; M.S. 1930. Columbia University, 1931–33. Bell Telephone Laboratories, Research Department, 1930–. Mr. Sykes has been engaged in the application of piezoelectric crystals to selective networks, and more recently in the use of coaxial lines as filter elements.