The Bell System Technical Journal

October, 1927

Television [1]

By HERBERT E. IVES

SYNOPSIS: The chief problems presented in the accomplishment of television are discussed. These are: the resolution of the scene into a series of electrical signals of adequate intensity for transmission; the provision of a transmission channel capable of transmitting a wide band of frequencies without distortion; means for utilizing the transmitted signals to re-create the image in a form suitable for viewing by one or more observers; arrangements for the accurate synchronization of the apparatus at the two ends of the transmission channel.

INTRODUCTION

THIS paper is to serve as an introduction to the group of papers following, which describe the apparatus and methods used in the recent experimental demonstration of television over communication channels of the Bell System. In that demonstration television was shown both by wire and by radio. The wire demonstration consisted in the transmission of images from Washington, D. C., to the auditorium of the Bell Telephone Laboratories in New York, a distance of over 250 miles by wire. In the radio demonstration, images were transmitted from the Bell Laboratories experimental station at Whippany, New Jersey, to New York City, a distance of 22 miles. Reception was by two forms of apparatus. In one, a small image approximately 2 in. by 2½ in. was produced, suitable for viewing by a single person, in the other a large image, approximately 2 ft. by 2½ ft., was produced, for viewing by an audience of considerable size. The smaller form of apparatus was primarily intended as an adjunct to the telephone, and by its means individuals in New York were enabled to see their friends in Washington with whom they carried on telephone conversations. The larger form of receiving apparatus was designed to serve as a visual adjunct to a public address system. Images of speakers in Washington addressing remarks intended for an entire audience, and of singers and other entertainers at Whippany, were seen by its use, simultaneously with the reproduction of their voices by loud speaking equipment.

[1] Presented at the Summer Convention of the A. I. E. E., Detroit, Mich., June 20–24, 1927.

Characteristic Problems of Television

The problem of television in its broad outlines is that of converting light signals into electrical signals, transmitting these signals to a distance, and then converting the electrical signals back into light signals. Given means for accomplishing these three essential tasks, the problem becomes that of developing these means to the requisite degree of sensitiveness, speed, efficiency, and accuracy, in order to re-create a changing scene at a distant point, without appreciable lapse of time, in a form satisfactory to the eye.

A convenient starting point for the discussion of television is the human eye itself. In this an image is formed upon the retina, a sensitive screen, consisting of a multitude of individual light-sensitive elements. Each of these elements is the termination of a nerve fibre which goes directly to the brain, the entire group of many million fibres constituting the optic nerve. A theoretically possible television system could be made by copying the eye. Thus a large number of photosensitive elements could be connected each with an individual transmission channel leading to a distant point, and signals could be sent simultaneously from each of the sensitive elements to be simultaneously used for the re-creation of the image at the distant point. The number of wires or other communication channels demanded in a television system of this sort would be impractically large. For practical purposes, reduction of the number of transmission channels is made possible by the fact that, while in vision all parts of the image on the retina are simultaneously and continuously acting to send nerve impulses, the inertia of the visual system is such that a sensation of continuity is obtained from discontinuous signals, provided these succeed each other rapidly enough. Due to the phenomenon of persistence of vision, it is immaterial to the eye whether the whole view be presented simultaneously or whether its various elements be viewed in succession, provided the entire image be traversed in a sufficiently brief interval, which for purposes of discussion may be taken as 1/16th of a second or less.[2] We thus have available in television the same artifice which is used in the much less exacting problem of transmission of pictures over a telephone line, that is, of *scanning*, or running over the elements of the image in sequence, instead of endeavoring to transmit all of the elementary signals simultaneously. The development of a television system therefore

[2] This figure of 1/16th of a second, commonly quoted in discussions of this sort, is a convenient one, although the frequency of image repetition necessary to extinguish "flicker" is actually proportional to the logarithm of the field brightness. A somewhat higher rate of image repetition was used in the final television apparatus.

necessitates, at an early stage, the design of some scanning system by which the image to be transmitted may be broken up into sequences of signals. In the simplest case, where one transmission channel is to be used, the whole image will be resolved into a single series of signals; if more than one transmission channel is to be utilized, the resolution may, by parallel scanning schemes, or their equivalent, be broken up into several series for simultaneous transmission.

Like the eye, an artificial television system must have some light-sensitive element or elements by means of which the light from the object shall produce signals of the sort which can be transmitted by the transmission system to be used. For a television system to operate over electrical transmission lines this means some photoelectric device. It is obvious that this photoelectric device must be extremely rapid in its response, since the number of elements of any image to be transmitted must be some large multiple of the fundamental image repetition frequency, that is 16 per second. The response should, of course, be proportional to the intensity of the light, and finally, the device must be sufficiently sensitive so that it will give an electrical signal of manageable size with the amount of light available through the scanning system. This latter requirement, that of sensitiveness, is one which, it was realized, from studies made with our earlier apparatus for the transmission of still pictures over wires,[3] would be extremely difficult to meet. In the picture transmission system a very intense beam of light from a small aperture is projected through a transparent film and on to a photoelectric cell. In practical television, the system must be arranged to handle light reflected from a natural object, under an illumination which would not be harmful or uncomfortable to a human being. Actual experiment showed that the greatest amount of light which could be collected from an image, formed by a large aperture photographic lens on the small scanning aperture of the picture transmission apparatus, was less by a factor of several thousand times than the light projected through it for still picture transmission purposes. Assuming the same kind of photoelectric cell to be used, the additional amplification required over that used in the picture transmission system, taking into account also the higher speed of response demanded, would bring us at once into the region where amplifier tube noise and other sources of interference would seriously affect the result. This indicated clearly that some more efficient method of gathering light from the object than the commonly assumed one of image formation by a

[3] "Transmission of Pictures over Telephone Lines," Ives, Horton, Parker and Clark. *Bell System Technical Journal*, Vol. IV, No. 2, April, 1925.

lens was required, unless some much more sensitive type of photoelectric cell should be found.

Assuming that means could be developed for producing an electrical signal proportional to the intensity of the light, of sufficient quickness to follow a rapid scanning device, and of sufficient strength either as directly delivered from a photosensitive device or as amplified, the next problem is that of its transmission over an electrical communication system. We may quickly arrive at an understanding of certain of the transmission problems by reviewing the requirements for the transmission of photographs. In the system of still picture transmission now in use by the American Telephone & Telegraph Company, a picture 5 in. by 7 in. in size, divided into the equivalent of 10,000 elements per square inch or 350,000 elements, is transmitted in approximately seven minutes. This requires the transmission of a frequency band of about 400 cycles per second on each side of the carrier frequency. If we plan, in the transmission of television, to transmit images of the same fineness of grain, it would mean that what is now transmitted in seven minutes would have to be transmitted in a 16th of a second, which in turn means that the transmission frequency range would have to be nearly 7000 times as great. That is, a band approximately 3,000,000 cycles wide would be required. Bearing in mind that wire circuits are ordinarily not designed to utilize frequencies higher than 40,000 cycles per second, and that with radio systems uniform transmission of wide signal bands becomes extremely difficult, it is seen at once that either an image of considerably less detail than that which we have been considering must suffice, or else some means for splitting up the image so that it may be sent by a large number of channels is indicated.

A further theoretical requirement must also be given consideration. This is that the complete television signal will consist of all frequencies up to the highest above discussed, and down to zero, that is, an essential part of the signal is the direct current component, furnished by those parts of the scene which do not change. The problem of handling the very low frequency components, presents difficulties both in the vacuum tube amplifier system adjacent to the photosensitive device, and in ordinarily available transmission channels.

In any case certain fundamental transmission requirements must be met. These are that the attenuation of the signals must be uniform over the whole frequency range and that the speed of transmission of all frequencies must be the same. Also, as in the transmission of sound signals, the amount of interference or noise must be kept down sufficiently not to impair the quality of the signal or picture.

Assuming the undistorted transmission of the signals to a distant point, the next fundamental problem of television is the reconstruction of the image, or the translation of the electrical signal back into light of varying intensity. Just as at the sending end we have seen that the production of a useful electrical signal with the amount of light available from a naturally illuminated object is a major problem, so at the receiving end the converse problem, that of securing an adequately bright light from the electrical signal, presents great difficulty. The nature of the problem may be understood by assuming that it is to be done by projecting the received image on a screen similar to an optical lantern projection screen. If the spot of light which is to build up this image scans the whole area in the same way that the object is scanned, we find that the amount of light which can be concentrated into a small elementary spot will, when distributed by the scanning operation over the whole screen, reduce the brightness of the screen in the ratio of the relative areas of the elementary spot and the whole screen. The amount of this reduction will, of course, depend upon the number of elements into which the picture is divided, but will in any event be a factor of several thousand times. It is doubtful whether any light source exists of sufficient intensity such that an image projected by it can be spread out by a scanning operation over a large screen and give an average screen brightness which would be at all adequate. It is possible to imagine optical systems by which such a thing as the crater of an arc could be projected upon the screen, but the motion of this image and its variation in intensity would involve the extremely rapid motion of lenses, mirrors and apertures of a size such as to render the operation mechanically impracticable. It appears from these considerations that the only promising means of reconstructing the image would be those in which a light source, whose intensity can be controlled with great rapidity, is directly viewed.

Another element of a television system upon whose solution success depends as much as any other is that of synchronization; the reconstruction of the image, postulated in the last paragraph, is only possible if the reconstructed elements fall in exactly the right positions at the right times, to correspond with the signals as generated at the analyzing end. The criterion for satisfactory synchronization will be expressed in terms of variation from identity of speed by figures which will depend on the fineness of grain of the image which it is planned to send. No element of the image must, of course, be out of place by a considerable fraction of the size of the element.

General Outline of Means Employed in the Present Television System

It has been pointed out above that if the goal which we set in television is the transmission of extended scenes, with a large amount of detail and hence made up of an exceedingly large number of elementary areas, we meet with the necessity for transmission channels of a character which are not now available. In the present development it was decided at the start to restrict our experiments to a size and grain of picture which, if the scanning and re-creating means were developed, would be capable of transmission over practical transmission channels, either wire or radio. This restriction fortunately leaves us with the possibility of meeting what was felt to be the typical problem of a Telephone Company, namely, the transmission of a human face in a television system used as an adjunct to a telephone system. Taking, as a criterion of acceptable quality, reproduction by the halftone engraving process, it is known that the human face can be satisfactorily reproduced by a 50-line screen. Assuming equal definition in both directions, 50 lines means 2500 elementary areas in all. 2500 elements transmitted in 1/16 second is 40,000 elements per second. The frequency range necessary to transmit this number of elements per second with a fidelity satisfactory for television cannot be calculated with assurance in advance. An approximate value can however be arrived at from a study of the results obtained in still picture transmission. In pictures transmitted by the system already referred to, individual faces contained in a square space ½ inch on a side are quite recognizable.[4] Taking the ratio of this area to the area of the whole picture, and using the frequency range figure already deduced, for a complete 5 in. by 7 in. picture, it appears that a band of 20,000 cycles would be sufficient to transmit such an image in 1/16 second.[5] These considerations led to the choice of a 50-line (2500-element) image as one which would be both satisfactory as to detail rendering, for our purposes, and as calling for frequency transmission requirements sufficiently low to give a good margin of safety in existing single communication channels.

As a method of scanning, the method which is probably mechanically simplest, namely, that of rotating disks with spirally arranged holes, proposed by Plotnow [6] in 1884, was chosen. In accordance with the

[4] Cf. Fig. 18 of the paper referred to (Reference 2).

[5] A factor which this analogy does not cover is that if the image is moving so that it falls on several discrete scanning elements in rapid succession a very material apparent increase in the fineness of the image structure results. This effect is similar to that by which the relatively coarse-grained individual images in a motion picture film fuse to give smooth appearing pictures.

[6] Plotnow, D. R. P. 30105, 6.1, 1884.

choice of grain above indicated, the disks were perforated with 50 apertures.

For the second element of the problem, the light-sensitive means, the alkali metal photoelectric cell was chosen as possessing the qualities of proportionality of response and quickness of reaction. The currents produced by it are at best quite small, but they lend themselves to the process of amplification by the three-electrode vacuum tube amplifier.

The problem of securing a large enough signal, which is intimately associated with that of securing enough light from the object, was, in our development work, postponed in the earlier stages, our first experimental work having been done by concentrating light through photographic transparencies.[7] The solution of the problem of securing adequate light was subsequently attained by reversing the light path and projecting a narrow beam of light through the scanning disk upon the object. By this means only the element of the object which was being scanned was illuminated at any one time, thereby reducing the average illumination enormously, and the problem of increasing the signal strength could be attacked by increasing the amount of photosensitive surface as well as by increasing the brightness of the scanning light.[8]

The problem of amplifying the photoelectric currents to sufficient value for transmission was solved by a practical compromise which at the same time met one of the transmission difficulties. This compromise consisted in amplifying and transmitting only the fluctuating or alternating current components of the signal, leaving the direct current component, which determines the general tone value of the image, for empirical reintroduction at the receiving end. By this scheme, stable amplifier constructions were made available, and the transmission channels, particularly the wire channels, could be utilized in their normal working form.

At the receiving end, the problem of securing a sufficiently bright image was solved, as indicated earlier, by the use of self-luminous surfaces of much higher intrinsic brightness than it is possible to secure by illumination of a surface by any light source which can be rapidly controlled as to its intensity. The self-luminous surfaces

[7] As one step in the development work moving picture film, projected by a commercial projector in synchronism with the scanning disks, was successfully transmitted.

[8] A still further advantage is obtained by limiting the scanning light to the region of the spectrum to which the photoelectric cells are sensitive (blue and violet). This is unnecessary where one-way transmission only is used but is of value where in two-way transmission a transmitted image is to be viewed by a person being scanned.

employed were glow lamps containing neon gas, the brightness of which changes with sufficient rapidity to follow the incoming signals.

The problem of synchronization was postponed in our earlier development work by mounting the scanning and receiving disks upon the same axle. It was later solved for the demonstration apparatus by the utilization of synchronous motors controlled by two frequencies, a low frequency, that of the image repetition period, and a high frequency, chosen of such a value that the fraction of the cycle through which transient phase displacements occurred amounted in angular displacement to less than half the angular extent of a single disk aperture. The synchronization control therefore called for the transmission of additional currents for synchronization purposes over and above the picture current.

In order to transmit and synchronize the image signals it is necessary to transmit three different frequency bands, one for the image, and two for the high and low frequency synchronization controls. In the demonstration of April 7, 1927, the images were sent in the wire demonstration over a high quality open wire line. The synchronization control was sent over two separate carrier channels of a second telephone line. In addition to these lines, another line was used for conveying the telephone conversation. In the radio demonstration two different wave-lengths were used respectively for the image signals and for the synchronization signals which were, as in the wire demonstration, carried on two different carrier frequencies. A third channel was used for the voice. In the case of both wire and radio transmission, it is quite possible to put all of the different signals upon the same transmission channel, using different carrier frequencies.

It will aid toward a clear understanding of the reasons for the success of the system of television described in the following papers if we summarize at this point the chief novel features to which that success is due. They may be listed as follows:

1. Choice of image size and structure such that the resultant signals fall within the transmission frequency range of available transmission channels.[9]

2. Scanning by means of a projected moving beam of light.

3. Transmission only of alternating current components of image.

4. Use of self-luminous surfaces of high intrinsic brilliancy for reconstruction of the image.

5. High frequency synchronization.

[9] As the succeeding papers show, the margin between the frequency range required by the scanning apparatus and that which could be made available was quite liberal. It appears in the light of our experience that apparatus with 60 or 70 scanning holes instead of 50 might be used with the transmission facilities which were at our disposal.

Applications and Future Developments

It is not easy at this early date to predict with any confidence what will be the first or the chief uses for television, or the exact lines that future development may take. It must be clearly understood that television will always be a more expensive service than telephony, for the fundamental reason that it demands many times the transmission channel capacity necessary for voice transmission. This expense will inevitably increase in proportion to the size and quality of the transmitted image.

The kinds of service which are naturally thought of upon consideration of the services now rendered in connection with sound transmission are: first, service from individual to individual, parallel in character to telephone service, and as an adjunct thereto; second, public address service, by which the face of a speaker at a distant point could be viewed by an audience while his voice was transmitted by loud speaker; third, the broadcasting of scenic events of public interest, such as athletic contests, theatrical performances and the like.

The first two types of service just mentioned lie within the range of physical practicability, with apparatus of the general type already developed. The third type, because of the uncontrolled conditions of illumination, and the much finer picture structure which would be necessary for satisfactory results, will require a very considerable advance in the sensitiveness and the efficiency of the apparatus, to say nothing of the greatly increased transmission facilities. For all three types of service, wire or radio transmission channels could be utilized, for while the problems incident to securing distortionless transmission over wide frequency bands, or multiple transmission channels, are different in detail in the two cases, they appear to be equally capable of solution by either means. However, the very serious degradation of image quality produced by the fading phenomena characteristic of radio indicates the practical restriction of radio television to fields where the much more reliable wire facilities are not available.

The Production and Utilization of Television Signals[1]

By FRANK GRAY, J. W. HORTON and R. C. MATHES

SYNOPSIS: The design of a television system, once the fundamental principles are understood, involves a detailed consideration of the methods by which the several important functions are to be performed.

(1) In the present system the initial signal wave is obtained by sweeping a spot of light over the subject in parallel lines completely scanning it once every 18th of a second. The light reflected is collected by large photoelectric cells which control the transmitted current. At the receiving station the picture current controls the brightness of a neon lamp from which the received image is built up by means of a small aperture moving in synchronism with the spot of light at the transmitting station. For presentation to a large audience television images may be produced by a neon lamp in the form of a grid having a large number of separate electrodes. A high frequency excitation controlled by the picture current is distributed to the successive electrodes in synchronism with the spot of light at the transmitting station.

(2) Space and time variations in the reflecting power of the subject are translated into time variations in signal strength. For design purposes these time variations are represented by component frequencies, a minimum band of which must be properly transmitted to insure an adequate reproduction of the image. Within this band there must be maintained a certain degree of uniformity in the efficiency of transmission of the separate components. Also, their phases must not be permitted to shift unduly in relation to each other.

(3) The design of the terminal amplifiers is based on the quantitatively determined characteristics of the photoelectric cells and of the neon lamps as well as on the limits imposed by the transmission study and by the characteristics of available transmission media, whether telephone line or radio system. The circuits employed at the transmitting station furnish an amplification such that the power delivered to the transmission medium is 10^{15} times the power received from the photoelectric cells.

SECTION I. APPARATUS FOR THE ANALYSIS AND SYNTHESIS OF THE IMAGE

THE introductory paper to this series of articles on television explained principles along which any television system must operate to transmit an image over a single pair of wires or other channel of communication. As the first step in such a transmission, the space variations in brightness from point to point in the view must be translated into time variations in an electrical current that can be sent over the channel of communication. This translation may be accomplished by a scanning process that operates on the view to produce the same effect as if the view were cut up into a single long strip and passed rapidly in front of a light-sensitive cell to generate an electrical current varying with the brightness along the strip. To eliminate flicker in the reconstructed image and also to follow moving

[1] Presented at the Summer Convention of the A. I. E. E., Detroit, Mich., June 20–24, 1927.

subjects in a view, the scanning process must be repeated and a new picture transmitted at least every sixteenth of a second.

Many purely theoretical methods could be, and have been, devised to accomplish such a scanning process and to translate a view into electrical currents or signals. Unfortunately, however, a practical system of television must operate with materials and conditions as they exist, and these practical limitations constitute the serious problems of television.

The high speeds and relatively large amplitudes with which any television scanning mechanism must move, and the necessity for synchronizing the transmitting and receiving apparatus lead to the use of synchronously rotating machines as apparently the only practical solution of the scanning and receiving problems. Consequently, the present television system has been designed to operate with continuously rotating mechanical parts.

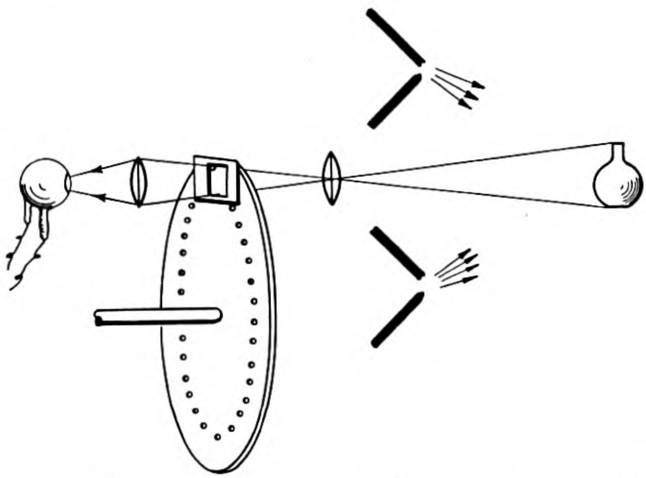


Fig. 1—Several light sources illuminate the subject; a lens forms an image which is scanned by a spiral of apertures, through which the light falls on a single photoelectric cell.

The efficiency that must be secured in the optical part of any scanning method is fixed by the three following factors—the amount of picture detail that is to be transmitted, the efficiency of the light-sensitive cell, and the practical limit to amplifier systems. The first of these factors decides the area from which light can be collected at any one instant. In the present case this was fixed in an initial survey of the entire television problem when it was decided to confine the first attempt to the transmission of pictures as if they were made up of 2500 small elemental areas; that is, to scan the view in a series of fifty parallel lines. The second factor is determined by the sensitivity of the potassium hydride photoelectric cell. This cell is, at the present time, the most efficient light-sensitive cell that can follow the rapid variations in light intensity without a time lag. The third factor, the limitation of amplifier systems, results from the extraneous currents that are present in metallic conductors and

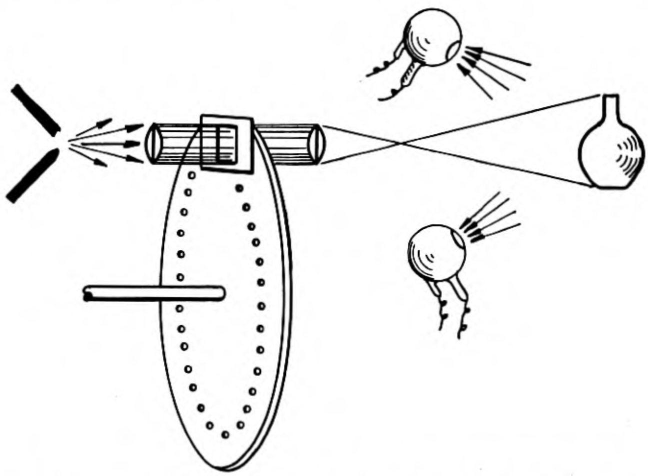


Fig. 2—Light from a single source is projected as a small moving spot on the subject; the reflected light is received by several photoelectric cells

amplifier tubes. The thermal agitation of the electrons in any input resistance generates such currents; and rapid variations in the number of electrons emitted from the hot filament of an amplifier tube also generate disturbing voltages. For successful amplification, the initial photoelectric current must be considerably larger than these extraneous currents. Consequently, the optical arrangement must be such that at any one instant it collects enough light from an elemental area of the view to generate this minimum permissible output current from the photoelectric cell.

The operation and advantages of the scanning method actually used in the present process for transmitting television images may be better understood by first considering a simple and analogous method illustrated by Fig. 1. The subject is illuminated by lights placed in front of it as shown. A lens forms an image of the subject on the rotating disk. This disk is pierced with a series of small holes or apertures arranged in the form of a spiral; and, as the disk rotates, the apertures trace across the image one after the other in a series of parallel lines. The frame limits the size of the image and prevents more than one aperture being in the image at one time. Light, passing through an aperture as it travels across the image, falls in the light-sensitive cell and generates a picture current proportional to the brightness of the image from point to point along strips taken one after the other across the image.

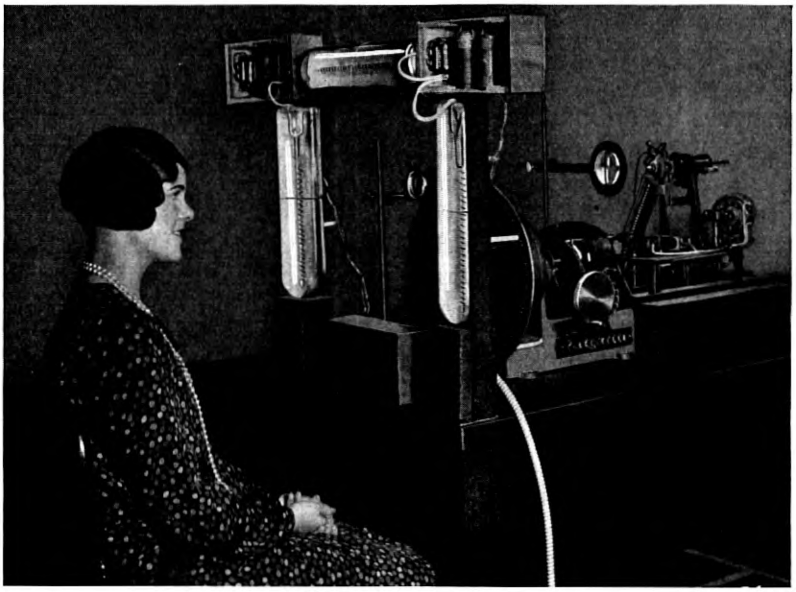


Fig. 3—Illustrative transmitting apparatus. Light from the arc lamp is condensed on the disk, which is driven by a high frequency synchronous motor. The disk carries a spiral of pin hole apertures, each of which in turn projects a moving spot of light on the subject. Reflected light is collected by three large photoelectric cells.

In any system such as that outlined above, which depends upon scanning an image of the view as formed by a lens, the efficiency of the system is ultimately limited, for any given size of image that can be scanned, by the ratio of aperture to focal length of the best lens that can be secured. Experiments show that, with the best lens available to form a one-inch-square image, it would be necessary to illuminate a subject with a 16,000-candle power arc at a distance of about four feet in order to secure an image bright enough for a photoelectric cell to give an output current above the noise level in an amplifier system. In other words, television would apparently be extremely inconvenient to the subject if it were to be carried out from an image formed by a lens.

In the system actually used for television transmission, this apparent limitation has been evaded by reversing the entire optical system of Fig. 1 and arranging it as shown diagrammatically in Fig. 2. Instead of scanning an image of the subject, the actual subject is scanned directly by a rapidly moving spot of light. An illustrative laboratory set-up, Fig. 3, shows the arrangement of parts in such a transmitting station. A fifteen-inch disk rotating approximately eighteen times per second carries a series of fifty small apertures arranged in the form of a spiral. A beam of light is condensed by a lens from a 40-ampere Sperry arc to intensely illuminate a limited area in the path of the moving apertures; and a slender, intense beam of light passes through each aperture as it moves across the illuminated area. A frame in front of the disk permits light to emerge from only one aperture at a time and the lens in front of the disk focuses an image of this moving aperture on the subject. As a result of this arrangement the subject is completely scanned in a series of successive, parallel lines by a rapidly moving spot of light, once for each revolution of the disk; and on account of the transient nature of the illumination the subject is scarcely aware that he is being exposed to it.

As the spot of light traces across the subject, light is diffusely reflected or scattered from the subject in all directions, and some of the light that is reflected forward passes into three large photoelectric cells placed just in front of the person who is being viewed. The current outputs from the three photoelectric cells operate in parallel into a common amplifier system. As the beam of light passes, for instance, across a person's eyebrow less light is reflected to the photoelectric cells, and as the beam passes across his forehead more light is reflected. Since the current output from the photoelectric cells is proportional to the received light, the current follows accurately the brightness of the various elemental areas of the subject's features as he is traced over by the scanning beam. This fluctuating current is unidirectional.

The actual operation of such an optical system, its influence on the lighting effects and quality of the reproduced image, may best be understood by noting that optically the system is identically the same as if all of the rays of light were reversed in direction to give an optical system equivalent to Fig. 1. The television apparatus sees the subject exactly as if rays of light came out of the photoelectric cells to illuminate the subject; the lens formed an image of the subject on the disk; and the apparatus scanned this image and reproduced it at the receiving end. The lights and shadows seen in the image are the same as if the subject were illuminated by three large lights in the positions of the photoelectric cells and looked at from the position of the lens. It also follows from the above considerations that, within its range of resolving power, this scanning method will not only reproduce a plane subject, such as a drawing, but that it will also faithfully reproduce three-dimensional figures with sharp edges and elevations and depressions, just as well as they could be reproduced in a photograph.

In addition, because the light passes in an approximately parallel beam through a disk aperture, the slender beams of light sweeping across the region in front of the transmitter just barely overlap each other even at a considerable distance from the apparatus. Consequently, it is not necessary that the subject be at the exact positions at which the small apertures are sharply focussed; and within wide limits no confusion results as the subject moves toward or away from the apparatus. The brightness as well as the size of the received image decreases as the subject moves away from the photoelectric cells; and for good transmission of the human features, which reflect very little blue light to which the photoelectric cells are sensitive, a person should not be more than a few feet away from the cells.

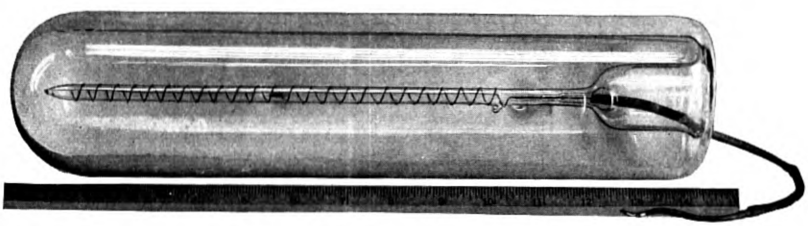


Fig. 4—Large photoelectric cell. The cell presents forty square inches of photosensitive surface to receive light reflected from a subject

This method of scanning permits two very large gains to be made in the amount of light available for producing photoelectric currents. The transient nature of the light permits a very intense illumination to be used without inconvenience to the subject. Furthermore, the optical efficiency of the system is not limited by the apertures of available lenses; but can be increased by using large photoelectric cells and more than one cell connected in parallel.

The photoelectric cells of the potassium hydride, gas-filled type used in the transmitting stations, were specially constructed for the purpose and are probably the largest photoelectric cells that have ever been made, Fig. 4. Three of these cells present an aperture of 120 square inches to collect the reflected light.

With this large collecting area and the strong light intensity that can be used for the transient illumination, the cells give an electrical output that, though still extremely small, is safely above the noise level of an amplifier system.

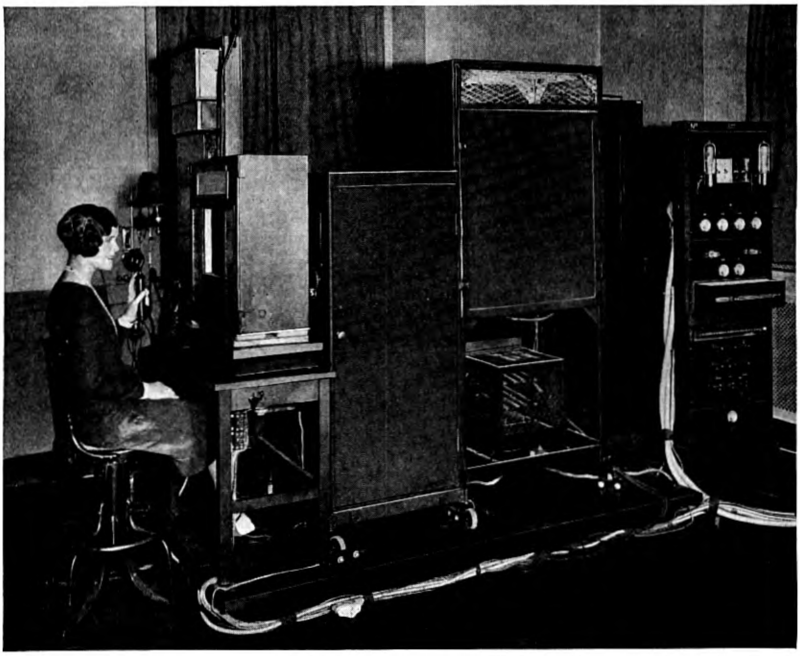


Fig. 5—Television transmitting apparatus. Sweeping beams of light pass out through the tunnel-like opening in the photoelectric cell case; light reflected from the subject is collected by three large photoelectric cells behind the screened openings.

A photograph, Fig. 5, shows the details of a television transmitting station as it is operated in the field. The arc, rotating disk and photoelectric cells are contained in separate cabinets and aligned as shown in the photographs. The three photoelectric cells and first stages of amplification are mounted in a shielded, sound-proof case. The slender, sweeping beam of light coming from the disk cabinet passes through the tunnel-like opening in the photoelectric cell case and scans the subject seated in front of it. The apparatus sees the person from light reflected back into the three large cells located just behind the screened openings in the case.

The variations of the feeble picture currents delivered from these photoelectric cells are highly amplified and transmitted over a wire or radio channel of communication by circuits described elsewhere in this series of articles. At the receiving station this current shape is reamplified, impressed on a direct current, and finally produces an image in the receiving apparatus.

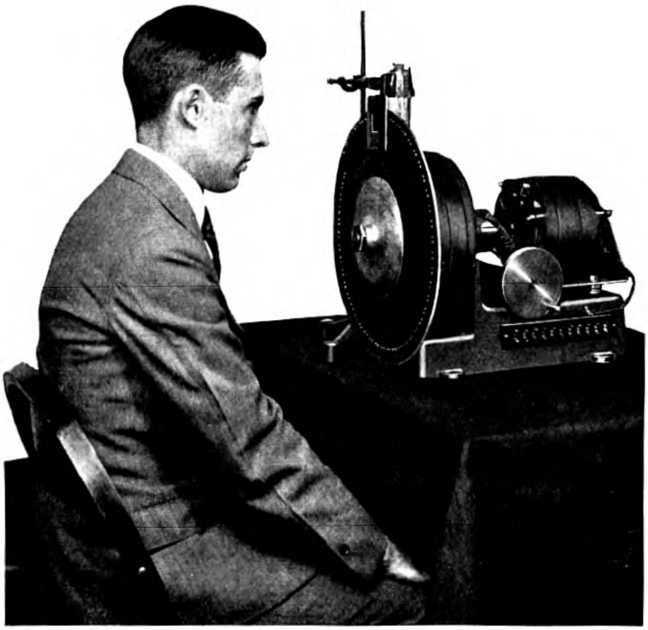


Fig. 6—Illustrative receiving apparatus. A neon lamp operated from the picture current illuminates a series of small apertures as they pass across the field of view; the observer sees an image reproduced in the frame.

A photograph, Fig. 6, shows an illustrative arrangement of the parts in one type of television receiver. An essential part of this type of receiver is a disk similar to the one at the transmitting station and also provided with fifty small apertures arranged in the form of a spiral. The driving motor rotates the disk in exact synchronism with the one at the transmitting station. The observer looks at a small rectangular opening or frame in front of the disk. This frame is of such dimensions that only one aperture can appear in the field of view at a time. As the disk rotates, the apertures pass across the frame one after the other in a series of parallel lines, each displaced a little from the preceding one until in one revolution of the disk the entire field has been covered. Beyond the disk is a special form of neon glow lamp shown in detail by Fig. 7. In this lamp, the cathode is a flat metal plate of a shape and area sufficient to entirely fill the field defined by the frame in front of the disk. The anode of the glow lamp is a similar metal plate separated from the cathode by only a very small space (about one millimeter). At the proper gas pressure this space between the plates is within the "cathode dark space" where no discharge can pass. As a consequence, the glow discharge develops on the outer surface of the cathode, where it shows as a perfectly uniform, thin, brightly glowing layer.

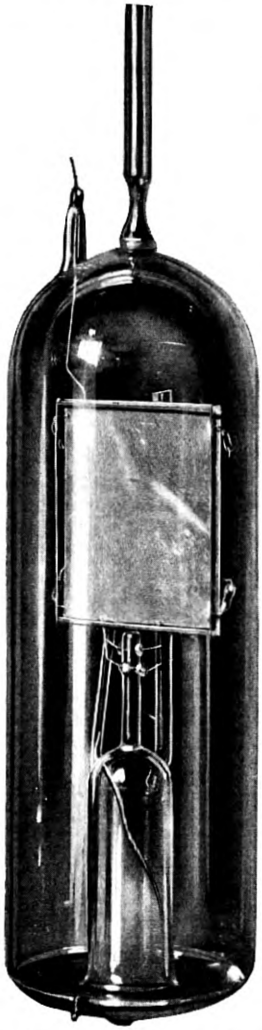


Fig. 7—Neon receiving lamp. The rectangular cathode is covered by a uniform layer of glow slightly larger than the field of view on a television disk

As an aperture in the disk moves across the field, the observer, looking through at the neon lamp behind the disk, sees the aperture as a bright point. When the disk is rotating at high speed, the observer, owing to the persistency of vision, sees a uniformly illuminated area in the frame, provided that a constant current is flowing through the lamp. (The line structure that would otherwise appear in the field is largely eliminated by using apertures that slightly overlap in their paths across the field.)

The brightness of the neon lamp is directly proportional to the current flowing through it; and when a picture is being received, the lamp is operated directly from the received picture current. As a result of the system just described, there is at any instant, in the field of view at the receiving station, a small aperture illuminated proportionally to the brightness of a corresponding spot on the distant subject. Consequently, the observer sees an image of the distant subject reproduced in the frame at the receiving station.

Fig. 8 shows the external appearance of the disk type of receiver in which the images appear. The disk rotates inside of a rectangular cabinet and the observer views the image through the shielding window. The largest disk, three feet in diameter, gives a 2 in. by $2\frac{1}{2}$ in. rectangular image. Each television receiver is also equipped with a telephone receiver and transmitter; and it is possible for the observer to both see and converse with a distant person at the same time.

Considering the limited number of picture elements, a surprising amount of detail can be transmitted with this television system. A distant person can be seen and easily recognized and his motions can be plainly followed as he talks into a transmitter, turns the pages of a magazine and goes through other similar motions. Large-sized pictures in a magazine can be seen as the subject turns the pages and looks at them himself.

Fig. 8—Disk receiving apparatus. The observer looks through the shielding window at a picture on the 36-inch disk

An auxiliary television receiving system also accompanies each transmitting set and enables the operator to see that he is sending a satisfactory picture current out over the channel of communication. This auxiliary or pilot picture is formed on the scanning disk itself. A small fraction of the outgoing picture current is tapped off and amplified to operate a neon lamp, which is placed behind the disk ninety degrees around from the scanning beam. An image of the subject may thus be seen on the scanning disk just as at a receiving station. To correct for the ninety-degree phase shift, the spiral of apertures on the transmitting disk is continued by additional apertures a quarter of a turn beyond the starting point. The first turn alone of the spiral is used for scanning; and the last turn alone, to form the pilot image; consequently, this image appears exactly in frame. A small mirror on the front of the motor cabinet reflects this image to the operator and enables him to see the character of the picture that he is sending out over the channel of communication.

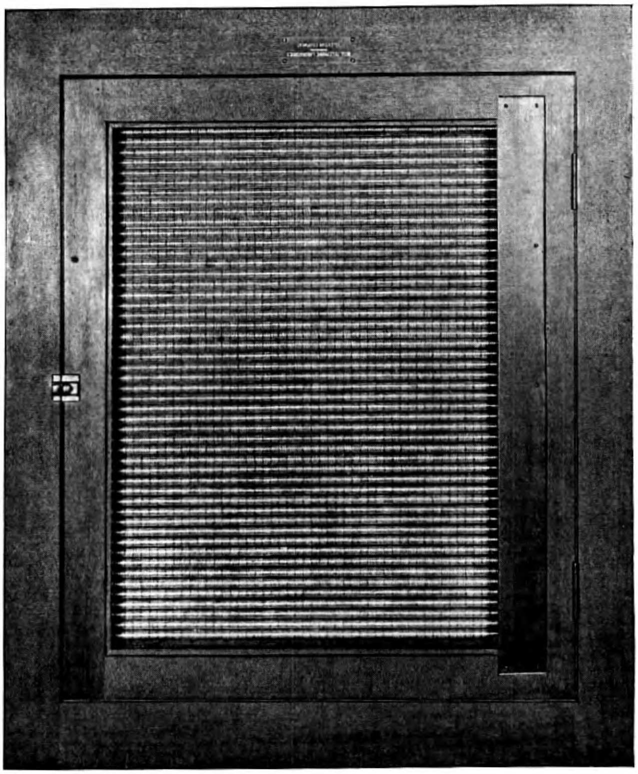


Fig. 9—Large grid. The large grid is a neon lamp with 2500 electrodes on a tube bent back and forth to form a luminous screen that is visible throughout a large auditorium.

When it is desirable to present television images to a large audience, a special grid type of receiver is used. The grid has the appearance of an illuminated screen and can be seen throughout a large auditorium. The image is not projected on the screen from a lantern like a moving picture; such optical projection would be inefficient and demand

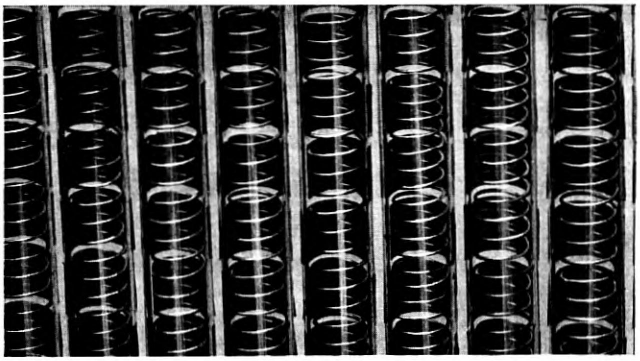


Fig. 10—Detailed structure of the grid. The exterior electrodes are pieces of metal foil cemented to the outside of the tube. The interior electrode is a long spiral of wire.

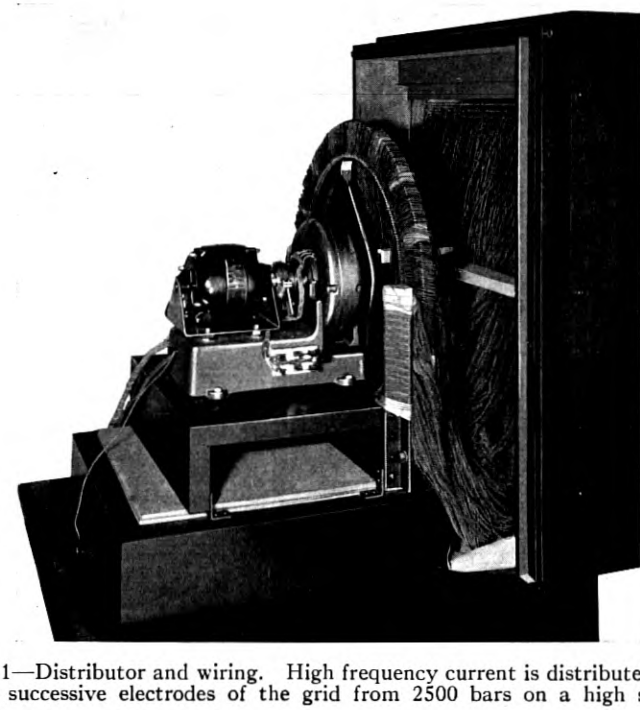


Fig. 11—Distributor and wiring. High frequency current is distributed by 2500 wires to successive electrodes of the grid from 2500 bars on a high speed distributor.

the electrical control of an impractical amount of light. The picture current itself is distributed by a commutator to successive elemental areas of a large neon lamp. This lamp, as shown in Fig. 9, consists of a single, long, neon-filled tube bent back and forth to give a series of fifty parallel sections of tubing. The tube has one interior electrode and 2500 exterior electrodes cemented along the back side of the glass tubing, Fig. 10. A high frequency voltage applied to the

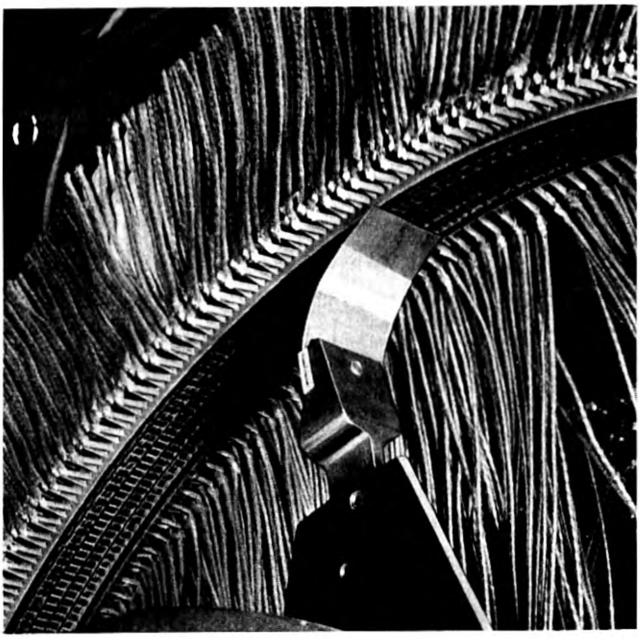


Fig. 12—Details of the distributor. The bars are arranged in four rows each displaced with respect to the other three. The sliding brush is a strip of thin sheet metal.

interior electrode and any one of the exterior electrodes will cause the tube to glow in front of that particular electrode. The glow discharge actually passes to the inside wall of the glass tubing and the high frequency current flows by a capacity effect out through the glass wall to the exterior electrode. The high frequency voltage is commutated to the electrodes in succession from 2500 bars on a distributor, Fig. 11, with a brush, Fig. 12, rotating synchronously with the disk at a transmitting station. Consequently, a spot of light moves rapidly and repeatedly across the grid in a series of parallel lines one after the other and in synchronism with the scanning beam at the transmitting station. With a constant exciting voltage the grid appears as a uniformly illuminated screen; but, when the high frequency voltage is modulated by the received picture current, an image of the distant subject is produced on the screen and his motions can be followed just as in the smaller images formed on a disk.

This method of presenting television images to a large audience permits a very efficient use of the available energy to reproduce a picture. The modulated current produces a glow discharge that exactly covers an elemental area of the picture on the screen and is viewed directly by the audience; consequently, there is absolutely no loss of energy after the picture current has been converted into light. In addition, each illuminated area of the screen responds to the picture current in the same manner; the exterior electrodes are exactly alike, and the use of a single tube assures the same pressure and purity of neon throughout the grid.

Fig. 9 shows such a screen set up for demonstration in an auditorium. A loud speaker is mounted just below the screen and it is thus possible for a large audience to both see and listen to a distant person at the same time.

Section II. The Television Signal Wave

So far it has been assumed that the electrical signal wave is perfectly transmitted between the conversion devices which transform the light variations into electrical variations and back again. Perfect transmission is, however, impossible with practical apparatus. There are certain requirements placed upon the generated signal wave by the characteristics of practical communication channels, and reciprocally certain demands are made upon a transmission system by the inherent nature of an adequate television signal. In addition to exploring these mutual requirements experimentally it is desirable to analyze them in such a way that, as far as possible, quantitative expression may be given to them. This expression in the case of the signal wave is best made by the methods of the Fourier analysis; considering the signal as made up of many sine wave components of various frequencies. The requirement on the signal may then be described in terms of these components and the requirements on the connecting transmission system in terms of attenuation and phase characteristics over a band of frequencies. These requirements will now be discussed as a basis for the subject matter of the succeeding section of this paper and of the following companion papers of this group on "Wire Transmission Systems for Television" and "Radio Transmission Systems for Television."

The problems to be discussed may be conveniently considered under three headings:

(*a*) The Character of the Television Signal.

(*b*) Requirements upon the Signal Wave Set by the Characteristics of Available Transmission Channels.

(*c*) Requirements which the Transmission Channels must meet in order to carry Television Signals.

(*a*) *The Character of the Television Signal.* As we have seen, the voltage produced across a resistance in series with the photoelectric cell is a fluctuating unidirectional potential. The generated signal therefore has frequency components beginning at and including zero frequency. The value of the voltage at any instant is roughly proportional to the average reflected illumination at that instant from an illuminated spot whose size depends upon the apertures in the scanning disk. At any point where there is a sudden change in the tone value of the subject there will also be a sharp change in the generated voltage. It will, therefore, be seen that but for the limits of speed of action of the photoelectric cell and its connected circuits the generated signals would tend to include components over the whole frequency range up to infinity. Since it is possible to effectively transmit but a limited range of these components, the width and location of the frequency band necessary for the acceptable reproduction of a given size and structure of image must be determined. It is convenient to consider first the low frequency end of the band.

In the early experimental work it was soon found that in attempting to amplify the lower frequencies by the use of direct current amplifiers, unstable conditions of operation were reached before sufficient amplification was obtained to operate the receiving apparatus. Experiments were then made with resistance-condenser coupled amplifiers which showed that, if the efficiency of such an amplifier at the frequency equal to the number of pictures sent per second was not more than about two T U below its average efficiency for the transmitted range, acceptable reproduction of the picture was secured together with stable operation of the amplifiers. When the low frequency cut-off of the amplifier was set much above this, spurious shadows were introduced into the picture. That there will be a critical lower frequency for the transmission of an unchanging scene is obvious since the Fourier series into which the signal may be analyzed starts with a constant term and the sine wave terms begin with the picture frequency and include a vast number of its harmonics. If the constant component (d-c.) is removed, the lowest frequency which remains to be transmitted is therefore the picture frequency.

The effect of removing the d-c. component of the signal can be qualitatively traced in a simple manner. Imagine three types of still pictures or scenes to be transmitted by the system. Let the first be quite dark in general effect and require fluctuations in the signal current of a certain average amount for its transmission. Such a picture would have a low direct current component. Let the second picture consist largely of medium grays and require about the same fluctuations in signal intensity for its delineation. Such a picture will have a medium direct current component. Let the third picture be very light in general effect with such difference in light and shadow as would require the same fluctuations in signal intensity as the other two pictures. Such a picture would have a relatively high direct current component. In passing through a resistance-condenser coupled amplifier, the signals for all three types of pictures would be changed from fluctuations superimposed upon direct current to alternating currents, all of about the same average value.

At the receiving end of the circuit the direct current component may be reinserted by superimposing the alternating current fluctuations upon a fixed value of direct current such as the steady state current in the last amplifier tube. This direct current component would give the best average results if it corresponded to that suitable for the gray picture, which would, of course, then be most nearly correctly reproduced. However, most of the detail of the dark and light scenes would also be reproduced though the tone values would be distributed about a medium gray. Fortunately a change in character of this kind has proven for the most part unimportant. Where it is important it can be taken care of very simply by providing, at the receiving end, means, either manual or automatic, for changing, in accordance with the type of scene being transmitted, the magnitude of the unidirectional current upon which the received alternating current is superposed, which amounts simply to the restoration of the direct current.

In the case of scenes which are changing, however, frequencies lower than picture frequency will in general be generated and their suppression may be expected to affect to some degree the perfection of the picture. In effect, these frequencies are analogous to changes in tone values in the case of still pictures and their elimination results in fluctuations in the apparent brightness of the image. This effect is not disturbing with many types of subjects, as for example in the reproduction of the face.

One remarkable result of not transmitting the direct current component of the signal in the case of the reflected beam method of scanning is that the television transmitting apparatus can be located and operated in a well-lighted room, for if this general illumination is constant it simply increases the direct current component of the signal. Similarly if the scene itself contains a source of steady light, this will be visible only in so far as it reflects the scanning beam.

Turning now to the upper part of the frequency range, experimental data on the highest necessary components were obtained by the use of circuits with low pass instead of high pass characteristics. With the television terminal apparatus operating at 17.7 pictures per second, it was found that a filter whose phase distortion had been corrected over practically all of its pass band of 15,000 cycles produced a degradation in image quality which was just detectable when the human countenance was being transmitted. Since the electrical terminal apparatus without the filters would efficiently transmit frequencies higher than this, the experiment showed either that frequencies higher than this were not present in the generated signal, that they were not effectively reproduced, or that they contribute little to the appearance of the image. This upper limit to the useful frequency range for this apparatus is rather lower than was anticipated from the initial survey, but because of psychological factors (decreased discrimination of tone values for fine details, apparent improved resolution when the subject is moving, etc.) it proves satisfactory for television purposes.

It is of importance, however, to know where the limitation in frequency range occurs in the apparatus and how it might be modified. Considerable information on this point is obtained by studying the nature of the distortion introduced by the aperture in the optical system and that introduced by frequency limitation in the electrical part of the system. It is convenient to consider them together as the type of distortion turns out to be similar for the two cases. This distortion may be considered most simply in relation to the type of signal corresponding to a sudden unit change in tone value at some point in the subject. With an ideal

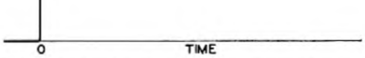


Fig. 13—Elementary signal change

television system in which the instantaneous values of signal current are at all times proportional to the tone values of the points being scanned, the resulting signal would be represented by the graph of Fig. 13. Such a consideration involves no real loss in generality as any signal shape may be considered as the result of infinitesimal abrupt changes in intensity.

It is readily seen that if a square aperture passes with uniform velocity over a part of the picture having an abrupt change from dark to light the result is that we get a signal from the photoelectric cell which, instead of building up instantaneously, builds up linearly during a time, T, Fig. 14, which is the time required for the aperture to pass a given point.[1] The net effect is an apparent sluggishness in the response of the system. The dotted curve of Fig. 14 shows the integrated illumination passing through a circular aperture of a diameter corresponding to the same time, T, for the condition of Fig. 13. Due to the simpler analysis the discussion will be carried out in terms of the square aperture though the sluggishness due to the round one is seen to be slightly less.

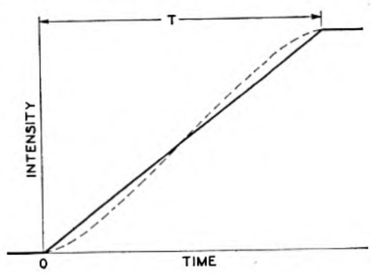


Fig. 14—Elementary signal change as distorted by a square aperture

Now this kind of sluggishness in response is quite similar to that introduced in the electrical part of the system when the upper frequencies are cut out or not transmitted as efficiently as the lower ones. The effect of frequency limitation can be investigated theoretically in a fairly simple fashion if we make the ideal assumption that all frequencies are transmitted without distortion up to a cut-off frequency, f_c, and extinguished beyond it. In Appendix I, it is shown how the signal of Fig. 14 is affected by a frequency limitation of this type. We can then plot a set of curves as shown on Fig. 15

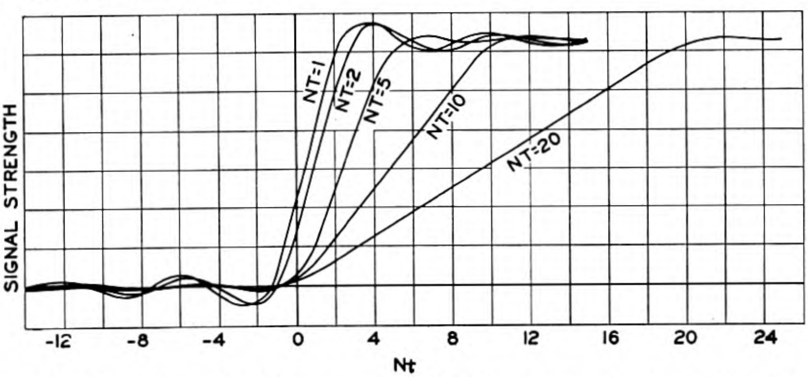


Fig. 15—Elementary signal change as distorted by a square aperture and by ideal frequency restriction

[1] This effect of aperture distortion was pointed out in the paper "Transmission of Pictures over Telephone Lines" by Ives, Horton, Parker and Clark, *B. S. T. J.*, April, 1925.

from which we can measure the total time of rise due to both the aperture and frequency limitation. The abscissa is the product of $N = 2\pi f_c$ and the time, t. Any one curve serves for a wide range of values of N and T as long as their product is the same. Call the new time of rise τ. Then we can plot a relation as on Fig. 16 between

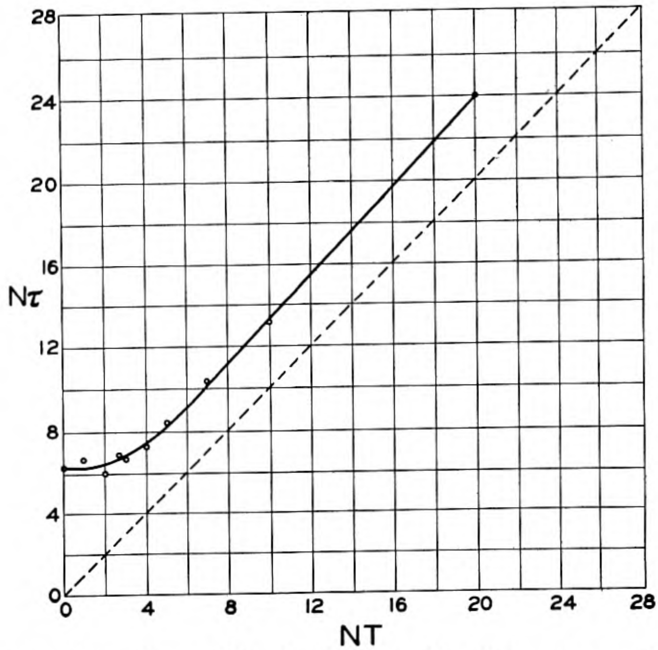


Fig. 16—Sluggishness due to distortion as a function of the aperture width and frequency restriction

$N\tau$ and NT from which we can draw conclusions as to the relative effects of aperture and frequency distortion.

Below the knee of this curve we have approximately

$$N\tau = 2\pi$$

$$\tau = \frac{1}{f_c}$$

and the frequency cut-off determines the whole distortion. Similarly above the knee

$$N\tau = NT + \pi$$

$$\tau = T + \frac{1}{2f_c}$$

and the controlling influence is that of the aperture.

Unless one effect is much more easily remedied than the other, the knee of the curve appears a reasonable point to select for operation. At the knee $NT_k = 2\pi f_c T_k = \pi$ and $T_k = 1/2f_c$. At this point the total lag is not much greater than that due to the frequency restriction alone and is $1/f_c$ or twice T_k. That is, at this point, the additional lag in the time of rise of signal due to the restricted frequency range is equal to that due originally to the aperture, though the additional lag due to the aperture is not much greater than that due to the frequency restriction alone. For a square aperture in a square picture of 2500 elements sent 16 times a second $T = 1/40,000$ of a second, and $f_c = 20,000$ cycles at the knee of the curve. The point on the curve where the effect of frequency restriction introduces a sluggishness in following light changes comparable to that introduced by a square aperture is the same frequency as that arrived at as the upper limit to useful frequencies by considerations from still picture transmission, in the introductory paper by Mr. Ives. Its value is equal to one half the number of picture elements.

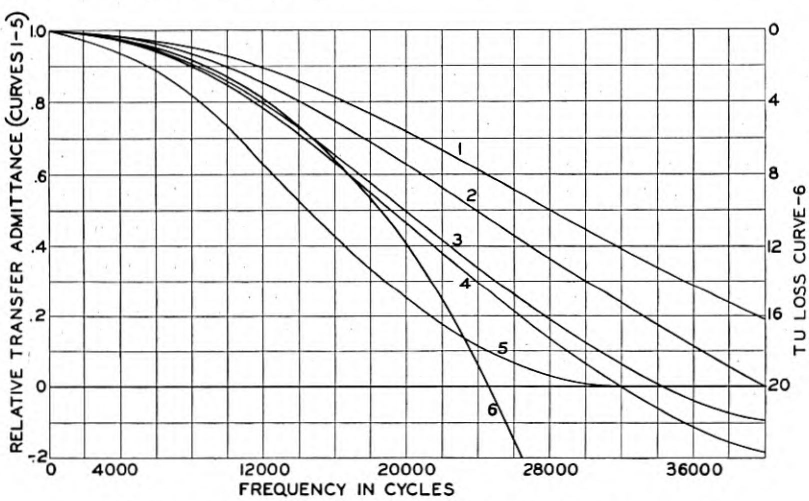


Fig. 17—Equivalent transfer admittance of various apertures

It has furthermore been found possible to determine ideal electrical transmission characteristics or equivalent transfer admittances of circuits which produce exactly the same distortions as various types of apertures. While it appears impossible at present to construct a physical circuit which will produce such characteristics over the whole frequency range, the problem is not difficult if we limit ourselves to the most important frequency band. This is of interest as it points out the possibility of compensating for the effect of the aperture by putting in an electrical network with frequency transmission characteristics the inverse of those so determined. Within the range of important frequencies it turns out that the effect of the aperture is the same as that of a network which changes merely the relative amplitudes of the frequencies into which the picture signal may be analyzed. Neglecting constant multiplying factors, the relative variation over the frequency range for a square aperture is given by the factor $\frac{\sin T\omega/2}{\omega}$ and for a round aperture by $\frac{J_1(T\omega/2)}{\omega}$, where, as before, T is the maximum time for the aperture to pass a given point and J_1 is the Bessel's function of the first order. The derivation of these factors is given in Appendix II. On Fig. 17, Curve 1 gives the relative values of the equivalent transfer admittance for the square aperture and Curve 2 for an inscribed circular aperture, both in case of a 50-line scanned picture which is square and sent 16 times per second. T then is equal to 1/40,000 sec.

In the system as set up for demonstration the image is rectangular with the vertical and horizontal dimensions in about the ratio 5 to 4. The circular aperture is about 1¼ times 1/50 of the vertical height and the scanning is done 17.7 times a second. T is then 3.53×10^{-5} seconds and Curve 3 gives the corresponding frequency characteristics. Curve 4 shows that a square aperture of the same area as the circular aperture for Curve 3 gives a fairly good approximation to Curve 3. Curve 5 gives the combined effect of the two circular apertures, sending and receiving, corresponding to Curve 3. Curve 6 is Curve 5 plotted in terms of T U on the right hand scale.

An inspection of this last curve indicates that this frequency attenuation characteristic of the aperture introduces a considerable loss at 15,000 cycles and leaves little of the signal components above 20,000 cycles. To see if an electrical circuit of characteristics inverse to those of the aperture would materially improve the resolution of the image, the circuit,[1] which, together with its frequency characteristics, is shown in Fig. 18, was inserted between the sending and receiving amplifiers. It was designed to compensate for most of the aperture distortion and its phase distortion was made small below 20,000 cycles. On the fan-shaped test pattern of Fig. 19 a noticeable improvement was observed, the black and white angles being resolved closer to the tip of the pattern. In the case of faces the improvement appeared to be very little but could be detected

[1] This is a constant resistance type of corrective network or equalizer. See Chap. XVIII, "Transmission Circuits for Telephonic Communication," K. S. Johnson.

in the slightly better definition of sharp narrow lines such as the frames of horn-rimmed spectacles. When a system of considerable attenuation is employed between the sending and receiving terminals, it would in general be preferable to split the equalizing between the sending and receiving ends to make the best use of the sending end power in riding over interference.

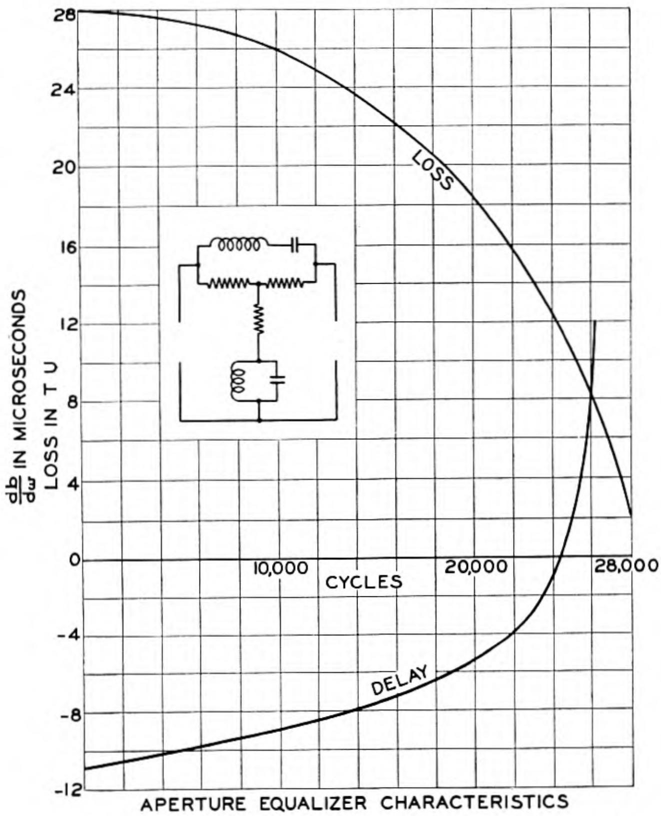


Fig. 18—Circuit for equalizing the aperture effect and its amplitude and phase characteristics

In arriving at the amount of electrical equalization which shall be adopted in any particular case it must of course be borne in mind that as the aperture is made narrower the amount of distortion introduced by it becomes less. As we narrow the aperture, however, the available illumination becomes less and the signal generated by the photoelectric cell becomes smaller. A limit is therefore soon reached at which the difficulties of amplification become greater than the difficulties of equalization and a minimum practical aperture width is thereby determined. If the distortion is corrected by narrowing the aperture, it is apparent that the apparatus will generate, at but little lower than the correct relative efficiency, frequencies much higher than those thought necessary from the more general considerations in Mr. Ives' introductory paper. Decision as to the desirable frequency transmission band for the connecting communication channel would be no different for either method of reducing the distortion due to the aperture.

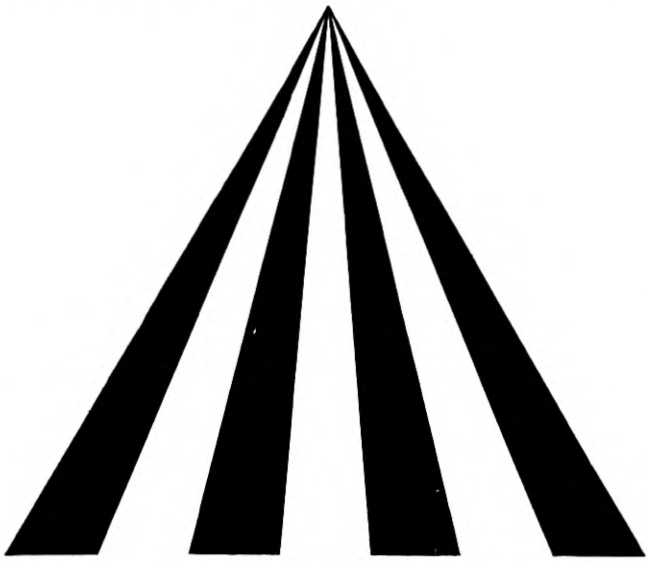


Fig. 19

In summary, then, we may say that experiment and theory show that the lowest frequency essential to satisfactory results is the picture frequency, and the highest frequency required is approximately one half the number of picture elements scanned per second.

(*b*) *Requirements upon the Signal Wave Set by the Characteristics of Available Transmission Channels.* The limitations upon the signal wave set by present available communication channels are:

1. The magnitude of the signal necessary to override the interference to which such channels are subject.

2. The frequency range which such channels can transmit.

The first of these is self-explanatory. It determines the required amplification and load capacity of the transmitting apparatus. In the companion paper on *Wire Transmission Systems for Television* are the data on interference and on permissible signal to noise ratio which were used in the design of the terminal transmitting amplifiers to be described in the latter part of this paper.

In considering the frequency range of lines, it was apparent in the beginning that the wire channel might include sections of cable. With existing loading systems for such cables a frequency range of not over 40,000 cycles appeared available. The terminal apparatus was therefore designed to deliver a generated signal whose essential components lay well within this limit, and the laboratory tests mentioned in the preceding section showed that this requirement was met.

A lower frequency limit was imposed by the necessity of a transformer for joining the transmission line to the terminal equipment. Fortunately it proved possible to design transformers as described in the final part of this paper in which this limit was at or below the essential low frequency limit found in the preceding discussion of the signal wave.

(*c*) *Requirements Which Transmission Channels Must Meet in Order to Carry Television Signals.* We have shown that a certain band width of frequency components is essential to the adequate reproduction of the image. This sets the frequency limits of the transmission channel which must be provided. It is essential, however, that within these transmission limits the channel should present a reasonably uniform attenuation, and that the phase relations should be fairly accurately maintained. The problem as presented to the transmission engineers of wire, radio and terminal equipment for the recent demonstration was to meet the following requirements:

First, transmission must be provided for frequencies between about 10 cycles and 20,000 cycles.

Second, the amplitude frequency characteristics within this range should be uniform to about $\pm$ 2 T U.

Third, the phase shift through the range should be maintained so that the slope of its characteristic as a function of frequency is constant to $\pm$ 10 or 20 micro-seconds over all but the lowest part of the frequency range. There, about 50 times this limit was considered the maximum permissible.

These requirements were arrived at by considerations based on theory and experiments on television and analogy to similar requirements in telephotography. The first requirement follows directly from the discussion of the essential frequencies in the signal. The following paragraphs are intended to illustrate the significance of the remaining requirements.

As we have as yet no quantitative measure of the goodness of reproduction of the image, the matter of the second and third transmission requirements on received amplitude and phase characteristics over the frequency scale is one which had to be decided largely on the basis of the experimental results and judgment based on general considerations. We have already seen that the removal of the very lowest frequencies simply changes the tone value of the whole picture. It may be similarly reasoned that departures from the average efficiency of transmission in the lower part of the frequency range would result in the appearance of diffuse shadows or high lights. Likewise, it may be concluded that broad deviations from the average efficiency of transmission in the uppermost part of the signal frequency range would result in the accentuation or the fading out of the finer detail of the scene. Steep slopes in the amplitude-frequency curve would result in the superposition of oscillations upon signals representing sudden changes in intensity. To reduce these effects every reasonable effort was made to keep the variations in the amplitude characteristic with frequency as slight as possible, aiming to hold these characteristics for the separate parts of the demonstration system to within $\pm$ 2 T U or better.

In addition to transmitting the component frequencies with the same relative efficiency as regards amplitude, it is also particularly essential in television to send them through the system with small relative phase shifts; that is, with constant velocity or what is equivalent, a phase shift proportional to frequency. It has long been known in optical theory that the envelope of a group of waves of nearly the same wave-length and nearly the same frequency may travel along with a "group velocity" somewhat different from the phase velocities of the component elements. If the system has but small departures from a flat amplitude-frequency characteristic and from a linear phase shift frequency characteristic, it can be shown that the time of group transmission or "envelope delay" is given by $db/d\omega^2$, the slope of the curve obtained by plotting the phase shift, b, for the system, against the angular velocity, $\omega = 2\pi f$. The time of transmission of a crest for any sine wave component of frequency $\omega/2\pi$ is, of course, given by b/ω. If $b = c\omega$, $b/\omega = c$ and $db/d\omega = c$. Then the phase and envelope times of transmission are equal and all frequencies as well as their group envelopes get over in the same time. If b is given in radians, $db/d\omega$ is given in seconds. In general a knowledge of b as a function of ω is necessary and sufficient to determine the phase distortion. A knowledge of $db/d\omega$ as a function of ω is not sufficient to determine all factors in signal distortion. It is, however, often easier to measure with the needed accuracy and in transmission systems such as have been used for still pictures and television has proven a useful index of phase characteristics.

After a preliminary estimate from experience with still pictures that the limit on $db/d\omega$ should be $\pm$ 10 microseconds, an electrical network consisting of five sections of a simple lattice structure was used for testing the effect of phase distortion with television apparatus. This network introduced negligible amplitude distortion and a drift in the value of $db/d\omega$ of 50 microseconds over the frequency range of 0 to 20,000 cycles. Its effect was perceptible in blurring the image of a face and it decidedly affected a sharp pattern of two parallel lines of such width and spacing as to be just within the resolving power of the apparatus. This variation of $db/d\omega$ was about $2\frac{1}{2}$ times greater than that postulated. Hence $\pm$ 10 microseconds was agreed on as a desirable limit for $db/d\omega$, though it was felt that this limit might be exceeded by a factor of two in restricted parts of the frequency band.

When this network was combined with a filter the slope of whose envelope delay curve was in the opposite direction so that over the greater part of the frequency range the combined delay of the two

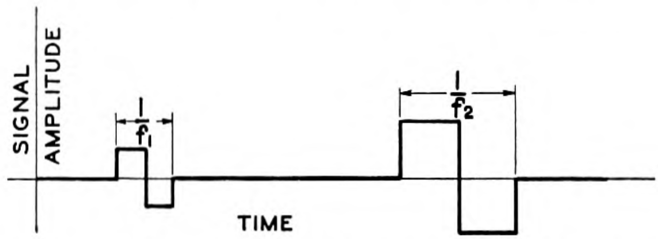


Fig. 20—Signal details of concentrated frequency spectrum for illustrating the effect of envelope delay

circuits was constant and equal to 140 microseconds, this time delay effect was very graphically brought out. Every time the combined circuit was cut in, the undistorted received image jumped to a new position a little over 10 per cent of the width of the picture to one side in the direction of scanning.

To see why $db/d\omega$ should be maintained at a constant value, consider two sharply defined details near together in the picture which would produce a variation in signal intensity with time as indicated in Fig. 20. Imagine each to be cyclically continued so that the small detail defines a frequency f_1 and the other defines a frequency f_2. It is then known from Fourier analysis that the frequency spectra of the two details are chiefly concentrated around the frequencies f_1 and f_2. If $db/d\omega$ is appreciably different at the frequencies f_1 and f_2 for any part of the system, the two details will be displaced relatively to each other along the line of scanning and, in most cases, if this shift is appreciable, some change in the shape of the signal wave defining each detail results with further increase in the distortion. The same relative shift would occur if the narrow detail were located upon the broader one, in which case such a shift would be more apparent. It would seem reasonable to expect then that differences in the envelope time of transmission comparable to a whole picture element (about 28 microseconds in the demonstration apparatus) would be noticeable.

In most images very few details will have signal shapes, as in this special case, in which the frequency components are concentrated in narrow frequency bands. An abrupt change in signal strength, for instance, is represented by components distributed over the whole frequency range. We can imagine these frequencies divided into any arbitrary number of groups, each of which determines a wave form. When these wave forms are added together, they will reproduce the original abrupt change in signal strength. If, however, they are sent through a system in which the envelope delays for the different groups are unequal, the individual wave forms will be relatively displaced and will no longer combine correctly. As a result the image is blurred. For some types of phase distortion the effect appears as an oscillatory transient following sudden changes in intensity.

It was furthermore found by experiment that the limit of $\pm$ 10 microseconds was not necessary for the lower frequencies. Reference to the delay characteristics of the transformers described in the latter part of this paper shows that in the lower part of the frequency scale deviations from the nearly uniform value of delay at the upper frequencies appear of magnitude greater than 100 microseconds. When the signal was sent through these transformers, however, there was no observable distortion of the image. The requirements are therefore much more lenient at the low frequencies.

In the terminal apparatus the problem of meeting the above outlined phase transmission requirements was not a very serious one. The circuits involved are such that when a flat amplitude-frequency characteristic had been secured the phase distortion was also negligible.

Section III. Terminal Circuits for Sending and Receiving Television Signals

The preceding sections have discussed the methods by which an object, the image of which is to be transmitted, is made to control the time variations in a light, thus giving a luminous signal wave, and the means by which the image may be reconstructed with the aid of an electric signal wave corresponding to this initial luminous wave in its relative instantaneous amplitudes. Certain important relations between the characteristics of the signal wave and the resulting image have been pointed out. There remains the question of obtaining an electric signal wave suitable for long distance transmission and of providing for the control of the illumination at the receiving terminal by the electric signal wave as delivered by the transmission medium.

In the use of wire lines for television it is fortunately true that a suitably prepared open-wire circuit possesses a frequency range sufficient for the transmission of all the essential components of the signal wave. Details regarding the characteristics of the wire circuits are given in a companion paper by Messrs. Gannett and Green, from whose work are obtained data essential to the design of the terminal equipment. These data fix the power level at which the signal should be delivered to the line and the power level which will be available at the receiving end. When the transmission is by radio it is, of course, necessary to effect a frequency translation in order to secure a wave suitable for radiation and transmission through the ether. In this case, however, the radio system, which is described in a paper by Mr. E. L. Nelson, when considered as a whole may be conveniently taken as a system capable of the transmission of a signal wave occupying the same frequency range as that supplied to the wire circuits. In fact the design of the radio system is such that it may be used interchangeably with the wire line in so far as the remaining electrical terminal equipment is concerned.

The terminal circuits, then, fall into two groups: first, those used at the transmitting terminal for building up the wave controlled by the time variations in light to the power level required by the line; and second, those used at the receiving terminal to bring the wave delivered by the line to the proper form for controlling the luminous sources from which the received picture is built up.

Transmitting Circuits

Starting with the photoelectric cell in which the initial luminous signal wave is converted to an electric signal wave, we are interested in the magnitude of various pertinent constants. The cell may be considered for our purposes as an impedance, the value of which is determined by the quantity of light reaching it. With no illumination at all this impedance is almost entirely a capacitance of the order of 10 m.m.f. When the cell is illuminated this capacitance becomes effectively shunted by a very small conductance which is roughly proportional to the square of the voltage between the electrodes.

For a fixed potential the magnitude of this conductance is nearly a linear function of the illumination. With a suitable potential in series with the cell, then, there is obtained a current the amplitude of which is proportional to the quantity of light reaching the cell.

In order to connect the photoelectric cell to the amplifier, there is introduced in series with the cell and its polarizing battery a pure resistance the voltage drop across which is used to control the grid potential of the first tube. It is desirable, of course, to make this resistance high in order to have available as much voltage as possible. Its value is, however, limited by two considerations. The added series conductance must not be so low that it appreciably disturbs the linear relation between the illumination and the total conductance of the circuit. The voltage drop must also be so small, in comparison with the total potential in the circuit, that the photoelectric cell operates at an approximately constant polarizing potential.

In view of the extremely small voltage of the electric signal wave as delivered by the photoelectric cell circuit, it is essential that great care be taken to prevent such interference as may enter the initial amplifier stages from approaching a comparable magnitude. The most troublesome sources of interference are electrostatic induction, electromagnetic induction, mechanical vibration, and acoustic vibration. By mechanical vibration is meant disturbances transmitted through the supports as the result of building vibrations and similar phenomena. By acoustic vibrations are meant impulses transmitted through the air which strike the several elements of the amplifier and cause motion which results in variations in their electrical constants. Electrical disturbances are reduced to a minimum by placing the amplifier as close as possible to the photoelectric cells, thereby keeping the leads short, which avoids electrostatic pick-up and also prevents the formation of closed loops of any appreciable size, thus avoiding electromagnetic induction. The amplifier is provided with a very complete electrical shield and both the shielded amplifier and the photoelectric cells are placed in a carefully shielded cabinet.

The tubes used, namely, the so-called "peanut" tubes, are, under ordinary conditions, remarkably free from any microphonic action. At the very low signal levels used, however, certain extra precautions have to be taken against this effect. In addition to lining the amplifier box with sound-absorbing material, the tubes themselves have been wrapped in felt and placed within a heavy lead case. This prevents such acoustic disturbances as reach the interior of the amplifier container from having any noticeable effect on the tube. The lead container is supported entirely by an elastic suspension and thus serves a dual function, as the heavy mass, supported in this way, is capable of little response to such mechanical vibrations as may be transmitted through the cabinet and the walls of the amplifier shield. With these precautions it has been found possible to make the effect of all external disturbances of about the magnitude of the thermal disturbances referred to in the first part of the paper.

A schematic diagram of the amplifier tubes directly associated with the photoelectric cell is given in Fig. 21. Attention has already been

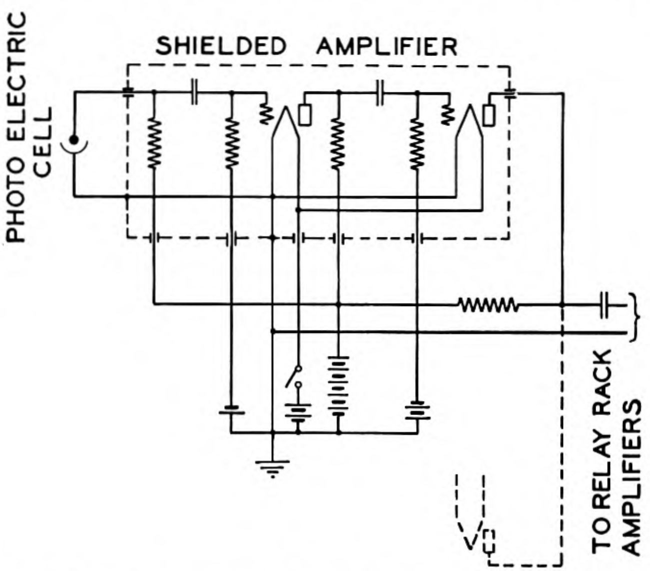


Fig. 21—Schematic of vacuum tube amplifier used with photoelectric cells

called to the fact that the initial signal, that is, the time variation of the light reflected from the scanned object, contains a direct current component. The amplification of this direct current component is, as has been stated, out of the question in any amplifier intended for continued operation over long periods of time. The requirements as to the range of frequencies to be transmitted, as discussed in the preceding section, make it necessary to provide a circuit having practically uniform efficiency from 10 cycles to above 20 kilocycles. The relative phase shift of the several components must also be kept very small. In view of the large amplification and consequent large number of stages necessary, it has been thought impracticable to use transformer coupling between all stages as the aggregate frequency and phase distortion might well be greater than could be tolerated. The so-called resistance capacitance coupling has therefore been used.

The arrangement of the several photoelectric cells in their cabinet, as shown in Fig. 3, is such that one amplifier can be connected directly to two of the cells leaving the third to operate a second amplifier. The outputs of these two amplifiers are then connected in parallel to the common battery supply equipment shown at the bottom of the two vertical cells.

By the use of two stages of amplification in the photoelectric cell amplifier, the signal is brought to such a level that it may be carried by suitably shielded leads to other amplifiers outside the photoelectric cell cabinet. This permits of using the convenient relay rack form of mounting. The signal level is, however, still low and may be adequately handled in amplifier units which differ but little from those used with the photoelectric cell.

The remaining requirements placed on the amplifiers at the transmitting terminal are those set by the telephone line. One of primary importance is that which determines the amount of energy needed. In order that the signal wave shall be of such magnitude that any interference present in the line may be negligible in comparison, it is desired that the alternating current delivered by the final amplifier stage shall be at least 4 milliamperes into an impedance of 600 ohms. The energy to be supplied is, therefore, approximately 0.01 watt, which determines the choice of the last amplifier stage. To build up the signal to a value sufficient to operate this output tube it has been found that eight stages of the small-sized tubes and one stage of greater load-carrying capacity must be used. The total amplification given by these ten stages is approximately 130 T U. It is through this known gain of the amplifiers that we get our only accurate quantitative data as to the magnitude of the initial signal wave. This comes out to be about 10^{-15} watts or, with a 100,000-ohm resistance in series with the photoelectric cell, the potential available at the first tube is roughly 10 microvolts.

The characteristics of the line also determine the means by which it shall be coupled to the final amplifier stage. In order to secure the proper impedance matching and to prevent the line from being unbalanced with respect to ground, it was felt desirable to use transformers if possible rather than to attempt the design of a tube circuit capable of meeting the requirements directly. The problem included both output and input transformers, and specified an amplitude-frequency characteristic constant to within $\pm$ 0.5 T U from 10 cycles to 25,000 cycles. The input coils intended for use at the receiving terminal had the additional requirement that a minimum of interference current should be induced in the secondary due to potentials between the line and ground. The success with which this problem has been solved is shown by the curves of Fig. 22. Curve 1 is the

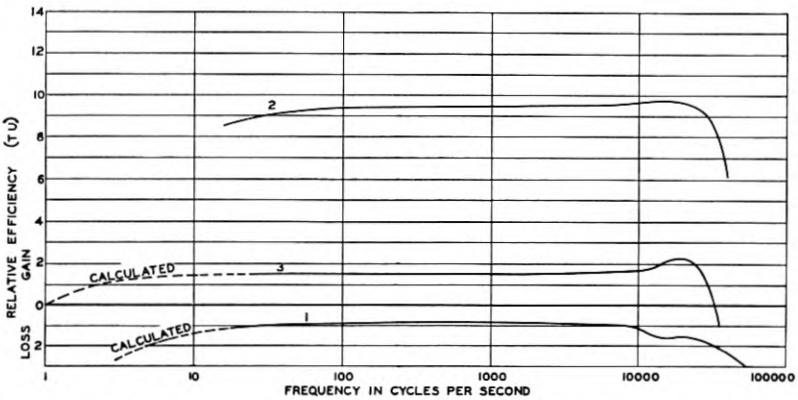


Fig. 22—Transmission characteristics of iron core transformers

1. Output transformer connected between impedances of 2000 ohms and 600 ohms.
2. Input transformer having voltage step-up of 6.5 connected between 600-ohm line and vacuum tube.
3. Input transformer having voltage step-up of 2.5 connected between 600-ohm line and vacuum tube.

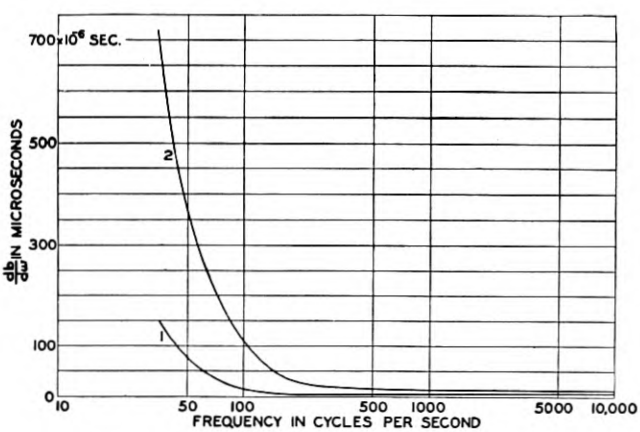


Fig. 23—Envelope delay characteristics of transformers

1. Output transformer
2. High ratio input transformer

transmission characteristic of the output transformer which is designed to work between impedances of 2000 ohms and 600 ohms when connected between generator and load circuits having these values.

Curves 2 and 3 show the effective transmission gain of transformers having voltage step-ups of 6.5 and 2.5 respectively, when used to connect the first stage of the vacuum tube amplifier to a 600-ohm generator impedance. The envelope delay curves for the output transformer and for the high ratio input transformer are given in Fig. 23. Photographs of the coils are given in Fig. 24. A large

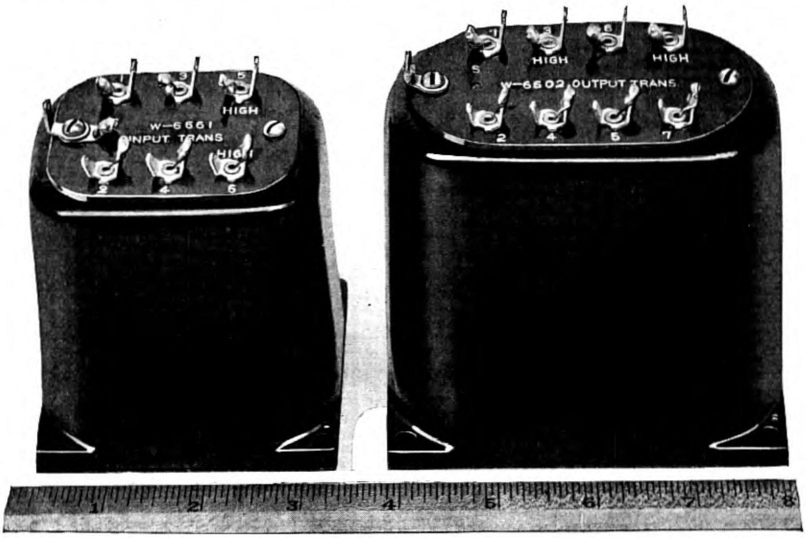


Fig. 24—Transformers used for coupling amplifier circuits to long distance telephone lines

factor in being able to get coils of this type lay in the availability of permalloy for the core material. The output transformer is connected to the amplifier through a blocking condenser in order to avoid possible saturation in the core due to the passage of direct current.

Measurements made on the several elements of the amplifier system have shown that its overall frequency characteristic is constant to within $\pm$ 2 T U from 10 to 20,000 cycles.

In an amplifier having as much gain as that just described it is apparent that a slight change in the potential of the power supply will cause a considerable change in the overall efficiency. Moreover, variations in the intensity of the light source used with the scanning system will cause corresponding changes in the intensity of the initial luminous signal wave. To insure that the energy level supplied to the line is at all times of the proper magnitude a level indicator has been provided to permit continuous observations of the output of the amplifier. This consists of an amplifier-rectifier circuit so arranged that the space current of the last tube is a function of the alternating current voltage impressed on the first, being roughly proportional to the square of its amplitude. By means of a direct current milliammeter, therefore, it is possible to keep a very accurate check on the amplitude of the signal delivered to the line.

Receiving Circuit

Coming now to the receiving terminal equipment we find that the signal wave which was delivered to the line at a power level of 10 milliwatts may, under some conditions, be reduced to a level 50 T U below this, or to 0.1 microwatt. It is, therefore, necessary first of all to provide amplification to bring the signal to a level where it may operate the circuits controlling the illumination from which the image is to be reconstructed. In view of the fact that several types of receiving equipment are to be operated and also since the signal may be derived from any of several sources, either wire line, radio or local transmitting station, it is desirable to fix some one energy level as a reference point and to bring all signals to this value so that they may be supplied interchangeably to the several receiving systems. A convenient reference level is that already set as the proper input to a telephone line, namely, 10 milliwatts. At the receiving terminal, therefore, amplifiers have been provided which are similar to the final stages used at the transmitting terminal. These include units containing the small-sized tubes and terminate in units identical with that supplying current to the line except that the output transformer is omitted. The first stage is, as mentioned in the preceding section, connected to the line through an input transformer. The amplifiers associated with the several incoming signals are each provided with a level indicator of the type already described. These terminal amplifiers and the several receiving circuits are all terminated in jacks, exactly like telephone circuits, and it is possible, therefore, to connect any receiving machine to any desired transmitting station simply by patching the proper jacks together, exactly as telephone circuits are connected at the central office.

Before describing the final stages of the amplifier circuits it is necessary first to examine the properties of the light source which is to be controlled. In the case of the disk receiving machines described in the first section of this paper it is recalled that a single neon lamp is used having a rectangular electrode the entire area of which glows at each instant with an intensity proportional to the intensity of the initial luminous signal. The current voltage characteristic of a typical neon lamp is given in Fig. 25. It will be seen that no current flows until the voltage across the lamp reaches the breakdown potential which, in the example shown, is about 210 volts. From this point on the current increases linearly with respect to voltages in excess of a value somewhat below the breakdown point. It will also be seen

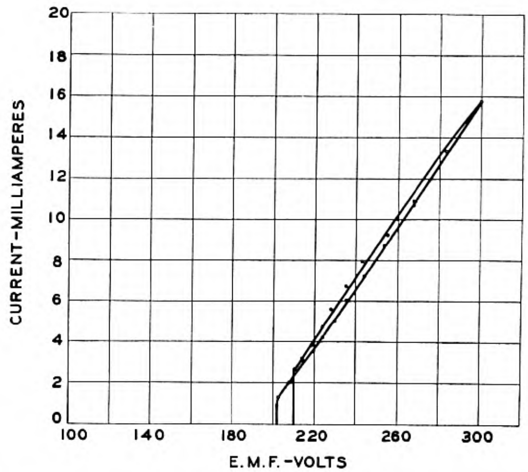
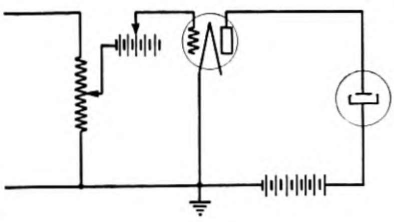
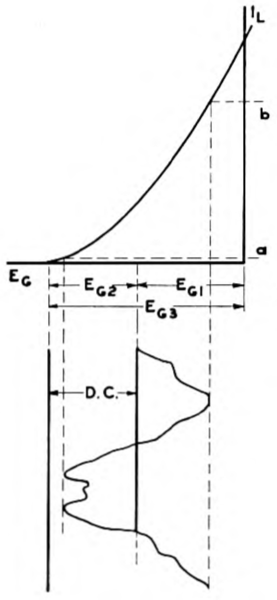


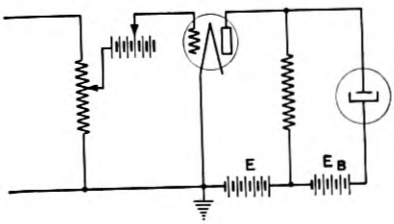
Fig. 25—Current voltage characteristic of typical neon lamp

from the curve that the value of current depends somewhat upon the direction in which the voltage is changing. In most cases, however, the function comes sufficiently close to being single valued for our present purposes. In view of the well-established linear correspondence between the intensity of the illumination resulting from the glow discharge and the current, it is required to so arrange the circuits that the current through the lamp is at all times proportional to the illumination at the transmitting terminal.

It will be recalled that the electric signal wave as transmitted through the various amplifier circuits differs fundamentally from the initial luminous wave in that the direct current component has been eliminated. It is necessary, therefore, to restore this component before the changes in light intensity at the receiving terminal will follow those at the transmitting terminal. The several factors entering at this point may perhaps best be examined in terms of an elementary circuit such as given in Fig. 26. In this case the neon lamp is connected in series with the plate circuit of a vacuum tube and its polarizing battery. The circuit may be considered for the present as equivalent to one in which the neon tube is replaced by an ohmic resistance and in which the potential of the polarizing battery is







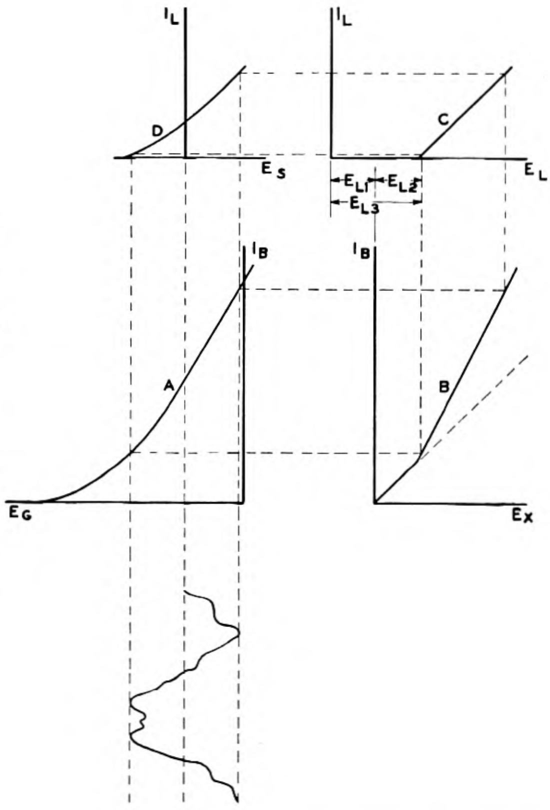


Fig. 26—Circuit schematic and operating characteristic of neon lamp amplifier

Fig. 27—Circuit schematic and operating characteristics of circuit arranged for linear operation of neon lamp

reduced by an amount corresponding to the back e.m.f. of the lamp. Under these conditions the relation between current—and therefore illumination—and the voltage on the grid of the vacuum tube is as shown by the curve given with the figure. This curve takes into account the change in potential between the plate and filament of the vacuum tube due to the voltage drop in the lamp resistance. If the reactances in the circuit are negligible, this curve may be taken as the dynamic characteristic of this portion of the system.

Let us assume that to properly build up the desired image at the receiving terminal the light is to be varied between the limits set by the two horizontal lines *a* and *b*. It is apparent that two adjustments are necessary in the grid circuit. The amplitude of the impressed alternating current must be such that the difference between its positive and negative maxima is equal to the difference between the grid voltages corresponding to these currents. This is taken care of by suitable adjustments of the amplification. It is further necessary that the bias introduced by the grid polarizing battery be such that the positive and negative peaks coincide with these same values of grid voltage. Under these conditions the grid battery must be looked upon as supplying two absolutely distinct biases, one the bias for the tube and the other the bias for the signal. For example, if the signal wave as delivered to the grid circuit contained the original d-c. component properly amplified, it would be necessary to adjust the system so that zero current would be obtained with no impressed signal. To accomplish this the tube would require the negative grid bias E_{G3}. Variations in signal voltage would then be considered as taking place about this value of grid potential as the origin. Thus E_{G3} is the operating bias of the tube. To properly locate the signal wave, however, it is necessary to add the positive bias E_{G2}. It will be seen from the curve that this bias corresponds exactly to the direct current component which is to be restored to the signal. The sum of these two biases, obviously, gives the actual bias, E_{G1}, with which the tube is operated.

In the circuit as shown the well-known curvature of the vacuum tube prevents us from obtaining a linear relation between the current through the neon lamp and the signal voltage. This condition may be overcome by a number of circuit modifications of which that shown in Fig. 27 is typical. Instead of connecting the neon lamp and the vacuum tube directly in series, a resistance is provided across which is set up a potential, E_X, proportional to the current through it. Across this resistance is shunted the neon lamp and a biasing battery, E_B. The adjustment of this circuit is indicated by the curves shown.

Curve A expresses the relation between the grid potential of the vacuum tube and its plate current. Curve B shows the relation between this same plate current and the voltage across the external resistance. When no current is flowing through the vacuum tube, the potential of the biasing battery is insufficient to break down the neon lamp and no current flows through the circuit containing the neon lamp and the plate circuit resistance. As the current through the vacuum tube is increased from zero, the total current flowing is that through the resistance branch. When, however, the potential drop across this resistance reaches such a magnitude that, together with the potential of the biasing battery, it is sufficient to break down the neon lamp, the latter will begin to draw current which thereafter increases linearly with further increases in the voltage, E_X, across the external resistance. The voltage across the neon lamp itself differs from that across the resistance by the amount of the battery E_B. The relation between the neon lamp current and the voltage across it, as given by Curve C, may therefore be plotted directly above the characteristic just discussed by displacing the vertical axis an amount corresponding to E_B. This amount is shown as E_{L1}. Here again we have two separate biases controlled by a single adjustment. The potential E_{L2} is fixed by the minimum plate current which can be taken from the tube without departing too seriously from the linear portion of the tube characteristic. It is, therefore, an operating bias of the circuit which is unaffected by any characteristic of the neon lamp. The latter, however, must be operated with a bias E_{L3} corresponding to its effective back e.m.f. As in the case of the grid circuit bias just considered, the bias E_{L1} actually introduced into the circuit is the difference between these two independently determined biases.

By projecting values of lamp current horizontally and plotting their intersections with vertical projections through the corresponding grid potentials on the vacuum tube characteristic we obtain Curve D, which expresses the relation between the instantaneous value of the signal and of the current in the neon lamp as derived from the characteristics of the several elements of the circuit. Inasmuch as the intensity of the illumination is proportional to the lamp current, it will be seen that we have approached the desired linear correspondence between the instantaneous values of the signal and of the light.

It will be noted that care has to be exercised to insure that the alternating current as impressed on the last vacuum tube is of the proper polarity. If it is not, the received image will be a negative instead of a positive. This may be controlled either by the connections to any one of the transformers or by the number of vacuum tube stages. With an even number of stages the polarity will be reversed from that given by an odd number. This is because an increase in negative potential on the grid of a vacuum tube causes a decrease in the space current and hence a decrease in the negative potential applied to the grid of the next tube.

In the case of the grid type of lamp with the individual external electrodes, the impedance to which energy must be supplied differs materially from that presented by the rectangular electrode lamp already described. For low voltages the impedance between any electrode and the central helix is effectively a capacitance of the order of 6 m.m.f. When, however, the voltage gradient in the interior of the tube becomes sufficient to break down the gas and cause a discharge to take place, the capacitance is increased to about 15 m.m.f. In fact, the tube may be looked upon as consisting of two capacitances connected in series. When the applied potential is sufficient to break down the gas and cause a glow discharge, that capacitance corresponding to the portion of the path inside the tube is effectively shunted by an ohmic resistance. The minimum discharge potential has been found to be independent of frequency over a wide range, but the current between electrodes is inversely proportional to the frequency because of the presence of the capacitance between the electrode and the glowing gas. Now, the brightness of the discharge is a function of the current sustaining it so that it becomes desirable to use high frequencies in order to get sufficient light without going to prohibitively high potentials. It is also desirable to operate at such a portion of the frequency scale that the percentage difference between the limits of the range shall be small, thus avoiding signal distortion due to the effect referred to above. There is, however, a definite upper limit to the frequency beyond which it would be impossible to operate because of the stray capacitances in the cable connecting the grid to the distributor. It has been found feasible to operate at a frequency of the order of a half million cycles.

The circuit problem, therefore, involves the production of a high frequency wave which varies in amplitude in accordance with the amplitude of the received picture signal. The solution has been conveniently obtained by using a radio broadcast transmitter the voice frequency circuits of which have been so modified that the extended range of frequencies required might be handled with minimum distortion.

The envelope of the 500-kilocycle wave modulated by the picture signal, as shown in Fig. 28, is proportional to the signal amplitude plus a direct current biasing component of such magnitude that when the envelope reaches 160 volts the tube fails to light. This corresponds to a black area in the picture. When no picture signal is being received, the amplitude of the unmodulated carrier wave causes the tube to light at average brightness, corresponding to the locally introduced d-c. component of the signal. It follows, then, that the amplitude of

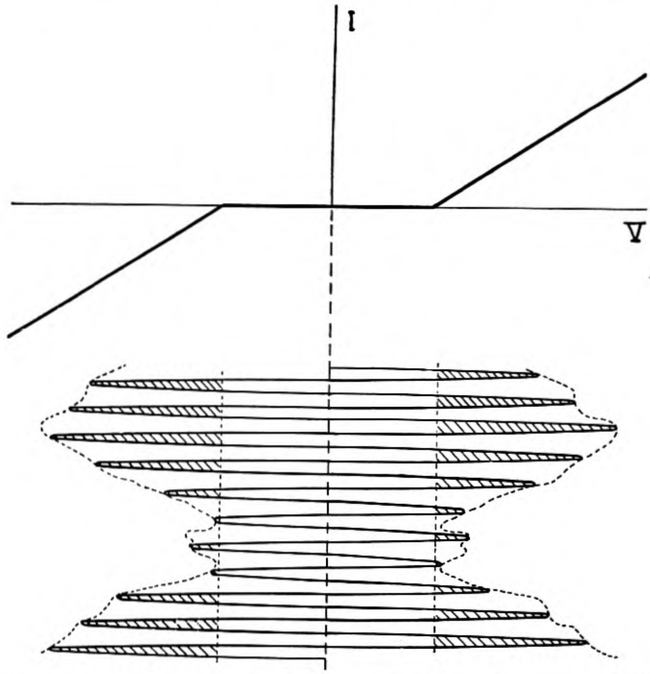


Fig. 28—Diagrammatic representation of relation between modulated high frequency wave impressed on grid type neon lamp and lamp characteristics. Intensity of glow is proportional to shaded area.

the unmodulated carrier is fixed, as in a previous example, by the joint requirements of two biases, that of the lamp and that of the signal bias.

There is a slight distortion inherent in this method due to the fact that the light, which is proportional to the shaded area of the curve of Fig. 28, is not strictly proportional to the amplitude of the envelope with respect to the 160-volt limit. This is, of course, because these peaks are portions of a sine wave and hence the time variation of the glow resulting from any given carrier cycle is a function of its amplitude. The effect is small, however, being most noticeable at low values of illumination.

In the case of the grid-lamp receiver the signal amplitude is adjusted, as for the disk receiver, by a potentiometer in the low frequency portion of the circuit. The carrier amplitude, however, is adjusted by varying the plate potential applied to the oscillating tube. The coupling to the lamp is made by connecting the central helix and the distributor brush across a portion of the condenser of the oscillating circuit.

The frequency-amplitude relation of the envelope has been made practically constant by employing resistance capacitance coupling in the signal input amplifiers, by providing extremely high inductance retard coils for the modulator—which is of the Heising type—and by inserting resistance in the oscillating circuit to provide sufficient damping. The relations between the original picture signal and the envelope of the high frequency wave, with respect to both amplitude and phase shift, were observed over the signal frequency range by means of a Braun tube and found to be satisfactory. The impedance of the connecting leads to the commutator was also measured and found to have a negligible effect on the frequency and damping of the oscillating circuit.

It has been found that there may be a lag between the time when the potential is applied to an electrode and the time when the gas breaks down. This is especially true following an interval during which there has been no discharge within the tube. Because of this those electrodes which are the first to be connected in any one of the parallel portions of the tube may fail to light. To overcome this effect a small pilot electrode is kept glowing at the left-hand end of each tube, thus irradiating the branch in such a way that the illumination of all electrodes follows immediately upon the application of potential. These pilot electrodes, which are obscured from view of the audience by the frame of the grid, are supplied by means of an auxiliary connection to the oscillator with a potential somewhat lower than that ordinarily impressed upon the picture segments.

Appendix I

The signal of Fig. 13 in the body of the paper may be represented as follows:

$$\left.\begin{aligned} f(t) &= 0 \qquad \text{for } t < 0 \\ &= \frac{t}{T} \qquad \text{for } 0 < t < T \\ &= 1 \qquad \text{for } t > T \end{aligned}\right\} \tag{1}$$

or by a Fourier integral in the form

$$f(t) = \frac{1}{\pi}\int_0^\infty d\omega \int_\infty^\infty f(\lambda)\cos\omega(t-\lambda)d\lambda, \tag{2}$$

where λ is an auxiliary variable of integration and ω is 2π times the frequency. To get the effect of sending this signal through a system which transmits all frequencies without phase or amplitude distortion up to a cut-off frequency f_c it is only necessary to replace the upper limit of the first integral sign by N where $N = 2\pi f_c$. Thus:

$$F(t) = \frac{1}{\pi}\int_0^N d\omega \int_{-\infty}^\infty f(\lambda)\cos\omega(t-\lambda)d\lambda.$$

Then from (1):

$$F(t) = \frac{1}{\pi}\int_0^N d\omega\int_0^T \frac{\lambda}{T}\cos\omega(t-\lambda)d\lambda + \frac{1}{\pi}\int_0^N d\omega\int_T^\infty \cos\omega(t-\lambda)d\lambda$$

$$= \frac{1}{\pi NT}\Big\{\cos Nt - \cos N(t-T)$$

$$+ Nt[Si(Nt) - Si(Nt-NT)]\Big\} + \frac{1}{\pi}\left[\pi + Si(Nt-NT)\right].$$

If we write $Nt = x$, $NT = z$, and $\pi F(t) = y(x)$, then

$$y(x) = \frac{1}{z}\Big\{\cos x - \cos(x-z) + x[Si(x) - Si(x-z)]\Big\}$$

$$+ \frac{\pi}{2} + Si(x-z),$$

where

$$Si(x) = \int_0^x \frac{\sin x}{x}dx.$$

A series of graphs of $y(x)$ for different values of the product NT is given in Fig. 15 in the body of the paper. These are generalized curves, the time scale depending on the particular value of cut-off frequency used. From these curves we can get the additional lag in the time, τ, in the rise of these curves over the original time T in Fig. 14.

Appendix II

Let $f(t)$ be the instantaneous intensity of the picture, and let it be represented by a Fourier integral:

$$f(t) = \int_0^\infty A(\omega)\cos[t\omega + \Phi(\omega)]d\omega. \tag{1}$$

Let T = time required for the aperture to pass a given point, Fig. 29.

Let $\varphi(t_1)$ be height of aperture at distance t_1 from its center.

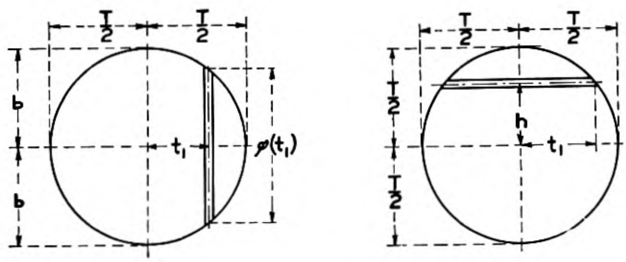


Fig. 29 Fig. 30

Analysis of the aperture

The instantaneous amount of light passing through the aperture is

$$\begin{aligned} F(t) &= \int_{t-T/2}^{t+T/2} \varphi(t_1) f(t_1) dt_1 \\ &= \int_{t-T/2}^{t+T/2} \varphi(t_1) dt_1 \int_0^\infty A(\omega) \cos [t_1\omega + \Phi(\omega)] d\omega \\ &= \int_0^\infty A(\omega) d\omega \int_{t-T/2}^{t+T/2} \varphi(t_1) \cos [t_1\omega + \Phi(\omega)] dt_1. \end{aligned} \tag{2}$$

In the case of the rectangular aperture

$$\varphi(t_1) = \text{a constant} \tag{3}$$

and, except for a negligible constant factor,

$$\begin{aligned} F(t) &= \int_0^\infty A(\omega) d\omega \int_{t-T/2}^{t+T/2} \cos [t_1\omega + \Phi(\omega)] dt_1 \\ &= \int_0^\infty A(\omega) \left\{ \frac{\sin [(t + T/2)\omega + \Phi(\omega)]}{\omega} - \frac{\sin [(t - T/2)\omega + \Phi(\omega)]}{\omega} \right\} d\omega \\ &= 2 \int_0^\infty A(\omega) \frac{\sin T\omega/2}{\omega} \cos [t\omega + \Phi(\omega)] d\omega. \end{aligned} \tag{4}$$

The transformation from $f(t)$ to $F(t)$ amounts merely to changing the relative amplitude of the Fourier components of $f(t)$ by a factor proportional to $\dfrac{\sin T\omega/2}{\omega}$.

In the case of the circular aperture we can divide the aperture up into narrow elements parallel to the direction of motion, as shown in Fig. 30. Elements at a distance h from the middle line of the strip have lengths

$$2t_1 = 2\sqrt{T^2/4 - h^2}. \tag{5}$$

Each element considered as an independent *rectangular* aperture has the frequency characteristic

$$\frac{\sin t_1\omega}{\omega} = \frac{\sin \omega\sqrt{T^2/4 - h^2}}{\omega}.$$

The mean of all of these elementary frequency characteristics is

$$\begin{aligned}\frac{1}{T}\int_{-T/2}^{T/2}\frac{1}{\omega}\sin\left[\omega\sqrt{T^2/4 - h^2}\right]dh &= \frac{2}{T\omega}\int_0^{T/2}\sin\left[\omega\sqrt{T^2/4 - n^2}\right]dh \\ &= \frac{1}{\omega}\int_0^{T/2}\sin\left[\omega T/2\sqrt{1 - \frac{4h^2}{T^2}}\right]\frac{2dh}{T} \\ &= \frac{1}{\omega}\int_0^1 \sin\left[T\omega/2\sqrt{1 - x^2}\right]dx \\ &= \frac{\pi}{2\omega}J_1(T\omega/2),\end{aligned} \tag{6}$$

where J_1 indicates a Bessel function of the first order. In place of the amplitude variation function $\frac{\sin (T\omega/2)}{\omega}$ for the square aperture, we have $\frac{J_1(T\omega/2)}{\omega}$ as such a factor. From the very nature of the physical processes under consideration it follows that this average value of the elementary frequency characteristics is effectively the frequency characteristic of the aperture as a whole.

Synchronization of Television[1]

By H. M. STOLLER and E. R. MORTON

SYNOPSIS: Synchronization of Television is the problem of holding two scanning disks so that their phase displacement is always less than four and one third minutes of arc. A 240-pole synchronous motor of the variable reluctance type is used as a basis. Coupled to it a direct current motor carries the steady component of the load. Hunting is eliminated by a condenser in series with the two synchronous motors whose capacitance is slightly less than that required to tune the circuit.

As the motor might lock into step in any of 120 possible angular positions, only one of which would give the proper phase relations, a two-pole motor, with only one locking position, was provided by tapping the armature of the direct current motor at two points and bringing out the leads to slip rings. This was used for synchronizing while the 240-pole motor, connected subsequently, held the close synchronism required. The disks rotate at 1062.5 r.p.m. which gives 17.7 cycles on the two-pole and 2125 cycles on the 240-pole motor.

For transmission the synchronizing current is attenuated to a level of .6 milliwatt and amplified at the receiving end. The 17.7-cycle current is an undesirably low frequency for transmission over telephone cables and so is used to modulate a 760-cycle current through a polarized relay. This is demodulated at the receiving end, where a polarized relay by interrupting a local battery current gives a rectangular wave which acts through vacuum tubes on the field of the direct-current motor.

THE problem of synchronization involved in television transmitting and receiving equipment is similar in principle to any synchronous motor problem but the requirements are of such a special nature that it is necessary to employ unusual features of motor design and control circuits to secure the required results.

GENERAL REQUIREMENTS

At the transmitting end a scanning disk is employed containing 50 holes spirally spaced around the periphery of the disk rotating at a speed of 1060 r.p.m.[2] It is desired to rotate a similar scanning disk at the receiving end so that the hole through which the observer is looking at a neon lamp will be in a position corresponding to the hole which is transmitting light at the same instant at the transmitting end. Since there are 50 holes in each disk, the holes will be spaced apart 7.2 degrees, thus 7.2 degrees of arc correspond at the receiving end to the width of the picture. Since the horizontal resolving power is approximately the same as the vertical (0.02 of the picture dimension), the arc occupied by a picture element is 0.02×7.2 or 0.144 degree. In order not to appreciably impair the quality of the picture,

[1] Presented at the Summer Convention of the A. I. E. E., Detroit, Mich., June 20–24, 1927.

[2] This speed was determined by transmission considerations and is discussed in the companion paper by Messrs. Gannett and Green.

it is necessary to hold the synchronization within approximately ½ of the width of one element. This gives 0.144 degree divided by 2 or 0.07 degree as the requirement within which synchronization should be held. By way of comparison it might be mentioned that the angular twist in a length of 6 ft. of 1-in. steel shafting operated at rated load is of about the same order of magnitude.

An ordinary four-pole synchronous motor when operating at full load, unity power factor, has an angular phase displacement of about 20 electrical degrees between the impressed and back e.m.f. This corresponds to 10 mechanical degrees since the motor has two pairs of poles. If this motor is operated at constant load and the line voltage is varied, the phase angle will decrease with increasing voltage, or when the voltage is held constant and the load is varied the phase angle will increase with increasing load. It is at once apparent therefore that the ordinary type of synchronous motor will not even approach the degree of precision required for the reason that any minute change in line voltage or load will cause variations in its phase angle of lag with respect to the impressed frequency of a far greater amount than 0.07 degree. Consider, however, a motor having 120 pairs of poles. Allowing 20 electrical degrees as the normal full load phase displacement, this would be equivalent to 20 divided by 120 or 1/6 degree mechanical phase displacement. Even this amount is over twice the required permissible displacement of 0.07 degree. Since the variation of the phase displacement is the important factor and not the absolute amount of displacement, it is evident that if the line voltage and load are held reasonably constant a synchronous motor with 120 pairs of poles should be sufficiently precise.

Another requirement in addition to close phase synchronization is regulation of the acceleration or deceleration of the generator at the transmitting end. Such regulation is required due to the fact that an appreciable time is taken for the transmission of the synchronizing current a distance of 220 miles (circuit length) between New York and Washington. The velocity of propagation over the cable was approximately 19,000 miles per second while that of the picture on the open wire of 285 miles circuit length was about 175,000 miles per second, the corresponding times of transmission being .0116 second and .0016 second, leaving a difference of .01 second approximately. Since the total permissible error in synchronization is .07 degree, it is reasonable to allow .02 degree as error due to acceleration regulation. Let a be the acceleration in degrees per second per second. Substituting in the formula $s = \frac{1}{2}at^2$ gives $.02 = \frac{1}{2}a(.01)^2$ or $a = 400$ degrees per second per second or a little over one revolution per second per second. For comparison consider a ¼-h.p. unregulated shunt motor. If the line voltage increases 10 per cent, it will cause an increase in speed from 4 per cent to 8 per cent depending on the magnetic saturation in its field circuit. This increase in speed will take place in a half second or more depending upon the moment of inertia of the load. Thus the acceleration in the case of a 1060-r.p.m. speed would be much greater than one revolution per second per second.

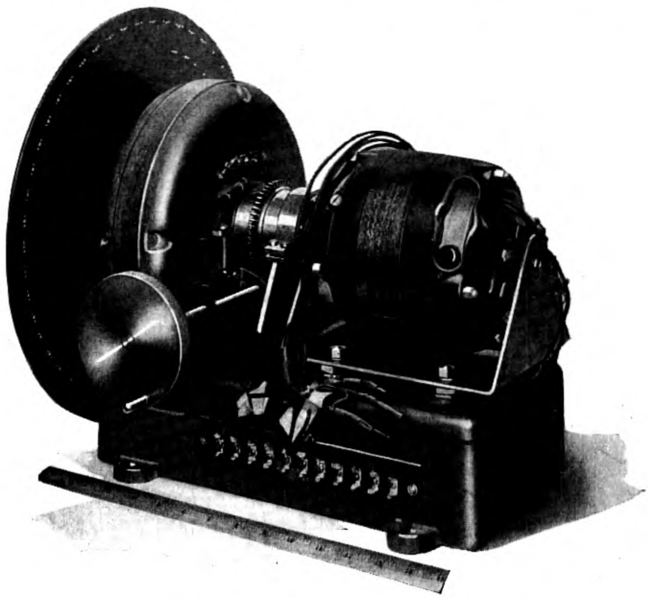


Fig. 1—Assembled motor

Since this problem of speed regulation is a separate one from that of the synchronization, the description of the regulating circuit is taken up later on.

Motor Design

In accordance with the phase displacement requirement as explained previously it was decided to build the synchronous motors with 120 pairs of poles, thus giving a frequency of 2125 cycles at 1062.5 r.p.m. which was the exact speed finally employed. For the sake of mechanical simplicity these machines were made of the variable reluctance type which gives one cycle per rotor tooth, thus requiring 120 teeth. The variable reluctance construction also simplifies the coil arrangement, the machine having only eight armature coils instead of a separate coil for each tooth. Fig. 1 shows a photograph of the assembled motor and Fig. 2 an inside view of the stator and rotor.

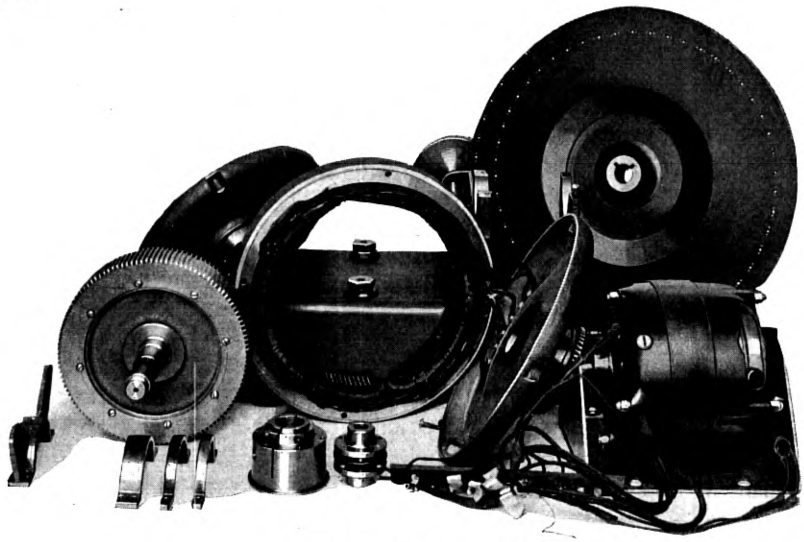


Fig. 2—Motor disassembled

In the preliminary experimental work two of these machines were directly connected (Fig. 3), permitting either machine to act as a

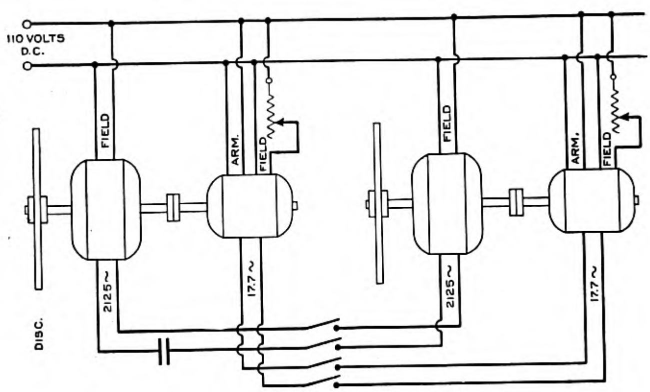


Fig. 3—Synchronization system over short wire line

synchronous motor loading down the other machine. Each machine was driven by a shunt d-c. motor having inherently poor regulation, the d-c. motors furnishing the power and the a-c. machines transferring the variations from one d-c. machine to the other to hold synchronism in a completely two-way system. As was to be expected, it was found that the motors hunted badly at a frequency of about four cycles per second. In other words, instead of holding within a fixed electrical phase angle of 20 degrees the receiving motor oscillated throughout a phase angle of about $\pm$ 20 electrical degrees. This, of course, made the picture wabble back and forth across the aperture and was therefore unsatisfactory.

The ordinary method of preventing hunting by means of copper bars embedded in the pole faces was not practical on account of the large number of poles and limited space. The hunting trouble was cured by employing a series condenser between the motors using a value of capacity somewhat less than that required to tune the circuit. A rigid analytical treatment of this anti-hunting circuit is beyond the scope of this paper but its operation depends in general upon the curvature of the tuning curve due to the variation of the inductance of the machine with phase displacement. Since the condenser operates on the total inductance of the circuit, it is desirable to make the natural periods of oscillation of the two motors different. Otherwise a decrease in the inductance of one machine may be accompanied by a simultaneous and equal increase in the inductance of the other, thus leaving the total inductance unchanged and preventing the condenser from functioning. This was done by making one disk substantially heavier than the other.

The series condenser also neutralizes the greater part of the internal reactance of the motors, thereby increasing the steady state torque.

Framing of Picture

There was still one unsatisfactory feature in this system in that the motor at the receiving end could interlock in any one of 120 different angular positions whereas in order to get proper framing of the picture it must be synchronized at a particular angular position. For example, if the disk at the receiving end is exactly 180 degrees out with respect to the disk at the transmitting end, the observer will see the lower half of the picture on top; a dark space representing the dividing line between pictures and the upper half of the picture at the bottom. Similarly, if the disk is 90 degrees out at the receiving end, the lower quarter of the picture will appear on the top and the upper three quarters of the picture on the bottom. The disk at the receiving end may be brought into correct angular position by providing means for turning the entire motor through the necessary angle. It was found, however, that the rate at which the motor can be turned was limited by the fact that if it were rapidly turned it would throw the motor out of step.

As an aid to framing, therefore, a second two-pole low frequency interlock was added to the system by providing the d-c. motors on each end with a pair of slip rings tapped to two opposite commutator bars. The d-c. shunt motors thus acted as converters furnishing 17.7 cycles at 1062.5 r.p.m. With this added feature on both the transmitting and receiving motors the process of synchronization was

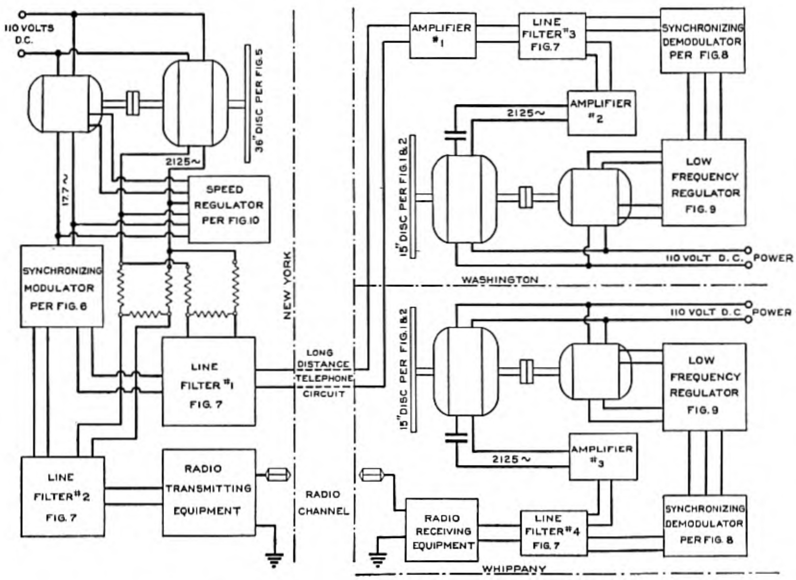


Fig. 4—Complete circuit of synchronizing system

first to close the 17.7-cycle circuit and adjust the field rheostat of the receiving motor until it came into step. Since this was a two-pole circuit there was only one angular position at which synchronization could occur. The high frequency synchronous machines were then connected together, thereby limiting the phase displacement to within .07 degree, as previously described. The high frequency motors in this system take the variation in load while the low frequency motor takes care of the steady constant component of load. Incidentally the addition of the low frequency synchronous motors greatly facilitated the synchronization of the high frequency motors inasmuch as it insured the proper initial speed. When the high frequency switch was closed there was merely a slight shift in phase angle to bring the receiving motor into step. The schematic circuit of the system thus far described is shown in Fig. 3.

Synchronization Over Long Lines

The above description explains the action of the synchronization system over lines of negligible impedance. In order, however, to secure similar results over a long distance telephone line or radio channel it is necessary to first attenuate the high and low frequencies to a power which can be safely applied to the transmitting end of the line and then amplify the power at the receiving end to restore it to the proper level. Fig. 4 shows the complete system employed.

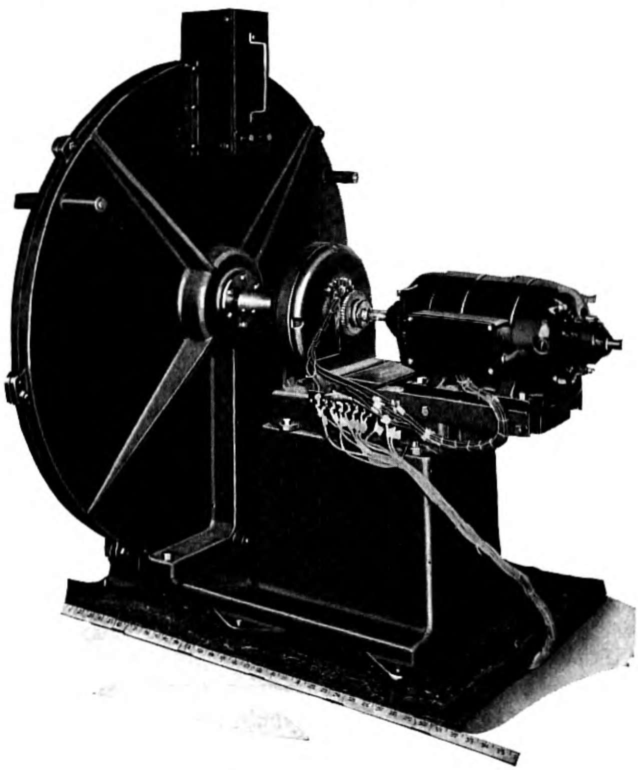


Fig. 5—Large scanning disc motor

While the high and low frequency machines on the transmitting end could have been designed so as to produce exactly the right power level, it was desirable, for the sake of interchangeability, to build the transmitting and receiving motor equipment of the same size. The output from the transmitting high frequency generator (shown in Fig. 2) when untuned was approximately 17 volts at 2125 cycles.

By means of a network this output was cut down to a level of 1 milliampere into 600 ohms impedance, the output impedance also being 600 ohms. This is a satisfactory level at which to transmit the high frequency, without inducing noise in adjacent wires in the telephone cables.

In the case of the low frequency interlock it was undesirable to attempt to transmit 17.7 cycles over a long distance line. The 17.7 cycles was therefore used to operate a polarized relay, the contacts of which modulated the output of a 760-cycle electro-mechanical oscillator [2] as shown in Fig. 6. In other words, the relay short-circuited

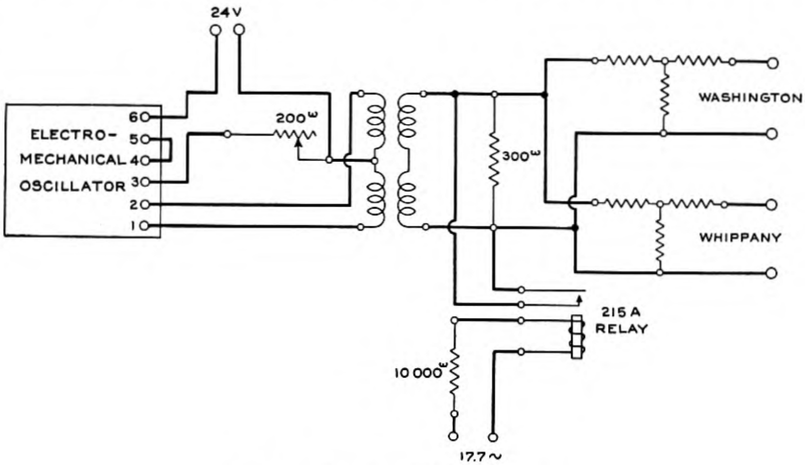


Fig. 6—Synchronizing modulator

the output of the oscillator alternate half cycles before application to the telephone line. Instead of using separate telephone pairs for the 2125-cycle and the modulated 760-cycle current, the two were combined by passing them through the line filter (shown in Fig. 7), thereby requiring only one pair for transmission of both frequencies. An identical network was employed for the radio channel. The problem of transmission of the synchronizing current is covered in the paper by Messrs. Gannett and Green and in the case of radio transmission in the paper by Mr. Nelson.

Receiving and Amplifying Circuits

Passing over this part of the problem, therefore, assume that the synchronizing currents have been obtained at the receiving end of the line. This power was delivered at a very low level, being about

[2] Described in the Bell Laboratories *Record*, March, 1927.

.3 of a milliampere into 600 ohms impedance, or 50 microwatts. It was then given a preliminary stage of amplification (amplifier No. 1, Fig. 4), passed through the line filter No. 3 (Fig. 7) and separated into 2125 cycles and 760 cycles modulated at 17.7 cycles. The 2125-cycle component was then amplified by two stages of amplification (amplifier No. 2) ending in push-pull 50-watt tubes and applied to

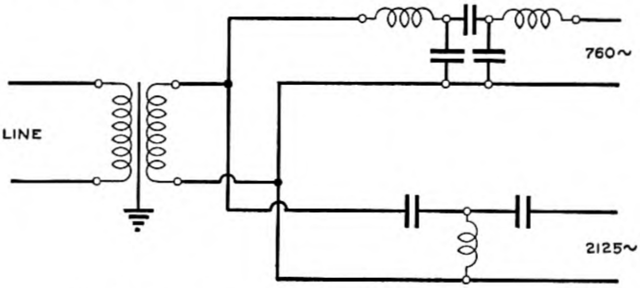


Fig. 7—Line filters for synchronizing frequencies

the high frequency motor. These amplifiers being of the standard type are not described. The terminal voltage on the output coil of the amplifier was made greater than that of the high frequency motor so that the power flow was normally from the amplifier to the motor.

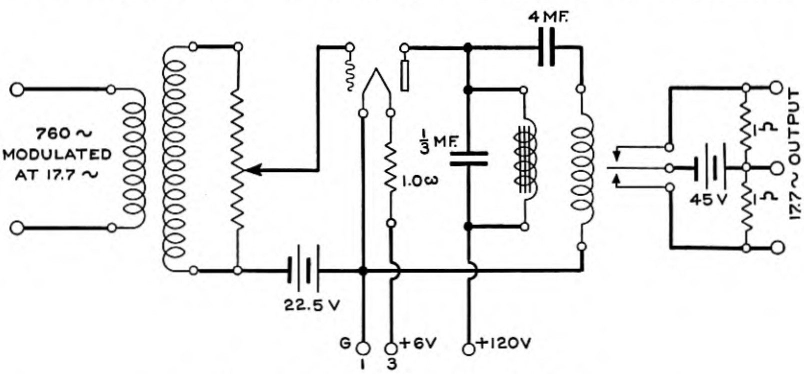


Fig. 8—Synchronizing demodulator

The anti-hunting condenser was retained between the amplifier and the motor.

In the case of the low frequency circuit the output from line filter No. 3 was received in the form of 760 cycles modulated at 17.7 cycles. This was passed through the demodulator (Fig. 8) which operated a polarized relay whose armature opened and closed its contacts at 17.7 cycles per second. The contacts of the relay provided square-wave low frequency current by interrupting power from a local battery source. On account of the limited power output which the vibrating contacts could safely handle without sparking, it became necessary to amplify this low frequency output. While this would have been possible by the use of ordinary amplifier circuits, it was found preferable from the standpoint of economy of apparatus to apply the low frequency regulation through a field circuit of the receiving motor. Referring to Fig. 9 it will be noted that the plate circuit of the reg-

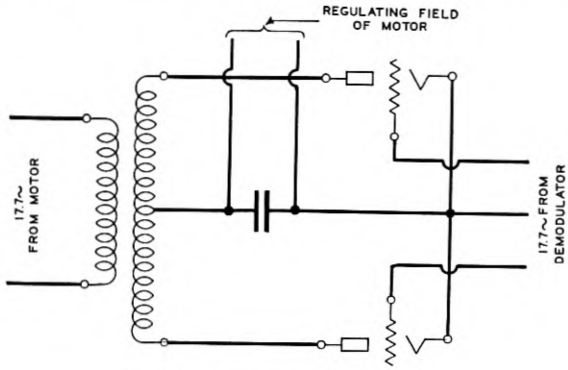


Fig. 9—Low frequency regulator

ulating tubes is supplied from the secondary of the transformer which is connected to the slip rings of the motor, while the grid circuit of these tubes is supplied with low frequency, low power 17.7 cycles from the contacts of the relay. As the motor is started up from rest the shunt field is weakened until the motor falls in step. At this point the frequency of the plate supply to the regulator tubes is identical with that supplied to the grids. If the phase relationship is such that the plates go positive at the same time that the grids are positive, then the space current of the tubes is increased and the regulating field (which is an aiding auxiliary field) is strengthened, thereby preventing a further rise in the speed of the motor. In other words, for each combination of load and line voltage there is an equilibrium phase position between the plate and grid voltages at which the corresponding regulating field current maintains the speed at the desired value.

Motor Operation

In actual operation the procedure was to first synchronize on the low frequency, and then on the high frequency circuit. The precise framing of the picture was then adjusted by rotating the motor by means of worm gearing through the necessary angle to center the image properly in the aperture. The high frequency current was of the order of 1.5 amperes at 2125 cycles with a terminal voltage of

100 volts at the high frequency motor. The power taken by the d-c. motor was approximately .8 ampere at 110 volts. The current through the regulating field controlled by the 17.7-cycle circuit was of the order of 20 to 40 milliamperes at 100 volts depending upon the phase position at which interlock occurred. It was found preferable to cut off the low frequency interlock feature after synchronization and framing had been obtained in order that irregularities in the time of contact closure of the relay might not produce changes in field strength of the d-c. motor which in turn would cause irregularities in power output. Such irregularities would give rise to phase shifts in the high frequency machine, thereby producing unsteadiness of the picture.

Operation on Radio Channel

In the case of transmission of the synchronizing current by radio instead of by wire the same apparatus is employed except that it was found necessary to use a much higher value of high frequency current in order to hold the high frequency motor in step, the current being approximately 4 amperes as compared to 1.5 amperes in the case of the other motors. This greater current was found to be necessary in order to hold the motor in step within the necessary phase angle of displacement, in spite of various types of interference picked up by the radio receiver, and associated circuits. This was mainly inductive interference from the picture and speech transmission sets arising from the fact that the synchronizing current was transmitted from New York to Whippany and picked up on a receiving set there, whereas the picture and voice current was transmitted from Whippany to New York. A certain amount of interference was also encountered from ship spark sets and static.

Speed Regulation of Transmitting Motor-Generator

As previously explained under "General Requirements" the essential requirement of the speed regulator at the transmitting end is to limit the acceleration to about one revolution per second per second, over intervals as small as .01 second. The ordinary type of centrifugally operated vibrating contact regulator keeps the motor continually accelerating and decelerating between an upper and lower speed limit and while such a system could theoretically be employed if the flywheel were made large enough, it was obviously preferable to employ a type of regulator in which the speed was inherently held constant without such acceleration and deceleration.

The regulating circuit employed is shown in Fig. 10. The complete theory of this regulating circuit is to be covered in another paper to be presented before the Institute. Briefly, the principle consists in employing a sharply tuned circuit as the primary speed-controlling element resonating at a frequency slightly less than the frequency at which the machine is operated. A voltage from the high frequency generator is applied to this tuned circuit and thence to a detector tube which in turn operates on the grids of a pair of push-pull regulator tubes; these tubes controlling an auxiliary regulating field winding on the motor. The circuit also contains anti-hunting means, the

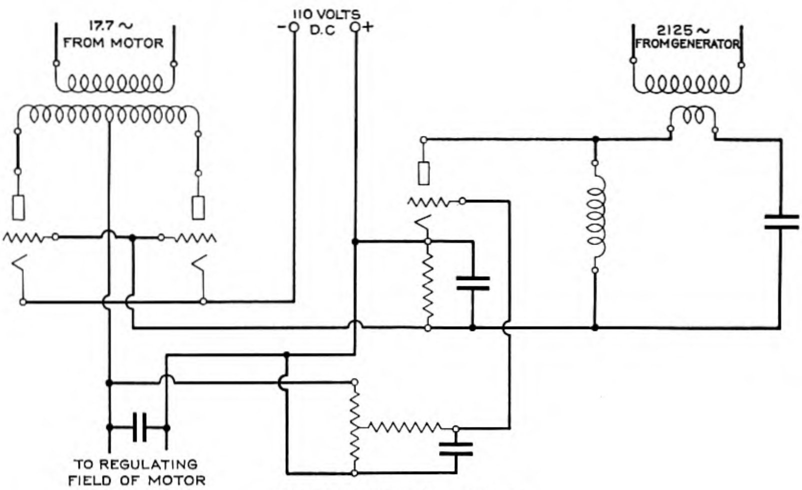


Fig. 10—Speed regulator

theory of which will be given in the later paper. Instead of applying this regulating circuit to the small 15-in. scanning disk motor shown in Fig. 3, it was decided on account of its greater flywheel effect to use the large 36-in. disk shown in Fig. 5 which was used for receiving the picture at New York. It therefore became the transmitter from the synchronizing standpoint for all of the other units although from the picture standpoint the big disk acted as a receiver.

Local Stations

In addition to the stations at Washington and Whippany there were three local stations in New York employing similar high and low frequency synchronous motors with 15-in. disks. These were controlled in the same manner except that first stage of amplification and the line filters were omitted. One station was employed for monitoring purposes, another operated a local transmitter, while the third operated the big grid receiver seen by the entire audience.

Wire Transmission System for Television[1]

By D. K. GANNETT and E. I. GREEN

Synopsis: This paper deals with the transmission problems which were met and solved in connection with providing wire circuits from Washington to New York for the television demonstrations which took place on April 7, 1927, and following. For transmission of the television images a single transmission channel was set up combining the frequency ranges usually assigned to telegraph, telephone and certain carrier channels. The special line requirements were met so successfully that the television images transmitted from Washington were indistinguishable from those transmitted locally.

Introduction

A SYSTEM of television, to be worthy of the real meaning of the name, must be capable of operation over a considerable distance. Spanning this distance, there must be a connecting medium suitable for faithfully transmitting the television currents. This paper describes how the connecting medium was provided between Washington and New York for the recent television demonstrations,[2] by adapting to this purpose existing wire facilities of the Bell System.

Fortunately, wire facilities of the type which were available between Washington and New York had been utilized for some time to transmit simultaneously many telephone and telegraph messages, involving a frequency range more than ample for the television requirements, so that the transmission characteristics of the lines throughout the necessary range of frequencies were well known. The matter of providing a suitable channel to carry the television currents consisted, therefore, in throwing together the frequency ranges which had heretofore been utilized for providing a number of separate telephone and telegraph channels. In addition to providing this very wide band communication channel it was necessary to apply special distortion-correcting networks so that the overall channel would possess proper characteristics and also to take care to avoid introducing disturbances due to such things as line irregularities, noise, etc.

Due to the perfection of the transmission methods which were utilized, it was found that when the circuit was first established, in accordance with the requirements which had been deduced, the television images transmitted from Washington were indistinguishable in quality from those transmitted locally, this result being secured

[1] Presented at the Summer Convention of the A. I. E. E., Detroit, Mich., June 20–24, 1927.

[2] "Television," H. E. Ives; "The Production and Utilization of Television Signals," F. Gray, J. W. Horton and R. C. Mathes; "Synchronization in Television," H. M. Stoller and E. R. Morton.

without any deviation from the adjustments which had been worked out in the original design.

Requirements

General. The ideal requirement for a transmission line for television, or for that matter any other purpose, is, of course, that it introduce no distortion whatsoever, in which case there could be no question but that the television images obtained in the receiving apparatus after transmission over the long distance line would be identical with the image obtained with the transmission only over a distance of a few feet. Practical transmission lines, however, tend to introduce a certain amount of distortion and the less the allowable distortion which is specified the greater will be the cost of providing a proper line. Before going ahead with the matter of engineering the line required to transmit the television currents from Washington to New York it was, therefore, first necessary that the requirements be set. The requirements were made more severe than strictly necessary in cases where they were easy to meet.

Frequency Range. In any system for the electrical transmission of intelligence, the required frequency range is, in general, proportional to the speed of transmission. In the case of picture transmission or television, the speed of transmission may be expressed in terms of the number of picture elements which must be transmitted per second, where a picture element is the smallest unit area which it is intended to be able to distinguish in the received picture from its neighboring unit areas.

When the picture currents are transmitted in the most efficient manner, the frequency range necessary is approximately equal to half the number of picture elements which must be transmitted per second. A simple way of seeing this is to realize that as the picture elements are transmitted in sequence, the greatest possible rate of variation of detail is obtained when alternate picture elements are black and white. A complete cycle corresponds in this case, therefore, to the time interval required to transmit two picture elements.

According to this relationship this particular television system in which about 40,000 picture elements per second are transmitted should require a frequency range of approximately 20,000 cycles. As a matter of fact it was found by a laboratory test that due to certain characteristics of the apparatus a frequency range as great as this was ample, just detectable distortion being introduced in the reproduction of the human face when the range was narrowed to about 14,000 cycles. In providing the line circuit, however, extending the frequency range to 20,000 cycles involved so little difficulty that it was decided to provide this very liberal frequency range.

In the particular television system which has been described the very low frequencies (below about 10 cycles) are suppressed. It was, therefore, not necessary that the line transmit these very low frequencies. The frequency range which the line should transmit was accordingly set as 10 cycles to 20,000 cycles.

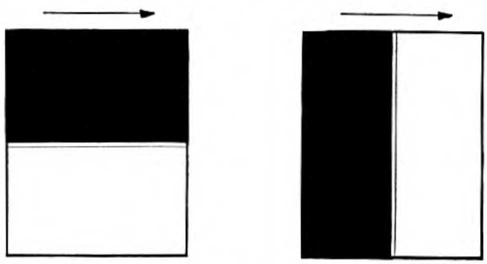
Attenuation. Referring to still picture transmission, it has been found that variations of attenuation with frequency of several transmission units do not appreciably impair the quality of the picture. Since no great difficulty was anticipated in meeting closer limits, however, it was decided to set the limits for the variation of attenuation with frequency at $\pm$ 2 T U within the frequency range of 10 to 20,000 cycles.

Phase Characteristics. A characteristic of wire lines, whose importance has been increasingly realized in recent years, is their phase characteristic. In speech transmission, transients due to unequal velocity of the different frequency components have been found to be an important consideration on some types of lines. In picture transmission and television, also, it is important that this phase distortion be controlled, as otherwise the image might be blurred due to the arrival of the various frequency components at different times. The type of transient which has been found to impair the quality of pictures is the type which is relatively rapid and the aim has been to make the phase characteristics such that those transients would be small.

The requirement with respect to phase for distortionless transmission is that β/ω be a constant where β is the phase change in radians for the entire circuit, and ω is equal to 2π times the frequency. β/ω is known as the "phase delay" or the steady-state time of transmission. $d\beta/d\omega$ is the time required for the transmission of the envelope of a wave whose components center closely about the frequency $\omega/2\pi$ and it will be referred to as the "envelope delay." Since it is more convenient to measure the envelope delay, the requirements were set up in terms of this quantity. When β/ω is constant, it is evident that $d\beta/d\omega$ is also constant. While the converse of this is not in general true, the conditions as actually encountered were such as to permit its use as a measure of the small variations involved.

The envelope delay characteristics of a number of circuits, which have been found to give varying degrees of transient on still pictures, have been measured. Also data were available from tests of picture transmission through filters and other networks whose delay characteristics were known. From these various data, the permissible deviations of the delay characteristic for still picture transmission were determined, and dividing these figures by 50, the ratio of the rate of transmission in picture elements per second in the two cases, the limits for the television circuits were obtained. In this way it was decided to attempt to keep within $\pm$ 10 microseconds, if possible, with outside limits of $\pm$ 20 microseconds. Check tests of these limits were made with the television apparatus in the laboratory by transmitting the currents through various known networks, and noting the effect on the received image.

Unlike the attenuation requirements, the delay requirements for television are not the same over the entire frequency range, but are much more lenient in the lower frequency range, as was shown by experiments in the laboratory. A physical picture of the reason for this may be obtained by reference to Fig. 1.



Arrows show direction of motion of light spot
A B
Fig. 1

Fig. 1A shows a picture placed in position before the sending machine, consisting of a piece of cardboard the same size as the image-area which can be transmitted, the upper half of the cardboard being colored black, while the lower half is white. As has been explained in the paper by Messrs. Gray, Horton and Mathes, the picture is scanned by a spot of light which moves from left to right in successive lines, tracing 50 horizontal lines across the picture in one sixteenth of a second. The first 25 of the lines lie on the black and the remaining 25 on the light part of the picture. The process is repeated 16 times per second, each repetition of 50 lines giving one complete cycle of black and white. The frequency components in this case are multiples of 16 cycles. A transient which blurs the picture outline over a given number, n, of picture elements (downwards) corresponds to a time interval equal to the time of tracing n lines, *i.e.*, $n/800$ second.

Now consider Fig. 1B. Here the picture has been rotated 90 degrees.

In this position, a complete cycle of black and white is obtained with each line instead of with each 50 lines. The frequency components in this case are multiples of 800 cycles and bear the same relations to 800 cycles as the components spoken of above bear to 16 cycles. A transient which blurs the picture outline n picture elements (horizontally, this time) corresponds to a time interval of n forty thousandths of a second. Evidently the delay requirements are 50 times more lenient in the former case than in the latter so that the delay requirement at the highest frequencies, which determine the fine detail in the direction of scanning, is 50 times as severe as at low frequencies, which determine the fine detail in a direction perpendicular to the direction of scanning.

In the still pictures referred to, the transients extended in the direction of travel of the light spot and there were no transients analogous to those discussed here in connection with Fig. 1A. For this reason the delay limits determined from still picture transmission are the ones which apply to the higher frequencies. For the lower frequencies the requirements are obtained by multiplying the high-frequency requirements by 50. For these reasons, together with the result of a Fourier analysis of the picture current, the limits were set at $\pm$ 10 or $\pm$ 20 microseconds from 400 to 20,000 cycles. Below 400 cycles, the departures from the constant delay were permitted to be $\pm$ 500 or $\pm$ 1000 microseconds.

Noise. Another important requirement is that relating to the ratio of the picture currents to the extraneous interfering currents which may arise in the line from power induction and other sources. Early experience with the television apparatus showed that considerably more noise was permissible in the case of television than in the case of still picture transmission so that in this case comparison with the still picture transmission would result in an unduly severe requirement. This is thought to be explained by the fact that in the case of television the pictures are flashed before the eye 16 times per second and the effects of the extraneous currents occur on successive flashes in different positions, so that defects of one flash are corrected on the next.

A set of experiments was performed from which it was determined that if the ratio of average picture currents to average noise currents exceeded about 10 the results were satisfactory. In order to assure considerable margin above this figure, it was decided to make the average television current to be transmitted into the line 4 milliamperes.

Echoes. If two paths exist by which the currents may travel from the sending point to the receiving point, the length of the two paths being different, a double image will be produced on the received picture, forming what may be termed visual echo. In the case of telephone lines, the echoes may exist on account of reflections between impedance irregularities in the circuit so that the currents arrive at the receiving point both by way of the direct transmission path and by way of a transmission path which includes an extra loop between two irregularities. If the echo is not greatly attenuated with respect to the main transmission, the result may be quite disturbing on the received picture. It has been found by experiment that the echo is too weak to be seen if it is more than 25 T U weaker than the main current and, accordingly, care was taken in setting up the New York-Washington circuit to avoid introducing echo paths of lower equivalent than this.

General Choice of Method

Two general methods are possible for transmitting the currents over the line circuits. One method is to transmit the currents directly without change of frequency. This method involves the transmission of the currents of the frequency range determined upon above, namely, from about 10 cycles to about 20,000 cycles per second.

The other general method is the carrier method, in which the television currents modulate a carrier current of suitable frequency and are thereby moved to another portion of the frequency spectrum prior to transmission over the line. At the receiving end of the line the carrier currents are then restored to the original frequencies of the television currents.

Several different schemes of carrier transmission are possible. The simplest is to modulate a carrier with the television currents and to transmit both side bands. This has the disadvantage of requiring the transmission of twice as wide a frequency range as that occupied by the original television currents. Another scheme is to transmit a single side band. A third possible scheme is to transmit both side bands for the lower frequencies and only one side band for the higher frequencies.

One advantage to be secured by the carrier method is that it lessens the severity of some of the line problems through avoiding the transmission of very low frequencies over the line circuit. At these frequencies the amount of noise found on lines is usually considerably greater than at the higher frequencies.

After weighing the relative merits of the carrier and direct transmission methods it was decided to make use of the latter because of its simplicity. An important factor in this decision was the successful development, for use in connecting the apparatus to the lines, of transformers providing adequate transmission of the entire frequency range from 10 cycles to 20,000 cycles.

Arrangements for Television Circuits

Line Layout between New York and Washington. The layout of the wires between New York and Washington is shown in Fig. 2. The circuit over which the waves actually carrying the pictures were transmitted (marked Picture Circuit) consisted principally of a pair of copper wires 165 mils in diameter. At a number of places on the route the circuits were carried in cable as indicated in the figure. The total length of the television circuits was about 285 miles, of which 8 miles consisted of cables and the remainder of open wire.

Transpositions. As the circuits employed were originally designed for voice-frequency operation only, except for a section at the New York end, it was necessary to add transpositions to them to prevent interaction with adjacent circuits at the high frequencies involved in the television transmission. The high-frequency currents were thus prevented from passing over into the adjacent circuits which would have resulted in irregularities in the attenuation, line impedance and phase shift characteristics of the circuit.

Incidental Cables—Loading. Any appreciable length of non-loaded cable included in an open-wire television circuit has certain very objectionable effects. The impedance irregularities introduced by the cable destroy the uniformity of the line attenuation, impedance and phase shift characteristics as a function of frequency, and tend to produce echoes as described above. Types of loading developed for use on incidental cables occurring in circuits employed for carrier telephone and carrier telegraphy operation [2] were employed to reduce these effects to a minimum. This carrier loading is designed so that when used on No. 13 A. W. G. cable circuits it provides an impedance which approximates very closely that of the open wire. With a spacing of about 930 feet between loading coils, this loading has a nominal cut-off of about 45,000 cycles, which corresponds to an effective transmission range extending up to about 36,000 cycles. In order to obtain a close match between the impedances of the open-wire and the cable pairs, thereby avoiding impedance irregularities, 13-gage pairs were selected for the television circuits in all of the cables.

The length of the submarine cable under the Hackensack River (about 1100 feet) was too great to permit the use of regular carrier

[2] "Development and Application of Loading for Telephone Circuits," T. Shaw and W. Fondiller, *Journal A. I. E. E.*, Vol. XLV, pages 253–263, March, 1926.

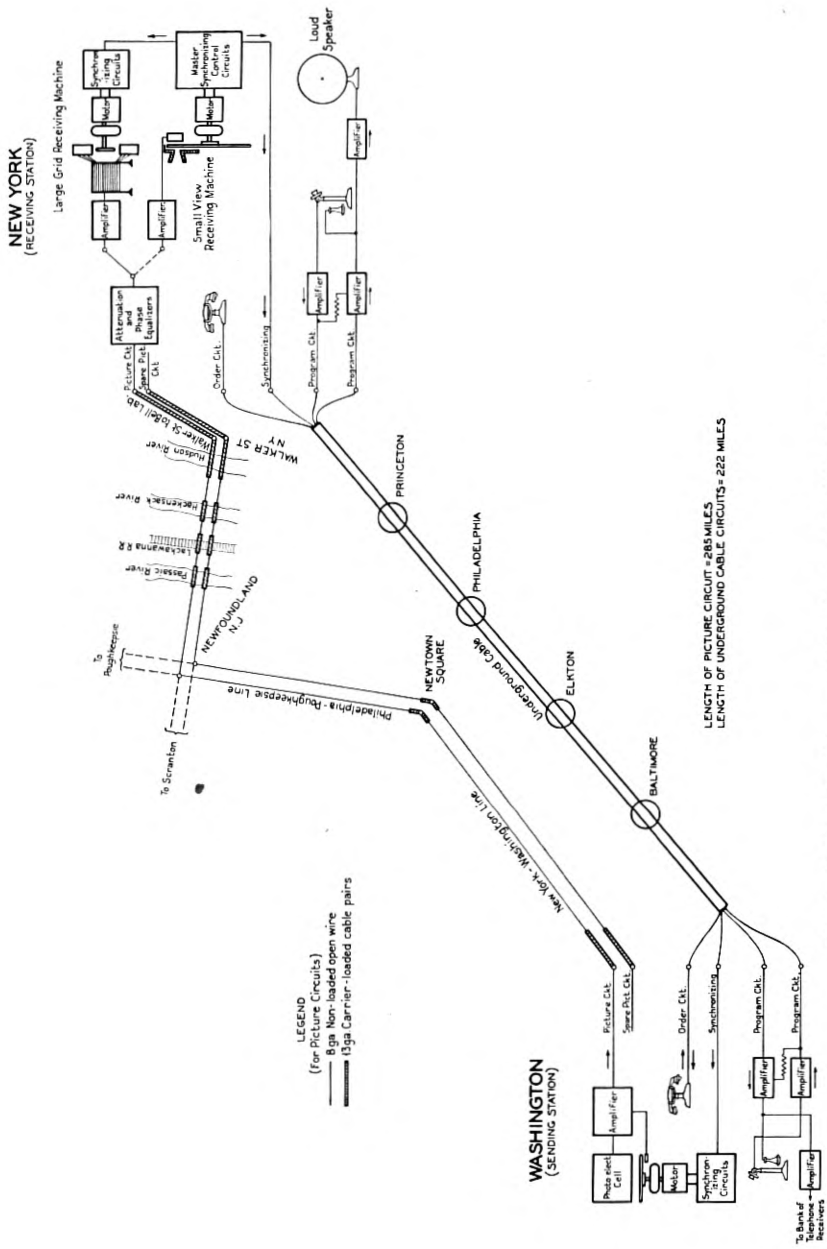


Fig. 2—Schematic diagram of circuits for television demonstration

loading, and a special loading arrangement having a slightly lower cut-off was, therefore, designed for this cable.

Equalization

Requirements. The requirements for the lines were stated earlier. In order to meet these overall requirements it was necessary to apply special forms of distortion-correcting networks.

Weather Changes. The above requirements applied, of course, to

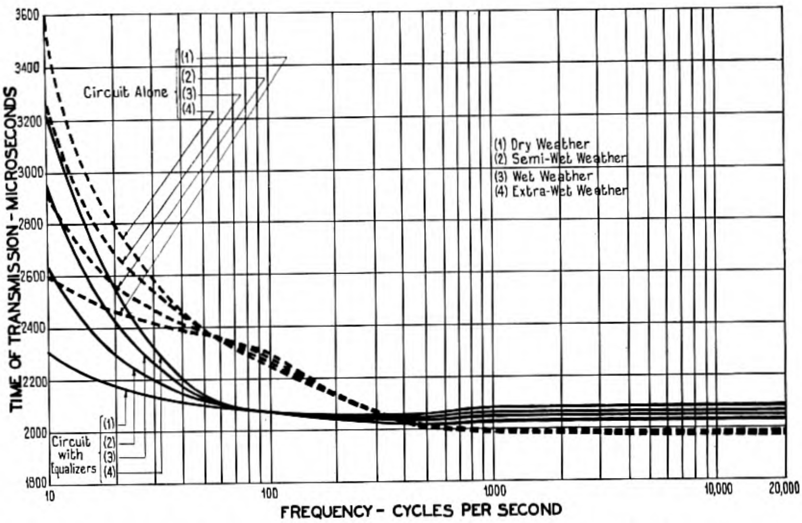


Fig. 3—Computed phase delay (β/ω) of television circuit with and without equalizers

all of the various weather conditions to which an open-wire circuit is subject. Due to the changes in the leakage conductance occurring at the insulators, the attenuation of an open-wire circuit varies with changing weather conditions. This change is particularly important at the higher frequencies. At 20,000 cycles, for example, the attenuation of a 165-mil open-wire pair may vary as much as 40 per cent for a change from dry weather to extra wet weather. For the circuit between Washington and New York this represents a possible attenuation change of about 10 T U, or a change of 10 to 1 in the magnitude of the received power. At 1000 cycles, the effect of wet weather is comparatively small, so that the net effect of the weather variations is to change the requirements for the attenuation equalizers. The phase shift introduced by an open-wire pair likewise varies to some extent with changes of weather, although the percentage variation is much smaller than in the case of the attenuation. In view of these variations in the line characteristics it was decided to provide basic networks which would equalize for dry weather conditions, and to make available, in addition, several steps of equalization which would compensate for changes in the direction of wet weather.

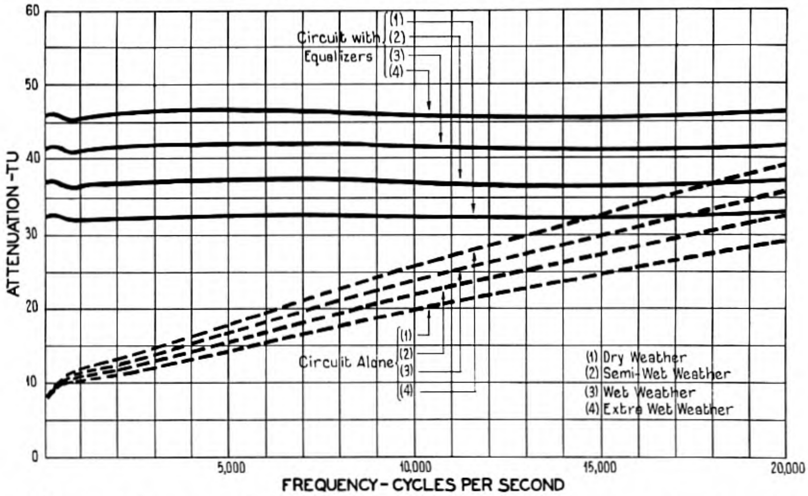


Fig. 4—Computed attenuation characteristics of television circuit with and without equalizers

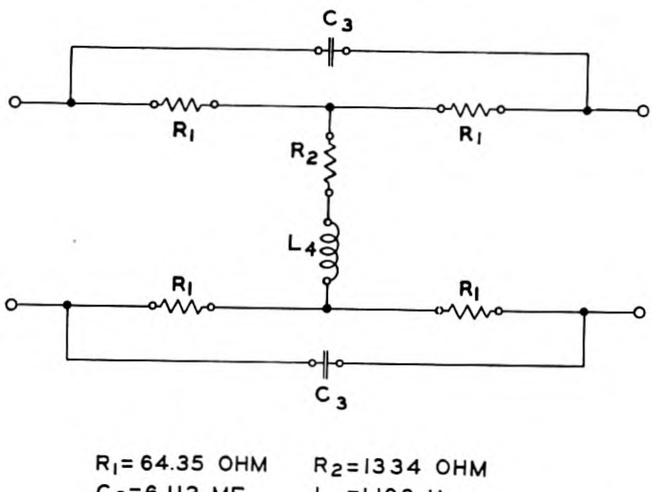
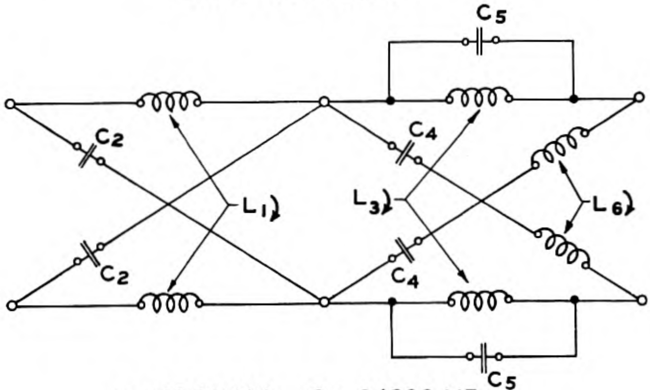


Fig. 5—Low-frequency equalizing network (dry and wet weather)

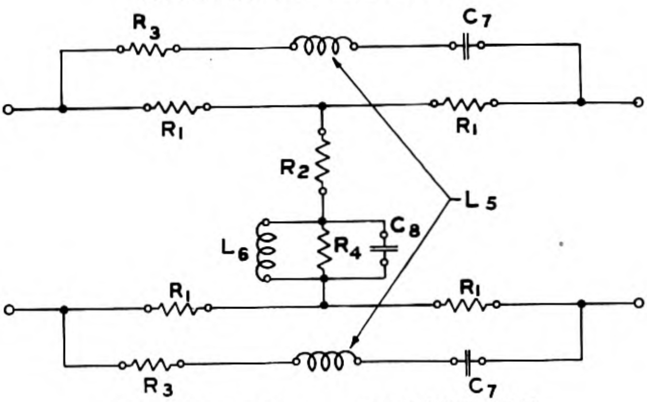
Low-Frequency Network. Computed curves of attenuation and phase delay for the overall Washington-New York circuit without correcting networks are shown in Figs. 3 and 4, respectively. The form of the dry weather attenuation curve suggested the use of two correcting networks, one for low frequencies, the other for high frequencies. The network which was designed to equalize the at-

PHASE CORRECTOR



L1=35.99 MH. C2=.04999 MF.
L3=17.51 MH. C4=.02432 MF.
C5=.02138 MF. L6=15.40 MH.

ATTENUATION EQUALIZER



R1=256.6 OHM R2=94.20 OHM
R3=111.9 OHM R4=1609 OHM
L5=9.668 MH. L6=1.829 MH.
C7=.010162 MF. C8=.02686 MF.

Fig. 6—High-frequency equalizing networks (dry weather)

tenuation at the lower frequencies is illustrated in Fig. 5. This network, in addition to equalizing the low-frequency attenuation, was made to provide sufficient correction for the low-frequency phase characteristic. It also proved satisfactory for all weather conditions.

High-Frequency Network for Dry Weather. The complete network for the correction at high frequencies under dry weather conditions was designed in two parts, an attenuation equalizer and a phaes corrector. These two structures are illustrated in Fig. 6. The

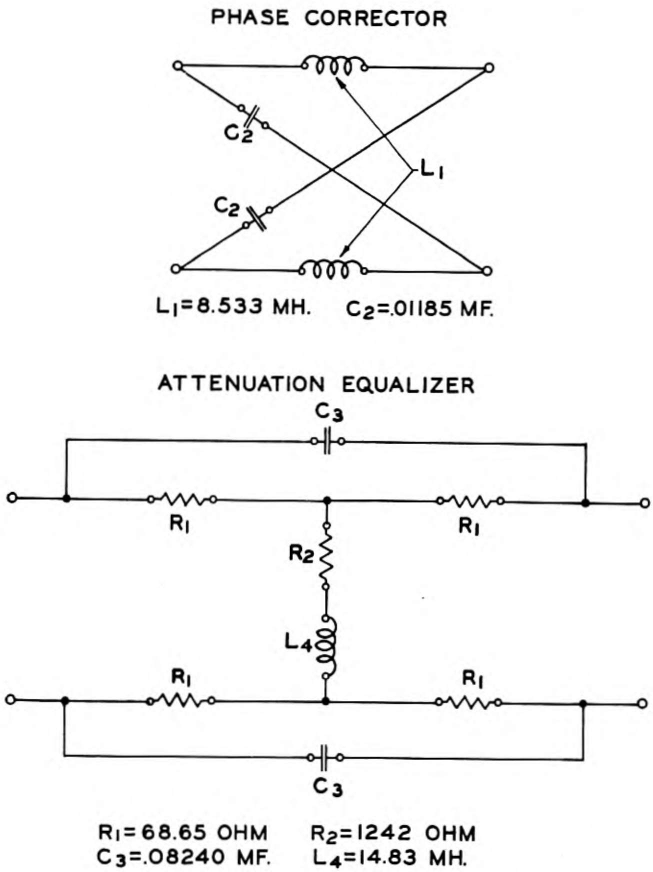


Fig. 7—Weather change equalizing networks

computed dry weather attenuation and phase delay resulting with the use of the combined low-frequency and high-frequency networks are illustrated in the curves of Figs. 3 and 4. It will be noted that the corrected attenuation curve is constant to within approximately $\pm$ 0.3 T U, while the corrected time of transmission falls well within the prescribed limits.

Weather Change Networks. Correction for the additional distortion introduced by changes from dry to wet weather was provided by three additional networks which were, for convenience, of identical design. The results obtained by using one, two or three of these networks were made to correspond, respectively, to three assumed weather conditions which may be designated semi-wet, wet, and extra-wet. These three conditions were determined upon the basis of the range of leakage conditions which exist on open-wire lines under different weather conditions.

The attenuation equalizing and phase correcting networks for one of these steps are illustrated in Fig. 7, while the computed attenuation and phase delay obtained by the use of the three different steps of weather correction are shown in Figs. 3 and 4.

The networks described above are of the "constant-resistance" type, whose characteristic impedance is a pure resistance at all frequencies.[3] These networks are designed to be connected in series. The methods used in the design of the networks involve a large amount of mathematical theory, a discussion of which is not necessary for the purposes of this paper.

Synchronizing and Voice Circuits

So far the discussion has dealt only with the problem of transmitting the television currents. In addition to this, there is required the transmission of voice currents and of synchronizing currents. It is entirely feasible to transmit these currents together with the television currents over a single circuit. However, for the purpose of simplification, separate facilities were employed in the television experiments for picture, voice and synchronizing currents.

The diagram in Fig. 2 shows the circuits which were actually provided for the demonstrations. It will be seen that in addition to the two picture or television circuits, there were provided a synchronizing circuit, a four-wire "program" circuit, and an order circuit.

The method of synchronizing the sending and the receiving machines has already been described in the paper by Mr. Stoller. It requires two currents, one having a frequency of about 18 cycles and the other about 2125 cycles. In order that an ordinary telephone circuit might be used for this purpose, the lower frequency was made to modulate by means of a telegraph relay, a carrier current having a frequency of about 750 cycles per second. An amplifier-detector at the receiving end of the synchronizing system demodulated the 750-cycle current, delivering 18 cycles to the television apparatus.

The requirements for the synchronizing circuit were that it must

[3] Partially described in U. S. Patent No. 1,603,305 to O. J. Zobel.

transmit a narrow range near 750 cycles, and the single frequency of 2125 cycles. These synchronizing frequencies are determined by the speed of the motors, which was chosen so that the frequencies would be suitable for transmission over two channels of a voice-frequency carrier telegraph system,[4] but later it was found more convenient to use a separate telephone circuit.

The circuits labeled "program" provided telephonic communication between the observer at New York and the person being viewed at Washington. A loud speaker was also connected to this circuit at New York to transmit the voice to the audience when the large grid receiving arrangement was employed. A special by-passing connection was provided between the amplifiers at the terminals of the circuit so that speech from the local microphone could be heard as well as speech from the distant city.

The order circuit was for the purpose of providing communication between the engineers operating the television apparatus.

Line Measurements

In order to determine that the circuits set up as outlined above were satisfactory, their overall characteristics were measured. Certain matters of interest in this work are noted below.

Measurements of Envelope Delay. In order to measure the envelope delay to an accuracy comparable to the requirements for the lines, it was necessary to develop special apparatus. Fig. 8 shows in schematic form the circuits of the apparatus designed for this purpose. The apparatus measures not the absolute envelope delay of a circuit, but the relative delay of one circuit at any frequency from about 600 cycles to 20,000 cycles or more with respect to the delay on the other circuit at a fixed frequency.

The functioning of the apparatus may be briefly described as follows: Simultaneously into each line there was transmitted a carrier current, each carrier being modulated by 250-cycle current from the same oscillator. The modulation was accomplished in push-pull vacuum tube circuits so that the undesired products of modulation were eliminated by balance. The carrier on the line under measurement was adjusted to the frequency at which a measurement was desired, and the carrier on the other circuit, used for reference, was kept at a fixed frequency of 5100 cycles.

At the receiving point identical circuits were provided for amplifying

[4] "Voice-Frequency Carrier Telegraph System for Cables," B. P. Hamilton, H. Nyquist, M. B. Long and W. A. Phelps, *Journal A. I. E. E.*, Vol. XLIV, pages 213–218, March, 1925.

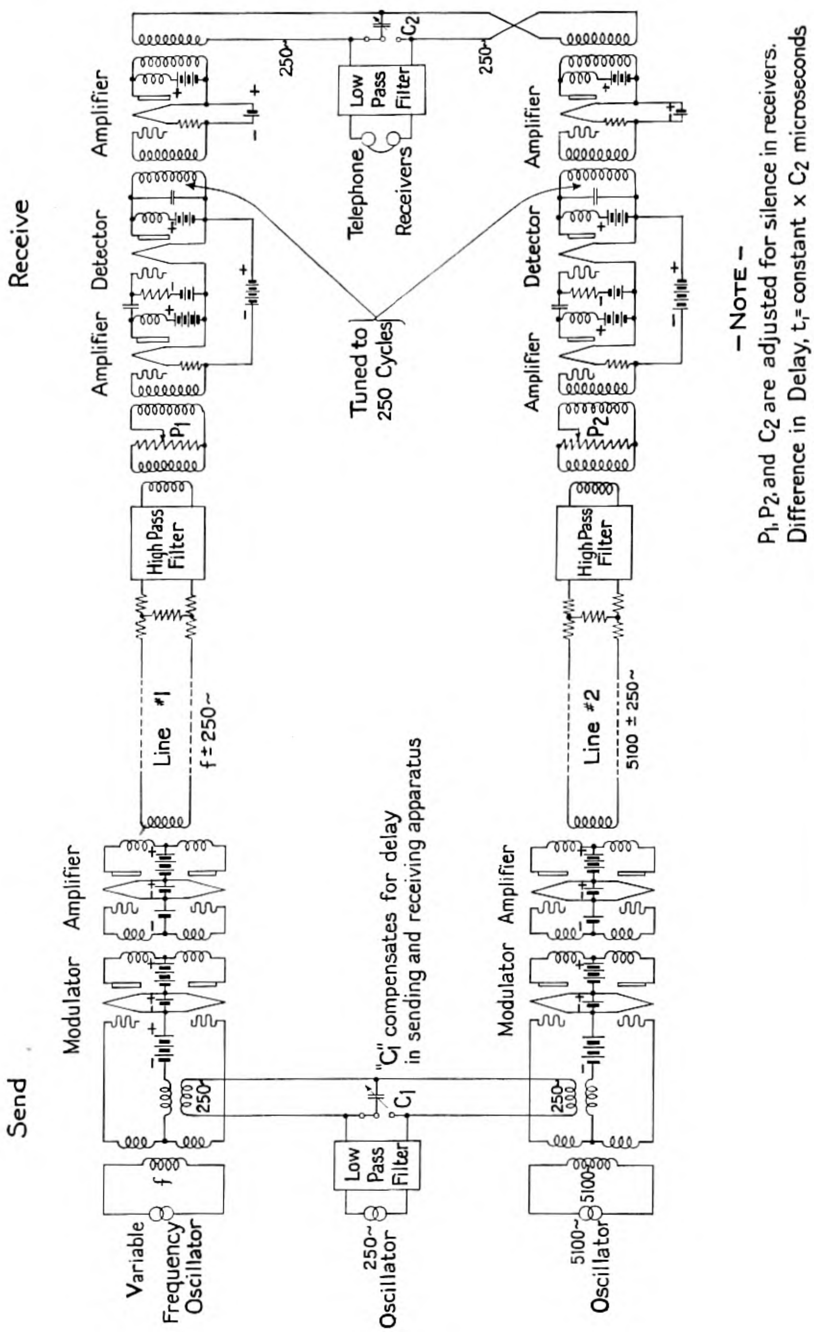


Fig. 8—Arrangements for measuring envelope delay of television circuits

and demodulating the received currents from the two circuits. The 250-cycle outputs from the two sets of receiving apparatus were connected in opposition to a pair of telephone receivers through a low-pass filter. Potentiometers P_1 and P_2 were provided for adjusting the relative intensities of the two 250-cycle output voltages and a condenser C_2 was arranged so that it could be used to change the phase of either of the 250-cycle voltages. It is evident, then, that

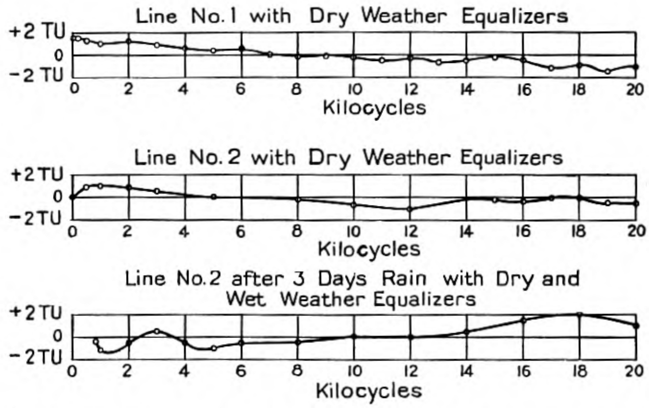


Fig. 9—Measured attenuation characteristics of line circuits plus equalizers

by making suitable adjustments the two voltages could be adjusted to exactly the same intensity and opposite phase so that no sound is heard in the telephone receivers. As long as the value of C_2 is small, the envelope delay of one line at the carrier frequency with respect to the delay of the other line at 5100 cycles is proportional to the value of C_2.

The condenser C_1 shown at the sending station is for the purpose of introducing a phase shift in the 250-cycle current of either channel relative to the other in order to compensate for the differences in delay of the apparatus itself at the two frequencies. The value of C_1 was determined by experiment before moving the sending apparatus to Washington and was adjusted to its calibrated value for each frequency when the oscillator frequency was adjusted.

The measurement of the phase shift of the 250-cycle current, which is transmitted by means of a carrier over a circuit as described above, is actually a measurement of the difference between the phases of the two received side-band currents situated 250 cycles either side of the carrier. The envelope delay is equal to $\Delta\beta/\Delta\omega$ where $\Delta\omega$ equals 2π times 500, and $\Delta\beta$ equals the measured difference in phase of the two side bands in radians.

Measurements and Performance. How well the requirements which were set up earlier were met by the lines and the distortion-correcting networks is shown in Figs. 9 and 10. The attenuation characteristics

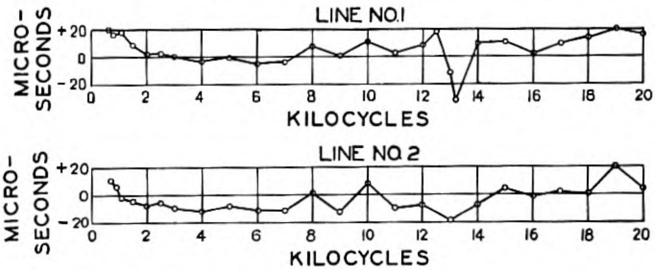


Fig. 10—Measured envelope delay of line circuits plus equalizers

are well within the established limits, and the phase characteristics show only a single slight departure for one circuit in a very narrow range of frequency. It is of interest, in view of the fact that the distortion-correcting networks were designed and built before any measurements were made on the lines they were to fit, that no changes or adjustments were found to be necessary in the networks, in order to obtain these characteristics.

Comparison of the television images obtained from transmission over the line with those obtained from transmission from one side of the room to the other, showed that no difference in quality could be observed.

Radio Transmission System for Television[1]

By EDWARD L. NELSON

SYNOPSIS: Starting from the general requirements imposed on the transmitting medium, this paper discusses the engineering of a radio system for television purposes and describes the radio facilities actually employed for the recent Bell System demonstration. The tests to which the system was submitted to determine its suitability are outlined and the measured frequency-response characteristics are shown. An interesting phenomenon due to multi-path transmission, the production of positive and negative secondary images, is reported. A brief series of experiments concerned with the transmission of both voice and image "on a single wavelength" is also described.

IN other papers of this symposium, the general nature of the television problem has been discussed, the scope of the recent Bell System demonstration has been outlined, terminal apparatus for television has been described, and the general requirements to be met by the transmitting agency have been formulated. This paper is concerned with the problem of engineering a suitable radio system for television purposes and with a description of the radio facilities actually employed for the demonstration.

Requirements Imposed on the Radio System

The radio experiments were conducted from the Bell Telephone Laboratories' Experimental Station 3XN at Whippany, New Jersey. Between this point and the main Laboratories Building at 463 West Street, New York City, some 22 miles distant, three separate communication channels were required—one for the picture, a second for synchronizing, and a third for speech and music. The demonstration being of a three-cornered nature involving New York, Washington and Whippany, it was deemed to be highly advantageous to transmit the necessary synchronizing currents for both the wire and radio systems from a master generating set located in the auditorium of the West Street Building. Hence the synchronizing channel was required to operate from New York to Whippany, while the picture and speech channels necessarily transmitted in the reverse direction.

From the radio standpoint, the problem presented for solution may be described as follows:

1. There is given television transmitting and receiving apparatus designed to work into and out of specified impedances at stated signal

[1] Presented at the Summer Convention of the A. I. E. E., Detroit, Mich., June 20–24, 1927.

energy levels. Signal components ranging in frequency from 10 to 20,000 cycles must be transmitted with as little discrimination with respect to either amplitude or phase as reasonable design practices will permit. It is required that a suitable radio system be designed to afford satisfactory transmission between terminals when operated under prevailing conditions with respect to static, other radio traffic, and local electrical disturbances. The maximum allowable "noise" level is probably somewhat arbitrary but it has been found that if the ratio of signal to interference current is 10:1 the results are satisfactory. The variation of amplitude with frequency should probably not exceed $\pm$ 2 TU at any point in the required signal band. The equivalent of the circuit must be substantially constant; in other words, no fading effects can be tolerated. In this respect a variation of perhaps 3 TU is the maximum allowable.[2]

2. For synchronizing purposes, a second channel must be provided to transmit 17.7 and 2125 cycles, the impedances and the signal energy levels at both ends of the circuit being known. The grade of transmission required in this case is probably considerably lower than that needed for the picture circuit but stable operation must be assured.

3. Arrangements must also be made for a high quality telephone channel to transmit speech and music for loud speaker reproduction.

4. All of these channels must, of course, be capable of operating simultaneously without mutual interference and without effect on established radio services.

Preliminary Survey

In the vicinity of New York, an assignment of this type is surrounded with unusual difficulty due to the serious congestion which exists in the ether. Operations were started, therefore, by undertaking a survey of available frequency bands at periods of the day during which transmission might be required.

The pioneering nature of the project and the character of the apparatus available led to an early decision to base the system on the transmission of the carrier and both sidebands. Since the upper limit for the signal was specified as 20,000 cycles, an interference-free band somewhat greater than 40,000 cycles in width was, therefore, required. The unusual width of this band indicated the desirability of fixing upon a relatively high carrier frequency. No readily available

[2] Definitely agreed on limits were essential to proper coordination of the various development activities and figures of the order mentioned were assumed for design purposes.

substitute for the ordinary type of tuned circuit was at hand and such circuits discriminate seriously against side frequencies differing by more than a few per cent from the frequency to which they are adjusted.

The results of the survey disclosed two bands somewhat wider than that required centering approximately about 1575 and 1750 kilocycles. It was also conclusively demonstrated that the operation of the synchronizing channel at a frequency above the broadcasting

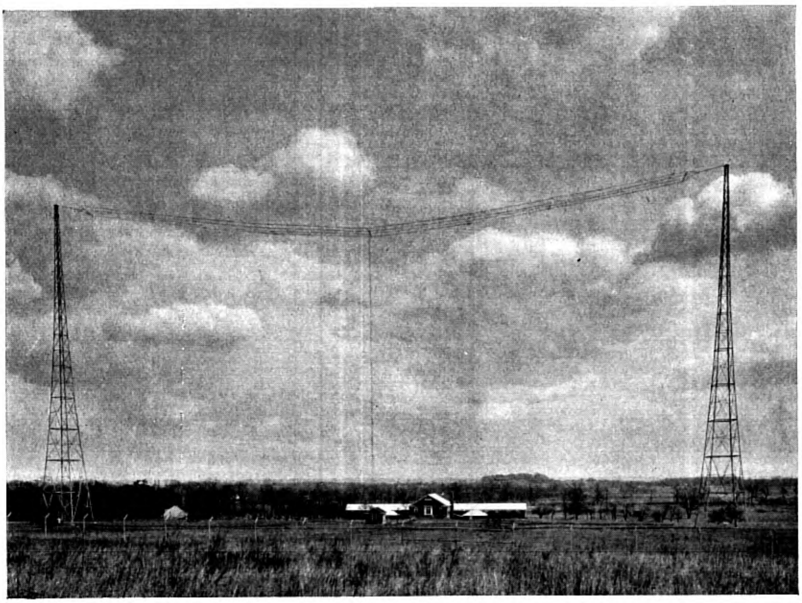


Fig. 1—General view of the Whippany Station, 3XN

band was entirely out of the question. With two broadcasting stations located in the immediate neighborhood, one producing a field strength of perhaps 50 millivolts per meter and the other several volts per meter, the operation of a third transmitter on an adjoining frequency with the maximum obtainable separation between antennae, resulted in an almost continuous interference spectrum. It was decided, therefore, to transfer the synchronizing channel to a frequency of the order of 185 kilocycles, which would be sufficiently remote to remove interference from this source, and to make further studies in the regions about 1575 and 1750 kilocycles based on transmission from Whippany. No difficulty was anticipated in making suitable arrangements for the speech channel on account of the narrower band required and the well-known nature of the problem.

The Whippany Station, 3XN

A general view of the station site at Whippany is shown in Fig. 1. The property consists of some 47 acres. The main laboratory building, which is located near its center, is a two-story structure affording approximately 18,000 square feet of floor space. The principal antenna system involves two 250-foot steel towers with a suitable buried ground system, which is placed some 500 feet out in front of the building in order that the latter may be clear of the denser portion

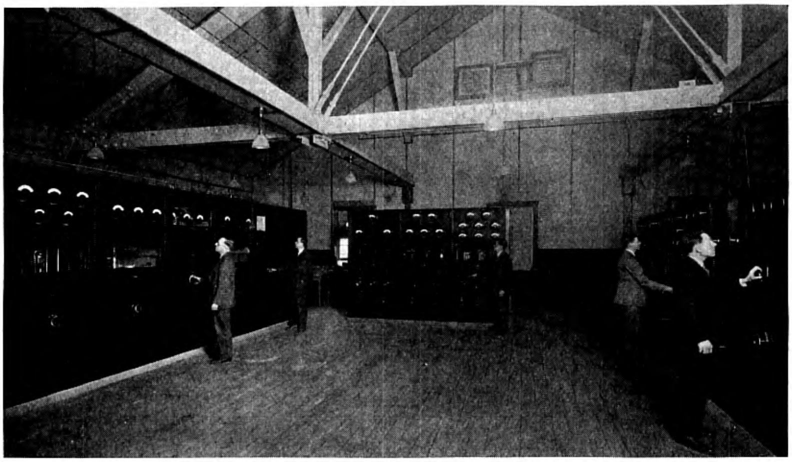


Fig. 2—Operating room at 3XN. Transmitter for television channel on the right. Power supply unit and radio transmitter for the speech channel in the center and on the left, respectively.

of the electric field. This antenna was assigned to the picture channel. For the voice channel, a separate structure located 500 feet in the rear of the laboratory building or approximately 1000 feet from the other was employed. The original supports in this second case were 60-foot wooden masts but subsequently metal topmasts were added, bringing the total height to 100 feet. Both antennae were energized by means of radio-frequency transmission lines. The antenna tuning and coupling apparatus was housed in small buildings placed under the center of each antenna, that for the larger structure having a copper roof which was securely connected to the ground network.

This type of installation is thought to afford a number of advantages. By separating the building and the antenna it becomes a much simpler matter to control the electrical factors which enter into the design of the latter. Removing the building from the field tends toward reduced dielectric and eddy current losses and consequently toward higher antenna efficiency. The resulting improvement may be expected to more than compensate for the slight loss in the line, which should not exceed 3 per cent. Removing the field from the building is equally advantageous in that it simplifies the precautions which normally have to be taken to prevent the radio-frequency energy from affecting the performance of audio amplifiers and other supple-

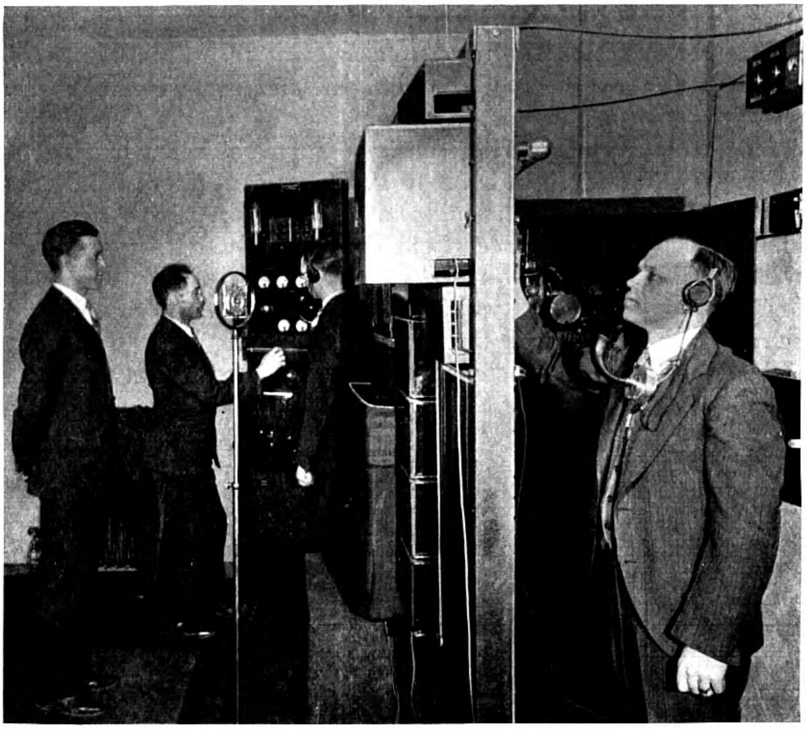


Fig. 3—Television transmitting apparatus in the studio at Whippany

mentary vacuum tube apparatus. The most serious disadvantages arise from the fact that the antenna must be tuned and the current in it measured at a point remote from the transmitting apparatus proper.

In spite of the fact that the station building was not directly under either antenna, some difficulty was anticipated from radio-frequency fields produced within the transmitting equipment due to the relatively high amplification employed with the photoelectric cells. In order to minimize trouble of this nature a special shielded studio was constructed in one of the wings of the building to house the television terminal apparatus. Walls, ceiling and floor were completely covered with No. 24 gage sheet copper lapped about one inch and carefully soldered. The windows were covered with fine copper gauze. The door was covered with sheet copper which was carried around the edges so that in closing it made a firm wiping contact with the surrounding frame. Circuits for lighting and miscellaneous power service were led in through two specially constructed transformers fitted with grounded copper shields between the primary and secondary windings. The picture circuits leading to the radio transmitter, the microphone circuits, and the necessary studio signal and control circuits were run in lead cable and in most cases were brought into the room through suitable radio-frequency filters enclosed in metal boxes attached to the copper sheathing. In order to avoid the possibility of the heavy current leads to the arc bringing in radio-frequency energy, and to eliminate the noise and heat from the arc, provision was made for mounting the latter in its metal cabinet outside of the room. The circular opening through which the light beam was projected into the room was protected by the lamp cabinet which was also grounded to the sheathing. Satisfactory acoustic conditions within the studio were obtained by applying celotex wall board over the copper and by the use of suitable floor coverings.

Transmitting and Receiving Apparatus

For the television channel, arrangements were made to install a standard Western Electric 5-B Radio Broadcasting Transmitter and to modify it for the purpose. This transmitter is a 5-kilowatt unit (carrier output without modulation) designed for high quality telephone transmission in the 500–1500-kilocycle band. It will transmit signal components ranging from 50 to 5000 cycles without noteworthy discrimination. At 30 cycles and at 10,000 cycles there is some loss in efficiency and beyond these points the characteristic curve falls rapidly. The necessary changes, therefore, involved both the radio and audio circuits, the latter phase of the problem being perhaps the more difficult.

The schematic circuit of the modified transmitter is shown in Fig. 4. The revised radio frequency circuits were very similar to the standard arrangement, the changes mainly affecting the magnitudes of various coils and condensers. The output circuits were, of course, redesigned to meet the conditions imposed by the transmission line. The circuit was of the master oscillator—modulating amplifier—power amplifier type. The master oscillator employed a 50-watt tube operating in a

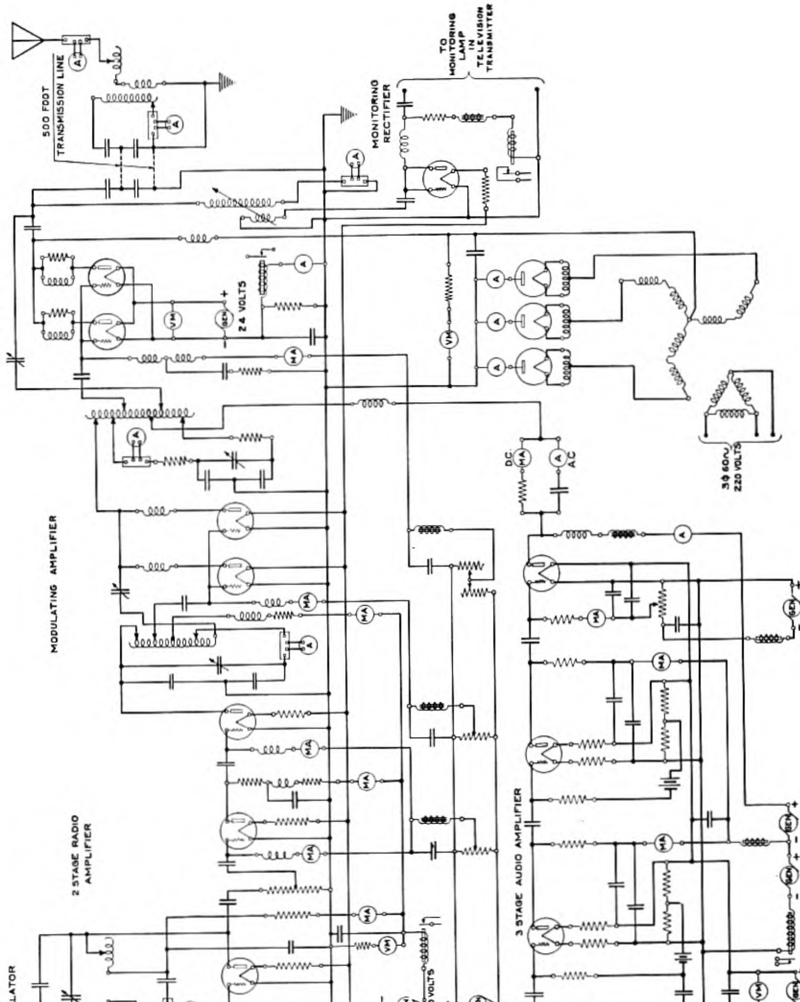


Fig. 4—Schematic of radio transmitter for television channel

circuit designed to afford a high degree of stability. This was connected to the input of the modulating amplifier through two radio-frequency stages, also employing 50-watt tubes. These two stages precluded the possibility of the oscillator frequency being appreciably altered by effects due to modulation. The modulating amplifier employed two 250-watt tubes in parallel and operated on the Heising system. In the standard equipment, the audio stages involve one 50-watt tube and two 250-watt tubes in parallel. To meet the more rigorous requirements of television with an ample factor of safety, this portion of the transmitter was removed from service and a specially constructed three-stage amplifier was substituted. As shown in the drawing, the latter consisted of two 50-watt resistance-coupled stages and a final power stage based on a 5-kilowatt water-cooled tube which raised the signal currents to a power level of approximately one half kilowatt.

In order that it might be possible to check the performance of the radio transmitter under all operating conditions, a suitable monitoring rectifier was constructed and coupled to the output circuit of the radio-frequency power amplifier. A circuit was run back to suitable switches on the television control panel so that either the output of the photo-electric cell amplifiers or the rectified output of the radio transmitter could be impressed on the pilot lamp of the television transmitter. By comparing the two images, it thus became a relatively simple matter to detect any serious maladjustment in the radio apparatus.

The problem of providing a suitable transmitter for the speech channel was rendered quite simple by the fact that at the time there was in process of development at Whippany a 50-kilowatt equipment intended for broadcasting applications. The detailed description of this transmitter is beyond the scope of the present paper. It may be said, however, that it consists of a piezo-electrically controlled master oscillator employing a 50-watt tube directly followed by a 50-watt modulating amplifier. Modulation is by the Heising system, employing one 50-watt and one 250-watt tube in the audio stages. The output of the modulating amplifier is amplified by three push-pull, neutralized, radio-frequency stages the last of which employs six water-cooled tubes at approximately 17,000 volts. This set is capable of delivering 50 kilowatts (unmodulated carrier) to the antenna and during modulation instantaneous peaks approaching 200 kilowatts are attained.

The radio receiver employed at Whippany for the reception of the synchronizing signals at 185 kilocycles presents no features of unusual interest. A double-tuned input circuit was used followed by three stages of radio-frequency amplification, a detector, and two audio stages of conventional design employing transformer coupling. No serious difficulty was encountered in obtaining ample selectivity to insure satisfactory operation in the face of the strong local signals but care was necessary in locating the receiver and in laying out the antenna in order to avoid the inductive type of interference which is almost always experienced in the immediate vicinity of a large radio station. The receiving antenna was located approximately 700 feet from the two transmitting radiating systems.

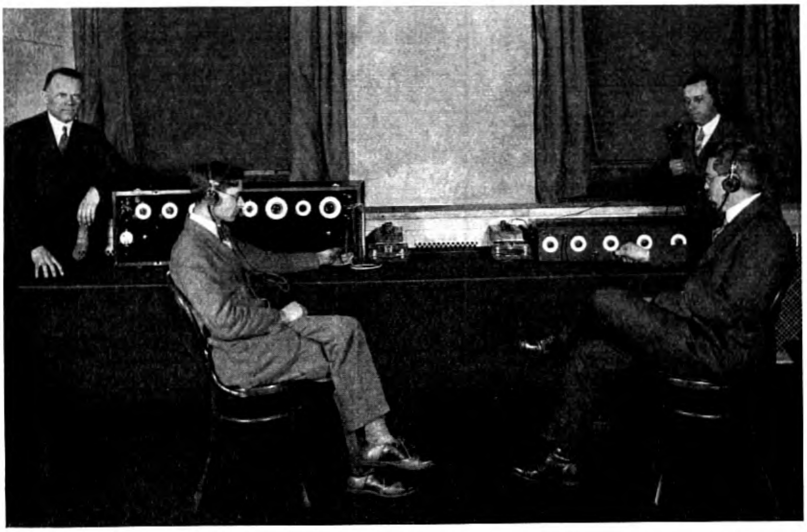


Fig. 5—Radio receiving equipment for the television and speech channels in the auditorium of Bell Telephone Laboratories, New York

The receiver employed at the New York terminus of the television channel presented a somewhat knotty problem on account of the relatively wide frequency band which it was required to pass while providing the maximum discrimination against interference. The width of the required band pointed very definitely toward the superheterodyne. This type of circuit is also very stable, permits of all the amplification that may be needed or that may be employed under ordinary noise conditions, and is very selective against interference immediately adjacent to the desired band. It is quite susceptible, however, to interference from components differing from the desired carrier frequency by an amount approximately equal to the intermediate frequency. If the interfering component lies in the neighborhood of the frequency of the oscillator, beats will be produced which may or may not pass the intermediate-frequency amplifier and the associated filters depending on their design. If the interfering component lies on the opposite side of the wanted carrier from the oscillator and differs from the former by the intermediate frequency, it will be passed by the receiver, subject only to the attenuation due to the radio-frequency circuits (the input circuits tuned to the wanted

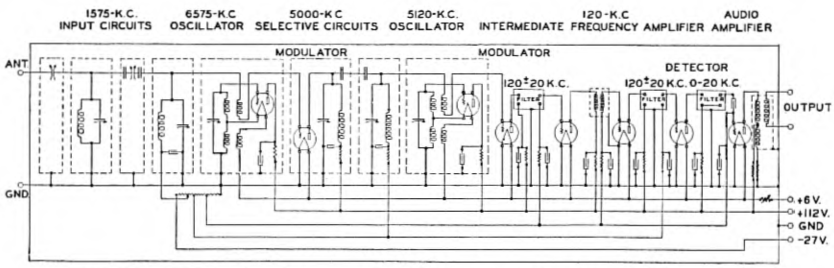


Fig. 6—Schematic of triple detection receiver for television channel

carrier). This characteristic must be given careful consideration in the design of selective receivers of the superheterodyne type and has led to the introduction of carefully designed, loosely coupled, input circuits or an initial tuned radio-frequency stage for this purpose. Neither of these expedients were possibilities in the television receiver, however, because of the extraordinary width of the required transmission band. Recourse was had, therefore, to a triple detection arrangement. Speaking somewhat in the vernacular, the desired signal was "beat up" to 5000 kilocycles where it was passed through sharply tuned coupled circuits, then "beat down" to 120 kilocycles, amplified, filtered and rectified, finally passing through a suitable low pass filter, audio amplifier and output transformer to the television reproducing apparatus.

The circuit arrangement is shown schematically in Fig. 6. Two tuned circuits with capacity coupling were connected to the input of the first detector or modulator. A relatively tight coupling was employed to produce the well-known double-peaked resonance curve capable of affording the required band width. The antenna was not tuned but was loosely coupled to the selective circuits by means of an adjustable capacity. The incoming radio signal was impressed upon the grid of the modulator tube along with a suitable voltage from an oscillator operating at 6575 kilocycles. The 5000-kilocycle components which resulted were selected by means of two carefully designed tuned circuits also capacity coupled. The purpose of this stage in the process will be evident if it is appreciated that at 1575 kilocycles, $\pm$ 20 kilocycles represents a 2.6 per cent band while at 5000 kilocycles the same side frequencies represent only a 0.8 per cent band. In the latter case, therefore, it is possible to employ materially sharper circuits without discriminating against the higher signal components. The 5000-kilocycle circuits connected to the grid of a second detector or modulator tube upon which suitable voltages from a 5120-kilocycle oscillator were impressed. The 120-kilocycle components in the output of this modulator were selected by means of a band-pass filter which worked into a two-stage intermediate-frequency amplifier. A second band-pass filter led to the third or final detector. A 20-kilocycle low-pass filter was employed in the plate circuit of the latter. This filter was designed for a low input impedance at 120 kilocycles in order to meet the necessary condition for efficient rectifier action and it also served as a coupling element for the audio stage which followed. A special output transformer with a permalloy core was provided to step down to the relatively low impedance of the line leading to the television apparatus proper.

A superheterodyne receiver of more conventional design was employed for the speech receiver. The circuit arrangement involved a double-tuned input circuit, one tuned radio-frequency stage, oscillator and modulator, two intermediate-frequency stages, detector and one audio stage. It was highly selective and afforded substantially distortionless transmission for signal frequencies ranging from 50 to 5500 cycles.

The transmitting equipment for the synchronizing channel consisted of a Western Electric 6-A Radio Broadcasting Transmitter modified to operate at 185 kilocycles. In order to avoid the necessity of transmitting directly the 17.7-cycle component required for synchronizing purposes, a 760-cycle carrier was modulated at 17.7 cycles by means of a relay and impressed upon the input of the radio transmitter together with the steady 2125-cycle component. At the receiving end, the 2125- and modulated 760-cycle components were separated by means of suitable filters, and the latter rectified to produce the desired 17.7-cycle current.

Tests of the System

As soon as the various apparatus units could be made ready for service, a comprehensive series of transmission tests was undertaken. In order to determine the relative suitability of the 1575- and 1750-kilocycle bands disclosed by the preliminary survey, transmissions from Whippany at intervals throughout the day were arranged. Field strength measurements were taken at the receiving point employing apparatus of the type described by Englund and Friis [3] and observations on the relative strength of the received signals were made by inserting a sensitive microammeter in the plate circuit of the third detector of the television receiver. These data indicated that the lower frequency band suffered considerably less attenuation and also afforded much more stable transmission. In spite of the comparatively short distance (approximately 22 miles), marked fading was experienced beginning with the sunset period and increasing in amplitude as the night advanced. The high frequency band proved to be particularly disadvantageous in this respect. It was decided, therefore, to fix upon the lower frequency band and to confine the demonstration to the afternoon when reasonably stable transmission conditions prevailed.

Following the choice of a definite operating frequency, a number of modifications were made in the transmitting antenna to improve its efficiency and increase the field strength at the receiver. This work finally resulted in a measured field strength of approximately 2500 microvolts per meter for an antenna input of 5 kilowatts.

Further consideration of the available data on transmission and traffic conditions and the performance characteristics of the apparatus units involved lead to a choice of 1450 kilocycles for the speech channel. In spite of an antenna input of approximately 30 kilowatts, the initial tests at this frequency were very unsatisfactory due to inadequate field strength at the receiver which necessarily resulted in an abnormally high noise level. The height of the antenna was, therefore, increased from 60 to 100 feet by installing iron pipe topmasts. This change brought the field strength at the receiver to approximately the same value as that obtained for the television channel (2500 microvolts per meter) which was considered to be satisfactory for the purpose.

In order to insure that the reproduction of the picture might not suffer from serious discrimination against essential frequencies at some point in the radio system, very careful tests were made on the individual units and on the system as a whole.

The frequency characteristic of the transmitter was determined by connecting a vacuum tube oscillator producing a relatively pure wave to its input terminals through a suitable network involving a thermal milliammeter and an adjustable artificial line. A rectifier of known characteristics and a second thermal meter protected against radio-frequency currents by means of a low-pass filter were coupled to the

[3] "Methods for Measurements of Radio Field Strengths," C. R. Englund and H. T. Friis. Presented to the Spring Convention A. I. E. E. at Pittsfield, Mass., May 25, 1927.

output circuit of the water-cooled tubes. Employing a frequency of 1000 cycles, the input was adjusted to produce normal modulation and the readings of the input and output meters noted. The oscillator frequency was then changed by a convenient amount while holding the input reading constant and the artificial line readjusted, if necessary, to produce constant output current. Under these conditions, any change in the setting of the artificial line indicates an equal variation in the transmission efficiency of the transmitter which is evaluated by this method directly in TU.

The characteristic of the receiver was determined in a similar manner. A low power transmitter of known characteristics was connected to it through a suitable attenuating network which, in so far as the receiver was concerned, simulated the receiving antenna. The radio-frequency input to the receiver was adjusted to approximately the normal value and a series of measurements taken with variable audio-frequency inputs as indicated above.

The overall measurements were also based on a similar procedure impressing a constant input on the 600-ohm input terminals of the transmitter through a suitable artificial line and adjusting the latter to give a constant current into a 600-ohm load at the output of the receiver, taking necessary precautions, of course, to preclude overloading at any point in the system.

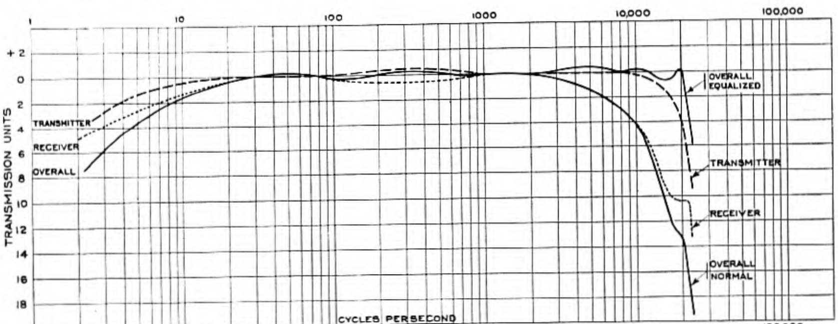


Fig. 7—Measured characteristics of television channel

The experimental characteristic curves thus obtained are shown in Fig. 7, where the abscissæ represent cycles per second and the ordinates departure from the 1000-cycle value in TU. As will be noted, at the lower frequencies exceptionally good performance was obtained, the overall characteristic being only 2 TU down (or deficient) at 10 cycles and only 6 TU down at 3 cycles. The results for the higher frequencies, however, were not so satisfactory, a loss of approximately

13 TU being observed at 20,000 cycles, probably due to the tuned circuits in the receiver. Since modification of these circuits to obtain a flatter characteristic would have been difficult and would have occasioned a noteworthy sacrifice in selectivity, a compensation network was designed for use in the 600-ohm output circuit of the receiver which introduced a negligible loss at 20,000 cycles, a substantially constant loss of 13 TU at frequencies below 2000 cycles, and for intermediate frequencies losses represented by the height of the "normal overall" curve above the horizontal line representing − 13 TU. With this network connected between the receiver and the television equipment, the average level throughout the band was, therefore, reduced some 13 TU but the resulting characteristic as measured beyond the network was that which has been designated "overall equalized." Above 20,000 cycles the characteristics all fell very rapidly which is an indication of the degree of selectivity attained. This was contributed to by the radio-frequency tuned circuits, the band-pass filters in the intermediate-frequency amplifier and the 20,000-cycle low-pass filter between the final detector and audio amplifier. The individual characteristics of the various filters were designed to be 60 TU down 20 kilocycles from the specified cut-off frequency.

Similar measurements were made upon the speech channel but a less thorough study was deemed sufficient in that case due to the existing background of experience.

Effects of Fading

With the system as outlined above, very satisfactory performance was obtained during the afternoon and early evening hours when reasonably stable transmission conditions were prevalent. Later at night, however, when marked fading became evident, some rather unexpected but easily explainable phenomena were observed which may be of sufficient interest to warrant brief mention.

When marked fading occurred, the normally clear reproduction was accompanied by "ghosts" or additional images which faded in and out in an erratic manner, sometimes appearing as positives and sometimes as negatives. The effect was most clearly observed when using one of the various types of test screens employed, a white card bearing a black diamond-shaped outline, approximately a square with its diagonals vertical and horizontal. With this simple type of pattern, it became evident that the secondary images were additional reproductions which were "out of frame" by a greater or less amount. In other words, each of these additional images consisted of a portion of two diamonds placed side by side with the corners just touching. Images "out of frame" along the vertical axis are frequently seen on the motion picture screen.

The explanation is fairly obvious. The present more or less generally accepted view of fading is that it is a manifestation of transmission along two or more paths, at least one of which is variable, producing a continually changing phase relationship between the components and a corresponding waxing and waning of the resultant signal. In the present case, the major image was probably produced by the so-called "ground wave." The secondary images probably resulted from components which were transmitted upward at a relatively sharp angle and turned back to the receiving station from the Heaviside layer, the difference in framing being due to the longer time of transmission.

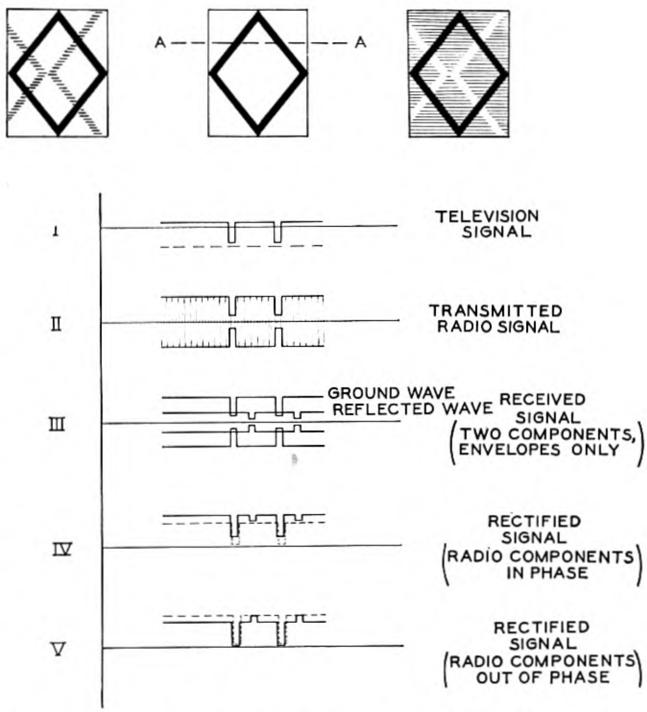


Fig. 8—Production of positive and negative secondary images due to multi-path transmission

The production of negative secondary images is a most interesting phase of the phenomena. This effect may be explained quite easily by means of a series of signal diagrams such as is shown in Fig. 8.

If attention is confined to the interval during which scanning takes place along the line AA, it is evident that the television signal will have the form shown. Amplitudes above the dotted line indicate the current through the photoelectric cell. Since transformer-coupled amplifiers are employed in the television apparatus, however, the direct component is eliminated and the zero axis for the input to the radio transmitter is the solid line. Sketch II shows the modulated output of the radio transmitter. The received signal, shown in III, is assumed to consist of two components, the larger due to the "ground wave," and the smaller due to reflected energy from the Heaviside layer. The latter lags somewhat because of the greater length of the transmission path. The resultant of these two components will necessarily depend on the relative phase of the two carriers at the receiving point. Two cases are considered: when the components are exactly in phase, and when they are exactly out of phase. The effect at intermediate positions may be readily evaluated from these examples. With the components in phase, the detector output is proportional to their sum which is shown in IV. It is evident that this will result in a major image and a secondary positive image. If the components are out of phase, the rectified signal shown in V results. It is simply a matter of subtracting amplitudes. This resultant consists of the desired signal with the amplitude somewhat reduced which will produce a gray background. The secondary image will be formed by the two small peaks shown and will be lighter than the background, in other words a negative.

A pattern frequently observed was the diamond with a cross through its center due to a secondary image. This represents a change in framing of approximately one half line. With 17.7 pictures per second and 50 lines per picture, this corresponds to a difference in transmission time of $1/17.7 \times 1/50 \times 1/2$ or 5.65×10^{-4} seconds. A rough computation of the height of the reflecting layer based on this figure and a distance of 22 miles between transmitting and receiving stations gives 100 kilometers, which is substantially in agreement with determinations made by other methods.

Transmission of Voice and Image with a Common Carrier Frequency

Following the demonstration, a brief series of supplementary tests was arranged to obtain some appreciation on experimental grounds of the problems involved in transmitting both voice and image with a single radio transmitter. The system employed may be considered as the extension of carrier current technique to radio, but has been described in various other terms: "multiplex radio," "double modulation," "the Hammond system," etc. The output of a 30,000-cycle oscillator was modulated with the speech signal. The resulting carrier and sidebands were selected by means of a suitable filter passing frequency components ranging between 25,000 and 35,000 cycles and impressed on the input terminals of the radio transmitter along with the 10 to 20,000-cycle signal from the television apparatus. A suitable low-pass filter was employed in the line to the latter in order to preclude "crosstalk" due to 25,000–35,000-cycle energy working back into the final amplifier stages. The input to the radio transmitter thus consisted of a band extending from 10 to 20,000 cycles together with a 25,000 to 35,000 band, with a particularly strong component at 30,000 cycles representing the low-frequency carrier.

In order that it might be capable of handling this wider band without discrimination, further modifications in the radio transmitter were required. In the case of some of the radio-frequency circuits, which were required to pass a 70,000-cycle band, it was found to be necessary to insert resistance to reduce the sharpness of resonance. On account of lack of time, it was not possible to obtain a complete series of characteristic curves for the transmitter under these conditions. Isolated measurements with a single-frequency input of 35,000 cycles indicated, however, that components of this order could be transmitted without serious loss and the subsequent performance of the system as a whole confirmed this conclusion.

It is well known that if a sinusoidal alternating current $i = I_0 \sin \omega t$ is modulated with a signal of frequency $f = \Phi/2\pi$, the resulting modulated current may be represented by the expression:

$$i = I_0 \sin \omega t + \frac{kI_0}{2} \sin (\omega + \Phi)t + \frac{kI_0}{2} \sin (\omega - \Phi)t,$$

where k is a fraction indicative of the degree of modulation. In other words, a modulated current, or wave, may be resolved into three components: (1) a steady component, known as the "carrier," which has the amplitude and frequency of the original unmodulated current, (2) an "upper sideband" which is equivalent to the signal spectrum with each individual frequency increased by an amount equal to the carrier frequency, and (3) a "lower sideband" which is an inverted reproduction of the signal spectrum, that is, each individual signal component is laid off in the downward direction from the carrier frequency, or subtracted from it. Hence, assuming a carrier frequency of 1575 kilocycles and a signal input to the radio transmitter of the type described above, the antenna current, or the transmitted wave, may be represented diagrammatically as shown in Fig. 9.

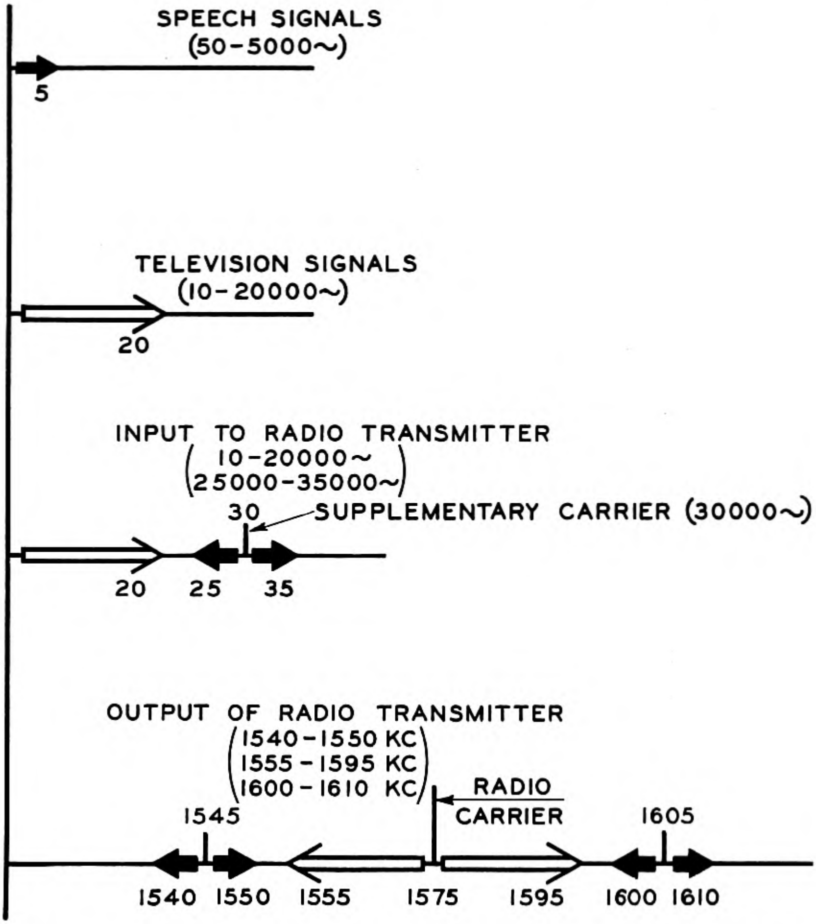


Fig. 9—Diagrammatical representation of frequency conversions in multiplex radio system

It is evident that this type of radio signal can be received by employing an arrangement which will accept the entire band and subject it to rectification in the usual manner. If this is done, the television signal and the 30,000-cycle supplementary carrier modulated with speech will appear at the output of the detector. Branch circuits with suitable filters will enable these two components to be separated and the television signal passed on to the reproducing apparatus. The other component must be rectified to derive the original speech signal, which may then be impressed on the loud speaker amplifiers.

The reception scheme actually employed during the experiments was somewhat different. The television signal was received separately by means of the triple detection set employed for the demonstration. The speech signal was received in a similar manner employing the set utilized for the speech channel during the demonstration. This latter receiver was tuned to 1545 kilocycles. That reception in this manner is feasible, is evident from the diagram. The 1540–1550-kilocycle zone contains two speech sidebands and a carrier of 1575 − 30 or 1545 kilocycles. It is quite possible, therefore, to demodulate in one step, instead of "beating" the various components against the main carrier (1575 kilocycles) to produce a 30-kilocycle supplementary carrier which must be rectified a second time to derive the speech signal. The 1600–1610-kilocycle band was ignored. The receivers were sufficiently selective that, with the 5-kilocycle interval which existed between the two bands, no noteworthy crosstalk was experienced.

The results obtained in this manner were not as satisfactory as those to be had with the other system described. This can be attributed to two factors, both concerned with the transmitting apparatus: (1) In order to transmit both signals with the same transmitter, that is, the same vacuum tubes, the individual current amplitudes had to be reduced to at least one half, resulting in too weak a radio signal to clear the prevailing noise levels in New York, (2) In spite of the reduced amplitudes, a certain amount of inter-modulation was experienced in the transmitter which resulted in "crosstalk" between the channels. Notwithstanding these deficiencies, however, it was possible to recognize the speaker and to understand his remarks; but a short time ago, the performance would have been considered a very noteworthy achievement.

Experiments of this nature, although not new, are of particular interest where television is concerned, since, as Dr. Ives has indicated, the logical trend of development is toward a finer picture structure involving the transmission of much wider frequency bands, or what is more likely, the use of parallel scanning schemes and multi-channel transmission. The work, while necessarily somewhat cursory, may, therefore, be of value in affording an indication of the significance of multi-channel radio transmission in this connection. From a popular standpoint, these tests have been described as the transmission of both voice and image "on a single wave-length." To what extent this statement falls short of actually representing the facts in the case is obvious from Fig. 9. It will be seen that a wider frequency band is actually employed with this system than was required for two separate channels. Furthermore, this wider band is much less effectively utilized. Two bands are required for the voice channel in place of one. At the receiver, one of these bands was disregarded. To have received both would have required apparatus accepting twice the band width and the gain in signal would have been offset by the corresponding increase in noise level. For all useful purposes, therefore, the energy radiated in the form of the second band is wasted.

To proceed further with a discussion of multi-channel radio transmission is beyond the scope of the present paper. Whatever the system employed, however, one conclusion illustrated by these experiments may be pointed to with confidence: television by radio requires a discrete and fairly wide frequency band. Hence the frequently predicted introduction of television as an adjunct to radio broadcasting without extensive changes in existing channel arrangements is extremely unlikely.

Contemporary Advances in Physics. XIV. Introduction to Wave-Mechanics

By KARL K. DARROW

IN a period when a limited domain of physical phenomena is exciting wide fervent interest and commanding intensive study, and continues for years to monopolize the attention of many brilliant theorists, sometimes it is the fortune of an ingenious mind to express or interpret or picture the already-discovered laws in a new way which makes so greatly favourable an impression, that in a moment it sweeps its rivals from the field. The new theory may not lead to more or better agreements with experience than did its predecessors; it need not make predictions which they were incapable of making; its mathematical processes may be identical with theirs, the old symbols reappearing with new names in the old equations. Contrariwise it may be born well endowed with these advantages which normally decide the contest between old theories and new, yet owe its victory not to them at all. It triumphs because it seems natural or sensible or reasonable or elegant or beautiful—words said of a theory which fulfils some deep-seated demand or evades some deep-rooted prejudice in the minds of its judges. Later its vogue may pass, not through the disclosure of any intrinsic defect, but because the physicists of the rising generation do not share the prejudices and the predilections of those who first applauded it. The kinetic theory of gases was welcomed by a generation which wished to believe in atoms; the electromagnetic theory by people prejudiced against the notion of action at a distance; the quantum-theory has always had to do battle against those who yearn for continuity in their images of Nature, and the theory to which these pages are devoted has captivated the world of physics in a few brief months because it seems to promise a fulfilment of that long-baffled and insuppressible desire.

Wave-mechanics being a new way of interpreting a vast field of well-known phenomena, it is unnecessary as indeed it would be impossible for me to recite in this place everything which the new theory is meant to explain. A few years hence, indeed, we may recognize in certain phenomena only newly or not yet discovered the securest basis for the new conceptions; but for the present, any adequate description of the facts on which Bohr's atom-model is based is nearly sufficient. I will recall only that the cardinal and dominant facts of the field which is the hunting-ground of the present generation of theorists are these: atoms exist in Stationary States—they emit or absorb radiation in passing from one of these States to another—the frequency of the radiation is proportional to the energy-difference between the two States from one to the other of which the atom passes. Moreover, for certain kinds of atoms and molecules there are empirical formulæ which express known interrelations among the energy-values of the various Stationary States. These in brief are the major facts to be explained.

Bohr proved that the energy-values of the Stationary States of the hydrogen atom could be reproduced by affirming, *first*, that the atom consists of an electron and a nucleus of known masses and equal and opposite known charges; *second*, that these revolve around their common centre of mass according to the classical laws of mechanics and without radiating energy; *third*, that among all the conceivable orbits which such particles might describe there are certain ellipses, distinguished by certain especial and peculiar features, which alone the particles are permitted to choose—to each "permitted" ellipse there corresponds a Stationary State, and each Stationary State may be visualized as a permitted ellipse.

The first of these assumptions has never since departed from the physicists' world-pictures. In wave-mechanics it is still implicit, though easily overlooked. The second and third have not so firm a foothold. As I have elsewhere remarked, they are and always will be as good as they ever really were. If we make the first two of Bohr's assumptions, then it follows as a matter of course that whichever Stationary State of the hydrogen atom we may wish to consider or may hereafter discover, we shall always be able to find an elliptic orbit with the proper energy-value to serve as its picture. Yet this alone is not an important fact; the serious question is, whether the family of all permitted elliptic orbits is set apart from the vast multitude of forbidden ones by some simple and striking distinction which they all share and none of the rest possesses, whether they rejoice in some intrinsic patent of aristocracy. At first it seemed so; now, however, it turns out that the distinctive feature which originally was supposed to ennoble just the orbits required to account for the Stationary States, and no others, is not perfectly suited to every case. This weakened the prestige of the elliptic orbits; and though the introduction of the Spinning Electron has done much to save the situation, it has not done enough to preserve them from the crescent disparagement of those who never really liked them.

With other atoms and with molecules, the situation is much the same. Bohr and his successors visualized atoms as groups of electrons surrounding nuclei; diatomic molecules, as paired nuclei surrounded by their jointly shared electron-family, capable of revolving like a dumbbell around their centre of mass and of vibrating like the two ends of a spring along their line of centres. These pictures persist in wave-mechanics; but the permitted vibration-amplitudes, the permitted rotation-speeds, and the permitted electron-orbits adduced to symbolize the Stationary States languish for the moment in the same discredit as the permitted elliptic orbits of the hydrogen atom.

Meanwhile, the humiliation of the electron-orbits accentuates the grave defect of the original atom-model of Bohr. That model offered nothing to interpret the fact that when an atom passes between two Stationary States of energy-values (let me say) E_i and E_j, it emits (or absorbs) radiation of the precise frequency $(E_i - E_j)/h$, the quotient of the energy-difference by the notorious constant of Planck. Neither in the initial State nor in the final State are the constituent parts of the atom-model vibrating with this frequency (except in occasional untypical cases). The frequencies of the waves streaming out from the atom do not agree with the frequencies of the motions assumed to exist inside the atom—a very uncomfortable idea, altogether discordant with all our experience of sound and electrical circuits.

If it should be found possible to incorporate into the atom-model something vibratory, having for its vibration-frequency the quotient of the energy-value of the then-existing Stationary State by Planck's constant: then in the foregoing case this "something" would be vibrating initially with frequency E_i/h and finally with frequency E_j/h, and the frequency of the emitted radiation would be the heterodyne or beat-frequency of these two. This is an agreeable idea; and wave-mechanics offers it. If then it should be found possible to arrive at the energy-values of the Stationary States by imposing conditions upon this vibrating entity instead of the electron-orbits, we should achieve as much as the electron-orbits enable us to achieve, and have the foregoing advantage also, and perhaps others as well. This is what wave-mechanics promises.

To this introduction I wish to join two warnings before plunging into the exposition. In the first place, wave-mechanics has several aspects, and may be approached from several directions; the one which I have chosen for this article is not the one which de Broglie elected nor the one which Schroedinger prefers.[1] In the second place, wave-

[1] I suspect that the method of exposition which I shall follow is the one which Schroedinger meant when he wrote "I had originally the intention of establishing the new formulation of the quantum-conditions in this more visualizable (*anschaulich*) way, but preferred a neutral mathematical form, because it makes the essence clearer." Schroedinger himself stresses the formal likeness between ordinary me-

mechanics is yet incomplete. It has been applied with success to many problems, but there are situations—those involving the Spinning Electron, for instance—in which the way to apply it is not yet clear, and many theorists are groping. The new theory is still plastic; many minds, perhaps the hands of many experimenters, have yet to work upon it before it is molded into its final shape.

Classical Mechanics and Wave-Mechanics

The underlying principles of "classical" or "Newtonian" mechanics may be stated in several alternative ways, each of which is especially well adapted to certain particular classes of problems. The most familiar of the statements is Newton's own. Unfortunately, it is another and less current which is the most expedient for the problems with which we have to deal. This formulation I will derive from Newton's, by imagining a particular extremely simple mechanical system and using Cartesian coordinates.

Conceive then a particle of mass m and charge e, moving in an electrostatic field of which the potential is a function $U(x, y, z)$ of the coordinates.[2] Its momentum is a vector of which the components are $m\dot{x}$, $m\dot{y}$, $m\dot{z}$. These are called the *momenta with respect to the coordinates* x, y, z, and are designated by p_x, p_y, p_z. The force upon the particle is the negative of the product of e into the gradient of the potential, a vector of which the components are dU/dx, dU/dy, dU/dz.

Newton's way of stating the underlying principles of mechanics then gives:

$$dp_x/dt = \dot{p}_x = -\,e dU/dx; \qquad \dot{p}_y = -\,e dU/dy, \qquad \dot{p}_z = -\,e dU/dz. \quad (1)$$

Multiplying the members of these three equations by $\dot{x}$, $\dot{y}$ and $\dot{z}$ respectively, and adding, we find:

$$\frac{d}{dt}\frac{1}{2}m(\dot{x}^2 + \dot{y}^2 + \dot{z}^2) = -\,e\left(\frac{dU}{dx}\frac{dx}{dt} + \frac{dU}{dy}\frac{dy}{dt} + \frac{dU}{dz}\frac{dz}{dt}\right). \quad (2)$$

On the left we have the rate of change of the kinetic energy T of the particle as it travels along its path. To interpret the right-hand

chanics and geometrical optics on the one hand and wave-mechanics and diffraction-theory on the other. I have not yet found this comparison helpful, and therefore cannot present it in a convincing manner.

I wish to acknowledge the valuable assistance of my colleague Mr. L. A. MacColl in preparing the mathematical portions of this paper.

[2] The reader will doubtless recognize that I am leading up to the case of the electron traveling in the field of a nucleus; I must therefore recall that in the case of the electron the charge e is intrinsically negative, and that according to the classical electromagnetic theory equation (1) should contain a term describing the reaction of the emitted radiation upon the electron—a term which is omitted in all contemporary atomic theories.

member, introduce a symbol V to designate the value of U at the locality where at any moment the particle actually is, multiplied by $+ e$; this is the potential-energy-function of the particle, and the right-hand member of (2) is its rate of change. Therefore:

$$\frac{d}{dt}(T + V) = 0,$$

$$T + V = \text{constant} = E. \tag{3}$$

The constant E is (by definition) the energy. As the behavior of the particle depends upon the field, the ensemble of particle and field should be considered as one entity, the *system*, of which kinetic energy T and potential-energy-function V and total energy E are properties.

To bring out the next feature, I take the still more specific case of a particle of charge e and mass m moving in the inverse-square central field of a "nucleus," an immobile point-charge equal in magnitude and opposite in sign to the electron-charge. Using Cartesian coordinates with the origin at the nucleus, we have $V = - e^2/\sqrt{x^2 + y^2 + z^2}$; using polar coordinates,[3] we have $V = - e^2/r$. It is obvious that polar coordinates permit a much simpler expression for V than do Cartesians; on the other hand, they entail a distinctly more complicated expression for T. The proper choice of coordinates is often a vital question. For a few paragraphs I will carry along the reasoning in both coordinate-systems. The underlying equation (3) becomes, in the one and in the other:

$$\tfrac{1}{2}m(\dot{x}^2 + \dot{y}^2 + \dot{z}^2) - e^2/\sqrt{x^2 + y^2 + z^2} = E, \tag{4a}$$

$$\tfrac{1}{2}m(\dot{r}^2 + r^2\dot{\theta}^2 + r^2 \sin^2 \theta \cdot \dot{\varphi}^2) - e^2/r = E. \tag{4b}$$

In these equations, we have the potential-energy-function expressed as a function of the *coordinates* (x, y, z or r, θ, φ) and the kinetic energy expressed in terms of the coordinates and the *velocities* ($\dot{x}$, $\dot{y}$, $\dot{z}$ or $\dot{r}$, $\dot{\theta}$, $\dot{\varphi}$). It is desirable to express the kinetic energy in terms of the coordinates and the *momenta*. We have already met the momenta in Cartesian coordinates, the quantities $m\dot{x}$, $m\dot{y}$, $m\dot{z}$. It is obvious that they are the derivatives of the expression for the kinetic energy with respect to the velocities, always in Cartesian coordinates:

$$p_x = dT/d\dot{x}; \qquad p_y = dT/d\dot{y}; \qquad p_z = dT/d\dot{z}. \tag{5}$$

The momenta in any other coordinate-system are defined in the same

[3] The equations of transformation are: $x = r \sin \theta \cos \phi$, $y = r \sin \theta \sin \phi$, $z = r \cos \theta$.

way; first the kinetic energy is expressed as a function of the velocities, then differentiated with respect to these. In polar coordinates

$$p_r = dT/d\dot{r} = m\dot{r}; \qquad p_\theta = dT/d\dot{\theta} = mr^2\dot{\theta}; \qquad p_\varphi = dT/d\dot{\varphi} = mr^2 \sin^2\theta\cdot\dot{\varphi}. \tag{6}$$

Expressing in the equations (4a) and (4b) the kinetic energy in terms of the coordinates and momenta, we have

$$\frac{1}{2m}(p_x^2 + p_y^2 + p_z^2) - e^2/\sqrt{x^2+y^2+z^2} = E, \tag{7a}$$

$$\frac{1}{2m}\left(p_r^2 + \frac{1}{r^2}p_\theta^2 + \frac{1}{r^2\sin^2\theta}p_\varphi^2\right) - e^2/r = E. \tag{7b}$$

Whenever in any problem the kinetic energy and the potential energy of the system are given as functions of coordinates and momenta, the problem is prepared for treatment by the methods of classical mechanics.

To make the next step, we consider the function $L = T - V$, the difference between the kinetic energy and the potential-energy-function of the particle, a function of the particle as it travels along its path in the force-field:

$$L = T - V = 2T - E \tag{8}$$

and the time-integral of this function

$$W = \int Ldt = \int 2Tdt - Et. \tag{9}$$

Into the expression for W, insert explicitly the expression for kinetic energy in Cartesian or in polar (or in any other) coordinates:

$$W = m\int(\dot{x}^2 + \dot{y}^2 + \dot{z}^2)dt - Et = m\int(\dot{x}dx + \dot{y}dy + \dot{z}dz) - Et, \tag{10a}$$

$$W = m\int(\dot{r}^2 + r^2\dot{\theta}^2 + r^2\sin^2\theta\cdot\dot{\varphi}^2)dt - Et = m\int(\dot{r}dr + r^2\dot{\theta}d\theta + r^2\sin^2\theta\cdot\dot{\varphi}d\varphi) - Et. \tag{10b}$$

From all of this it follows that

$$p_x = dW/dx, \qquad p_y = dW/dy, \qquad p_z = dW/dz, \tag{11a}$$

$$p_r = dW/dr, \qquad p_\theta = dW/d\theta, \qquad p_\varphi = dW/d\varphi, \tag{11b}$$

and in general, *the momenta belonging to any coordinate-system are the derivatives of the function W with respect to the coordinates.*

Into the fundamental equation (7a) substitute these expressions for the momenta, and obtain:

$$\frac{1}{2m}\left[\left(\frac{\partial W}{\partial x}\right)^2 + \left(\frac{\partial W}{\partial y}\right)^2 + \left(\frac{\partial W}{\partial z}\right)^2\right] + V(x, y, z) = E \tag{12}$$

or, seeing that the quantity $\sqrt{(\partial W/\partial x)^2 + (\partial W/\partial y)^2 + (\partial W/\partial z)^2}$ is the magnitude $|\nabla W|$ of the *gradient* of the function W, the gradient of a function being a vector well known in vector-analysis and denoted by prefixing the sign ∇ or the abbreviation *grad* to the symbol of the function:

$$|\nabla W|^2 = 2m(E - V). \tag{13}$$

This equation governs the space-derivatives of the function W; it is complemented by the equation derived from (9) which governs the time-derivative of W:

$$\partial W/\partial t = -E. \tag{13a}$$

At this point the procedure of classical mechanics and the procedure of wave-mechanics diverge from one another.

Were we to follow the classical procedure, we should perform certain integrations and other processes, and arrive in the end at equations describing trajectories or orbits—in the particular case of an inverse-square central force-field, at equations describing elliptical orbits. The particular elliptical orbit to which the reasoning would conduct us would be determined by the value which had originally been assigned to the energy E, and the values which we attributed to the various constants of integration supervening in the course of the working-out. The function W, having served its purpose, would have vanished from the scene, leaving with us the electron swinging in its orbit within the atom or the planet in its orbit across the heavens.

The procedure of wave-mechanics, however, is based upon the observation that the equations (13) and (13a) together are *the description of a family of wave-fronts, traveling with the speed* $E/\sqrt{2m(E - V)}$ *through space.*

To display this aspect of the equation, let it be supposed at some prescribed time-instant t_0 the function W has a certain prescribed constant value W_0 at every point of a surface S_0; for instance, that at time $t_0 = 1$ it is equal to unity all over the sphere of unit radius centered at the origin. It is to be shown that at a slightly later instant $t_0 + dt$ there is again a surface everywhere over which the value of W is W_0, this not however being the same surface S_0, but another—a surface S_1 so placed that from any point P_0 on S_0 the shortest line to S_1 is perpendicular to S_0 and its length is $(E/\sqrt{2m(E - V)}dt)$.

This is easily shown. Imagine a vehicle[4] which at the instant t_0

[4] I use this word instead of "particle" lest this entity be confused with the moving electron to which the foregoing equations relate. The electron does travel along a curve normal to the surfaces of constant W, not however with the speed u about to be defined, but with a different speed related to u in a curious and significant way (cf. the allusion on p. 695).

is traveling through P_0, along the line normal to S_0, with a speed to be designated by u. At the instant $t_0 + dt$ it occupies a locality where the value of W is given by the formula:

$$\begin{aligned} W_0 + dW &= W_0 + |\nabla W| ds + \left(\frac{\partial W}{\partial t}\right) dt \\ &= W_0 + u|\nabla W| dt - E dt, \end{aligned} \tag{14}$$

for in the time-interval dt it travels over a distance $ds = udt$ along the normal to the surface S_0, and along this normal the slope of the function W is equal to $|\nabla W|$, and meanwhile at each point of space W is varying directly with time by virtue of the term $-Et$ occurring in the equation (9) which defines it. Now if the imaginary vehicle happens to be moving with just the speed defined by the equation

$$u = E/\nabla W = E/\sqrt{2m(E - V)}, \tag{15}$$

the coefficient of dt in equation (14) vanishes; that is, the vehicle as it moves outward keeps up with the prescribed value of W; but this is the same thing as saying that the value given for u in (15) is the speed of the wave-front.

At this (if not an earlier) stage of the argument, one begins to wonder what W "really is"; one turns back to seek the original definition of this artfully constructed function, so suddenly advanced from an auxiliary to the central rôle of the theory; one tries to grasp it, to form an image of it. I can do little to satisfy this very human craving. I can point out that W is that quantity "action" with which the Principle of Least Action has to do; this feature scarcely makes it more conceivable, but at least enhances its prestige. I can point out that since no one has ever seen what moves or is inside an atom, the conception of waves in an intangible medium curling and flowing around a centre is no more far-fetched than the conception of intangible particles sailing in ellipses around a nucleus. (To this one can reply that the planets in their courses supply a visible analogue for the notion of revolving electrons, but no one has seen in the sky the wave-fronts of the function W.) I can point out that for some important purposes, notably the prediction of the Stationary States, it makes no difference what the function W "really is"—no more difference than it makes to the solver of a quadratic equation whether the variable be called x or t, whether in the mind of the propounder of the equation it stood for distance or for time. One might in fact begin with the forthcoming equation (20) as foundation, laying it down without introduction or apology; yet there must be deep-lying interconnections between the classical mechanics and the new, which such a procedure might mask. I can refer the reader to Schroedinger's own attempts to interpret W, some of which will figure in the last section of this article; or I can invite him to grow his own conception of W. This last in fact is what I will do.

Now if it is proposed to regard the fundamental dynamical equation (13) as the description of a family of wave-fronts perpetually wandering through space with the speed $E/\sqrt{2m(E - V)}$—and this is precisely what is proposed—then the description is obviously incomplete; for it omits to state the wave-length of these waves or the frequency of whatever be the vibrating thing which manifests itself by the waves, and indeed if the frequency were separately stated there would be no place for it in such an equation as (13). That equation, in fact, may be compared with the bare statement that the ripples traveling over the water of a pond from the place where a stone fell in are circles expanding at a given speed, or that the sound-waves proceeding through air from a distant source are plane waves traveling about 340 metres per second. To describe the ripples or the sound-waves completely it is essential to discover some ampler equation; a like extension is necessary here.

In treating familiar vibrating mechanical systems, stretched strings and tensed membranes and the like, it is customary to employ the general Wave-Equation

$$u^2\left(\frac{d^2\Psi}{dx^2} + \frac{d^2\Psi}{dy^2} + \frac{d^2\Psi}{dz^2}\right) \equiv u^2\nabla^2\Psi = \frac{d^2\Psi}{dt^2} \tag{16}$$

in which ∇^2 stands for the Laplacian differential operator (page 671); Ψ stands for the sidewise displacement of the string or distortion of the membrane or whatever it is that is transmitted as a wave; and u for the speed of propagation of the wave. It is furthermore customary to supplement this by the equation

$$d^2\Psi/dt^2 = -4\pi^2\nu^2\Psi, \tag{17}$$

in which ν stands for the frequency of the vibration; combining which with (16), one obtains

$$\nabla^2\Psi + k^2\Psi \equiv \nabla^2\Psi + \frac{4\pi^2\nu^2}{u^2}\Psi \equiv \nabla^2\Psi + \frac{4\pi^2}{\lambda^2}\Psi = 0, \tag{18}$$

in which $\lambda = u/\nu$ stands for the wave-length of the wave-motion.

All of these matters will be developed at length in the following section. At this point it is necessary only to return to the description of the wave-motion partially but only partially described by (13), and complete it by the assertion—not an inevitable nor a self-evident assumption, but an original and daring hypothesis—that it is indeed a wave-motion endowed with a frequency, and this the frequency

$$\nu = E/h. \tag{19}$$

This manner of introducing into every mechanical system a vibration-frequency linked with its energy by the vital quantum-relation (19) was the invention of Louis de Broglie.

The wave-equation to which this hypothesis leads us then is:

$$\nabla^2\Psi + \frac{8\pi^2 m}{h^2}(E - V)\Psi = 0. \tag{20}$$

This is a particular form of the wave-equation of de Broglie and Schroedinger. It is the form which I will use throughout this article, for it is adequate to the first steps in the processes of atom-design—adequate, for instance, to supply a theory of the major features of the spectrum of atomic hydrogen, though not of its fine-structure; adequate also to interpret the data of the experiment of Davisson and Germer, and sufficient for an introduction to the ways of thinking which constitute wave-mechanics. Nevertheless it is certainly not the general wave-equation, for it is subject to at least two limitations.

The first of these is, that equation (20) is based upon Newtonian, not upon relativistic mechanics. We should therefore expect it to be valid only for slow-moving particles, to be the limiting form of a relativistic wave-equation appropriate to all velocities. Such an equation, indeed, was the first propounded by de Broglie. The past history of atomic theory suggests that we should need it when embarking upon the enterprise of explaining the fine-structure of the hydrogen spectrum. The latest developments in that history, however, indicate that the mere replacement of equation (20) by its relativistic analogue would not suffice for that enterprise; due allowance must be made in addition for the "spin" of the electron.[5] Wave-mechanics being yet too young to have furnished an answer to this twofold problem, the relativistic equation still wants what may in the end turn out to be its main experimental support. Yet it can scarcely be doubted that relativity must figure in the general wave-equation.

The second limitation upon equation (20) is due to its origin in

[5] For the application of the relativistic equation to the hydrogen atom without allowance for the spinning electron, see V. Fock, *Zs. f. Phys.*, **38**, pp. 262–269 (1926). See also the first footnote on p. 688.

equation (13), and to a peculiar feature of that equation—to the fact that in it the magnitude of the gradient of W stands equated to a function of the coordinates. This indeed is the feature which rendered it possible to imagine flowing waves. Now this feature occurs because the system to which equation (13) relates—the particle voyaging in a force-field—has a *kinetic-energy-function which is the sum of the squares of the momenta* (multiplied by a constant). Had we presupposed a system possessing a kinetic-energy-function not capable of being so expressed—two particles of different masses voyaging in a force-field, or a rigid rotating body of irregular shape, for example—the equation which we should have obtained in lieu of (13) would not have had the peculiar feature aforesaid; the wave-picture would not have offered itself, much less the equation (20) which was superposed upon the wave-picture. It is precisely at this obstacle that the mode of thought known as *non-Euclidean geometry* proves itself useful. It proposes equations of a general type, which can be written down for every system of which the kinetic-energy-function is preassigned, and which for the single particle floating in a force-field become the equation (13) and (20). In the language of non-Euclidean geometry, even the words and the symbols for *wave* and *wave-speed* and *gradient* and *Laplacian* are preserved; but whether they are advantageous to anyone not already versed in this subject may well be doubted. Suffice it to say, that non-Euclidean geometry provides a general equation [6] of which (20) is a special case, and that the general equation has already justified its existence by its successes in dealing with certain atom-models and molecule-models such as the rigid rotator used in the study of band-spectra. But the question as to what the waves "really are" becomes in these cases all the darker and more perplexing.

One further step, and we attain to the idea on which the calculation of the energy-values of the Stationary States reposes.

It is very well known that a medium capable of transmitting waves, and *bounded* in certain ways, may develop what are variously known as standing waves—stationary wave-patterns—the phenomena of resonance. Air enclosed in a box, a string pinched at the ends, a membrane clamped around its circumference, the mobile electricity in

[6] Let the kinetic-energy-function of the system, expressed in terms of the co-ordinates and velocities, be written

$$T = \sum_i \sum_j Q_{ij}\dot{q}_i\dot{q}_j$$

and let Δ stand for the Laplacian operator in the non-Euclidean configuration-space of which the metric is $ds^2 = \Sigma\Sigma Q_{ij}dq_idq_j$; then the general wave-equation of de Broglie and Schroedinger is:

$$h^2\Delta\psi + 8\pi^2(E - V)\psi = 0.$$

a tuned circuit—each of these vibrates in a wave-pattern of "nodes" and "loops" if the frequency of vibration imposed upon it conforms to one of its own "natural frequencies" or "resonance frequencies." To each of these natural frequencies corresponds a particular pattern of loops and nodes; when one of them is impressed upon the medium, its corresponding wave-pattern springs into existence, and would continue forever were it not for friction internal or external. When any frequency not agreeing with one of the resonances is imposed upon the bounded medium, the resulting motion is very much more complicated. The calculation of these natural frequencies, the mapping of these vibration-patterns, is performed by using the methods of one of the great divisions of mathematical physics—the methods underlying the Theory of Acoustics.

May the Stationary States, then, of a natural atomic system be visualized as stationary wave-patterns such as these, and their energy-values as the products of the natural frequencies by the constant of Planck? Are the problems of atomic theory to be solved by devising atom-models imitated after familiar resonant bodies or tuned circuits, and applying to these "acoustic models" the mathematical technique of the Theory of Acoustics? This idea was developed by E. Schroedinger.[7]

Familiar Examples of Stationary Wave-Patterns

To display the laws governing wave-patterns, I will develop three examples: the stretched string, the tensed membrane, the ball of fluid confined in a spherical shell. The first of these is the simplest and most familiar of all instances; excursions into the theory of vibrating systems commence always at the wire of the piano and the string of the violin. Physically, this is a case of one dimension (distance, measured along the length of the string); mathematically, it is a case of two variables (that distance, and the time). The example of the tensed membrane is not unfamiliar in the practice of telephony, though many of the diaphragms of actual instruments are too thick to be considered such; for a membrane is, by definition, infinitely thin. It is a case of two dimensions and three variables. It will reveal to us the desirability of choosing for each specific problem its appropriate set of coordinates; and we shall observe what happens when one of the chosen coordinates is cyclic, being an angle which for all practical purposes returns to its original value when increased by 2π; and we

[7] Since the present article is based henceforth chiefly on Schroedinger's publications, I wish to make particular reference here to works embodying de Broglie's contributions: his own *Ondes et mouvements* (Paris, Gauthier-Villars, 1926) and article in *Jour. de Phys.* (6), **7**, pp. 321–337 (1926); L. Brillouin, *ibid.*, pp. 353–368.

shall encounter functions not so widely known as the simple sine and cosine which suffice for the case of the stretched string. The little-known example of the ball of fluid, with its three dimensions and four variables, will repeat these lessons, and will serve as the final stepping-stone to the wave-motions imagined by de Broglie and by Schroedinger. To proceed to these, it will suffice to imagine strings and fluids not uniform like those of the simple theory of vibrating systems and sound, but varying from point to point in a curious and artificial way.

Example of the Stretched String

Imagine a stretched string, infinitely long, extended along the x-axis of a system of coordinates. Designate the tension in the string by T, the (linear) density of the string by ρ. To derive the differential equation governing the motion, conceive the string as a succession of short straight segments (Figure 1). Each segment exerts upon its neighbors a force, which is the tension in the string. When the string lies straight along the axis of x, each segment lies in equilibrium between the equal and opposite forces which its neighbors exert upon it. When however the string is drawn sidewise (remaining, we shall suppose, in the xy-plane) the neighbors of each segment are oblique to it and to one another, the forces which they exert upon it have components along the y-direction. These components are in general unequal, and their algebraic sum is a force urging the segment along the y-direction. Denote by dx the length of such a segment, by y its lateral displacement, by θ the angle between it and the axis of x; so that $dy/dx = \tan\theta$, and ρdx stands for the mass of the segment. The resultant force upon the segment is given by:

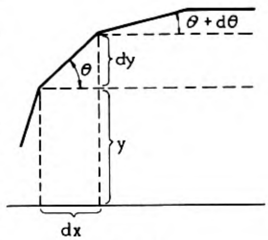


Fig. 1

$$\begin{aligned} F &= T[\sin(\theta + d\theta) - \sin\theta] = T[\tan(\theta + d\theta) - \tan\theta] \\ &= T\cdot d(\tan\theta/dx)dx = T(d^2y/dx^2)dx \end{aligned} \tag{101}$$

to the degree of approximation to which the difference between $\sin\theta$ and $\tan\theta$ may be neglected.[8]

Equating this to the product of mass by acceleration, we obtain:

$$\rho d^2y/dt^2 = T(d^2y/dx^2) \tag{102}$$

[8] This is the degree of approximation all but universal in the theory of vibrating systems and sound. The conclusions from this theory are therefore strictly valid only in the limit of infinitesimal displacements or distortions.

or using dots to symbolize differentiation with respect to time, and dashes to represent differentiation with respect to space:

$$\ddot{y} = \sqrt{T/\rho}\cdot y''. \tag{103}$$

This equation, a linear combination of a second derivative with respect to time and a second derivative with respect to space, is the first and simplest of our wave-equations.

It is called a wave-equation, because it may represent—it does not necessarily represent, but it may—a shape or a figure or a distortion of the string (whichever one may choose to call it) which travels continually and indefinitely along the string with a constant speed.

To illustrate this possibility, let us suppose that at the time $t = 0$ the string is distorted into a sinusoidal curve described by the equation:

$$y = A \sin mx \quad \text{at} \quad t = 0 \tag{104}$$

and that its points are moving parallel to the y-axis with speeds described by the equation:

$$\dot{y} = nA \cos mx \quad \text{at} \quad t = 0. \tag{105}$$

At any other moment t, the configuration of the string is described by the equations:

$$y = A \sin (nt + mx), \qquad \dot{y} = nA \cos (nt + mx), \tag{106}$$

for these satisfy the differential equation which underlies the whole theory, and they satisfy also the "initial conditions" specified by (104) and (105). They satisfy these equations, that is to say, provided that a certain relation is fulfilled among the constants n and m, and the quantities T and ρ which describe the physical nature of the stretched string; this relation being:

$$n/m = \sqrt{T/\rho}. \tag{107}$$

If this relation is fulfilled, the condition of the string throughout all time is described by the equations (106).

Examining these equations, we perceive that they signify that the values of displacement and speed, which at the time $t = 0$ existed at any point x_0 on the string, are at any other time t to be found at the point $x_1 = x_0 - (n/m)t$. These values are moving steadily along the string; the whole configuration of the string, its sinusoidal shape and its transverse velocities, is slipping steadily lengthwise in the direction of decreasing x—the shape of the string is being transmitted as a wave, with the ratio of the constants n/m for its speed of propagation u:

$$u = n/m = \sqrt{T/\rho}. \tag{108}$$

This result justifies the title *wave-equation* for the differential equation (103), and the meaning *speed of propagation* for its coefficient $\sqrt{T/\rho}$.

The reader will scarcely have failed to notice, however, that the result was obtained only by prescribing very sharply defined physical conditions. The string was supposed infinitely long; it was supposed distorted into the form of a sine-wave; the transverse speeds of its successive particles at the instant $t = 0$ were preassigned as rigorously as their positions. Were we to alter this last specification, we should arrive at very different results. If for instance we should make the assumption that at $t = 0$ the string is distorted into a sine-wave *and is stationary*, the equations (106) would not be adequate to describe what happens. We should then be forced to have recourse to a more general solution of the differential equation:

$$y = C \sin (nt + mx) + D \sin (nt - mx) \tag{109}$$

and to adjust the constants C and D so as to conform to the newly prescribed initial conditions, which are:

$$y = A \sin mx, \qquad \dot{y} = 0 \quad \text{at} \quad t = 0. \tag{110}$$

The adjustment is attained by making $C = D = \frac{1}{2}A$, whereupon we get:

$$y = A \sin nt \cos mx, \tag{111}$$

an equation which describes not a wave advancing perpetually along the string, but a stationary oscillation with nodes and loops of vibration, like those which a violin-string properly bowed exhibits, those in the air-column of Kundt's tube which the hillocks of dust reveal. One would hardly detect by instinct in this stationary wave-pattern the superposition of two oppositely gliding wave-trains each traveling with the speed $u = n/m = \sqrt{T/\rho}$. Yet the one is always equivalent to the other, and in the equation (111), the coefficients n and m are linked to one another through the wave-speed characterizing the string, and the equation may be written

$$y = A \sin umt \cos mx, \qquad u = \sqrt{T/\rho}. \tag{112}$$

Although the tension and the density of the string thus determine n when m is preassigned (or vice versa), nothing so far brought upon the scene compels any limitations upon the coefficient m. The infinitely long wire can sustain vibrations of any wave-length, or vibrations of two or any number of wave-lengths simultaneously, with any inter-relation whatever among their several amplitudes and phases. On this fact rests our freedom to impose any initial conditions whatsoever on such a wire (subject to the usual restrictions of continuity and finiteness). For, if it be demanded that at $t = 0$ the displacement y shall vary along the wire according to any totally arbitrary function $f(x)$, and the transverse speed y according to any totally arbitrary function $g(x)$, then we have only to expand these functions f and g into Fourier series, or if need be, Fourier integrals; and each term in such an expansion corresponds to such a solution as (109), with a specific value of m and such specific values of C and D as the initial conditions require; and the configuration of the wire forever before and after is described by the sum of all these solutions. In such a case we should not see an unchanging distortion of the wire slipping steadily along its length with a constant speed, nor a stationary pattern of nodes and loops. All the obvious features of wave-motions would be blotted out; and yet the infinitely complicated and variable figure of the string would be equivalent, in the last analysis, to a multitude of sinusoidal wave-trains perpetually gliding to and fro with the same uniform speed.

As soon, however, as we impose *boundary-conditions*, the vibrations which the string can execute are severely restricted.

As a simple and familiar example of boundary-conditions, I will assume that the string is clamped at the points $x = 0$ and $x = L$, and concern myself only with the finite length of string, L, comprised between these two fixed extremities.

As a preparation for future developments, it is advisable to restate the underlying differential equation, and solve it *ab initio*. We have:

$$\ddot{y} = u^2 y'', \tag{113}$$

in which u stands for the speed of propagation of a sine-wave along an infinite wire. We essay a tentative solution, in the form of a product of a function of t only by a function of x only:

$$y = g(t) \cdot f(x). \tag{114}$$

The differential equation subjects the functions g and f to the condition:

$$f''/f = \ddot{g}/u^2 g = -m^2, \tag{115}$$

for, since the first member of this triple equation does not depend on t, and the second does not depend on x, each of the two must be independent of both t and x, and equal to a constant which (for the sake of consistency with prior notation) I denote by $-m^2$. Solutions of these differential equations into which the underlying one was broken up are these:

$$f = A\cos mx + B\sin mx, \qquad g = C\cos mut + D\sin mut. \tag{116}$$

So far, there is no limitation upon m.

Now come the boundary-conditions, formulated thus:

$$f(0) = f(L) = 0. \tag{117}$$

We have now encountered, in its simplest example, the peculiar and characteristic problem of the Theory of Acoustics, which is also the peculiar and characteristic problem of the type of Atomic Theory which is inherent in wave-mechanics. This is not the question which we meet in the theory of moving particles, where we are asked what path a particle will follow through all future time if its position and velocity at a single moment are given. A similar question will indeed presently be asked and answered; but this peculiar problem intrudes itself at the beginning.

To adjust the function $f(x)$ to the boundary conditions, it is evident that we must set $A = 0$ and $\sin mL = 0$; therefore we must assume that m has one of the values:

$$m = k\pi/L \qquad k = 1, 2, 3, 4 \cdots. \tag{118}$$

The boundary-conditions have compelled the coefficient m to choose among a rigidly defined series of values. The wave-lengths, and consequently the frequencies, of the permitted vibrations are strictly determined.

The permitted values of m are known in German as the *Eigenwerte* of the differential equation for the boundary-conditions in question. The English term would be "characteristic values"; but it is long and has many meanings, and I think it preferable to borrow the German word as a foreshadowing of the application which Schroedinger has made peculiarly his own. To each *Eigenwert* of m there corresponds a value of the vibration-frequency $mu/2\pi$, which in German is called an *Eigenfrequenz;* but here we may as well keep to the English term *natural frequency.*

To each *Eigenwert* there corresponds a solution of the differential equation, an *Eigenfunktion.* In the present instance the *Eigenfunktion* corresponding to the *Eigenwert* $m = k\pi/L$ is:

$$y_k = \sin\frac{k\pi}{L}x\left(C_k\cos\frac{k\pi u}{L}t + D_k\sin\frac{k\pi u}{L}t\right). \tag{119}$$

It represents a sinusoidal stationary oscillation of the wire, with nodes at the ends and at $(k - 1)$ points spaced evenly between the ends—a case not difficult to realize with a violin-string, if k be not too great. The constants C and D specify the amplitude of the oscillation, and its phase at any given instant.

It is of course not necessary that the motion of the wire should conform to a single *Eigenfunktion*. Any number of *Eigenfunktionen*, corresponding to different permitted values of m—different integer values of k—might coexist simultaneously, each with its particular values of C_k and D_k; the actual distortion of the wire would be the superposition of all. It would in fact be necessary to adjust the initial distortion of the wire and the initial velocities of its points with infinite accuracy, to cause its future motion to conform exactly to a single *Eigenfunktion*. On the other hand, any choice whatever of initial distortion and initial velocities would entail a future motion compounded out of the various *Eigenfunktionen* with suitable values of C_k and D_k, which could be computed. This process corresponds to that of determining the future orbit of a particle of which the position and the velocity at a given instant are preassigned.[9] Both in acoustics and in wave-mechanics it is, as a rule, much more laborious than the determination of natural frequencies; and happily it is often less important, though not always to be neglected.

Example of the Tensed Membrane

The differential equation of the tensed membrane is:

$$\nabla^2 z = \frac{d^2 z}{dx^2} + \frac{d^2 z}{dy^2} = \frac{1}{u^2}\frac{d^2 z}{dt^2}\,. \tag{120}$$

The coordinate-axes of x and y lie in the equilibrium plane of the membrane, and z stands for the displacement of any point of the membrane normally from this plane. The symbol u stands for the speed of a sine-wave traveling in an infinite membrane of the same tension T and surface-density ρ as the actual one, and is determined by the equation:

$$u^2 = T/\rho, \tag{121}$$

which is derived by an obvious extension of the method employed in deriving the like equation for a stretched string. In an actual bounded membrane the motion may be tremendously complicated, but it can

[9] Inversely, the imposition of quantum-conditions upon orbits corresponds to the determination of natural frequencies; here is the bridge between the atom-models with electron-orbits and the atom-models of wave-mechanics.

be analyzed into a multitude of wave-trains traveling to and fro with the speed u.

The symbol ∇^2 (to be read *del* or *nabla-squared*) stands for the Laplacian operator which in rectangular coordinates is d^2/dx^2, or $(d^2/dx^2 + d^2/dy^2)$, or $(d^2/dx^2 + d^2/dy^2 + d^2/dz^2)$, according as we are dealing with one, two or three dimensions. In other coordinates than rectangular, it naturally assumes other forms. Now in these problems of two and three dimensions, the choice of coordinate-system and the imposition of boundary-conditions are two decisions which cannot be separated from one another. Were we to decree that the membrane should be square or rectangular with its edges clamped, the suitable coordinate-system would be the rectangular. The problem would then be extremely simple (the reader can easily solve it for himself by using the method adopted for the stretched string, and will arrive at very similar results) but not so instructive to us as the problem of the circular membrane with clamped edge. For this we must adopt polar coordinates (with the origin at the centre of the membrane, naturally). In these, the Laplacian operator assumes the form:

$$\nabla^2 = \frac{d^2}{dr^2} + \frac{1}{r}\frac{d}{dr} + \frac{1}{r^2}\frac{d^2}{d\theta^2}\,. \tag{122}$$

We restate the fundamental differential equation (120) in this fashion; we essay a tentative solution in the form of a product of a function $f(r)$ of r exclusively, a function $F(\theta)$ of θ exclusively, and a function $g(t)$ of t exclusively; and we discover as before that each of these functions is subjected to a differential equation of its own. The procedure is like that already used in the case of the stretched string clamped at its ends. First we have

$$\frac{1}{f}\frac{d^2f}{dr^2} + \frac{1}{rf}\frac{df}{dr} + \frac{1}{r^2F}\frac{d^2F}{d\theta^2} = \frac{1}{u^2g}\frac{d^2g}{dt^2} = -\,m^2, \tag{123}$$

for, since the first member of this triplet does not depend on t, the second not on r nor on θ, both must be independent of all three variables and equal to a constant which, as before, I denote by $-\,m^2$. The differential equation for the factor dependent on t has the solution:

$$g(t) = A \cos mut + B \sin mut. \tag{124}$$

Our experience with the stretched string suggests that m will be restricted to certain *Eigenwerte*, derived from the boundary-conditions; and this is true; but before arriving at these, we must attend to the differential equation governing the functions f and F. This assumes the form:

$$\frac{r^2}{f}\frac{d^2f}{dr^2} + \frac{r}{f}\frac{df}{dr} + m^2r^2 = -\frac{1}{F}\frac{d^2F}{d\theta^2} = \lambda^2, \tag{125}$$

both members of the equation being, by the familiar reasoning, equal to a constant which I denote by λ^2. It follows that the function $F(\theta)$ is of the form:

$$F(\theta) = C \cos \lambda\theta + D \sin \lambda\theta \tag{126}$$

and the coefficient λ thus far seems to be unrestricted. But it carries its own restrictions in itself; for the coordinate θ is a cyclic coordinate, like longitude on the earth; whenever it is altered by 2π, we are back at the same place. The function $F(\theta)$ must therefore repeat itself whenever θ is altered by 2π; but this will not occur, unless λ is an integer:

$$\lambda = 0, 1, 2, 3 \cdots. \tag{126a}$$

These are the *Eigenwerte*, and the functions (126) with one or another of these values assigned to λ are the *Eigenfunktionen*, of the equation (125). In this case we have obtained *Eigenwerte* for the parameter and *Eigenfunktionen* for the solutions of a differential equation, not out of boundary conditions but out of the simple fact that the independent variable is by its nature cyclic. Such cases will occur in the undulatory mechanics.

We arrive at the third and last step of the problem: the determination of the function $f(r)$. It is governed by the differential equations:

$$\frac{d^2f}{dr^2} + \frac{1}{r}\frac{df}{dr} + \left(m^2 - \frac{\lambda^2}{r^2}\right)f = 0, \tag{127}$$

a distinct equation for each of the permitted integer values of λ. As the solution of such an equation as (115) is a sine-function of the variable mx, so the solution of such an equation as (127) is a function of the variable mx; not however a sine-function, but a Bessel function. For the values 0, 1, 2, $\cdots$ of λ, the solutions of (127) are the Bessel functions of order 0, 1, 2, $\cdots$, denoted by $J_0(mr)$, $J_1(mr)$, $J_2(mr)$, and so forth.

Like the sine-function of mx, the Bessel functions of mr oscillate back and forth between negative and positive values as their variable increases from zero to infinity, and pass through zero at an infinite number of discrete values of mr. These do not lie at equal intervals, as do the values of mx at which $\sin mx$ vanishes. Their values may be found in the tables; I shall designate them as b^1, b^2, b^3, $\cdots$ in order of increasing magnitude, using the superscripts not as exponents, but as ordinal numbers so that I may reserve the subscripts to distinguish the various Bessel functions from one another. The function

$$Z = J_\lambda(mr)(C \cos \lambda\theta + D \sin \lambda\theta)(A \cos mut + B \sin mut) \quad (129)$$

represents a stationary oscillation of an infinitely extended membrane, in which λ lines intersecting one another at the origin are nodal lines, and an infinity of concentric circles centred at the origin are nodal circles. These lines and circles are motionless while the sections of the membrane which they delimit vibrate with the frequency $mu/2\pi$. The λ lines are spaced uniformly in angle; the radii r_1, r_2, $\cdots$ of the infinity of circles are obtained by dividing m into the roots b_λ^1, b_λ^2, b_λ^3, $\cdots$ of the Bessel function of order λ, $J_\lambda(mr)$.

How then does the boundary-condition upon the finite membrane enter in? Obviously, if a membrane of radius R be clamped at its edge, and if it is vibrating in the manner described by (129), then the edge must coincide with one of the nodal circles; the radius R must be equal to one of the quantities b_λ^i/m. Or rather, since the nodal circles are to be adjusted to the size of the diaphragm and not the size of the diaphragm to the nodal circles, the coefficient m must conform to one of the equations:

$$m = b_\lambda^1/R, \quad \text{or} \quad b_\lambda^2/R, \quad \text{or} \quad b_\lambda^3/R, \quad \cdots. \quad (130)$$

These equations define Eigenwerte of the parameter m *in the differential equation of the tensed membrane.* There is a double infinity of these—an infinite series of them for each of the *Eigenwerte* of the parameter λ. To each corresponds a natural frequency of the membrane, and to each corresponds an *Eigenfunktion*, the one written down in (129) with the proper value of m taken from (130). The constants A, B, C, and D in the *Eigenfunktionen* specify the amplitude of the oscillation, the phase of the vibrations at any given instant, and the orientation of the nodal lines with respect to any given axis. Any number of *Eigenfunktionen* may coexist simultaneously; the actual distortion of the membrane will be the superposition of all. Any initial conditions imposed on z and $\dot{z}$ (and not involving discontinuities or infinities) could be satisfied by adjusting the constants.

Example of the Ball of Fluid

Among the familiar vibrating systems the ball of fluid presents the closest analogy to the atom-model for the hydrogen atom in wave-mechanics, the wave-patterns in the two cases being strikingly alike.

In three dimensions and in polar coordinates (those appropriate to the boundary-conditions which we shall impose) the wave-equation assumes the somewhat alarmingly intricate form:

$$\ddot{\Psi} = u^2\nabla^2\Psi = u^2\frac{\operatorname{cosec}\theta}{r^2}\left[\frac{d}{dr}\left(r^2\sin\theta\frac{d\Psi}{dr}\right) + \frac{d}{d\varphi}\left(\operatorname{cosec}\theta\frac{d\Psi}{d\varphi}\right) + \frac{d}{d\theta}\left(\sin\theta\frac{d\Psi}{d\theta}\right)\right]. \tag{131}$$

The argument Ψ can no longer be visualized as a displacement perpendicular to the equilibrium-position of the undistorted medium, since all three dimensions are already used up. The reader may visualize it, if he will, as a condensation or a rarefaction, after the fashion of sound-waves. Perhaps not to visualize it at all would be a better preparation for the study of wave-mechanics.

In the familiar way, we essay a solution in the form of a product of a function of time $g(t)$, a function of radius $f(r)$, a function $\Phi(\phi)$ of the longitude-angle ϕ and a function $\Theta(\theta)$ of the colatitude-angle θ. As before, we find that the time-function is of the form:

$$g(t) = A\cos mut + B\sin mut \tag{132}$$

and, as before, we shall find that the boundary-conditions confine the coefficient m and the frequency $mu/2\pi$ to certain "permitted" values. The angle-functions and the radius-function are governed by the differential equations:

$$\frac{1}{f}\left[\frac{d}{dr}\left(r^2\frac{df}{dr}\right) + m^2r^2f\right] = -\frac{1}{Y}\operatorname{cosec}\theta\left[\frac{d}{d\varphi}\left(\operatorname{cosec}\theta\frac{dY}{d\varphi}\right) + \frac{d}{d\theta}\left(\sin\theta\frac{dY}{d\theta}\right)\right] = \lambda, \tag{133}$$

in which Y stands for the product of Θ and Φ, and λ for a constant which seems to be arbitrary, but as a matter of fact is constrained by the same circumstance as arose in the case of the membrane; for, whenever φ is altered by 2π and θ by π, we are back at the same place as before, and the function Y must have the same value as before; and this will occur only if

$$\lambda = n(n+1), \qquad n = 0, 1, 2, 3, \cdots, \tag{134}$$

these being the *Eigenwerte* for the differential equation in (133) for the angle-function.[10] The corresponding *Eigenfunktionen* are spherical

[10] This and the following statements about the functions Y_n are proved by writing Y in the second of equations (133) as the product of a function of θ and a function of ϕ, and so dissolving the equation into two in the manner which I have already

harmonics. To each value of n belongs a "spherical harmonic of order n," which itself is a sum of $(2n + 1)$ terms, each multiplied by a constant which is at our disposal and can be adjusted to fit initial conditions or to emphasize particular modes of vibration. These terms are products of sine-functions of φ by peculiar functions, the *Legendrian functions* $P_{n,\,s}$, of the variable θ; so that the *Eigenfunktion* for a permitted value $n(n + 1)$ of the parameter λ has this for its most general form:

$$Y_n(\theta, \varphi) = a_{n,\,0}P_{n,\,0}(\cos\theta) + \sum_{s=1}^{n} a_{n,\,s}\cos(s\varphi)P_{n,\,s}(\cos\theta) + \sum_{s=1}^{n} b_{n,\,s}\sin(s\varphi)P_{n,\,s}(\cos\theta). \tag{135}$$

Each term by itself describes a particular mode of vibration of the fluid; the sum represents a superposition of divers modes of vibration. If we isolate one of these modes by giving to n some particular value n_1, and to s some particular value s_1, and causing all the constants a and b in (135) to vanish except $a_{n_1,\,s_1}$ and $b_{n_1,\,s_1}$; we then find that Y, and consequently Ψ, and consequently the motion altogether, vanishes at s_1 values of φ and at $n_1 - s_1$ values of θ. If we draw a sphere centred at the origin, we find that its surface bears s_1 nodal meridian-circles, and $n_1 - s_1$ nodal latitude-circles, along which there is perpetual rest. If we consider all the spheres at once—if, that is to say, we consider the entire volume of the fluid medium—we see that when the fluid is vibrating in the mode distinguished by the integers (I had almost said "quantum-numbers"!) n_1 and s_1, it is divided into compartments by s_1 nodal planes intersecting along the axis $\theta = 0°$, and $n_1 - s_1$ double-cones having that axis for their axis and the origin for this apex.

We have not yet considered the dependence of the wave-motion on the radius r; but the close analogy between this and the corresponding stage of the problem of the tensed membrane will make the task easy. The differential equation (133) for $f(r)$ resembles Bessel's equation (127), and has the somewhat similar solution

$$f(r) = \frac{1}{\sqrt{r}} J_{n+\frac{1}{2}}(mr). \tag{136}$$

used five or six times; the values of the constant s in equation (135) are the *Eigenwerte* of the latter of these two. I thought it desirable not to overload the exposition by carrying through all stages of the process of solution, especially as the splitting of $Y_n(\theta, \phi)$ into the two functions is of secondary importance in the atom-model to which all this leads up; nevertheless the reader may find it advantageous to supply the lack.

This function vanishes, entailing the vanishing of the wave-motion, at an infinity of discrete values of the variable mr:—the roots of the function, which I denote in order of increasing magnitude by B^1, B^2, B^3, $\cdots$. In an infinite medium we could assign any value whatever to r, and then there would be an infinity of nodal spheres, their radii given by B^1/m, B^2/m, B^3/m, $\cdots$. If the medium is bounded by a rigid spherical wall of radius R, the coefficient m must possess one of the values B^i/R, so that one of the nodal spheres may coincide with the wall. These are the *Eigenwerte* of the constant m, and the natural frequencies of the corresponding vibrations are given by $B^iu/2\pi R$. The *Eigenfunktionen* are given by the equation (136) with the various values B^i/R substituted for the parameter m.

The *Eigenfunktionen* of the fundamental differential equation for the fluid sphere are, therefore, each a product of a radius-function given by (136), with a "permitted" value for the constant m determined by the boundary-condition; an angle-function given by (135), with "permitted" values for the constants n and s, determined by the fact that the angles are cyclic variables; and a time-function given by (132), with a "permitted" vibration-frequency determined by the boundary-condition. Each *Eigenfunktion* with the indices m, n, s describes a mode of vibration, in which the fluid sphere is divided into compartments by s meridian planes, $(n - s)$ double-cones, and a certain number of spheres, upon each of which the fluid is perpetually at rest; within the compartments, it vibrates with a prescribed frequency.

Atom-Models in Wave-Mechanics

Case of a "String" for which the Wave-Speed is Variable, or even Imaginary

Thus far I have used the images of the stretched string, the tensed membrane, and the elastic fluid to illustrate the behavior of the differential equation

$$u^2\nabla^2\Psi = d^2\Psi/dt^2, \tag{151}$$

when the coefficient u^2 is a positive constant. In these examples u^2 is interpreted as the ratio of the intrinsically positive quantities "tension" (or "pressure") and "density," and turns out to be equal to the square of the speed of propagation of sine-waves in the string, membrane, or fluid. In certain problems of undulatory mechanics we encounter just such an equation. In some of the most important applications of Schroedinger's theory, however, one meets with differential equations of the type of (151), in which however the coefficient u^2 depends on the coordinates and even assumes negative values! Such equations need not be more difficult to solve than the conventional wave-equation in which u^2 stands for a positive constant; but the image of the elastic medium becomes unsatisfying. In the one-dimensional case, so long as u^2 remains a positive function of x, one can visualize a string of which the density varies along its length; but when u^2 passes through zero and becomes negative, the wave-speed attains zero and is superseded by an imaginary quantity. One may speak, in such a case, of a "string" or a "fluid" characterized by an "imaginary wave-speed." So speaking, one comes perilously close to the verge of using words devoid of physical meaning; but otherwise, there is no verbal language with which to relieve the monotony of the procession of equations.

The differential equation of the type of (151), with a constant negative value of the coefficient u^2, is not a difficult one. Confining ourselves to one dimension, we find for one of the solutions of the equation for a "string with constant imaginary wave-speed" this expression:

$$\Psi = (A \cos mUt + B \sin mUt)(Ce^{mx} + De^{-mx}), \tag{152}$$

in which U stands for the (real) square root of $-u^2$. This is a much less tractable function than the product of sine-functions which serves when u^2 is positive. One cannot, for instance, find *Eigenwerte* for the constant m whereby the function can be made to vanish at all times at two distinct points upon the "string"; or rather, one can find only the value $m = 0$, which fulfils this familiar boundary-condition by destroying the function. Similarly, one cannot force Ψ to remain finite everywhere except by annulling either m or else both A and B, again destroying the function. Vibrations which are sine-functions of time are, however, permitted by the differential equation.

Consider now the equation

$$d^2y/dx^2 = (a - bx^2)d^2y/dt^2, \tag{153}$$

which may be regarded as the wave-equation of a string of which the wave-speed varies with x along its length as the function $(a - bx^2)^{-1/2}$, being therefore real over the central part from $x = -\sqrt{a/b}$ to $x = +\sqrt{a/b}$, and imaginary from each extremity of this central range outward to infinity. In the usual way, we derive the equations:

$$y = f(x)g(t), \qquad g = A \cos \nu t + B \sin \nu t,$$
$$d^2f/dx^2 + \nu^2(a - bx^2)f = d^2f/dx^2 + (C - x^2)f = 0, \tag{154}$$

and it is incumbent upon us to solve the equation [11] for $f(x)$.

[11] The constant $\nu^2 b$ has been equated to unity, which entails no loss of generality.

Essay a solution in the form of a power-series, multiplied by $e^{-(1/2)x^2}$:

$$f(x) = e^{-(1/2)x^2} \sum_{n=0}^{\infty} a_n x^n. \tag{155}$$

Substitute this into the differential equation, and group all the terms involving the same power of x. For each such group, we have

$$a_{n+2}(n + 1)(n + 2)x^n - a_n(2n + 1 - C)x^n, \tag{156}$$

and equating each group separately to zero, we arrive at the relation

$$a_{n+2}/a_n = (2n + 1 - C)/(n + 1)(n + 2). \tag{157}$$

Put $a_0 = 0$, thus causing all the even-numbered coefficients to vanish; assign any arbitrary value to a_1, and calculate the odd-numbered coefficients a_3, a_5, a_7, and so onward. Or, put a_1 and all the odd-numbered coefficients equal to zero, assign any arbitrary value to a_0, and calculate the even-numbered coefficients a_2, a_4, a_6, and so onward. Either way we shall get a solution of (154), whatever the value of the parameter C; but there are certain specific values of C which admit a peculiar sort of solution. It is, in fact, evident from (156) that we shall arrive at two entirely distinct results, according as C is or is not equal to some value of $(2n + 1)$—according, that is to say, as C is or is not an odd integer. For, if C is equal to an odd integer $(2n + 1)$, the chain of coefficients will come to an abrupt end at the member having that particular value of n; it and all the succeeding members will be zero; the power-series in the tentative (and adequate) expression (155) for the unknown function $f(x)$ will consist of a finite number of terms. But, if C is not equal to an odd integer, the power-series will go on forever.

Here we have a new kind of *Eigenwert*. If the parameter C, in the differential equation for the curious kind of "string" which I have just defined, has for its value one of the numbers:

$$C = 2n + 1, \qquad n = 0, 1, 2, 3, 4, \cdots, \tag{158}$$

the equation enjoys a special sort of solution. If the parameter does not have one of these *Eigenwerte*, the solution of the differential equation is altogether different.

Let us see what difference these *Eigenwerte* make in the general solution (155) of the differential equation. If the parameter C has some other value than one of these, the series $a_n x^n$ goes on forever; and as x approaches infinity, the value of its summation increases at such a rate as to overwhelm the steadily declining factor $e^{-(1/2)x^2}$, so that the function $f(x)$ is infinite at both ends of the range $-\infty < x < \infty$.

If however C is equal to one of the *Eigenwerte*, the series $a_n x^n$ comes to an abrupt end; and as x approaches infinity, the decline of the factor $e^{-(1/2)x^2}$ overpowers the increase of the summation, and $f(x)$ remains finite at infinity. The values 1, 3, 5, $\cdots$ of the constant C are therefore the *Eigenwerte* which permit solutions *which remain finite all through the range of values of the independent variable from positive to negative infinity.* This condition replaces the boundary-conditions applied to the ordinary stretched string.

The *Eigenfunktionen* are:

$$f_m(x) = e^{-(1/2)x^2} H_m(x), \tag{159}$$

the symbol $H_m(x)$ standing for the finite series $\sum a_n x^n$ constructed according to the rules of the foregoing paragraphs, and terminating at the mth term. These are known as the *polynomials of Hermite.*[12]

Interpretation of the Simple-Harmonic Linear Oscillator by Wave-Mechanics

The foregoing section contains all that is necessary to Schroedinger's theory[13] of the linear simple-harmonic oscillator—an object, or a concept, famous in the history of the quantum-theory; for it was the linear oscillator which Planck first "quantized"—of which, that is to say, Planck first proposed that it be endowed with the power of receiving and retaining and disbursing energy only in fixed finite amounts; thereby arriving at an explanation of the black-body radiation-law, and founding the quantum theory.

Conceive a particle of mass m, constrained to move along the x-axis, attracted to the origin by a force $-k^2x$ proportional to its displacement, and consequently prone to oscillate to and fro across the origin with frequency $\nu_0 = k/2\pi\sqrt{m}$. Its potential energy is the following function of x:

$$V = \tfrac{1}{2}k^2x^2 = 2\pi^2 m\nu_0^2 x^2. \tag{160}$$

The wave-equation assumes the form

$$\frac{d^2\Psi}{dx^2} + \frac{8\pi^2 m}{h^2}(E - 2\pi^2 m\nu_0^2 x^2)\Psi = 0. \tag{161}$$

A simple change of variable ($q = x\cdot 2\pi\sqrt{m\nu_0/h}$) transforms this into the equation (154):

$$d^2\Psi/dq^2 + (C - q^2)\Psi = 0; \qquad C = 2E/h\nu_0. \tag{162}$$

[12] The first five are written down by Schroedinger, *Ann. d. Phys.*, **79**, p. 515 (1927). An arbitrary numerical multiplier remains at disposal.

[13] Schroedinger, *Ann. d. Phys.*, **79**, pp. 514–519 (1926); for the general case in which the restoring-force is not supposed to vary as the displacement, consult H. A. Kramers, *ZS. f. Phys.*, **39**, pp. 828–840 (1926).

According to Schroedinger *the Stationary States of the linear oscillator are distinguished by the energy-values which cause this equation to have a solution finite at all values of the variable, infinity included.*

These are the values of the constant C which cause the parameter C to take one of the *Eigenwerte* set down in (158).

The energy-values of the Stationary States should therefore be

$$\begin{aligned} E_n &= \frac{h\nu_0}{2}(2n+1) \\ &= \frac{1}{2}h\nu_0,\ \frac{3}{2}h\nu_0,\ \frac{5}{2}h\nu_0,\ \cdots. \end{aligned} \tag{163}$$

The successive permitted energy-values of the linear simple-harmonic oscillator of frequency ν_0, the energy-values of its consecutive Stationary States, are therefore specified by wave-mechanics as the products of the fundamental factor $h\nu_0$ by the consecutive "half-integers" 1/2, 3/2, 5/2, and so onward.

The linear simple-harmonic oscillator thus furnishes an instance of "half-quantum-numbers." In most of the earlier theories it was either assumed or inferred that this "Planck" oscillator displayed "whole quantum-numbers"—that its permitted energy-values were the products of $h\nu_0$ by the successive integers 1, 2, 3, 4, $\cdots$. However, in the interpretation of certain features of band-spectra by the assumption that the two atoms of a diatomic molecule vibrate as linear oscillators along their line of centres, the half-quantum-numbers sometimes led to better agreement with experience than did the whole-quantum-numbers.

The *Eigenfunktionen* corresponding to the consecutive Stationary States are these:

$$\Psi_n(x) = \text{const}\cdot e^{-2\pi^2 m\nu_0 x^2/h} H_n(2\pi x\sqrt{m\nu_0/h}). \tag{164}$$

The first five of these *Eigenfunktionen* are exhibited in Fig. 2. These curves may be regarded, if the reader so chooses, as the stationary-wave patterns of "loops" and "nodes," exhibited by five resonating strings along which the wave-speed varies according to the five laws obtained by assigning the first five values given by (163) to the constant E in the equation:

$$u = \frac{E}{\sqrt{2m(E - 2\pi^2 m\nu_0{}^2 x^2)}}\cdot \tag{165}$$

The various Stationary States of a linear oscillator are therefore imaged not as the fundamental and the overtones of one and the same string, but as the fundamental (and exclusive) nodes of vibration of distinct strings. It is important to realize this. Schroedinger's way of thinking provides not a single atom-model for each sort of atom, but as many distinct models as there are Stationary States.*

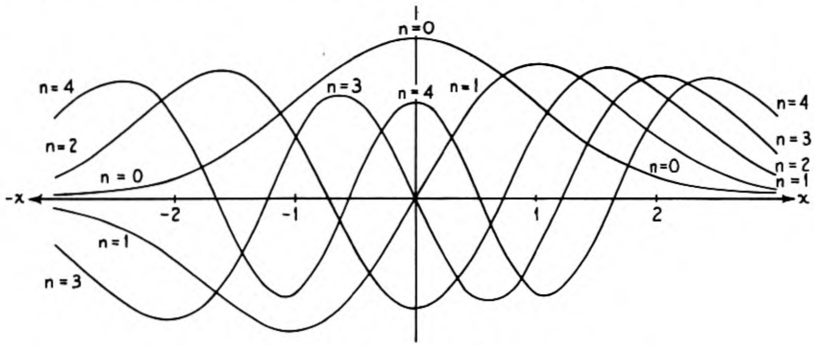


Fig. 2 (after Schroedinger).

Interpretation of the Hydrogen Atom by Wave-Mechanics

The hydrogen atom is conceived as a system endowed with the potential energy $V = -e^2/r$. This form for the potential energy, I recall, is obtained by imagining an electron and a nucleus, or more precisely two point-charges $+e$ and $-e$, separated by a distance denoted by r. The image of the electron and the nucleus does not come over explicitly into the new theory; but in spirit it does come over, for the potential-energy-function derived from that image is the basis for the new theory.

Polar coordinates for the wave-equation are imperiously suggested by a potential-energy-function of this form, and consequently it is thus expressed:

$$\frac{E^2}{2m(E + e^2/r)} \nabla^2 \Psi = \frac{d^2\Psi}{dt^2}, \tag{171}$$

and putting E/h for the vibration-frequency, we attain

$$\nabla^2 \Psi + \frac{8\pi^2 m}{h^2} \left(E + \frac{e^2}{r} \right) \Psi = 0. \tag{172}$$

The resemblance of these equations to those laid down for the ball of fluid is as unmistakable as the resemblance of the wave-equation for a linear oscillator to that of a stretched string. Here we have the case of a fluid in which the wave-speed varies from point to point, according to the law

$$u^2 = E^2/2m(E + e^2/r), \tag{173}$$

* Some may find satisfaction in conceiving, as my colleague Dr. T. C. Fry suggests, a "string" so constructed that the speed of propagation of waves along it is a function of their frequency.

and we meet the problem of finding modes of vibration and stationary wave-patterns.

If E is supposed positive, the wave-speed is everywhere real. Boundary-conditions of the usual sorts (*e.g.*, the prescription that the fluid shall be confined within a rigid spherical wall of given radius) might be imposed, and then *Eigenwerte* of the constant E could be calculated, and from these the wave-patterns and natural frequencies of the fluid. If no such boundary-conditions were prescribed, the equation (172) could be solved with any value of E.

If E is supposed negative, the whole state of affairs is changed. The wave-speed is now real within the sphere of radius $-e^2/E$, zero over this sphere and imaginary beyond it. This recalls the case of the "string" proposed as an analogy for the linear oscillator, for which the wave-speed was real along its central segment and imaginary from each end of its central segment onwards to infinity. There are important differences: in the present case, the variable r assumes positive values only, and the wave-speed at $r = 0$ is infinite though real.

In the case of the "string" with variable and at some points imaginary wave-speed, we found that the law of variation of wave-speed could be so chosen that the "string" enjoys a natural mode of vibration with a stationary wave-pattern and a natural resonance-frequency. This was done by selecting any of a series of *Eigenwerte* for a parameter of the differential equation. Here we shall do likewise.

Essaying for the function Ψ in (172) a solution in the form of a product of a function of θ and φ exclusively by a function of r exclusively, we arrive in the familiar way at differential equations:

$$\operatorname{cosec}\theta\left\{\frac{d}{d\theta}\left(\sin\theta\frac{dY}{d\theta}\right)+\frac{d}{d\varphi}\left(\operatorname{cosec}\theta\frac{dY}{d\varphi}\right)\right\} = -\lambda Y. \quad (175)$$

$$\frac{d}{dr}\left(r^2\frac{df}{dr}\right)+\frac{8\pi^2 mr^2}{h^2}\left(E+\frac{e^2}{r}\right)f = +\lambda f, \quad (174)$$

The equation (175) is the identical one which we encountered in the case of the ball of fluid. Here, as there, the fact that the variables θ and φ are cyclic requires *Eigenwerte* of the constant λ:

$$\lambda = l(l+1), \qquad l = 0, 1, 2, 3, 4, \cdots \quad (176)$$

Equation (174), however, is not the same as the corresponding equation (133) of the prior case; here we find the difference between the fluids of actual experience and the "imaginary fluid" which is to serve as material for the atom-model supplied by wave-mechanics for hydrogen.

If in that equation (174) one were to assign an arbitrarily chosen negative value to the parameter E, one would in general not be able to find a solution which is finite both at the origin and at infinity. This is the same situation as occurred in the theory of the linear oscillator, where an arbitrary choice of a value for the parameter there called C would in general have led to a solution implying infinite amplitude at both ends of the "string."

Schroedinger however discovered [14] that there is a series of *Eigenwerte* for the parameter E, each of which (subject to a limitation to be introduced below) entails a solution which is single-valued, continuous and finite over the entire range of the variable r.

These *Eigenwerte* are the following:

$$E_n = -2\pi^2 me^2/h^2n^2; \qquad n = 1, 2, 3, 4, \cdots. \tag{177}$$

The consecutive permitted energy-values of the system of potential-energy-function $-e^2/r$, *the Stationary States of the model for the hydrogen atom, are therefore specified by wave-mechanics as the quotients of the fundamental factor* $-2\pi^2 me^4/h^2$ *by the squares of the consecutive integers from unity onward.*

These agree with experiment. The formula (177) is in fact the renowned formula of Bohr, from which the whole contemporary theory of spectra sprang; a formula so successful that it is scarcely conceivable that any alternative theory should ever win acceptance unless by presenting the identical equation over again.

Schroedinger's models for the hydrogen atom in its various Stationary States thus are imaginary fluids each pervading the whole of space, and in each of which the wave-speed depends on the distance r from a centre, according to a peculiar law—the law obtained by inserting into the formula (173) the appropriate value for E, chosen from the sequence given in (177). If into (173) we were to put any value chosen at random for the energy-constant E, we should be inventing an imaginary fluid; but, in general, this fluid would not be capable of sustaining a continuous stationary wave-pattern of finite amplitude. Only when one of Bohr's sequence of energy-values is chosen do we get a fluid able to resonate as a ball of actual physical substance can.

The next task is to enquire into the wave-patterns in the imaginary fluids corresponding to these various permitted energy-values. This is much more difficult than the same problem for the imaginary strings corresponding to the various permitted energy-values of the linear oscillator, and the new complexities are not altogether due to the fact that we now have three dimensions to deal with instead of one; they

[14] Schroedinger, *Ann. d. Phys.*, **79**, pp. 361–376 (1926). For an alternative method of proof see A. S. Eddington, *Nature*, **120**, p. 117 (1927).

are due chiefly to the fact that the system is mathematically "degenerate." Owing to this circumstance there are more than one possible mode of vibration, more than one stationary wave-pattern, for each (except the first) of the permitted energy-values. To describe these it is necessary to consider both of the equations (174) and (175).

Since the equation (175) is identical with the corresponding equation derived for balls of actual physical fluids, the modes of vibration for Schroedinger's atom-model are identical with the modes of vibration of actual fluid spheres *insofar as the dependence on angle is concerned.* The imaginary fluid is divided into compartments by nodal planes, nodal double-cones and nodal spheres; and the division by planes and double-cones is identically such as we should find in the corresponding mode of vibration of an actual fluid ball; it is only the division by nodal spheres which differs.

To the first *Eigenwert*, E_1 ($n = 1$) there corresponds a single *Eigenfunktion* of equation (174); to the second, E_2, a pair; to the third, three; and so forth. This multiplicity is linked with the limitation upon the *Eigenwerte* which was foreshadowed above. In the expression for the function Ψ as a product of functions of the individual variables

$$\Psi(r, \theta, \varphi) = F(r) Y_l(\theta, \varphi), \tag{178}$$

if we assign an *Eigenwert* E_n to the parameter E in the first factor according to (174), we have still a choice of values to assign to the parameter l in the second factor according to (176). This choice however is limited. We must not take any value of l as great as or greater than the value adopted for n; otherwise the value of E_n would not be an *Eigenwert* in the sense adopted. Thus for $n = 1$ we are restricted to the choice $l = 0$; for $n = 2$ we have the alternative of $l = 0$ or $l = 1$; for $n = 3$ the option of $n = 0, 1$, or 2, and so forth. Each *Eigenwert* E_n thus admits $(n - 1)$ distinct spherical harmonics $Y_1(\theta, \varphi), Y_2(\theta, \varphi) \cdots Y_{n-1}(\theta, \varphi)$ as solutions of equation (175); and to each of these there corresponds, with each of these there goes, a distinct *Eigenfunktion* $F_{n,\,l}(r)$ of the equation (174), which is expressed as follows in terms of a variable $\rho = \dfrac{2\pi\sqrt{-2mE_n}}{h}\, r = \dfrac{4\pi^2 me^2}{nh^2}\, r \equiv \dfrac{1}{na_0}\, r$ instead of r to make the function seem less intricate:[15]

$$X_{n,\,l}(\rho) = \text{const. } \rho^{+l} e^{-\rho} \sum_{k=0}^{n-l-1} \frac{(-2\rho)^k}{k!} \binom{n+l}{n-l-1-k}, \tag{180}$$

[15] The factor in parentheses in equation (180) stands for the "number of combinations of $(n + l)$ quantities taken $(n - l - 1 - k)$ at a time," which is the $(n - l - 1 - k)$th coefficient in the binomial expansion of $(a + b)^{l+n}$.

The function $X_{n,l}(\rho)$ has $(n - l - 1)$ roots, so that the corresponding mode of vibration has $(n - l - 1)$ nodal spheres. To each permitted energy-value E_n there consequently correspond n different solutions of the general equation (172), differing from one another in respect of the number of nodal spheres:

$$\Psi_{n,l}(r, \theta, \phi) = X_{n,l}(\rho) Y_1(\theta, \phi); \qquad l = 0, 1, 2 \cdots (n - 1). \quad (181)$$

Each of these describes a permitted class of modes of vibration, owing to the subdivision of the spherical harmonic Y_1 into terms according to (135).

Allowing for the subdivision of the spherical harmonics, there are $(1 + 2 + 3 + \cdots n) = n(n + 1)/2$ modes of vibration for the nth permitted energy-value E_n.

The equation (181) exhibits the various modes of vibration of which our imaginary "fluid," the model for the hydrogen atom, is capable. It would be possible to describe these with a wealth of verbal detail. I hesitate to do so; for vast amounts of industry and ink have been expended during the last twelve years in tracing and describing electron-orbits, which are now quite out of fashion; and who dares affirm that in another five years the vibrating imaginary fluid will not be *démodé*? Yet it is altogether probable that for some years to come, if not for all time, the image of the vibrating fluid will furnish the customary symbolism for expressing the data of experiment. Therefore let me point out some features of the vibrations corresponding to the first (or "lowest," or "deepest") three states of the hydrogen atom:

Normal State, $n = 1$. One *Eigenfunktion*, $X_{1,0}(\rho)$; an exponential function of r, decreasing steadily from the origin to infinity, with no nodal spheres. Corresponding spherical harmonic $Y_0(\theta, \varphi)$,—a constant. The vibration consequently is described by

$$\Psi(r) = \text{const. } e^{-r/a_0} \quad (a_0 = h^2/4\pi^2 me^2) \qquad (182)$$

and is endowed with perfect spherical symmetry.

First Excited State, $n = 2$ (the state into which the atom relapses after emitting any line of the Balmer series). Two *Eigenfunktionen* $X_{2,0}$ and $X_{2,1}$; the first represents a vibration with a single nodal sphere, the second a vibration diminishing steadily in amplitude from the origin outward. The first is to be multiplied by $Y_0(\theta, \varphi)$ to obtain the complete description of the vibration; Y_0 being a constant, this mode is endowed with perfect spherical symmetry. The second is to be multiplied by Y_1, which is a combination of terms written out in equation (135); the permissible mode, or rather modes, of vibration involve nodal planes and double-cones, which the reader may work out for himself with the aid of (135).

Second Excited State, $n = 3$ (the state from which the atom departs when it emits the line H-alpha). Three *Eigenfunktionen* $X_{3,0}$, $X_{3,1}$ and $X_{3,2}$. The first corresponds to a vibration with two nodal spheres, and perfect spherical symmetry. The second and third correspond to vibrations with one nodal sphere, and with a steady diminution of amplitude from the centre outward, respectively; but being multiplied with the spherical harmonics Y_1 and Y_2, they describe modes which are not endowed with spherical symmetry, and involve nodal double-cones and nodal planes.

Generally: the state distinguished by the numeral n enjoys n distinct *Eigenfunktionen*, describing vibrations having respectively $0, 1, 2, 3 \cdots (n - 1)$ nodal spheres; to the *Eigenfunktion* with the maximum number of nodal spheres corresponds a single mode of vibration which is spherically symmetric, to the others various modes with varying members of nodal double-cones and planes.

If this is destined to be the "language of the future" for describing the data of experiment, it will be necessary to have dictionaries for translating it out of (or into) the "language of the present," the vocabulary of the Bohr-Sommerfeld atom-model in which Stationary States are represented by electron-orbits. They will contain definitions such as these: the numeral n is the total-quantum-number of the electron-orbits—the numeral l is one unit smaller than the azimuthal quantum-number k of the electron-orbit—the numeral $(n - l - 1)$, to which the number of nodal spheres is equal, is the radial quantum-number of the electron-orbit. To elucidate these "definitions" of the future dictionary, I recall that the Bohr-Sommerfeld atom-model provided, for the hydrogen atom in its state of energy-values E_n, a family of n distinct electron-orbits, of which one is circular while the other $(n - 1)$ are ellipses of varying degrees of eccentricity.[16] These ellipses were selected by laying down the conditions, that the integral $\int p_{\varphi} d\varphi$ of the angular momentum p_{φ} around the orbit shall be equal to the product of h by some integer k equal to or less than the prescribed n; and the integral $\int p_r dr$ of the radial momentum $\int p_r dr$ shall be equal to the product of h by the integer $(n - k)$; so that the sum of the integrals $\int p_{\varphi} d\varphi$ and $\int p_r dr$ shall be equal to the product of h by n. The quantities n, k and $n - k$ were given the names *total, azimuthal, radial quantum-number*. "Defini-

[16] The introduction some twenty months ago of the "spinning electron" caused a modification of this picture; for those who accept the modification, it is the "language of antiquity" which is compared in this paragraph with the "language of the future."

tions" such as those above (which are not necessarily the only self-consistent nor the best ones) make it possible to translate orbits of the orbit-model into modes of vibration of the wave-model, and vice versa; and to devise definitions for these three kinds of quantum-numbers from the qualities of the vibrations themselves.

Perturbations

Inasmuch as the wave-mechanics indicates n different *Eigenfunktionen* with n different collections of nodal spheres (not to speak of the still more greatly varied possibilities of nodal planes and double-cones) for the Stationary State having the *Eigenwert* and energy-value E_n, one may well ask whether there is any chance of distinguishing which of these, or which linear combination of these (for the differential equation will permit any) is actually adopted by a hydrogen atom.

Translating into the language of the Bohr-Sommerfeld atom-model, we find the question in this form: is there any way of distinguishing which of the n permitted elliptical electron-orbits is actually adopted?

When the question was asked in this form, it was answered by pointing out that if the force exerted upon the electron were not the pure inverse-square force ascribed to the nucleus, but the sum of this and a *perturbing force*, the energy-values of the n permitted ellipses would cease to coincide exactly. If for instance the atom under examination were composed of a nucleus of charge $11e$, a group of ten electrons very close to it and an "outer" electron relatively far out (the conventional model for a sodium atom in certain states); then the group of ten inner electrons would act upon the outer one with a perturbing force, and the n permitted ellipses of the outer one would be endowed with distinct energy-values—the single Stationary State of the outer electron would be dissolved into n distinguishable states. Even in the hydrogen atom, the dependence of the mass of the electron upon its speed should separate the energy-values of the various ellipses which but for this fact would share a common energy E_n, and produce the fine-structure of the hydrogen lines.

The very same thing occurs in wave-mechanics; and from the effect of a perturbing force, allowance for which is made in the potential-energy-function introduced into the wave-equation, we may expect to be able to distinguish the different modes of vibration attributed to a single *Eigenwert* and a single Stationary State of the unperturbed hydrogen atom.[17]

[17] In the language of the mathematicians, the perturbing forces remove the degeneracy of the problem; some kinds of perturbation remove it completely, others in part.

The results are, in fact, just like those obtained with the Bohr-Sommerfeld atom-model; and this is somewhat embarrassing. For, in order to perfect the Bohr-Sommerfeld model and establish a complete analogy between (for instance) the sodium spectrum on the one hand and the fine-structure of the hydrogen lines on the other hand, it was necessary to introduce a new feature—the "spinning electron." Something of the sort must evidently be done again—the "spinning electron" must be imported into the undulatory mechanics; but the exact way to do it seems as yet to elude the virtuosi of mathematical physics.[18]

In one case—when the perturbing force is an impressed electric field—the results obtained by the method of Bohr and Sommerfeld and those obtained by the method of Schroedinger agree to first approximation with each other and with the data of experience, without the introduction of a "spinning electron." As this case of the "Stark Effect" furnishes a convenient transition to the last section of the article, I will quote the results.[19]

The Stark Effect

Imagine a hydrogen atom, upon which an electric field F parallel to some arbitrary direction which we call the z-direction is acting. Owing to this field, the electron at the point x, y, z and the nucleus at the origin (we are still using the concept of the nucleus and the electron!) possess a potential energy composed of the "intrinsic" term $-e^2/r$ and the "perturbation" $+eFz$. The wave-equation takes the form:

$$\nabla^2\Psi + \frac{8\pi^2 m}{h^2}\left(E + \frac{e^2}{r} - eFz\right) = 0. \qquad (183)$$

Paraboloidal coordinates are indicated for this problem. Instead of the planes, double-cones and spheres of the polar coordinate-system which we earlier used, it is desirable to employ planes and two families of paraboloids of rotation; the planes intersect one another along the line through the nucleus parallel to the field (hitherto called the z-axis), and the two families of paraboloids have their common foci at the nucleus and their noses pointing opposite ways along that axis. The transformation is made by the equations:

$$x = \sqrt{\xi\eta}\cos\varphi, \qquad y = \sqrt{\xi\eta}\sin\varphi, \qquad z = \tfrac{1}{2}(\xi - \eta) \qquad (184)$$

[18] Unless the problem has been solved by C. F. Richter (cf. preliminary note in *Proc. Nat. Acad. Sci.*, **13**, pp. 476–479; 1927).

[19] Schroedinger, *Ann. d. Phys.*, **80**, pp. 457–464 (1926); P. S. Epstein, *Phys. Rev.* (2), **28**, pp. 695–710 (1926).

and the wave-equation appears in this guise:

$$\frac{d}{d\xi}\left(\xi\frac{d\Psi}{d\xi}\right)+\frac{d}{d\eta}\left(\eta\frac{d\Psi}{d\eta}\right)+\frac{1}{4}\left(\frac{1}{\xi}+\frac{1}{\eta}\right)\frac{d^2\Psi}{d\varphi^2} \\ +\frac{2\pi^2m}{h^2}[E(\xi+\eta)+2e^2-\tfrac{1}{2}eF(\xi^2-\eta^2)]\Psi=0. \tag{185}$$

Essaying as tentative solution a product of a function of φ by a function of ξ and a function of η, we obtain as usual three differential equations, involving E and two other parameters, to which specific *Eigenwerte* must be assigned either because the variable φ is cyclic, or because for values other than these *Eigenwerte* the solutions become infinite for certain values of the variable.

Suppose that we set $F = 0$, and ascertain these *Eigenwerte*, and insert them into the equations: we then find the imaginary fluid vibrating in a stationary wave-pattern, oscillating in compartments divided from one another by nodal planes and by nodal paraboloids pointing up or down the field. To each of the energy-values E_n there correspond $(1 + 2 + 3 + \cdots n)$ distinct wave-patterns, each having a distinctive number k_1 of nodal paraboloids of the one family, a distinctive number k_2 of nodal paraboloids of the other family, and a distinctive number s of nodal planes; the values of k_1 and k_2 and s are limited by the conditions that they must be integers, that they cannot be less than zero nor greater than n, and that their sum must be equal to $(n - 1)$: that is,

$$k_1 + k_2 + s + 1 = n. \tag{186}$$

(Translating into the language of the electron-orbits, we find that s becomes the equatorial quantum-number which represents the angular momentum of the electron around the direction of the field (in terms of the unit $h/2\pi$) and k_1 and k_2 become the parabolic quantum-numbers.)

Introducing now the impressed electric field F, we find that among the $(1 + 2 + 3 + \cdots n)$ modes of vibration which originally shared the energy-value E_n, those for which $k_1 = k_2$ retain this energy-value, while the rest are displaced by varying amounts given by the celebrated Epstein formula:

$$\Delta E = \frac{3Fh^2n}{8\pi^2me}(k_1 - k_2). \tag{187}$$

The Stationary State of energy-value E_n is thus "resolved" or "split" into several—not, however, into the full number $(1 + 2 + 3 + \cdots n)$ corresponding to the total number of modes of vibration, for some of these still share identical energy-values. The line resulting from the transition between two States, E_3 and E_2 for instance, is thus resolved into a set of lines lying close together. These individual "Stark-effect components" testify to the individual existence of the several distinct modes of vibration which, when there is no impressed electric field, should share a common energy-value E_n and be indistinguishable from one another.[20]

In the closing section we shall consider another aspect of these Stark-effect components. At this point I wish only to allude to a quaint little paradox which may already have disconcerted the reader. I have just said that the imaginary "fluid" executes stationary vibrations in which it is divided into compartments by nodal planes and nodal paraboloids, even when the impressed field F is made equal to zero; but earlier I said that the "fluid" representing the unperturbed hydrogen atom executes vibrations in which it is divided into compartments by nodal planes, nodal double-cones and nodal spheres. There is no actual contradiction between these two assertions; for a mode of vibration of the one kind can be obtained by superposing two or more modes of vibration of the other kind, with a proper distribution of relative amplitudes. Take the specific case of the "first excited state" of the hydrogen atom, $n = 2$. By the earlier process, we find three wave-patterns: (a) with one nodal sphere, (b) with one nodal double-cone, (c) with one nodal plane. By the later process, we find three wave-patterns: (α) with one nodal paraboloid facing one way; (β) with one nodal paraboloid facing the other way; (γ) with one nodal plane. The wave-patterns (c) and (γ) are evidently the same, while either (α) or (β) can be reproduced by superposing (a), (b) and (c) with the proper relative amplitudes.[21] If the field F acting upon a hydrogen atom in the first excited state were to be gradually reduced to zero, it would leave the atom, or to speak more carefully the "imaginary fluid," vibrating in a manner which would be one of the modes (α), (β) or (γ), hence a cleverly adjusted superposition of the three modes (a), (b) and (c). Suppose however that a very small field F were to be applied to a hitherto unperturbed atom; why should it necessarily find ready for it a vibration with precisely the proper relative adjustment of the modes (a), (b) and (c)? or if it did not, if it should find the atom vibrating say in mode (α), how would it persuade the "fluid" to change over into the manner of vibration suitable for its own operations to begin?[22]

[20] A couple of "contour maps" of the wave-patterns for two of these paraboloidal modes of vibration are given by F. G. Slack (*Ann. d. Phys.*, **82**, pp. 576–584; 1927).

[21] I have not actually proved this, but believe that it must follow from Schroedinger's general theorem.

[22] This same curious thing occurs in a somewhat different guise when the electron-orbit theory is adopted.

Interpretation of the Rotator by Wave-Mechanics

The rotator or rotating body, the "spinning-top" as the Germans often call it, is a very important object in the workshop of the builder of atom-models. It is the accepted molecule-model used in theorizing about the polarization of gases by electric and magnetic fields, and the basis of the accepted molecule-model used in explaining the band-spectra of diatomic and polyatomic gases. Most models devised for the latter purpose combine the features of the rotator and the linear oscillator; but for the present purpose it is sufficient to view these separately, conceiving the rotator as a perfectly rigid whirling body.

The treatment of the rotator by wave-mechanics is in one respect admirably simple, but eventually we are led into the mazes of the General Equation with its non-Euclidean geometry. One can however avoid the complexity long enough to benefit by the intelligible feature, by considering first a rotator such as was invented more than fifty years ago to account for the specific heats of diatomic gases such as hydrogen—a dumbbell not permitted to spin around its own axis-of-figure or line-of-centres, but revolving around some axis passing through its center-of-mass perpendicular to its line-of-centres. The orientation of such a dumbbell is specified by the angles θ and ϕ which define, in a polar coordinate-system, the direction in which its axis-of-figure is pointing. The energy is exclusively kinetic, so that the term containing V vanishes from the wave-equation, a circumstance which is very helpful. Representing by A the moment of inertia of the dumbbell about the axis of revolution, we find the wave-equation in the form:[23]

$$\nabla^2\psi + \frac{8\pi^2 EA}{h^2}\,\psi = 0. \tag{190}$$

In this equation the Laplacian operator is to be expressed in the polar coordinates θ and ϕ, as it was expressed in equation (131), but without the terms involving the third and missing coordinate r. We have before us, therefore, the second of equations (133), with a specific value for the constant there called λ:

$$-\operatorname{cosec}\theta\left[\frac{d}{d\phi}\left(\operatorname{cosec}\theta\frac{d\psi}{d\phi}\right)+\frac{d}{d\theta}\left(\sin\theta\frac{d\psi}{d\theta}\right)\right]=\frac{8\pi^2 EA}{h^2}\,\psi. \tag{191}$$

Here, as there, the function ψ must repeat itself whenever θ is altered by any multiple of π and ϕ by any multiple of 2π; for then we are back at the same place, i.e. at the same orientation of the rotator.

[23] Schroedinger, *Ann. d. Phys.*, **79**, pp. 520–522 (1926).

Here, as there, this necessity imposes of itself a certain condition upon the coefficient of ψ in the right-hand member, which is tantamount to this condition imposed on E:

$$E=n(n+1)\frac{h^2}{8\pi^2A}=(n+\tfrac{1}{2})^2\frac{h^2}{8\pi^2A}+\text{const.},\qquad n=0,1,2,3\cdots. \tag{192}$$

The energy of the rotator is thus *by the mere fact that the variables are cyclic* limited to a single sequence of permitted values, furnishing incidentally another example of half-quantum-numbers.

The *Eigenwerte*, the permitted energy-values, are thus for the rotator determined by an exceptionally lucid condition; yet the complications of the General Equation already appear on the horizon. Equation (190) differs from the wave-equation which I have hitherto used by virtue of the substitution of moment-of-inertia A for mass m. This replacement seems sensible enough; one might rely on intuition in this particular case; but strictly it is caused by the form preassumed for the General Wave-Equation and by the specific form of the kinetic-energy-function for this specially restricted kind of rotator. If now we remove the restriction, and permit the rotator to spin about its axis-of-figure at the same time as it whirls about some axis normal to that—if we pose the general problem of the rigid rotator unrestricted save by the conditions which the wave-equation imposes, it is necessary to invoke the General Equation with the non-Euclidean geometry. The problem is soluble, and has been solved;[24] the utility of the results for the interpretation of band-spectra gives valuable support to the form selected by de Broglie and Schroedinger for the General Equation.

The polarization of a gas by an electric (or magnetic) field may be treated by supposing that each atom is an electric (or magnetic) doublet. The treatment is simplest if we may assume that the electric (or magnetic) axis of the doublet coincides with the axis-of-figure of a dumbbell-molecule, not allowed to spin around its axis-of-figure. Let M stand for the moment of such a doublet, and suppose the field H to be parallel to the direction from which the angle θ of the foregoing paragraphs is measured. The field supplies the potential-energy term to be added to the left-hand member of equation (190); this new term is:

$$-V\psi=(MH\cos\theta)\psi. \tag{193}$$

It is easy to see that the wave-equation has *Eigenwerte*, so that the atoms are in effect limited to certain "permitted" orientations in the

[24] F. Reiche, *ZS. f. Phys.*, **39**, pp. 444–464 (1926); R. de L. Kronig, I. I. Rabi, *Phys. Rev.* (2), **29**, pp. 262–269 (1927).

field—a conclusion from the earlier atomic theory, which for magnetic fields has become a fact of experience through the experiments of Gerlach and Stern and others. To calculate the polarization of a gas, it is necessary to make a further assumption concerning the relative probabilities of these various orientations in a gas in thermal equilibrium; having done so, one obtains a formula for the polarization, or the dielectric constant, or the susceptibility of the gas as function of applied field and temperature. The customary assumption leads to a formula which, in the limiting case of high temperature and low field, agrees with the celebrated equation of Langevin for the polarization of a paramagnetic gas by a magnetic field:[25]

$$\text{Susceptibility} = I/H = NM^2/3kT. \tag{194}$$

Interpretation of the Free Electron in Wave-Mechanics

We now depart from the calculation of *Eigenwerte* and Stationary States, and return to the original ideas of de Broglie.

For a free electron moving in a field-free region—or any particle moving in a region where no force acts upon it—with a constant speed v along a direction which I will take as the x-direction, the (non-relativistic) wave-equation assumes the form:

$$\frac{d^2\psi}{dx^2} + \frac{8\pi^2 mE}{h^2}\psi = 0 \qquad (E = \tfrac{1}{2}mv^2). \tag{195}$$

This equation admits a sine-function as its solution whatever the value of the constant E and consequently does not restrict the energy-values which the electron is allowed to take (a contrary result would have been hard to swallow!). Assigning the value E/h to the frequency of the sine-wave and the value $E/\sqrt{2mE}$ to its speed, we obtain for the wave-length of the wave-train, "associated with" a free electron (or free particle) of mass m and speed v, this value:

$$\lambda = \frac{E/\sqrt{2mE}}{E/h} = \frac{h}{\sqrt{2mE}} = \frac{h}{mv}\,\cdot \tag{196}$$

For electrons of such speeds as ordinarily occur in discharge-tubes, these wave-lengths are of the magnitude of X-ray wave-lengths; for instance, a 150-volt electron is associated with a wave-length of very nearly one Angstrom unit.

[25] C. Manneback, *Phys. ZS.*, **28**, pp. 72–84 (1927); J. H. Van Vleck, *Phys. Rev.* (2), **29**, pp. 727–744; **30**, pp. 31–54 (1927). For the classical derivation of formula (194) and meaning of the symbols cf. my article "Ferromagnetism," this Journal, **6** (1927), pp. 351–353.

This coincidence makes one wonder whether, if a stream of such electrons were projected against a crystal such as is used for diffracting X-rays, there would be a semblance of diffraction. Nothing yet said about the waves leads inevitably to such an inference. On the contrary, it might well be argued that we have no greater justification for expecting to observe them in the ordinary world of space and time than for expecting the x and the y of an algebraic equation to come to life before our eyes. It might forcibly be pointed out that while in this instance and the instance of the hydrogen atom the "waves" are defined in ordinary space, there are other instances—supplied for instance by rotators—in which the wave-equation is formally similar to (195) and the theory quite as effective, and yet the alleged "waves" exist only in the configuration-space and indeed in non-Euclidean configuration-space, which is much the same as saying that they do not exist at all. Nevertheless it appears that there is indeed a diffraction of electrons by crystals, and that the wave-length indicated by the diffraction-angles is in accordance with the value given by de Broglie! The first evidence for this amazing and portentous effect will be narrated by its discoverers Davisson and Germer in the following issue of this Journal.[26]

Notice that the speed of the associated wave-train is not the same as that of the flying particle; it is $\sqrt{E/2m}$, that of the particle is $\sqrt{2E/m}$. It is, however, the wave-length of the wave-train which is measured by the diffraction-experiments; not the speed, and not the frequency. This is important; for it is the wave-length which is exempt from the consequences of the essential uncertainty in the value of E. In classical mechanics, energy-differences alone are definite, but the absolute values of the "energy" of a system are not defined; the definition of energy involves an arbitrary additive constant. If now we were to add an arbitrary constant to the kinetic energy of the free electron, and call E the sum of the two, we should alter the frequency and alter the speed assigned to the wave-train; but we should not alter the wave-length, for the wave-length is strictly equal to $h/\sqrt{2m(E-V)}$ with V standing for the potential energy of the free electron, and the added constant would enter into V and vanish by subtraction. Returning to the preceding sections of this

[26] Consult meanwhile their note in *Nature*, **119**, pp. 558–560; 1927. The prediction was first published by W. Elsasser (*Naturwiss.*, **13**, p. 711; 1925). For additional intimations of evidence for undulatory qualities in matter cf. G. P. Thomson, A. Reid (*Nature*, **119**, p. 890; 1927); T. H. Johnson, *Nature*, **120**, p. 191 (1927); E. G. Dymond, *Phys. Rev.* (2), **29**, pp. 433–441 (1927). Schroedinger's elegant treatment of the Compton effect is based upon the conception of electrons as wave-trains (*Ann. d. Phys.*, **82**, pp. 257–264; 1927); for a more elaborate treatment of Compton effect cf. W. Gordon (*ZS. f. Phys.*, **40**, pp. 117–133).

article, we see that Schroedinger calculated the energy-values of the Stationary States by conditions imposed upon the wave-lengths of the waves, not upon their frequencies; the wave-patterns depend only upon the wave-lengths, and the frequency of the light which an atom emits in passing between two Stationary States depends only on the difference between their energy-values. In relativistic mechanics, energy is defined absolutely, and this difficulty never even threatens to arise; yet it is worth while to note that the ambiguity of the concept "energy" in classical mechanics does not interfere with, nor is it resolved by, anything which has been observed in Nature and interpreted by wave-mechanics.

In relativistic mechanics, the wave-equation for the free-flying particle assumes the form:

$$\frac{d^2\psi}{dx^2} + \frac{4\pi^2}{h^2c^2}(E - m_0^2c^4) = 0 \qquad \left(E = \frac{m_0c^2}{\sqrt{1 - v^2/c^2}}\right). \tag{197}$$

The wave-length has the value $h\sqrt{1 - v^2c^2}/m_0v = h/mv$; the frequency is $m_0c^2/h\sqrt{1 - v^2/c^2}$; the speed of propagation of the waves is c^2/v, superior to the speed of light.

I can no more than allude to the strangely suggestive fact, that in general as well as in this special case the speed of the particle and the speed of the associated waves are related to one another in the same way as group-speed and wave-speed in ordinary optics.

Attempt to Find a Meaning for the Symbol Ψ

Thirty-three years ago the Earl of Salisbury, invited by reason of his eminence as a statesman to be the President of the British Association for the Advancement of Science, observed the physicists of his day involved in their fervent speculations over the nature of the æther; and of an address brilliant with felicitous phrases the best-remembered words are those by which he described the outcome of their travail: *The main, if not the only, function of the word* aether *has been to furnish a nominative case to the verb 'to undulate.'* Quite the same thing could now be said of the symbol Ψ, insofar as it serves to determine the energy-values of the Stationary States of the systems devised to represent atoms. When it is used for this purpose, it vanishes just as the final triumph is achieved. Like the variable under the sign of integration in a definite integral, it drops out of sight when the calculations which it proposes are actually performed. Indeed it might be discarded altogether during the process of calculating *Eigenwerte* and energy-values; one might speak exclusively of the "differential operator" $\nabla^2 - 8\pi^2m(E - V)/h^2$; many mathematicians do so.

Schroedinger however conceived the daring, the admittedly tentative and still incomplete but very alluring, idea of seeking in Ψ some measure of the density of electric charge. Specifically, it occurred to him to define the square of the amplitude of the oscillations of Ψ, which the *Eigenfunktionen* prescribe as function of the coordinates—to define this squared amplitude as the density of the electric charge, spreading the electron as it were throughout space.

Let us examine this idea, and see whither it leads.

To avoid avoidable complexities as far as possible, I take the simplest of all cases: the linear oscillator, represented by the imaginary "string" stretched along the x-axis, possessed of a wave-speed varying as $\sqrt{1 - x^2/L^2}$, real from the origin both ways as far as the points $x = \pm L$ and imaginary thenceforward. I will also refer to the still simpler "actual" case which served as an introduction to this one: the problem of the stretched string, clamped at its extremities at $x = \pm L$, possessed of a uniform real wave-speed u at all points between.

In both these cases of the imaginary and the real string, the search for the *Eigenwerte* and the *Eigenfunktionen* leads us to diverse natural modes of vibration, executed with various frequencies $\nu_0, \nu_1, \nu_2, \nu_3 \cdots$ and displaying stationary wave-patterns described by the *Eigenfunktionen:*

$$y_i = f_i(x)(A_i \cos 2\pi\nu_i t + B_i \sin 2\pi\nu_i t); \quad i = 0, 1, 2, 3 \cdots \qquad (201)$$

For the real string the functions $f_i(x)$ are sine-functions; for the imaginary strings which are the model of the linear oscillator, they are given by (155). I recall once more that in the latter case we have, not distinct modes of oscillation of one string, but the fundamental modes of as many strings as there are Stationary States.

When the real string is vibrating in the ith mode, or when we are dealing with the ith imaginary string, the function $f_i(x)$ is proportional to its vibration-amplitude. The form of equation (201) shows that this amplitude at any fixed point remains constant in time.

If the square of the vibration-amplitude is to be regarded as the density of electric charge along the string, it follows that when the oscillator is in one of its stationary states, and the string therefore vibrating in one of its modes, the density and the distribution of charge remain everywhere constant. There would be no to-and-fro motion of charges, and no tendency to radiation.

Suppose now that the real string is vibrating simultaneously in two modes, the ith and the jth; or that we have both the ith and the jth imaginary string coexisting (this is where the model is clumsiest!).

The vibrations are described by the following formula (I have put $A_i = A_j = 1$ and $B_i = B_j = 0$, which simplifies without injury to the generality of the result):

$$y = y_i + y_j = f_i(x) \cos 2\pi\nu_i t + f_j(x) \cos 2\pi\nu_j t, \tag{202}$$

which is easily transformed thus:

$$y = C \cos (2\pi\nu_i t - \alpha), \tag{203}$$

in which

$$C^2 = f_i^2 + f_j^2 + 2f_i f_j \cos 2\pi(\nu_i - \nu_j)t, \tag{204}$$

and α = a constant not important for our purpose.

Here we have a vibration in which the amplitude at any fixed point varies with time; the square of the amplitude is the sum of a constant term and a sine-function of time, and the frequency of the sine-function is the difference between the frequencies of the two coexisting modes of vibration.

Identifying the square of the amplitude with the density of electric charge, we see that this charge-density varies at each point with the frequency $(\nu_i - \nu_j)$. We might therefore expect radiation of this frequency.

Now the vibration-frequencies ν_i and ν_j corresponding to the modes of vibration, that is to the Stationary States i and j having energy-values E_i and E_j, are respectively E_i/h and E_j/h.

If therefore—to speak in a vague but suggestive fashion—the linear oscillator were simultaneously in two Stationary States, their energy-values being E_i and E_j, then the square of the amplitude of the oscillations of Ψ would be fluctuating at each point of the "imaginary string" with the frequency $(E_i - E_j)/h$; and if this squared amplitude were to be identified with charge-density, then the system might be expected to emit radiation of the frequency $(E_i - E_j)/h$.

Transition between two states would then signify *coexistence* of the two states.[27]

We proceed a step further in the development of this idea, by forming the following integral:

$$M = \int_{-\infty}^{\infty} xC^2 dx = \int_{-\infty}^{\infty} xf_i^2 dx + \int_{-\infty}^{\infty} xf_j^2 dx + \left\{ 2\int_{-\infty}^{\infty} xf_i f_j dx \right\} \cos 2\pi(\nu_i - \nu_j)t. \tag{205}$$

This integral represents the *electric moment* of the supposed distribu-

[27] I should again recall that in the picture we have, not two distinct coexisting modes of vibration of the same elastic string; but the fundamental (and solitary) modes of vibration of two distinct elastic strings.

tion of "electric charge" along the imaginary string, relatively to its centre at the origin. If it should turn out zero, there would be equal amounts of charge to left and to right of the centre; if it should turn out positive or negative, there would be more charge to the right of the centre than to the left, or more to the left than to the right; if it should turn out variable, if for instance the coefficient of the cosine-term should differ from zero, there would be a *surging of the charge* to and fro across the origin.

The functions $f_i(x)$ have been written down in equation (155), near which it was shown that they are alternately even and odd functions of x; $f_0, f_2, f_4 \cdots$ are even, $f_1, f_3, f_5 \cdots$ are odd. Their squares are consequently even, the products of their squares by x are odd, functions of x; and the first two integrals in the expression (205) for M vanish. As for the integral $\int_{-\infty}^{\infty} x f_i f_j dx$, its integrand is an odd function of x if i and j are both even or both odd, and in either case it vanishes; so that if two wave-patterns corresponding both to even-numbered or both to odd-numbered Stationary States coexist, there is no surging of charge to and fro, and the electric moment of the system remains constant. If i is even and j odd, or vice versa, the conclusion is not so immediate. It follows however from the general properties of the Hermite polynomials [28] that the integral $\int_{-\infty}^{\infty} x f_i f_j dx$ always vanishes unless i and j differ by one unit, so that in every case but this the electric moment is continually zero. This leads us to the rule:

If two modes of vibration i and j coexist, the electric moment of the "imaginary string" representing the linear harmonic oscillator varies sinusoidally with the frequency $(\nu_i - \nu_j)$, if and only if $i = j \pm 1$; otherwise the electric moment is and remains zero.

This may be interpreted as meaning physically that radiation occurs only when two "adjacent" states—two states for which the quantum-number differs by one unit—coexist; that *transitions are possible only between adjacent states.*

This is a Principle of Selection. It is the principle of selection predicted for the linear harmonic oscillator in the earlier versions of atomic theory, and sustained by observations on those features of band spectra which are attributed to simple-harmonic vibrations of molecules.

Thus in the case of the linear oscillator, the idea of interpreting the square of the amplitude of the Ψ-vibrations as density of electric charge is twice successful. When the oscillator is in one of its sta-

[28] Courant-Hilbert, *Methoden der math. Physik*, p. 76.

tionary states, the distribution of "charge" along the imaginary string which represents it is stationary; when the vibrations corresponding to two distinct Stationary States coexist, the distribution of the "charge" fluctuates, with precisely the frequency which experiment teaches us to expect from a transition between the states in question; when and only when the two coexisting stationary states are adjacent in the ordering, when and only when experiment teaches us to expect transitions, the fluctuation assumes the emphatic character of a bodily surging of the charge to and fro across the centre of the string.[29]

One further desirable result of identifying square of amplitude of Ψ with density of electric charge appears when from one dimension we go over to systems of two or three dimensions. As illustration I take the example of an hydrogen atom exposed to an electric field, represented by an imaginary fluid in three dimensions, the stationary wave-patterns of which correspond to the stationary states of the perturbed atom. If two of these stationary wave-patterns coexist, there may be a bodily surging of charge to and fro, with the frequency belonging to the transition between the stationary states which the wave-patterns represent. If in particular two wave-patterns sharing a common value of the quantum-number s (the "equatorial quantum-number," equation 186) coexist, there is a surging of the "charge," and its to-and-fro motion is parallel to the applied electric field; there is no component of the motion normal to the field. With this result agrees well the fact of experience, that the light emitted by virtue of transitions between stationary states differing only in the quantum-numbers k_1 and k_2, and sharing the same value of s, is polarized with its electric vector parallel to the field. Again, if two wave-patterns for which the values of s differ by one unit coexist, the resultant surging of the charge is perpendicular to the electric field; and it is a fact of experience that the light due to transitions between stationary states having values of s one unit apart is polarized with electric vector normal to the field. Finally, if two wave-patterns for which the values of s differ by two or more units coexist, there is no far-sweeping dis-

[29] Schroedinger has shown that if we conceive a great number of Stationary States with high values of i and artfully chosen relative "amplitudes" (i.e. values of A_i and B_i in equation 201) to exist simultaneously, we find the "charge-density" concentrated into a small region, a sort of knot or bundle which oscillates back and forth across the centre of the string with frequency ν_0 and with approximately the same amplitude of vibration as the original particle (the particle with mass m and restoring-force $-4\pi^2 m\nu_0^2 x$ of which the string is our image in wave-mechanics) would have if its energy were the same as that of the Stationary State which was made most prominent in the summation (*Naturwiss.*, **14**, pp. 664–666; 1926). This is a promising result, suggesting as it does that atoms in highly excited states may be groups of particles which, as the system returns to normalcy, spread out into a sort of fluid haze. The idea can be generalized widely, and merits a thorough analysis.

placement of charge; and in the spectra, the lines which such transitions should cause are missing.

Thus in the case of the hydrogen atom exposed to an electric field, and in other two- and three-dimensional systems as well, the identification of the square of the amplitude of the Ψ-vibrations with density of electric charge is thrice successful. In the picture, we see the electric charge stationary when the system is in a stationary state, fluctuating with the proper frequency when two states coexist; we see it surging back and forth *en masse* when the coexisting states are two between which a transition is "permitted," and otherwise not; we see it surging back and forth along the proper direction to explain the polarization of the light which results from the transition. As a device for picturing the radiation-process, Schroedinger's model is certainly unrivalled. In the earlier atom-models, even the frequencies of the emitted rays of light and the frequencies of the intra-atomic vibrations did not agree. Here at last they do, and when a tube full of hydrogen atoms is pouring out the light of the red Balmer line with its frequency of $4.57 \cdot 10^{14}$, it is permissible at last to imagine each of them as a mechanism, within which something is vibrating $4.57 \cdot 10^{14}$ times in a second.

Even the relative intensities of spectrum lines may fall within the scope of wave-mechanics. We have seen that in the case of the linear oscillator, the vanishing of the integral $\int x f_i f_j dx$ for all pairs of Stationary States for which i and j differ by more than one unit entails the non-occurrence of the corresponding transitions, the inability to emit or absorb the corresponding radiation. May it not be that the intensity of the radiation emitted by reason of the transition between any two states of any system, and polarized parallel to any direction x, is governed by the value of the integral $\int x \psi_i \psi_j dx$ involving the *Eigenfunktionen* ψ_i and ψ_j of the states in question? To develop this idea more assumptions must be introduced than I have yet mentioned, since every *Eigenfunktion* which I have thus far written down might be multiplied by any constant factor without ceasing to be an *Eigenfunktion*, and some rule must be laid down to fix these constant factors. To predict the relative intensities of the components into which certain hydrogen-atom lines are split by electric field, Schroedinger made a simple and natural assumption about these factors; and the results turned out to be in good agreement with the data.[30] I cannot enter further into this topic, except to remark that the point of contact between wave-mechanics and the matrix-mechanics of Heisenberg lies here; for the integrals in question figure as matrix-

[30] Schroedinger, *Ann. d. Phys.*, **80**, pp. 464–478 (1926); *Phys. Rev.* (2), **28**, pp. 1049–1070 (1926).

elements in the latter theory, which indeed appears to be an alternative way of thinking to reach the same conclusions as emerge from the speculations of de Broglie and Schroedinger.[31]

Nevertheless the image is still far from perfect; there is certainly something still lacking, something still to be discovered and added. Radiation may flow forth from the atom when two stationary states coexist, but it does not flow forever; one or the other of the wave-patterns must therefore die out, soon after the radiation commences; yet no agency has thus far been provided to effect the extinction of either. It may not be difficult to insert such an agency into the theory, in the form perhaps of an interaction between the Ψ-waves and the outflowing electromagnetic waves. It may be much more troublesome to extricate ourselves from the paradox into which the identification of square-of-amplitude-of-the-Ψ-vibration with density-of-electric-charge has led us. All of the numerical agreements between this theory of the hydrogen atom and the features of the hydrogen spectrum are obtained by putting $-e^2/r$ for the potential-energy-function of the atom-model. This is the potential-energy-function for a point-nucleus and a point-electron. If we dissolve the electron, spread it out like a cloud in space around the centre of the atom, how can we consistently retain the potential-energy-function derived from the picture of a point-charge? How is it defensible to define electric charge in one way in order to lay the cornerstone of the new theory, and then redefine it in a contrasting way in order to raise the superstructure?

Wave-mechanics, striking as are the pictures which it offers of certain of the processes within the atom, still abounds in conceptual difficulties of which the last is a fair instance; and those who share the view of Lessing that it is more desirable to be approaching truth perpetually than ever to attain it may still find satisfaction in physics. Wave-mechanics still is tentative, not definitive; a plan of campaign, rather than a conquest. The outcome cannot now be foreseen. Yet we may reflect that twenty-five years ago it was universally supposed that light possesses only the qualities of a wave-motion; and then experiment was piled upon experiment which showed that in addition it behaves in many situations as though it were a stream of corpuscles. Perhaps we stand at the beginning of an equally imposing series of experiments, which will show that matter with equal inconsistency partakes of the qualities of particles and of the qualities of waves.

[31] Schroedinger, *Ann. d. Phys.*, **79**, pp. 734–756 (1926); C. Eckart, *Phys. Rev.* (2), **28**, pp. 711–726 (1926).

Power Plants for Telephone Offices

By R. L. YOUNG

SYNOPSIS: The present paper gives a brief discussion of some of the more important problems connected with the supplying of power to telephone offices, and developments which are being perfected to bring about economies. Among the subjects discussed are the use of commercial types of charging generators together with appropriate filters, power factor correction, complete power unit assemblies for small installations, and the development of more nearly automatically controlled power installations with the object of reducing supervision.

I. THE POWER PROBLEMS

The purpose of the telephone power plant is to furnish energy of the required character in proper amount and available 100 per cent of the time. An elaborate telephone system, comprising buildings, central office equipment, outside plant lines and substation apparatus, together with a staff of operators, is rendered useless if the supply of power fails. No conversations can be held. No calls can be made and none received. In a way, the power plant might be termed the "heart" of the system, since every line and connection will be "dead" the moment the supply of power is interrupted.

Continuity

In order to meet the vital need of ever-ready power it is necessary in telephone power plants to arrange for some primary power source which is usually a commercial electric service from outside. The services are investigated with care to determine their reliability and, wherever possible, two services connected to different generating stations or systems are brought into the telephone building. In those cases where a single service only can be secured, a local means of charging such as an engine-generator set may be provided as a reserve on this service.

Even with the best commercial power services short interruptions are experienced, so that it is necessary to provide another source which shall be available at all times to operate the central office during temporary failures of the outside service. This is accomplished by the use of a storage battery of sufficient capacity to carry the load of the office during failure of the sources of power supply, the battery being continuously connected to the circuits so no interruption occurs. Common practice and experience have resulted in batteries of certain sizes being provided, these sizes being sufficient to carry the exchange load for intervals ranging from a few hours to several days, depending upon conditions. The present practices have been successful in maintaining continuous power supply, and central offices generally throughout the country have been ready to serve, even during periods of storm, fire or other calamities.

Type of Power Needed

Power as furnished by the public service companies is not of the sort suitable for operating telephone power plants, but must be converted from a relatively high voltage alternating or direct current to a lower

Fig. 1—Incoming direct-current power for large telephone building. About 1,000 h.p. of this is provided to drive motor-generators for reserve central office use, the regular power being alternating current. Both direct- and alternating-current services are duplicated. This panel provides four feeders direct to substation and four to network, capacity 3,480 kw.

voltage direct current for talking, supervisory and signaling purposes and to alternating current of various voltages for signaling. This conversion is commonly made by means of motor-generator sets or some type of rectifier, of either the mercury arc, hot cathode or other types. Since it is impossible to use outside power as furnished, suitable reserve machine equipment must be provided capable of replacing the regular machines before the reserve energy in the central office battery is exhausted.

The low voltage charging generators furnishing the bulk of the power must be electrically quiet so that they will not cause disturbing noises in the telephone circuits. It is, of course, economical to furnish most of the energy required by the telephone equipment directly from the motor-generator sets rather than from the reserve battery, since the conversion efficiency is substantially greater and the battery investment is much less. While various direct-current voltages are required, 24 volts and 48 volts predominate.

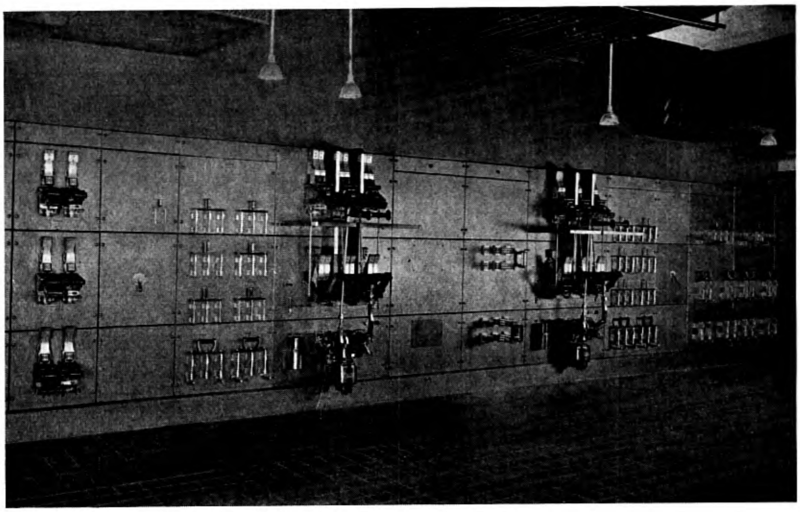


Fig. 2—Building switchboard to distribute incoming power shown in Fig. 1. The 5,000-ampere circuit breakers switch the important load circuits from this panel to a similar reserve panel fronting on a different street.

The signaling machines and batteries, while of relatively small output, are subject to rather exacting performance limitations. Twenty-cycle alternating current of approximately sine-wave form, at nominal voltages of 105, 100, 85, 77, etc., is needed for ringing on different types of circuits. For four-party selective ringing, positive and negative direct currents are superimposed upon the alternating current to secure wave shapes especially suited to the operation of biased ringers. For machine ringing, the 20-cycle current is divided into one or two second ringing periods separated by silent intervals during which direct current is provided for operating the tripping relay and stopping the ringing when the called subscriber answers.

For ringing over composited toll lines a higher frequency, which will not interfere with telegraph operation, is required and 135 cycles is provided. For other types of toll circuits "voice-frequency" ringing at 1000 cycles must be furnished.

Message registers in manual offices use direct current at 39 volts, coin collect and refunding operations require positive 110-volt and negative 110-volt direct current, while "tones" of approximately 160 and 480 pulsations per second are needed for giving various signals to the operators and the subscribers. A graduated tone, like a siren, is required for the "howler" used to call the attention of a subscriber to a telephone receiver left off the hook. Various flashing signals and combined tones and flashes are also used, such for example as the "busy" signal.

Operation and Maintenance

In addition to being designed for furnishing power of the required characteristics, the machines and apparatus must operate for long periods with a reasonably small amount of operating attention and maintenance. Due to the narrow requirements being placed by the circuits upon the power equipment and the more frequent readjustments required, it is becoming necessary in many cases to furnish automatic voltage regulation. As the cost of labor increases, it will become still more desirable to provide equipment which will largely run and regulate itself.

Sizes of Power Equipment

It has been stated at different times by people connected with the telephone companies and also by outsiders that the amount of power required to carry on telephone conversation is microscopically small, if not negligible. This perhaps is true when considering merely the small amount of alternating current which travels over the line and operates the diaphragms of the receivers. The great sensitivity of this instrument permits operation on very small energy.

There is, however, a large amount of equipment in the central office, including relays, lamps, and other apparatus, which must function in order that this small talking current may be provided and may go from the subscriber who wishes to talk to the subscriber he desires to reach. When this apparatus is multiplied for the thousands or even hundreds of thousands of conversations per day which may be supplied from a power plant serving two, three or perhaps more central offices, the size of equipment needed becomes quite substantial. In these multi-office power plants several of the largest charging generators each driven by an 80 h.p. motor, as well as a number of smaller charging sets may be required, while two batteries of the largest storage battery cells manufactured may be used in parallel to give the necessary battery reserve.

Some of the large telephone buildings house several central offices and, in addition, administrative, engineering, commercial and other departments. A joint incoming power service is often provided for such a building, of which the initial telephone power plant requirements may approximate 500 h.p. with an estimated energy consumption approximating 1,000,000 kw.-hrs. per year. Provision for double this demand in the ultimate may be made.

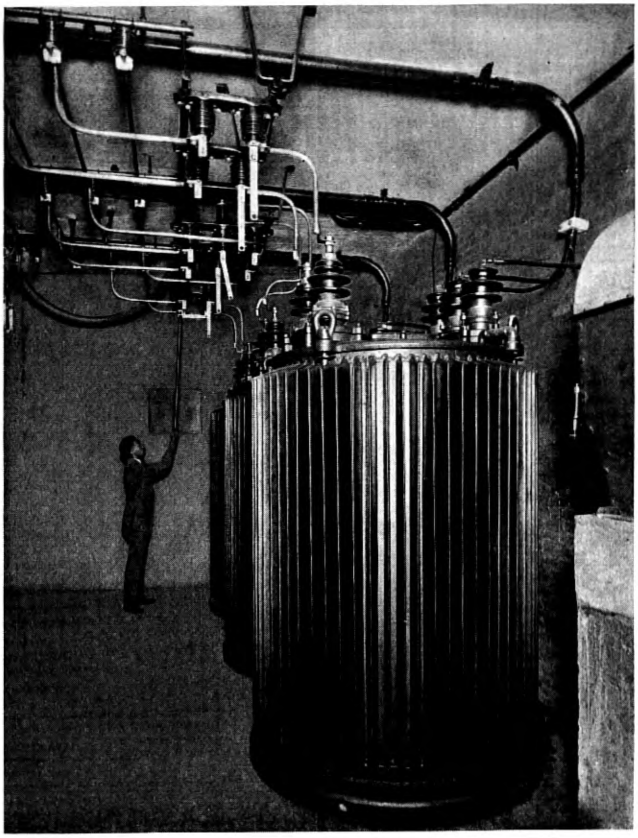


Fig. 3—Typical large transformer installation for breakdown service when two types of alternating current are furnished. Three 333-kv.-a., 11,000/2,200-volt transformers supply frequency changer, these being in addition to the 60-cycle transformers for regular service.

The range of sizes is very great, varying from the above down to the small "magneto" office which operates largely on dry cells and other primary batteries, and may also take 1/8 h.p. to run a magneto ringing machine from the power service. In such offices without electric power supply, all the equipment must be operated from primary batteries. The "magneto" office, so-called, is one which serves "local-battery" subscribers each of whom has dry cells and a hand ringing generator or magneto. In the larger "common-battery" systems all the power for both talking and signaling is provided from sources at the central office common to all subscribers.

In dial offices it is evident that more power equipment is required, since the processes of connecting through the circuits are performed by machine instead of by operators.

Cost of Power

The cost of power as purchased from the public service companies varies largely, depending upon the location, the amount purchased and to some extent upon the characteristics of the load. In large cities, power is billed at from 2 to 5 cents per kw.-hr. Usually a sliding

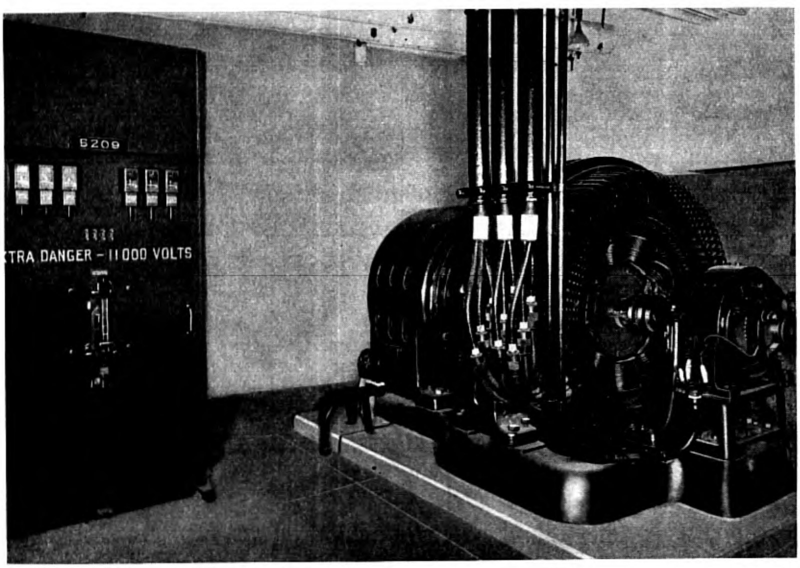


Fig. 4—Frequency changer set for breakdown service, 650-h.p., 25-cycle, 2,200-volt to 600-kv.-a., 60-cycle, 220-volt. Emergency power controlled by 11,000-volt truck switch.

scale is offered and the lower figures apply to purchases of alternating current in large quantities. Cities near sources of soft coal supply or near large water power developments get cheaper rates, in some cases being nearer 1 cent than 2. In small offices, power usually costs between 5 and 10 cents per kw.-hr., running as high as 15 cents in a small percentage of outlying rural offices. For studies involving the use of power furnished from a telephone power plant, it will, of course, be necessary to consider the cost of the machine and battery equipment, of the floor space and of the operating attendance in addition to the cost of the "raw material" power as purchased. A fair overall figure, including these charges, might approximate 30 cents per kw.-hr. for a typical dial office, or 40 cents for a typical manual office, the higher charge for manual offices, in general, being accounted for by the fact that the quantities purchased and used are less, involving somewhat higher purchase price and overhead. It should, of course, be appreciated that the amounts will vary considerably with local conditions including the type of equipment used and the "load factor," or the distribution of load throughout the day and night. In most telephone power plants this factor is unfavorable for low cost power since most of the traffic is concentrated within a few hours of the twenty-four. The cost of energy varies also during the life of the same office, being higher during the early years and lower when load on the power equipment more nearly approaches capacity.

II. Some Developments to Meet the Power Plant Problems

The objectives toward which development work is directed are improved service, reduced cost, simplification of installation and decreased maintenance. Under these headings one of the most important developments at the present time is the use of commercial type charging generators.

Commercial Type Charging Generators

Charging motor-generator sets furnish most of the energy used in telephone power plants. Up to the present time "telephone generators" have been built to give an electrically smooth direct-current output which will not cause interference with conversations when furnishing current to the telephone circuits. They have also been made mechanically quiet so as not to interfere with nearby testing. They are quite special in construction, including smooth core armature and brass gauze brushes, and are subject to certain limitations which make them larger and considerably more expensive to build than ordinary machines of the same capacity.

Filters consisting of choke coils and high capacity electrolytic condensers have been developed, and with these filters commercial type charging generators can be used to float or charge the central office battery, and this type of generator is now being made available. The purpose of the filter is to make the current from the discharge leads of the power plant sufficiently quiet for talking battery supply. It is also possible to use a somewhat higher speed machine which is smaller than the present type. The usual slotted mica commutator construction and self-lubricating carbon brushes are employed. While the mechanical noise tends to be greater because of the higher speed and the carbon brushes, it has been found that this is not a factor of importance under present conditions where power plants can usually be located more or less by themselves and well removed from the Wire Chief's testing equipment.

Filters

A new type of choke coil has been developed foı the filter used with commercial type generators. It is of the enclosed shell type design having short air gaps, using the materials more economically and having a higher inductance than the coils which have been available

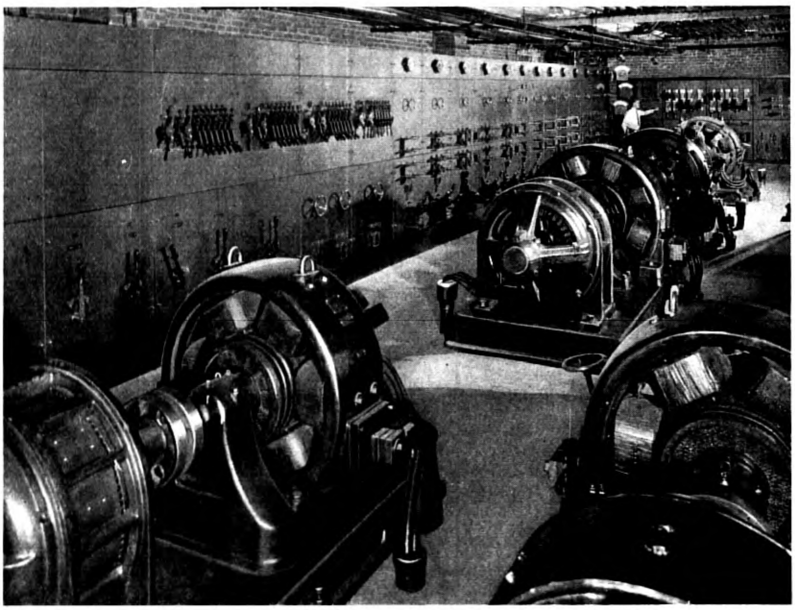


Fig. 5—Part of power room for two large panel units as provided five years ago. Twelve motor-generators for both alternating-current and direct-current service with power switchboard at left and battery control board in background. This was an alternate arrangement to use of large motor-generator to make both kinds of power available.

heretofore. Associated with this coil is a group of electrolytic condensers each of which has upwards of 1,200-mf. capacity on 24 volts, or roughly half this amount on 48 volts.

For one of the larger motor-generator sets the cost of the first set, including a filter, is about the same as that of the former "telephone generator" type outfit. The cost of the additional sets used in the power plant, which under present conditions are not provided with individual filters if a common filter is placed in the talking discharge circuit, is about half that of the present type sets.

Power Factor Correction

In connection with the new generators, synchronous motors are being made available for use where it is desired to improve the power factor of the load. The synchronous motors will be arranged to give

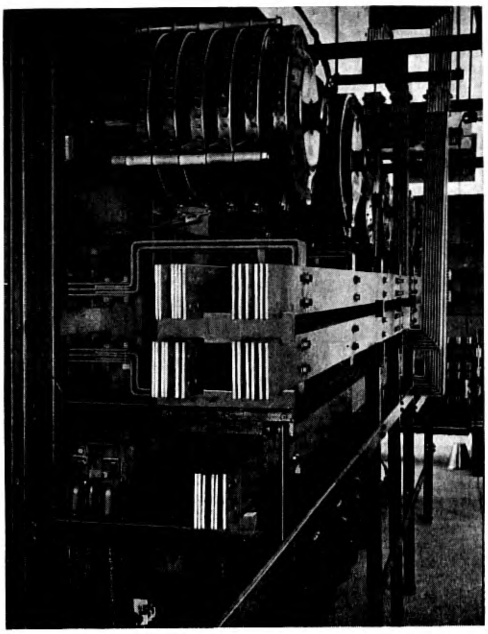


Fig. 6—Generator end of control switchboard, as provided five years ago, rear view. Six 1,000- and 1,500-ampere generators, bus bars terminated for growth when additional units required.

0.8 leading power factor, so all motors in the plant will not need to be of this type. The standard induction motors which are cheaper will be retained, both types thus being available to meet all conditions. For existing installations requiring a moderate amount of correction to avoid the imposition of penalties and where no new motor-generator sets are to be added, static condensers are available to improve the power factor and thereby reduce the excess charges on the power bill.

The addition of these is more economical than new synchronous motor-generators unless the replaced sets can be used to good advantage elsewhere.

Power Switchboards and Bus Bars

Two recent developments are now in use which will reduce the cost of control equipment and will have other advantages. Long power switchboards for large power plants had to be designed in detail and built individually. The use of unit control panels in power boards and battery fuse panels now permits layout of a simple schematic from which a power board can be assembled, using units which may be stocked as demand warrants. Considerable engineering expense is thus avoided since it will be unnecessary to work up the rather elaborate detailed drawings previously required for each major installation.

A further development of this idea is the "semi-remote control system" recently adopted. With this system most of the control equipment for each motor-generator set is located upon unit panels mounted at the set, thus reducing the main power board to small dimensions and giving increased flexibility which is particularly useful in connection with additions. Overhead bus bars and conduit are employed which are not installed till needed. The flexibility also aids in utilizing improvements and changes in the art occurring between the initial equipment installation and the additions made from time to time as the growth of the load requires.

From a production basis it is anticipated that this unit panel design will be easier to manufacture and to stock and that it will also be simpler to install than the earlier arrangements.

Complete Power Unit Assemblies for Small Applications

Where small amounts of power are required, the provision of storage batteries and associated charging equipment has been relatively high in cost of material and of installation. An appreciable cost reduction has been secured by the design of small power plant units complete with batteries mounted in cabinets and assembled with associated charging or floating equipment.

Crosstalk Reduction

In a common battery telephone office all subscribers are furnished with power from a single central office storage battery. There is a tendency toward "crosstalk," that is, mixing of conversations, so that fragments from one conversation might be overheard in another. This tendency is limited by so designing the battery and wiring common to all circuits that it will have very low impedance, particular attention being given to arrangement of cables and use of large conductors. This imposes certain limitations upon the location of equipment and may involve considerable cost for copper in the larger power plants. However, by means of the electrolytic condensers, previously mentioned, located at battery fuse panels, crosstalk on talking feeders can be reduced to very low values, the limitations on floor plan arrangement can be largely removed and substantial savings in copper can be made.

Battery Reserve

In order to insure continuous telephone service in spite of failure of the primary sources of power, it has been customary, as already mentioned, to provide storage battery capacity sufficient in itself to operate

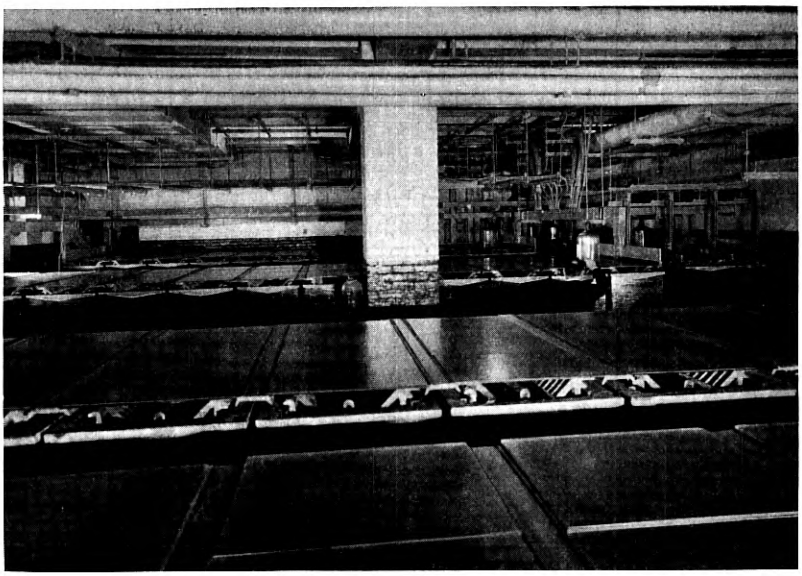


Fig. 7—Battery room for two large panel units.

the central office equipment for a considerable period. The amount of battery reserve provided depends upon the reliability of the regular outside power service and on the reserve source provided. This battery reserve may range from about three net busy hours for offices in large metropolitan districts to several days in small outlying offices. This reserve in the past has been successful in preventing suspension of telephone service due to failures of the power. With the greatly extended plant and the increasing reliability of the public service supply companies, however, the allowance of battery reserve in some cases can be safely decreased, permitting advantage to be taken of appreciable savings in the cost of battery equipment.

Simplified Installation

The developments just discussed will, it is believed, simplify the work of installation, although certain of these developments will not necessarily decrease the amount to be done, as some of the work has

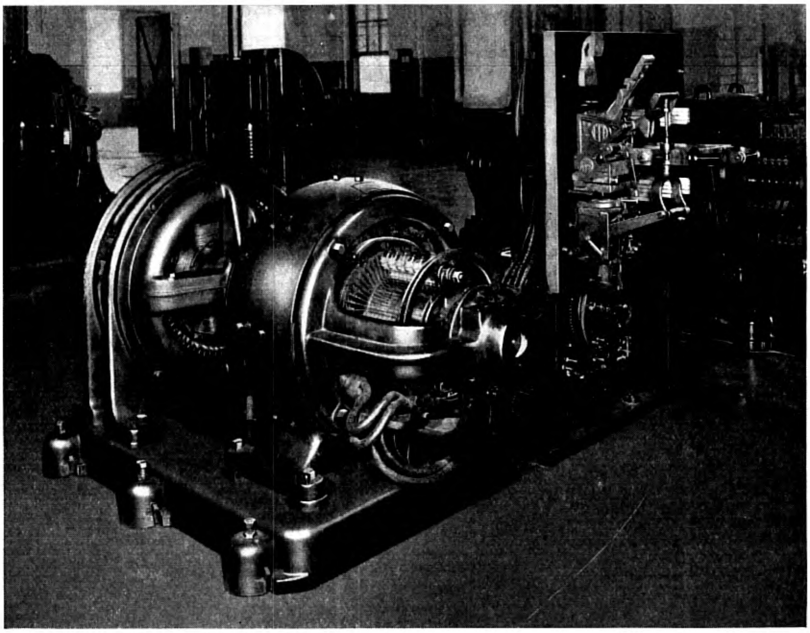


Fig. 8—Commercial type charging generator driven by synchronous motor. 52-kw. output compares with 33 kw. for the larger sized set on the right. 52-kw. "M" type generator in left background belted to gas engine.

been shifted from installer to factory and some in the reverse direction. The unit ringing control panel assemblies, for example, are being furnished wired in the shop with rows of terminal punchings to which the installer connects the wires from the generators. For charging equipment, the panels at the machines will be connected to a common overhead bus bar system, bus bars being shipped in stock lengths and cut by the installer as needed.

For small repeater and similar installations a power plant has been developed which can be placed upon shelves on a rack and connected to the power source and to the distributing bus bars. This compares with the former system of installing a number of separate units and wiring them up upon the job.

Floor plans suitable for the majority of offices are available and reasonably standardized layouts of power equipment to operate certain types of offices are found practicable.

Improved Service

Improved operation of telephone equipment is being made possible by more rigid requirements placed upon the power plant. Automatic voltage regulators for ringing generators have been in use for some time

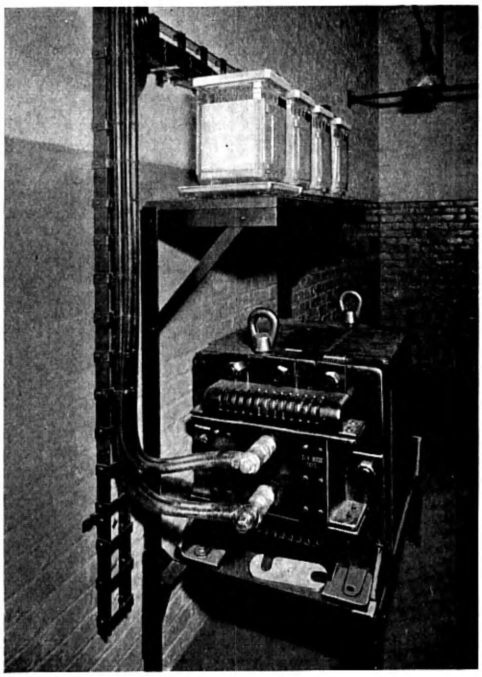


Fig. 9—24-volt discharge lead filter. 800-ampere coil, four 1,200-mf. condensers for use in suppressing generator and other equipment noises.

and the new alternating-current—direct-current system is reducing service troubles. Automatic voltage control equipment for charging generators has been developed and is being introduced. This, together with the full floating system of operating batteries, is capable of holding the main power supply for the central offices at much closer voltage limits than have been found practicable in the past, thus further stabilizing transmission and contributing to even more reliable operation of relays and supervisory equipment.

Reduction of Maintenance

It will be evident that many of the foregoing developments will reduce maintenance of power plant equipment as well as its first cost. The automatic regulating devices substantially reduce or eliminate the attention which would otherwise be required to readjust machines to compensate for load conditions. When this regulation is applied to generators which are floating batteries, it may also result in substantially increasing the life of the batteries, thus deferring replacements. The commercial generators are designed for and equipped with carbon brushes, and will require a minimum of attention.

The introduction of the enclosed type of small battery and the improvements in operating methods of large batteries are decreasing evaporation and spraying, thus reducing additions of water and the repainting of exposed equipment in battery rooms. The new methods also reduce the number of periodic overcharges or eliminate them entirely.

Combining Objectives in Signaling Machine Development

In designing new equipment it is, of course, desirable to accomplish as many objectives as possible. In this connection, mention might be made of a combination machine which may properly lay claim to attaining four important objectives, namely: improved service, reduced cost, simplified installation and reduced maintenance.

Several years ago it was the practice to secure ringing current for subscribers' bells and also for various tones and signals from a small motor-generator set, subject to generator voltage variations amounting to 35 volts as the load on the machine changed and the supply line voltage varied within stated limits. Each large central office unit required these motor-generators, one driven by a line motor and a reserve set driven by a battery motor. Direct current at + 110 and − 110 volts for controlling coin box telephones was furnished by two sets of dry cells or storage cells. In either case a third or spare battery was provided.

To replace the ringing sets and coin control batteries a combined ringing and coin-control motor-generator set has been developed and is being used except in the smallest offices, eliminating the cost and the maintenance of the separate batteries, giving closer voltage and frequency regulation for ringing, and automatically continuing service in spite of outside power failures. A description of the features of this equipment showing what it will do may be of interest as this represents a typical development.

Associated with the generator is a transformer, the primary winding acting as a balance coil for a three-wire direct-current system and the secondary winding having taps to provide one or more of the four alternating-current voltages used with 20-cycle ringing. The generator can thus serve a toll installation at 105 volts, a dial or manual office

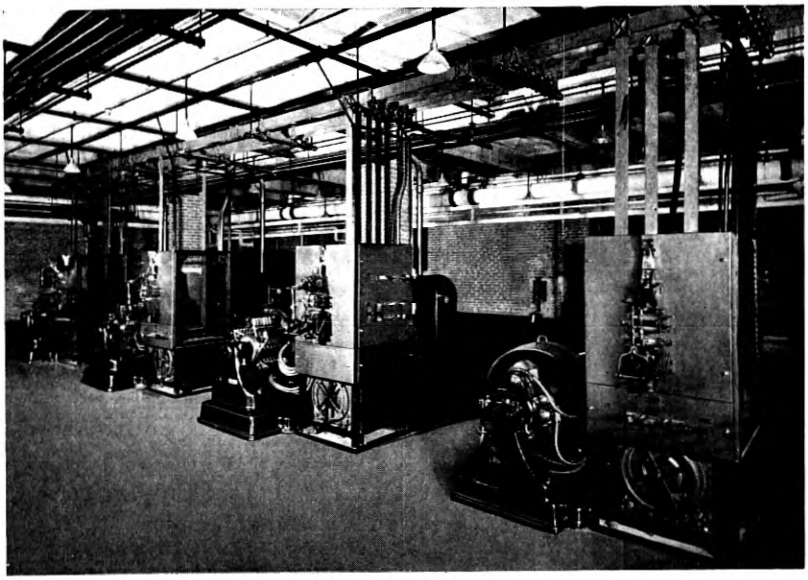


Fig. 10—Charging machines with remote controlled equipment at generator. Overhead bus bar system.

using 100 volts with the alternating-current—direct-current system and two offices, having "superimposed ringing" for party lines, one using 85 volts and the other 77 volts, the voltage used depending upon the type of subscriber sets installed in the district. Positive and negative superimposed currents are obtained by small storage batteries connected in series with the 77- or 85-volt tap of the transformer. All four types of office ringing can be secured from one machine simultaneously, though more than two is unusual. The voltage is controlled automatically within close limits, regardless of load or of normal variations in the voltage and frequency of the supply power.

In addition to ringing, the generator supplies approximately +110 volts direct-current for collecting coins and −110 volts direct-current for refunding coins, the two voltages in combination also exciting the generator field at 220 volts.

Brushes bearing on sectional and solid rings mounted on the generator shaft interrupt battery current and provide a high tone of 480 and a low tone of 160 pulsations per second which are used for various signaling purposes.

Through a 120/1 worm gear reduction an auxiliary shaft is run at 10 r.p.m. Attached to one end of this shaft is a "low-speed inter-

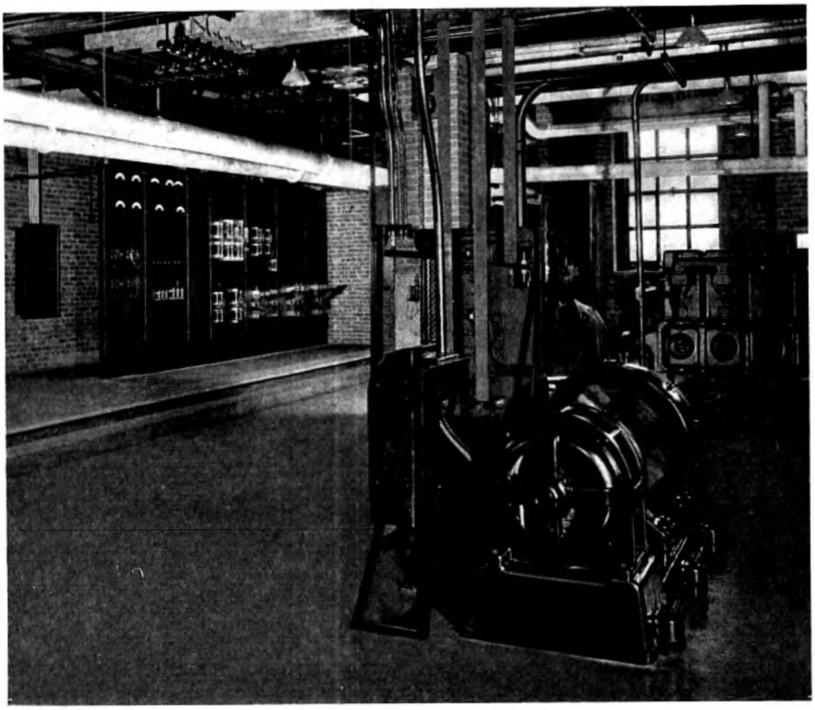


Fig. 11—Master control board for all charging sets and main storage batteries. Charging motor-generator in foreground, emergency engine-alternator set in background. Emergency lighting cabinet to left of control board.

rupter" which provides flashing or tone signals for "busy" and other uses and may provide half a dozen or more different signals. To the other end of this shaft is usually attached a ringing interrupter which divides the constant ringing current from the generator into machine ringing intervals such as 2 seconds ring, 4 seconds silent, or 1 second ring, 1 second silent, 1 second ring, and 3 seconds silent. This interrupter also controls battery current for tripping during the silent interval, and a "pickup" circuit the purpose of which is to prevent ringing the wrong party on party lines.

The generator and all the interrupters are regularly driven by an alternating-current line motor operating upon the outside power supply. In addition to this, however, the set includes a direct-current motor designed to operate on current from the central office battery but normally not connected to the battery. Automatic relays and magnetic switches close the connection to the battery when the regular power fails so the set continues to operate without interruption

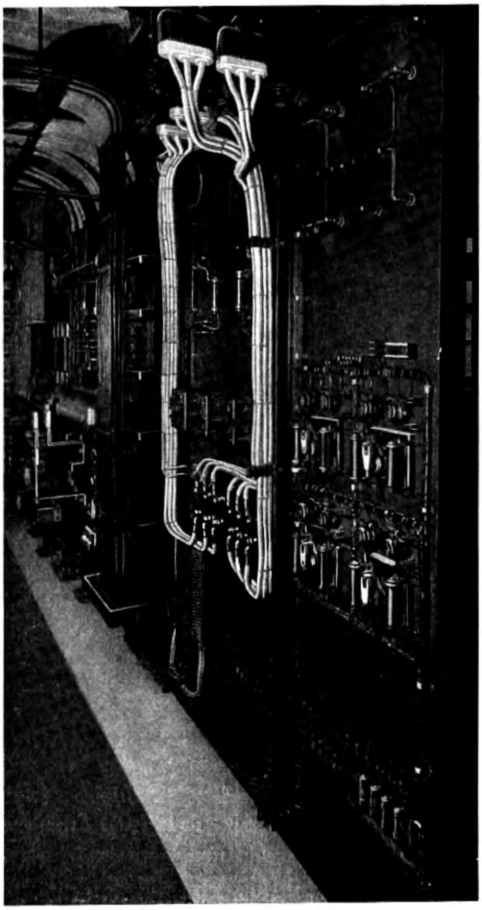


Fig. 12—Rear of control panel for charging sets and batteries.

or change in output. This feature avoids the delay of a few minutes which would otherwise result after a power failure, while the attendant started up a reserve motor-generator and transferred the load circuits, which may run from 17 circuits to twice that number for ringing, coin control, tones and signals. It also avoids accumulation of machine ringing calls which occur when a ringing generator stops during a busy period and which may overload the motor sufficiently to blow the protective fuses and prevent restarting.

The battery motor is equipped with an automatic speed regulator which keeps the generator frequency within two per cent of rated speed throughout the range of the battery motor supply voltage.

It is obvious that a combination which will do so many things at once costs more than a simpler type of machine. The fact, however, that it will operate several central office units and will replace coin control batteries makes it cheaper than the equipment formerly required to do the work. With fewer machines and no batteries, except those for superimposing, installation is simplified. The closer voltage and speed regulation reduce relay and ringing troubles and, in conjunction with the continuity of operation, improve service from the subscribers' viewpoint, as well as reduce the amount of maintenance required of the attendants.

III. The Future Telephone Power Plant

It may be of interest to consider the direction toward which developments in prospect are leading, that we may learn what the future telephone power plant may be like. It seems probable that further progress will be made in the application of unit panels and unit

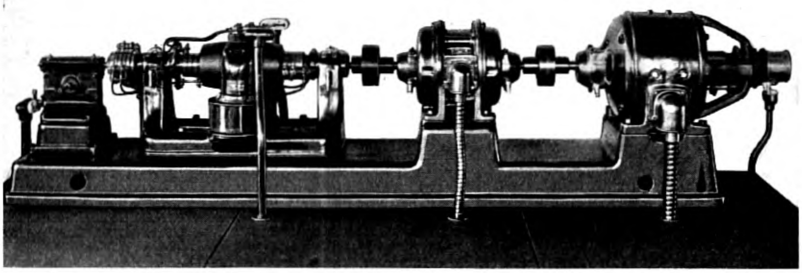


Fig. 13—One type of signaling machine—20-cycle ringing and direct-current 110-volt coin control machine, with reduction gearing for low-speed signals. Driven by a.-c. line motor with reserve battery motor automatically energized upon power failure. Speed controller on battery motor.

assembles or combinations of machines and control equipment. With the better characteristics, much of this equipment, including storage batteries, can be mounted with the circuit apparatus on standard racks, making self-contained units.

Because of the advantages to be obtained in circuit operation from closer voltage regulation and because of the higher cost of manual attendance, automatic regulators for machines and batteries will be used more extensively. Further attention will be given to the automatic operation of power plants, successfully accomplished for private branch exchanges and small offices.

Entirely automatic power plants for large offices could, it is believed, be developed without great difficulty somewhat along the lines of the automatic substations in use by some of the power and traction companies. These should, in general, require attention only periodically for cleaning, replacing worn parts, adjusting, etc., except during a failure of equipment or some other abnormal condition which would be indicated by an alarm. Since full automatic control would probably cost more than that requiring a limited amount of supervision, a study is required to determine how nearly automatic the equipment should be made for offices where an attendant will be required in any case for some of the equipment.

As for machines generally, the tendency will be towards greater use of more nearly commercial designs, construction and finish, eliminating as many as possible of the special features formerly necessary but not now required with changed conditions and the supplementary apparatus which recent developments have made available.

A more extended use of filters in power plant circuits may be expected.

In the direction of power supply, efforts have for some time been applied with some measure of success toward increasing the reliability of the service from outside, which work usually consists of cooperation with the electric supply companies in investigating conditions under which independent duplicate power services can be secured. The securing of reliable duplicate services permits elimination of a local emergency generating plant such as the engine-generator sets. As these efforts become more successful and the public service systems increase in extent and in reliability with the increase in interconnection, it should be found possible to reduce the amount of storage battery reserve in telephone power plants. Experimental introduction of low-voltage alternating-current networks similar to the direct-current networks used in the central parts of some large cities is being watched with interest and some installations are in progress. Although this might be classed as one electrical system, the safeguards against failure and the duplication of equipment is often such as to warrant entire dependence upon this power without a separate emergency source in the building.

With regard to types of batteries, the enclosed type in glass jars will be increasingly introduced wherever suitable because of the lower cost of installation and the subsequent reduced maintenance. The further extension of continuous floating systems of operation also makes it practicable in some cases to use batteries of the pasted plate construction in hard rubber jars, which are cheaper, particularly in first cost. On the general subject of battery operation, the use of the "continuous floating system" is being encouraged where practicable since this usually gives more efficient operation and always results in longer life for the storage batteries and in smaller sizes for equivalent reserve. As an alternate plan a "constant voltage charge system" is in process of adoption for general use where, for any reason, continuous floating is impossible or uneconomical.

The size and cost of power plants is largely controlled by the circuit and apparatus requirements, and improvements in these, such as reductions in current drains for dial equipment and for repeater tubes, are immediately reflected in the telephone power plant which will decrease in size and cost in almost direct proportion.

Quality Control

By W. A. SHEWHART

Introduction

A MANUFACTURER is interested in producing a controlled product—one in which the deviations about the average level of quality are no larger than can be accounted for as a result of chance. The present paper gives simple detailed methods for determining from inspection data whether or not a product is being controlled in the sense of indicating the presence of assignable causes of variation. Naturally the inspection data constitutes a sample of the effects of the manufacturing causes and hence the interpretation of these data in terms of what may be expected in the future is a statistical problem.

A controlled product is defined as one for which the frequency of deviations from the expected quality can be estimated by probability theory. To make such estimates, however, it is necessary to characterize or specify the distribution of quality which the manufacturer wishes to maintain. These specifications of the desired quality must be arrived at by methods customarily used in setting engineering standards, but when once they have been established the statistical methods amplified in this paper make possible the most economical control of this quality.

The limits within which quality may be controlled with a given amount of inspection depend upon the standards adopted for the quality to be maintained.

This paper interprets quality specifications in terms of five different types of constant systems of manufacturing causes. The five types chosen are sufficient to cover the entire range and it is believed that only five types are necessary because sampling theory indicates that little practical advantage would be derived by endeavoring to subdivide one or more of these. It is shown that quality control can be maintained with the fewest number of measurements and within the closest limits through the adoption of Type V.

Specification of Control

One of the principal objects of inspection is the detection of lack of control of manufactured product, that is, the detection of the presence of assignable causes of variation in the quality. A recent paper in this *Journal*[1] describes a quality control chart designed to attain this ob-

[1] Shewhart, W. A., "Quality Control Charts," October 1926.

ject and some of the results obtained through the application of the chart have also been presented.[2] In general the detection of the existence of assignable causes of variation leads to their elimination at a minimum of cost.

As a basis for this chart we start with the conception of a constant system of causes as being one such that the probability of a unit of product having the quality X within the range X to $X + dX$ is independent of time. For convenience in the present discussion we may represent this probability dP as a function f of the quality X and m parameters. Thus

$$dP = f(X, \lambda_1, \lambda_2, \cdots \lambda_m)dX. \tag{1}$$

The present paper presents different ways of specifying the constant system of causes and of detecting lack of control upon the basis of the different specifications principally by setting sampling limits on the parameters. In this way it is shown that the best control can be secured when all of the parameters together with the function f in Eq. 1 are specified. We shall assume, in what follows, one set of specifications after another for the constant system of causes and then show for each set how sampling limits may be established. Nomograms are presented to make the determination of the limits possible without the use of even a slide rule. We shall start with the simplest specification, usually referred to as Type I, which has found extensive use.

Type I often gives a satisfactory basis of control although it makes use, as we shall see, of only a fraction of the information given by the data used in connection with Specification Type V, which is the ideal set wherever the manufacturer is warranted economically in trying to secure the highest degree of control. The choice of specification to be adopted in a given case depends entirely upon the economic advantage attainable through the detection and elimination of assignable causes of variation. In particular the use of Type V specification in the initial stages of the development of the manufacturing process is almost always warranted, because it materially assists in arriving at a controlled process with a minimum of labor.

Specification Type I: The probability of the production of a defective piece of apparatus shall be p'.[3]

To set limits in this case is very simple indeed, particularly if we choose the probability P associated with the limits to exceed .9. It

[2] Jones, R. L., "Quality of Telephone Materials," *Bell Telephone Quarterly*, Vol. 6, pp. 32–46, January 1927.

[3] The primed notation is used throughout to denote parameters of the universe as contrasted with the estimates of these determined from the sample.

has been found satisfactory in many cases to take $P \doteq .99$ and so, upon this basis, we shall present the method of setting limits upon the expected fraction defective in a sample of size n. It is well known that the probability of an observed value of p lying within the limits $p' \pm 3\sigma_{p'}$ is approximately equal to .997 provided the fraction defective p' is approximately equal to the fraction non-defective q', and n is large. It can be shown, however, that irrespective of the magnitudes of p' and n the value of P so determined lies between .95 and 1.00 and for most cases met in practice P does not differ from .997 by as much as 1 per cent.

It is obvious, therefore, that, if we construct an alignment chart on which we may read directly the standard deviation $\sigma_{p'}$ when p' and n are given, then the average p' plus or minus three times the standard deviation $\sigma_{p'}$ gives the corresponding values of the limits.

Let us consider a practical problem, see how the question of whether or not a product is controlled really arises and see how control limits can be found from the alignment chart of Fig. 1 to answer this question.

Table 1 represents the observed fraction found defective over a period of 12 months for two kinds of product designated here as Type A and Type B. The table gives for each month the sample size n, the number defective m and the fraction defective $p = m/n$. The average fractions defective for the 12-month period are $\bar{p}_A = .0109$ and $\bar{p}_B = .0095$. Subject to later consideration we shall assume $p'_A = \bar{p}_A$ and $p'_B = \bar{p}_B$.

TABLE 1

Apparatus Type A				Apparatus Type B			
Month	n No. Insp.	m No. Def.	$p = \frac{m}{n}$ Fraction Def.	Month	n No. Insp.	m No. Def.	$p = \frac{m}{n}$ Fraction Def.
Jan.	527	4	.0076	Jan.	169	1	.0059
Feb.	610	5	.0082	Feb.	99	3	.0303
Mar.	428	5	.0117	Mar.	208	1	.0048
Apr.	400	2	.0050	Apr.	196	1	.0051
May	498	15	.0301	May	132	1	.0076
June	500	3	.0060	June	89	1	.0112
July	395	3	.0076	July	167	1	.0060
Aug.	393	2	.0051	Aug.	200	1	.0050
Sept.	625	3	.0048	Sept.	171	2	.0117
Oct.	465	13	.0280	Oct.	122	1	.0082
Nov.	446	5	.0112	Nov.	107	3	.0280
Dec.	510	3	.0059	Dec.	132	1	.0076
Average	483.08	5.25	.0109		149.33	1.42	.0095

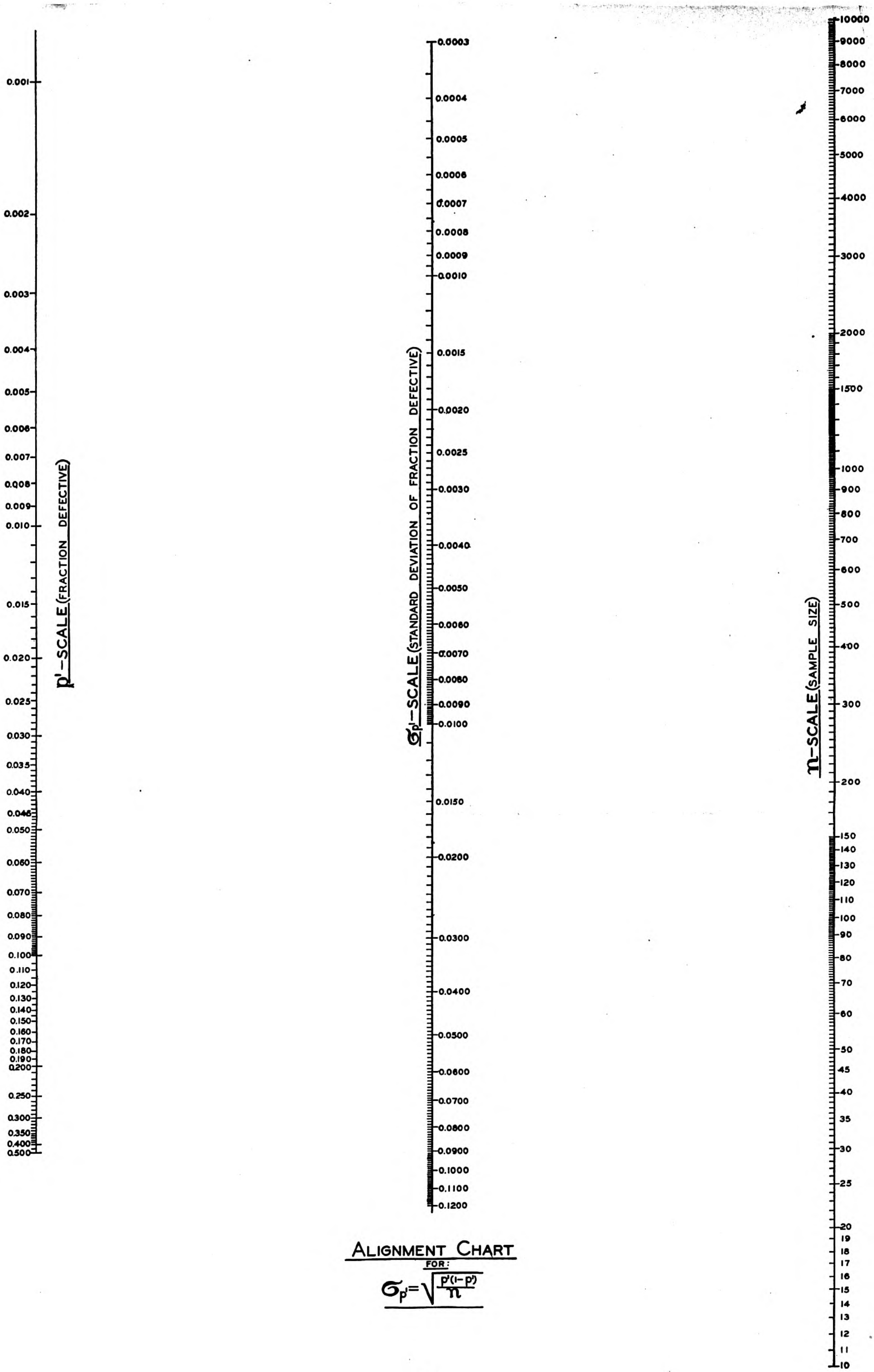

p'-SCALE (FRACTION DEFECTIVE)
0.001 0.002 0.003 0.004 0.005 0.006 0.007 0.008 0.009 0.010 0.015 0.020 0.025 0.030 0.035 0.040 0.045 0.050 0.060 0.070 0.080 0.090 0.100 0.110 0.120 0.130 0.140 0.150 0.160 0.170 0.180 0.190 0.200 0.250 0.300 0.350 0.400 0.500
σp'-SCALE (STANDARD DEVIATION OF FRACTION DEFECTIVE)
0.0003 0.0004 0.0005 0.0006 0.0007 0.0008 0.0009 0.0010 0.0015 0.0020 0.0025 0.0030 0.0040 0.0050 0.0060 0.0070 0.0080 0.0090 0.0100 0.0150 0.0200 0.0300 0.0400 0.0500 0.0600 0.0700 0.0800 0.0900 0.1000 0.1100 0.1200
n-SCALE (SAMPLE SIZE)
10000 9000 8000 7000 6000 5000 4000 3000 2000 1500 1000 900 800 700 600 500 400 300 200 150 140 130 120 110 100 90 80 70 60 50 45 40 35 30 25 20 19 18 17 16 15 14 13 12 11 10
ALIGNMENT CHART
FOR:
$\sigma_{p'} = \sqrt{\frac{p'(1-p')}{n}}$

Fig. 1

Is there any indication that the observed fluctuations in the fraction defective p could have been produced by other than chance causes? In other words, were apparatus Type A and apparatus Type B con-

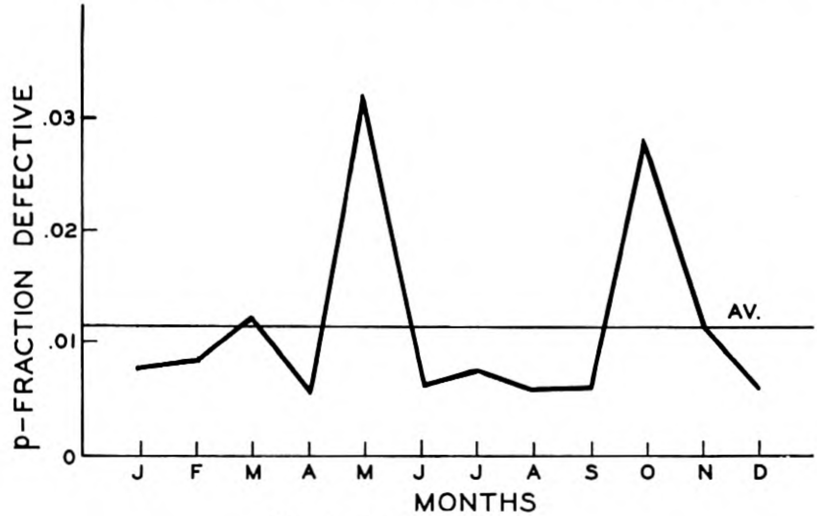


Fig. 2*a*. Apparatus Type A

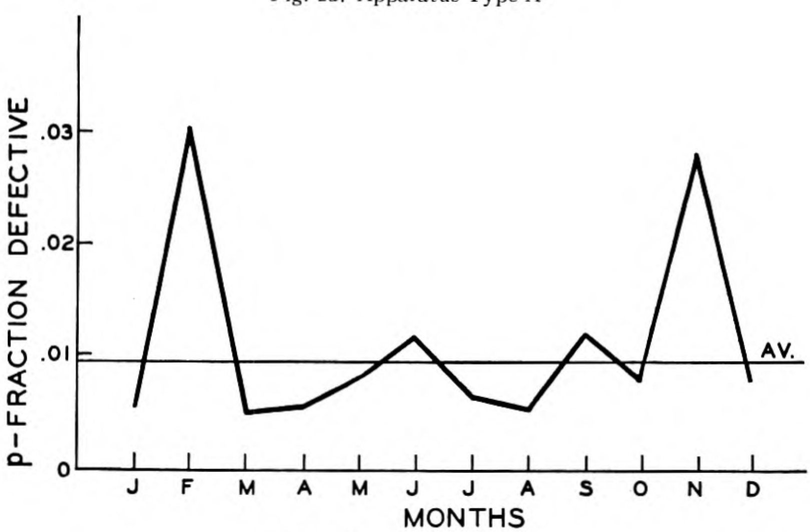


Fig 2*b*. Apparatus Type B.

trolled over the given period? Furthermore, is there any indication that the product could have been improved during this period without changing the process of its manufacture?

To better visualize the fluctuations in p, the data of Table 1 are shown graphically in Fig. 2*a* and Fig. 2*b*. It may appear that during the months of May and October there existed some assignable cause of variation in the production process of Type *A* apparatus. The same may appear to be true for Type *B* apparatus during the months of February and November.

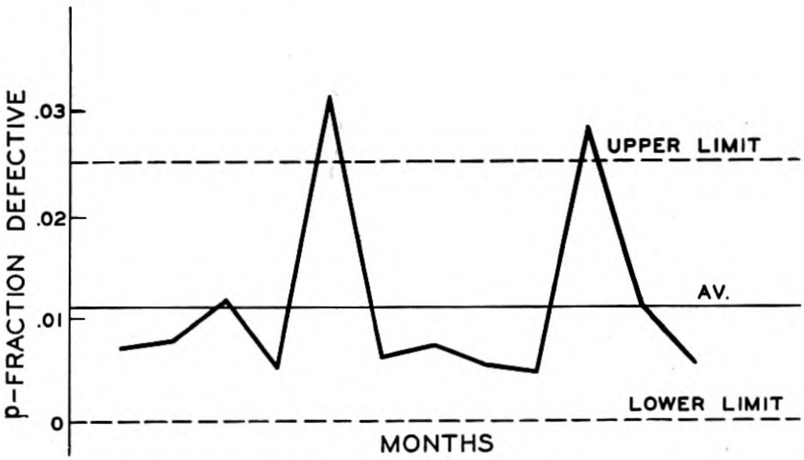


Fig. 3*a*. Apparatus Type A

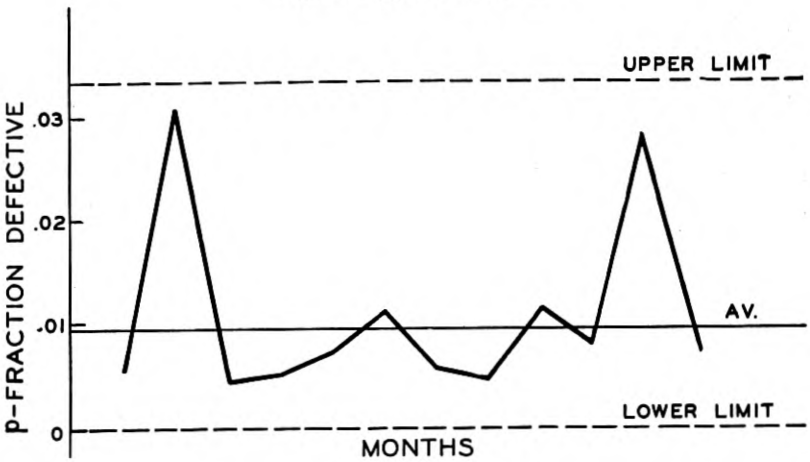


Fig. 3*b*. Apparatus Type B

We shall see upon investigation that there is evidence of lack of control of apparatus Type *A* but not any evidence of lack of control of apparatus Type *B*.

Taking n equal to the average sample size (483 for Type A), we connect by a straight line the point 483 on the n scale of Fig. 1 with the point .0109 on the p' scale. We read the intersection of this straight line with the $\sigma_{p'}$ scale as .0047. Hence the upper limit for p' is $.0109 + 3\sigma_{p'} = .0250$, a value which is exceeded during the months of May and October; and the lower limit is $.0109 - 3\sigma_{p'} = -.0032$. Of course negative values of p have no significance; hence we take the lower limit as zero. Following the same procedure for Type B apparatus, we get limits 0 and .0332.

We see that twice during the year Type A apparatus appears to have been out of control whereas at no time during the year can we say this of Type B.[4]

Now, we shall take up successively the method of finding limits corresponding to specifications involving:

a. Only one parameter (Type II).
b. Only two parameters (Type III).
c. Two parameters and a restriction on the function f over a certain range (Type IV).
d. Four parameters and a specific function f (Type V).

We shall find that the limits become progressively smaller in the above order. In fact for Specification Type II no limits can be set and for Specifications Type III and IV the limits are so large as to be in most instances impractical.

Specification Type II: The expected or average quality shall be $\overline{X}'$.

There is an indefinitely large number of constant systems of causes which would meet this requirement. Associated with each constant system of causes there are specific sampling limits. Sufficient information, however, is not called for in the Specification Type II to fix sampling limits on the quality of a single unit or on the expected or average quality.

In other words, Specification Type II is useless insofar as it does not provide for the detection of lack of control in the sense now under discussion.

Specification Type III: The expected or average quality shall be $\overline{X}'$ *and the standard deviation shall be* σ'.

Again there is an indefinitely larger number of different cause systems which would satisfy this requirement. However, it is re-

[4] Strictly speaking statistical theory only shows that two of the observed deviations in p_A are highly improbable upon the assumption that the product had been controlled about p'_A. It should be noted, of course, that the sample size is not the same from month to month and hence that the limits for a given month should really have been based upon the sample size for that month. However, in the present instance, this method of procedure leads to the same conclusion as given above and hence was not introduced because of necessary complications.

markable, even though this be true, that the work of Tchebycheff[5] makes it possible for us to give a lower bound to the probability that a unit of product will be produced with a quality X lying within the range $\bar{X}' \pm L_1$ and also therefore to the probability that an observed average quality of a sample of n units will lie within any given range $\bar{X}' \pm L_n$.

Taking $L_1 = c\sigma'$ $(c > 1)$, the probability $P_{c\sigma'}$ that the constant system of causes Type III will produce a unit of product having a quality X within the range $\bar{X}' \pm L_1$ is given by the expression

$$P_{c\sigma'} \geq 1 - \frac{1}{c^2}. \tag{2}$$

Expression 2 also defines the probability that the average quality of n units of product coming from the constant system of causes Type III will lie within the range $\bar{X}' \pm L_n$ where

$$L_n = \frac{c\sigma'}{\sqrt{n}}.$$

Let us illustrate the method of finding the limits under these specifications. Assume that the specified average resistance $\bar{X}'$ of a relay is 150 ohms and the standard deviation σ' is 5 ohms. What is the range within which we may expect 90 per cent of the product (i.e. $P_{c\sigma'} = .90$) to lie, assuming no assignable causes of variation in product? What is the similar range for the average of 1000 relay windings?

Turning to the nomogram of Fig. 4, we connect by a straight line the point $P_{c\sigma'} = .90$ and the point A near the center of the chart. The point on the c scale fixed by the intersection of the straight line so determined with the c scale is 3.15. The required values of L_1 and L_{1000} are therefore $L_1 = 3.15 \times 5 = 15.75$ ohms and $L_{1000} = \dfrac{3.15 \times 5}{\sqrt{1000}} = .50$ ohm. Hence the limits are 150 ± 15.75 ohms and $150 \pm .50$ ohms.

To avoid the slide rule computations in obtaining $c\sigma'$ and $c\sigma'/\sqrt{n}$ we can use the nomogram of Fig. 5. We enter this nomogram by the value $c = 3.15$ and find a point on the $c/\sqrt{n}$ scale which lies on a straight line with the point $c = 3.15$ on the c scale and $n = 1$ on the n scale. Connecting the point thus fixed on the $c/\sqrt{n}$ scale with the point $\sigma' = 5$, we read on the L scale 15.75 ohms. Carrying through the same procedure, but starting with $n = 1000$ instead of $n = 1$, we read on the L scale .50. These values give the limits found above.

[5] Tchebycheff, *Liouville Journal*, 1867. "Des valeurs moyennes," *Journal de Mathematiques* (2), Vol. 12, pp. 177–84.

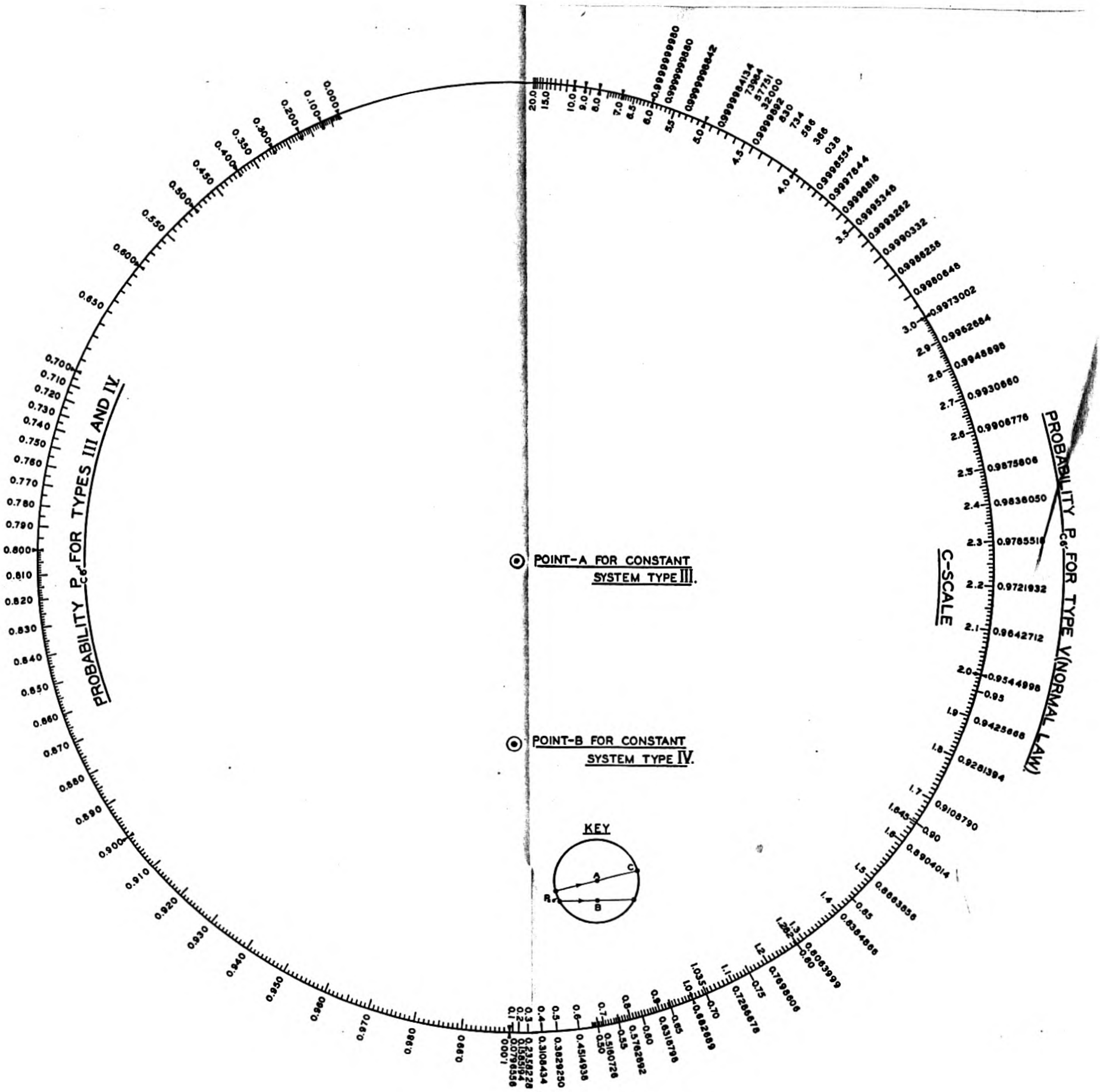

PROBABILITY P_C6 FOR TYPES III AND IV
PROBABILITY P_C6 FOR TYPE V(NORMAL LAW)
C-SCALE
POINT-A FOR CONSTANT SYSTEM TYPE III.
POINT-B FOR CONSTANT SYSTEM TYPE IV.
KEY
A
B
C
R

Fig. 4

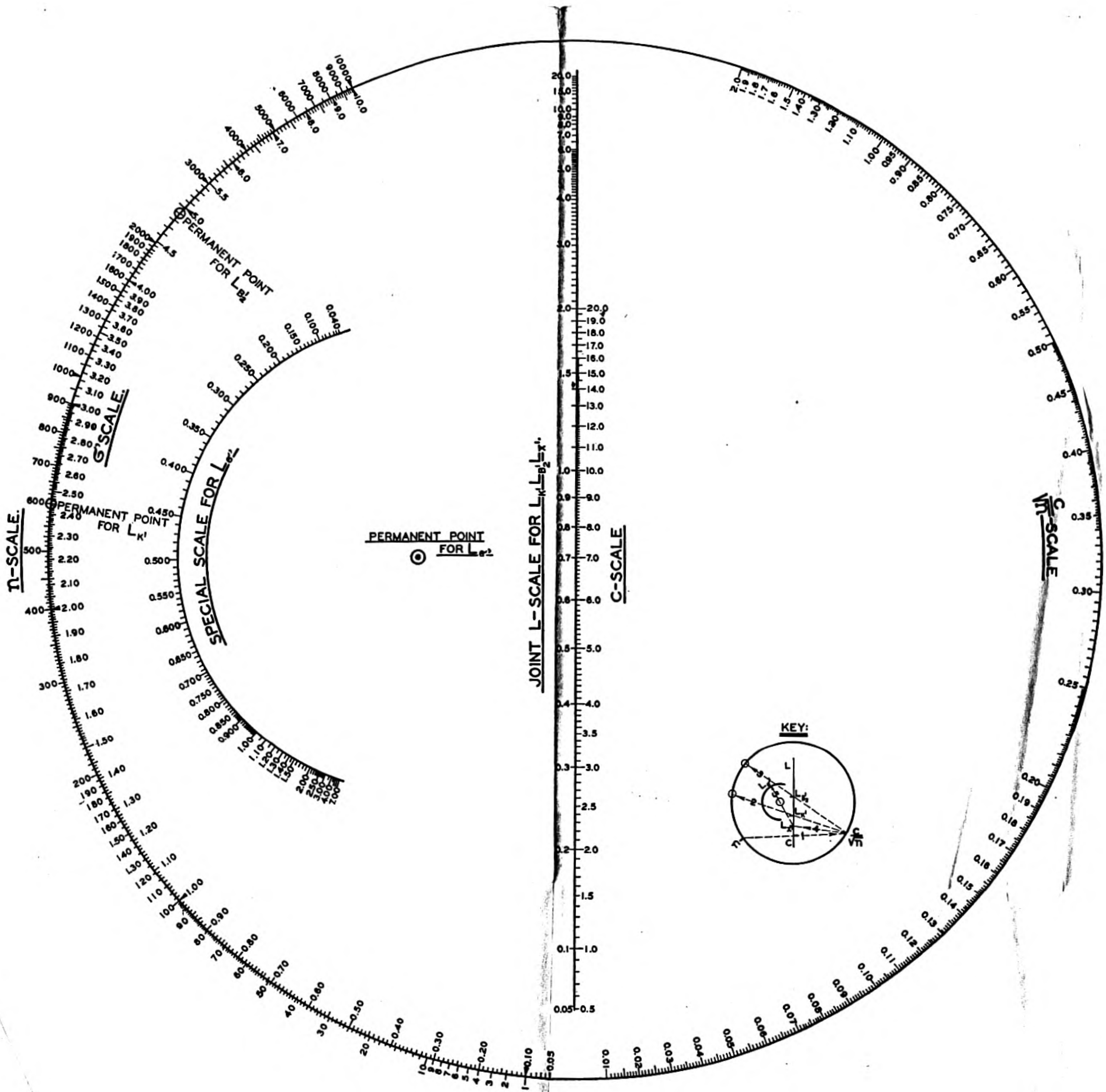

n-SCALE.
G'SCALE.
PERMANENT POINT FOR $L_{B_2'}$
PERMANENT POINT FOR $L_{K'}$
SPECIAL SCALE FOR $L_{a'^2}$
PERMANENT POINT FOR $L_{a'^2}$
JOINT L-SCALE FOR $L_{K'}L_{B_2'}L_{x'}$
C-SCALE
$\frac{C}{\sqrt{n}}$-SCALE
KEY:

Fig. 5

Specification Type IV: The expected or average values of quality and standard deviation shall be $\overline{X}'$ and σ' respectively. The expected modal and average qualities shall coincide and the probability function for the constant system of causes shall be monotonically decreasing for all values of x where x is measured from the mean.

In this case the lower bound to the probability $P_{c\sigma'}$ is given by the expression [6]

$$P_{c\sigma'} \geq 1 - \frac{1}{2.25c^2}. \tag{3}$$

The limits can be obtained just as in case of Type III except that we use point B in Fig. 4 instead of point A. It may be easily verified by this nomogram that the Type IV values of L_1 and L_n for the special problem considered for Type III are $L_1 = 10.4$ ohms and $L_n = 0.33$ ohm respectively.

This shows that the additional requirements placed upon Type IV over those of Type III make for better control in the sense that the associated sampling limits are thereby decreased. By going further in adding restrictions upon the cause system, we gain even more marked improvements in the condition for control. In fact it is the system now to be described that has been found to be the most useful practical standard where the quality is measured as a variable.

Specification Type V: The system of causes shall yield a product distributed according to the Gram Charlier series [7] with arithmetic mean $\overline{X}'$, standard deviation σ', skewness k' and kurtosis β_2'.

With the use of the four parameters we can detect lack of control of product through the failure of the observed value of any parameter determined from a sample of size n to fall within its sampling limits. It may happen that lack of control will be indicated by deviation beyond the sampling limits for only one of the four parameters. This case has already been illustrated in the article referred to in footnote 1. We shall now present, however, a method of setting these limits which is very easily applied.

As a specific example, let us assume the following expected values:

$$\begin{aligned} \overline{X}' &= 0, \\ \sigma' &= 1, \\ k' &= 0, \end{aligned}$$

[6] Camp, Burton H., "A New Generalization of Tchebycheff's Statistical Inequality," *Bulletin of the Amer. Math. Soc.*, December 1922, pp. 427–432. Eq. 3 is a special case of the general theorem of Camp. This theorem may be extended to determine lower bound to the probability of an error of the average as is done in this paper.

[7] Of course we might use certain other functions involving the same parameters.

and

$$\beta_2' = 3.$$

Also let us assume that the size of the sample n for which the limits are to be established is 1000 and that we wish to establish limits upon the basis of a probability $P \doteq .997$.

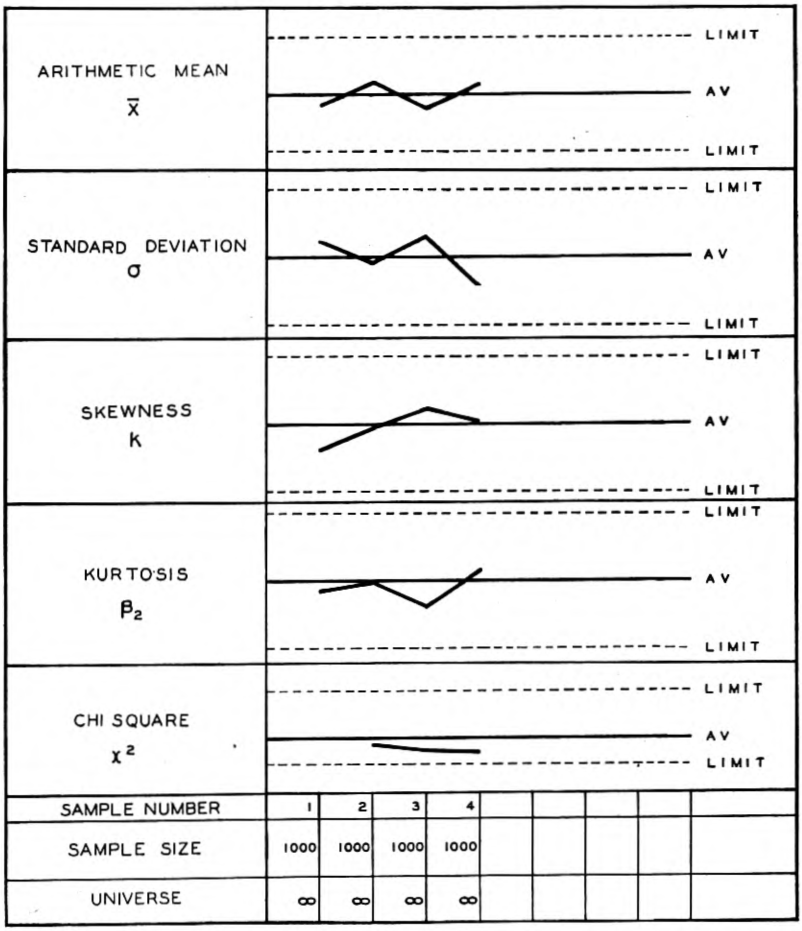


Fig. 6

The nomogram of Fig. 4 gives us immediately that $c = 3.0$ for $P = .997$. Hence we enter the nomogram of Fig. 5 on the c scale $c = 3.0$. The best way in which all four limits can be found by using the nomogram of Fig. 5 is then as follows, where the limits are set in the order $L_{k'}$, $L_{\beta_2'}$, $L_{\bar{X}'}$, and $L_{\sigma'}$. Join the point $n = 1000$ on the n scale and $c = 3.0$ on the c scale by a straight line and thus find a pivot point on the $c/\sqrt{n}$ scale. Holding the ruler on this pivot point, join it successively with the permanent points of $L_{k'}$ and $L_{\beta_2'}$ and with σ' taken all on the σ' scale and read accordingly $L_{k'}$, $L_{\beta_2'}$, and $L_{\bar{X}'}$, on the L scale. After reading $L_{\bar{X}'}$, release the pivot point and turn the ruler around the $L_{\bar{X}'}$ point so as to join it with the permanent point for $L_{\sigma'}$. Then read the intersection of the ruler with the inner circular scale $L_{\sigma'}$, hereby obtaining the limit for σ'. Thus in five movements of the ruler we find all four limits: [8]

$$\begin{aligned} 0 \pm L_{k'} &= 0 \pm .23, \\ 3 \pm L_{\beta_2'} &= 3 \pm .46, \\ 0 \pm L_{\bar{X}'} &= 0 \pm .095, \\ 1 \pm L_{\sigma'} &= 1 \pm .067. \end{aligned}$$

Figure 6 presents the graphical representation of the limits thus determined together with limits on χ^2 assuming that the theoretical frequency distribution was broken up into 13 cells.[9] The irregular lines show the fluctuations in the estimates of these parameters determined from four samples of 1000 each drawn under conditions satisfying the specifications just described for $\bar{X}' = 0$, $\sigma' = 1$, $k' = 0$ and $\beta_2' = 3$. Incidentally it should be noted that in every case the observed fluctuations in the estimates of the parameters are well within the sampling limits. This was to be expected because every effort was made in the sampling process to come as close as practicable to the ideal case where no assignable causes of variation were present. In this respect the data of Fig. 6 form an interesting contrast to the data of Fig. 4 of the article referred to in footnote 1, where evidence of lack of control was found.

Figure 7 makes it possible for us to set limits about the average or expected χ^2 corresponding to a probability of either .98 or .80. Thus for the data of Fig. 6 the limits for χ^2 corresponding to probability .98 are approximately 3 and 26 respectively as read from this chart. If limits corresponding to any other probability are desired, they can be readily obtained from tables for goodness of fit.[10]

We are now in a position to consider more in detail the advantages

[8] In case the given data bring the readings on the extreme points of the scale (where $\sigma' > 10$) it is advisable to take $\sigma'/10$ and multiply the final results obtained by ten. It is also helpful to remember that the L-scale on the nomogram of Fig. 5 can be considered as a regular scale of the product of two factors read on σ' scale and $c/\sqrt{n}$ scale.

[9] For the significance of χ^2 as here used, see paper, footnote 1.

[10] Elderton's Tables for Goodness of Fit reproduced in Pearson's "Tables for Statisticians and Biometricians" and also R. A. Fisher's "Tables for Goodness of Fit" given in his recent book "Statistics for Research Workers" will be found very helpful in the construction of curves similar to those of Fig. 7.

from a control viewpoint of Type V specification over the other suggested specifications. We have seen in Fig. 4 of the previous article on the control chart, footnote 1, that evidence of lack of control may be obtained through deviations in one parameter and not in

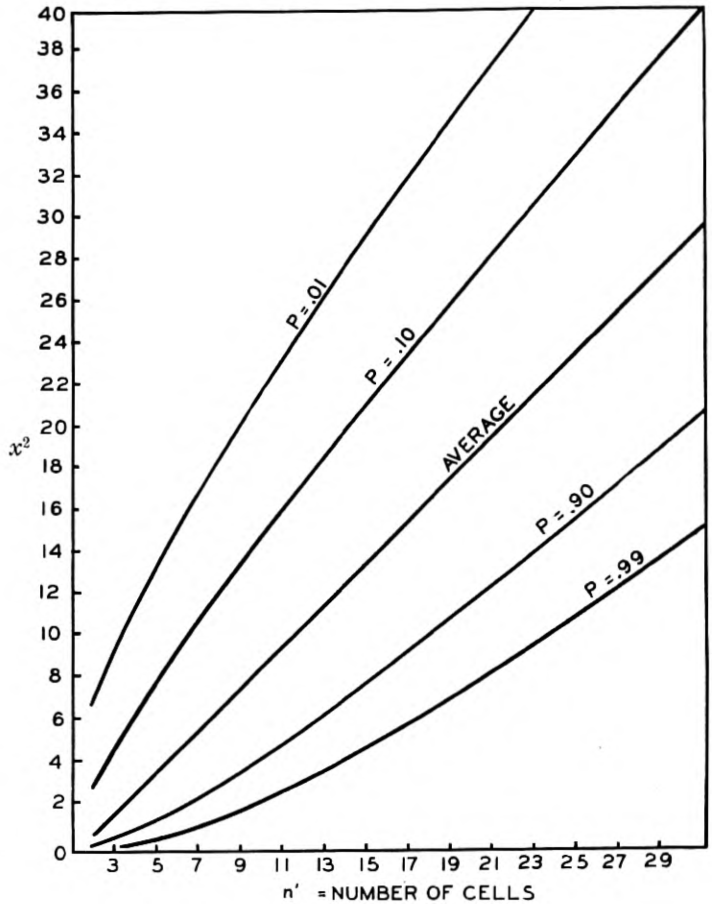


Fig. 7

another. For example, in this figure the per cent defective part of the chart corresponded to Specification Type I. Only 4 of the 12 points on this chart were outside the control limits whereas more than 4 points were outside the control limits for every other parameter and for the χ^2 part of the chart every one of the points was outside the control limits. Of course it is to be expected that the χ^2 test would be much more stringent than the test applied under Type I specification because the control limits established under the Type I specification are merely the limiting case of the limits set on χ^2 for the case of two cells. We see, however, when samples are actually drawn from a constant system of causes, as was done as nearly as possible in obtaining the data for Fig. 6 of the present paper, all of the estimates of the parameters remain well within the sampling limits at least the expected proportion of the time.

To show that the limits set by means of Specification Type V upon the expected or average value of the data in Fig. 4 of the article on control charts just referred to are much smaller than could have been set by means of either Specification Type III or Specification Type IV, Fig. 8 is given. The limits based upon Specifications Type III and IV

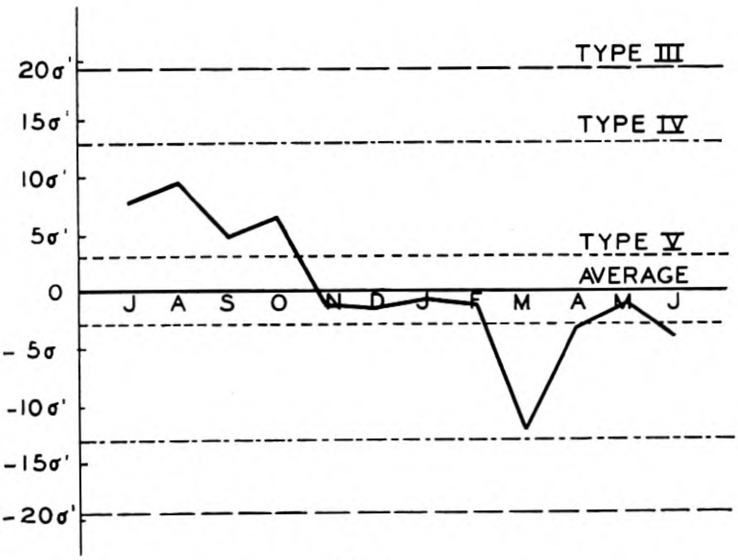


Fig. 8

were obtained directly from the nomogram of Fig. 4. The magnitudes of L_n stand in the order 19.3, 12.8 and 3.0. We see at a glance that lack of control, not indicated at all upon the basis of either Types III or IV, appears probable upon the basis of Type V.

Of course the use of the nomogram of Fig. 5 involves certain assumptions which now should be considered. The sampling limits are based upon the assumption that the sample is drawn from a normal universe. Even under these conditions the distributions of the values of estimates of the four parameters considered above are skew with the exception of that of the average, but approach normality as the size of the sample is increased. Theoretical and practical considerations lead us to believe, however, that satisfactory limits can be established by the method just described making use of the nomogram of Fig. 5 provided the following restrictions as to the size of the sample are made.

(*a*) The expected distribution of the averages of samples of any size n is normal about the expected value $\overline{X}'$.

(*b*) Comparatively small error [11] will be made in fixing the limits on the parameter σ' by means of the nomogram of Fig. 5 provided n is 25 or more.

(*c*) For a sample of size n of 500 or more the nomogram of Fig. 5 may be used in fixing limits on all four parameters.[12]

These limitations do not require necessarily that the distribution of the estimate of a parameter must be normal for n as large or larger than specified; instead they merely require that it may be represented by the first few terms of the Gram Charlier series for which the normal law integral over a range equally divided by the expected value of the parameter is a close approximation to the integral of the Gram Charlier series over the same range.

Fixing the Parameters

There are various ways of arriving at the values of the parameters to be accepted as the basis for quality control. Sometimes they may be fixed by the economics of the problem. Such is the case for the Type I specification when the economic standard fraction defective or p' is known. At other times the parameters are fixed by technical considerations such for example as in the case of an induction coil whose inductance must lie within well-defined limits in order to obtain a proper functioning of the entire circuit, for this would effectively fix $\overline{X}'$ and σ'. In most practical instances the technical considerations tend to fix only the average and standard deviation. At other times we may empirically choose the observed estimates of these parameters determined from the data obtained within the fixed interval of time wherein we have reason to believe the quality has been produced under essentially the same conditions. Irrespective, however, of what period is chosen as a base in fixing p' or any other parameter, the control chart serves to show whether or not the product has been controlled over this period. In any case the parameters are accepted at least as

[11] Pearson, Karl, "On the Distribution of Standard Deviations of Small Samples," *Biometrika*, Vol. X, Part IV, May 1915, pp. 522–529.

[12] Pearson, Karl, and others, "On the Probable Errors of Frequency Constants," *Biometrika*, Vol. XIV, 1903, p. 273 seq., Vol. IX, 1913, p. 22 seq. Isserlis, L., "On the Conditions under Which the Probable Errors of Frequency Distributions Have a Real Significance," *Proceedings of the Royal Society*, Series A, Vol. XCII, 1915, pp. 23–41.

temporary standards. In every case the choice of the fixed values calls for the exercise of engineering judgment. The statistical problem enters after these standards have been fixed. It is to determine whether or not the observed fluctuations in the observed estimates of the parameters are explainable upon the basis of chance. In general, the method of fixing the limits closely corresponds to that whereby a manufacturer sets up specifications for any kind of product.

It should be noted that from a statistical standpoint the control charts are based upon a priori reasoning. The type of cause system specified by the engineer is taken as a standard a priori system which is accepted as an ideal which the manufacturer hopes to maintain. The control chart thus makes it possible to differentiate between deviations in quality which can reasonably be accounted for on the basis of sampling and those deviations which cannot be so accounted for.

It will have been noted that the limits are a function of the size of the sample n. The question is therefore often raised: How large a sample shall be chosen?

So long as we are willing to risk our engineering judgment that the system of causes is controlled, we need take no samples. If, however, we have reason to believe that the quality has not been controlled or at least wish to make sure that it is being controlled to the extent that the deviation introduced by the assignable cause shall not escape detection if greater than some chosen value, it is necessary for us to take a sample of sufficient size to reduce the limits of sampling fluctuations in the particular parameter under study to just less than this same value.

In those cases where customary practice calls for the inspection of a certain number of units of product for reasons other than control, these data may be used in the manner outlined above to indicate the degree of control. In many instances the number of units of product to be inspected is so fixed as to insure with a known degree of probability that the apparatus passing from one stage of the manufacturing process to another meets a given tolerance for defects. This practice serves to fix the number to be inspected in order to maintain a given quality of apparatus as it passes through the stages of the manufacturing process. The use of the data so obtained in the form of a control chart serves to fix attention upon the assignable causes of variation in the quality. The presence of these causes having been detected, it generally becomes a comparatively simple matter to find and eliminate them. In this way we can secure a controlled product usually requiring less inspection and hence involving the lowest cost of manufacture.

I am indebted to Mr. V. A. Nekrassoff for the construction of the nomograms presented in this paper.

The New York-London Telephone Circuit

By S. B. WRIGHT and H. C. SILENT

SYNOPSIS: This paper discusses the special provisions which are in use on the transatlantic telephone to compensate for the variability of the wire and ether paths, for the radio noise, and for the fact that two-way transmission is effected upon a single wave-length. So-called technical operators are in attendance at each end of the radio path and are equipped to adjust the magnitude of the speech currents entering the radio transmitters to such a value as to load these transmitters to capacity. The amplification introduced at the radio receivers can also be adjusted to compensate for changes in the transmission efficiency of the radio paths. Finally, voice-operated relays together with suitable delay circuits are provided which so control the apparatus that at any given time it can transmit in but one direction. By this arrangement, a speaker's voice upon leaving his transmitting station cannot operate his own receiver although this is tuned to the transmitting wave-length.

TO the telephone subscribers who use the New York-London circuit the procedure of making a call and carrying on a conversation is as simple as that of any long distance telephone call. Even to the telephone operator who establishes a transatlantic connection there is little to differentiate the New York-London "wire-radio-wire" circuit from the hundreds of other circuits which appear as mere jacks on the switchboard in front of her. Beyond this point, however, there is an organization of physical plant, personnel and procedure very much different from the usual long distance telephone circuit.

Without going into any description of the radio portion of the New York-London circuit, which has been adequately treated in previous articles, this paper describes some of the interesting features of the circuit, including the method of electrical operation which has been worked out for making possible two-way talking in the usual way, in spite of difficulties introduced by "static," transmission variations and difficulties brought about by the use of the same "frequency band" for transmission in both directions. The method of operation involves manual adjustments of controls at the radio stations and at the circuit terminals, and automatic switching by means of vacuum tube-operated relays controlled by the voice currents of the telephone subscribers. The voice-operated relay system is particularly interesting, and is, therefore, rather fully described.

Before the operation of the circuit is described a brief general picture of the system will be given. Fig. 1 shows its geographical layout, and gives an idea of the relative lengths of wire and radio circuit involved.

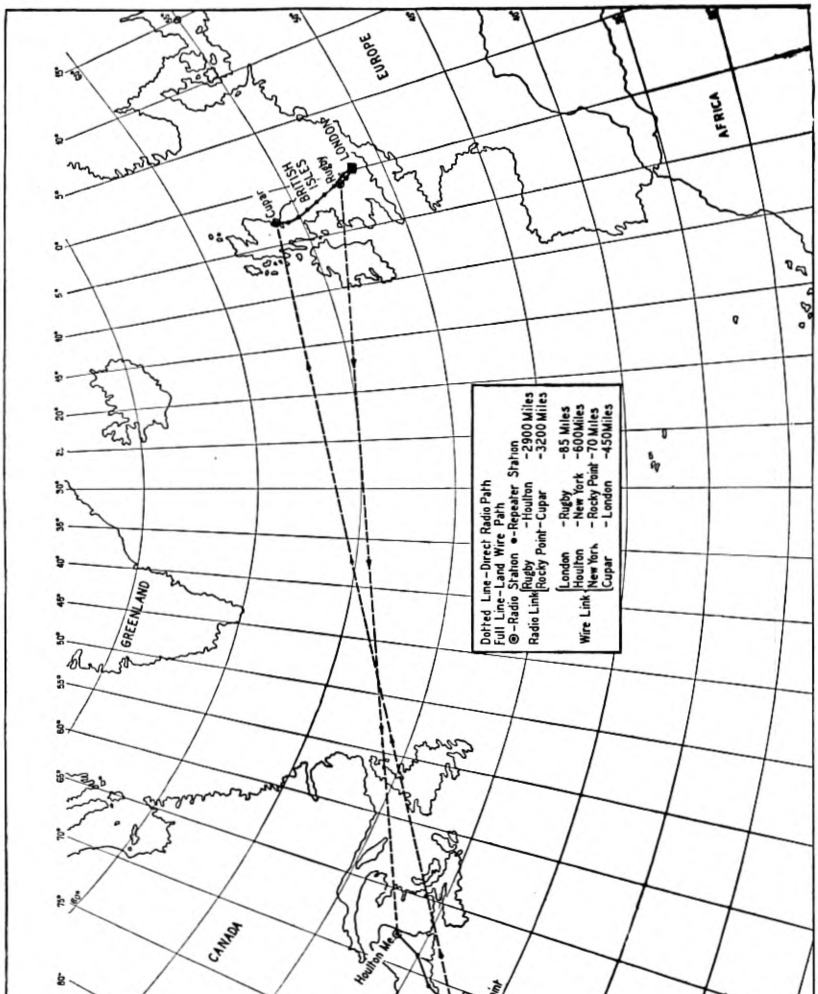

Dotted Line-Direct Radio Path
Full Line-Land Wire Path
⊙-Radio Station •-Repeater Station
Radio Link Rugby -Houlton -2900 Miles
Rocky Point-Cupar -3200 Miles
Wire Link London -Rugby -85 Miles
Houlton -New York -600 Miles
New York -Rocky Point -70 Miles
Cupar -London -450 Miles
EUROPE
AFRICA
BRITISH ISLES
Cupar
Rugby
LONDON
GREENLAND
CANADA
Houlton Me.
UNITED STATES
NEW YORK
Rocky Point

Fig. 1.

Fig. 2 (*A*) is a schematic circuit diagram emphasizing the land wire sections to permit showing the locations of intermediate repeater stations and terminal apparatus. This figure shows that the transatlantic is similar to a long "four-wire" land telephone circuit in which speech travels over different paths in the two directions. These two branches are combined into a single circuit at the terminals where special apparatus, including automatic switching devices, and specially trained men known as "technical operators" are stationed. The usual long distance girl operator establishes connections to subscribers.

As will appear shortly, the duties of a technical operator have nothing to do with the setting up of connections but require him to be continuously attentive to the electrical operation of the circuit, and to make adjustments of the amplification in the wire lines whenever the strength of voice currents bound for the radio transmitter changes. He is enabled to do this by watching the indicating needle of a sensitive vacuum tube-operated meter, called a "volume indicator." The volume indicator shows the strength or weakness of the electrical speech waves in the line. Alongside of this meter are located the dials with which he controls the amplification. Fig. 3 shows a technical operator watching the meter at the New York terminal. The apparatus shown on the panels in this picture includes the necessary terminal amplifiers and devices for adjusting and maintaining various parts of the wire and radio system.

Fig. 2 (*B*) shows the relative strength of voice waves or "electrical volumes" at various points in the circuit when a telephone subscriber in England is talking to one in the United States. The broken lines in this diagram indicate the magnitude of variations in the electrical volumes delivered to the circuit and received from it, as well as transmission variations in the radio section or "link." The relative values of electrical volume in both directions of transmission are, of course, essentially similar. The voice currents require about 1/15 of a second to travel from either terminal to the other over the circuit. It is interesting to consider that only about one fourth of this time is occupied in traversing the radio link, although radio constitutes about 85 per cent of the total length of the circuit, the remainder being in the wire lines and terminal apparatus.

It is important to note from Fig. 2 (*B*) that the ratio of the strongest to the weakest electrical volumes sent into the circuit at a terminal may be as much as 1,000 times. This is indicated at (*a*) in the figure. The variation is due partly to the different ways in which the subscribers talk, and partly to the variation in losses in the lines which connect the subscribers to the circuit. The technical operator adjusts

NEW YORK-LONDON TELEPHONE CIRCUIT

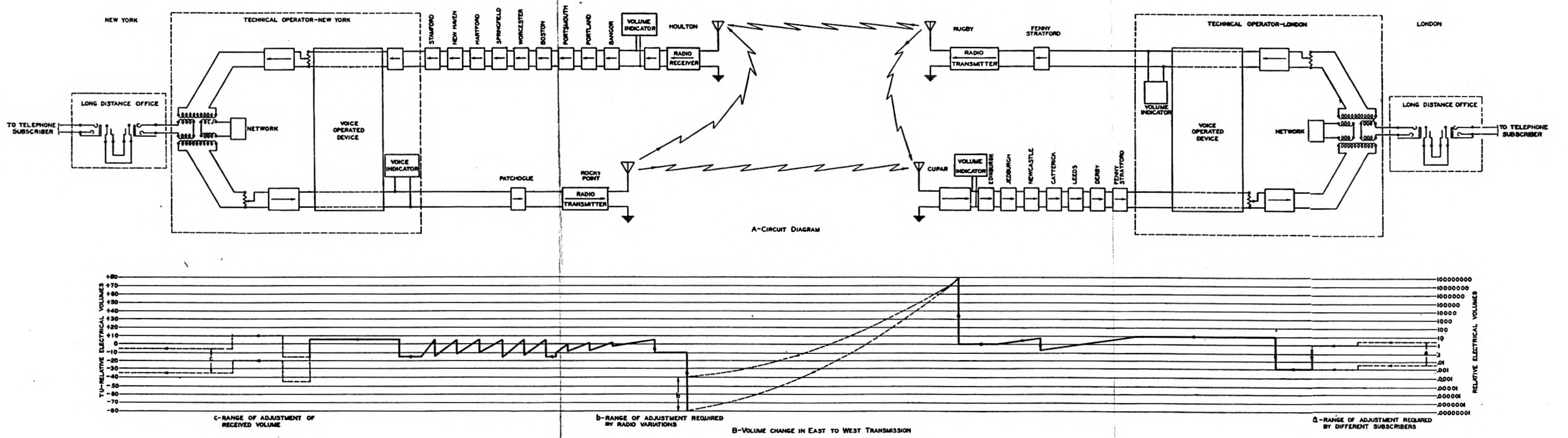


Fig. 2.

the amplifiers so as to keep the electrical volumes reaching the radio transmitter at a predetermined value. The technical operator also adjusts the received volume over the range indicated at (*c*) to give the best operation under different conditions of static, and for different types of connections.

An interesting fact to be observed here is that the voice power is amplified about 100,000,000 times in the radio transmitter and anywhere from 30,000 to 300,000,000 times at the radio receiver and associated amplifiers, depending on the loss in the radio path at any particular time as indicated at (*b*). Including the amplification which it is necessary to use on the wire "links," the total power amplification in either direction is approximately 10^{40}. Although more amplification is used in this circuit at a single point, such as at the radio transmitter, than at a single point in any other commercial telephone circuit, the total power amplification is less than in one of the telephone cable circuits from New York to St. Louis, where it is approximately 10^{50}.

Having in mind the foregoing facts, one can appreciate the difficulties which had to be faced in the way of operating this circuit and which have been successfully overcome. The more important may be summarized as follows:

(1) The transmission losses through the ether in the radio links vary from time to time in an irregular manner at intervals which preclude the possibility of making predetermined or systematic compensating adjustments of the amplification at the radio receivers.

(2) The radio links are frequently more noisy than wire circuits. This noise consists principally of stray electric waves (static) and varies considerably from time to time.

(3) The tendency for strong echo currents to exist in this circuit is considerably greater than in ordinary wire circuits. This is due partly to the methods employed for overcoming the difficulties brought about by (1) and (2), and partly to the fact that radio transmission in the two directions is carried out in the same frequency band.

These difficulties have been overcome by the following means:

(1) To overcome the variations in the transmission efficiency of the radio links, adjustments are made from time to time of the amplification in the radio receivers. Radio operators are in constant attendance at the receiving stations in order to make these adjustments.

(2) To assist in overcoming the effect of radio noise, adjustments are

made of the amplification in the wire links so that the radio transmitter is fully loaded up. This permits radiation of full power regardless of how loudly or weakly the subscriber talks, and regardless of the length of the circuit between the subscriber and the transatlantic terminals. This keeps the radio speech waves as large as possible compared to the noise at all times. These adjustments are made by the technical operators under the guidance of the "volume indicators."

(3) To suppress echo effects, a system of voice-operated switching relays has been devised whose function is to interrupt, when not in use, any transmission path which may double back to its source and give rise to echoes or singing in the circuit.

The manual adjustments of controls required in (1) and (2) should require no further explanation.

Before describing the voice-operated switching system of (3), it will be desirable to explain what this system is required to do. As previously stated, the adjustments employed to eliminate the two radio effects—namely, variability and noise—tend to increase the severity of echo effects. This follows from the fact that such adjustments result in a net transmission loss from terminal to terminal which is not constant as in ordinary telephone circuits, but which varies from time to time depending on the loss in the ether path and the strength of the voice currents which are delivered to the circuit terminal. The overall transmission of the circuit may vary from a loss to a considerable gain. If means were not taken to prevent it, this gain would set up between the two subscribers, circulating currents of rather large amplitude producing either severe electrical echo effects or the totally inoperative condition known as "singing."

A further echo difficulty was brought about through the use of a common frequency band or group of wave-lengths for transmitting in both directions. This was highly desirable to reduce the amount of frequency space occupied in the ether since there is but a limited suitable frequency space available. The radio waves at the frequencies used (namely, 58.5 to 61.5 kilocycles) cannot be directed to a definite point or confined to a single path. The radio receiver cannot, of course, when both transmitters are operating at the same frequency, select one transmitter from the other by any ordinary tuning means. Referring to Fig. 1, since the distance from the receiver at Houlton, for example, to Rocky Point is much less than the distance from Houlton to Rugby, the antenna at Houlton is exposed to a signal from the transmitter in America which is much stronger than the signal from

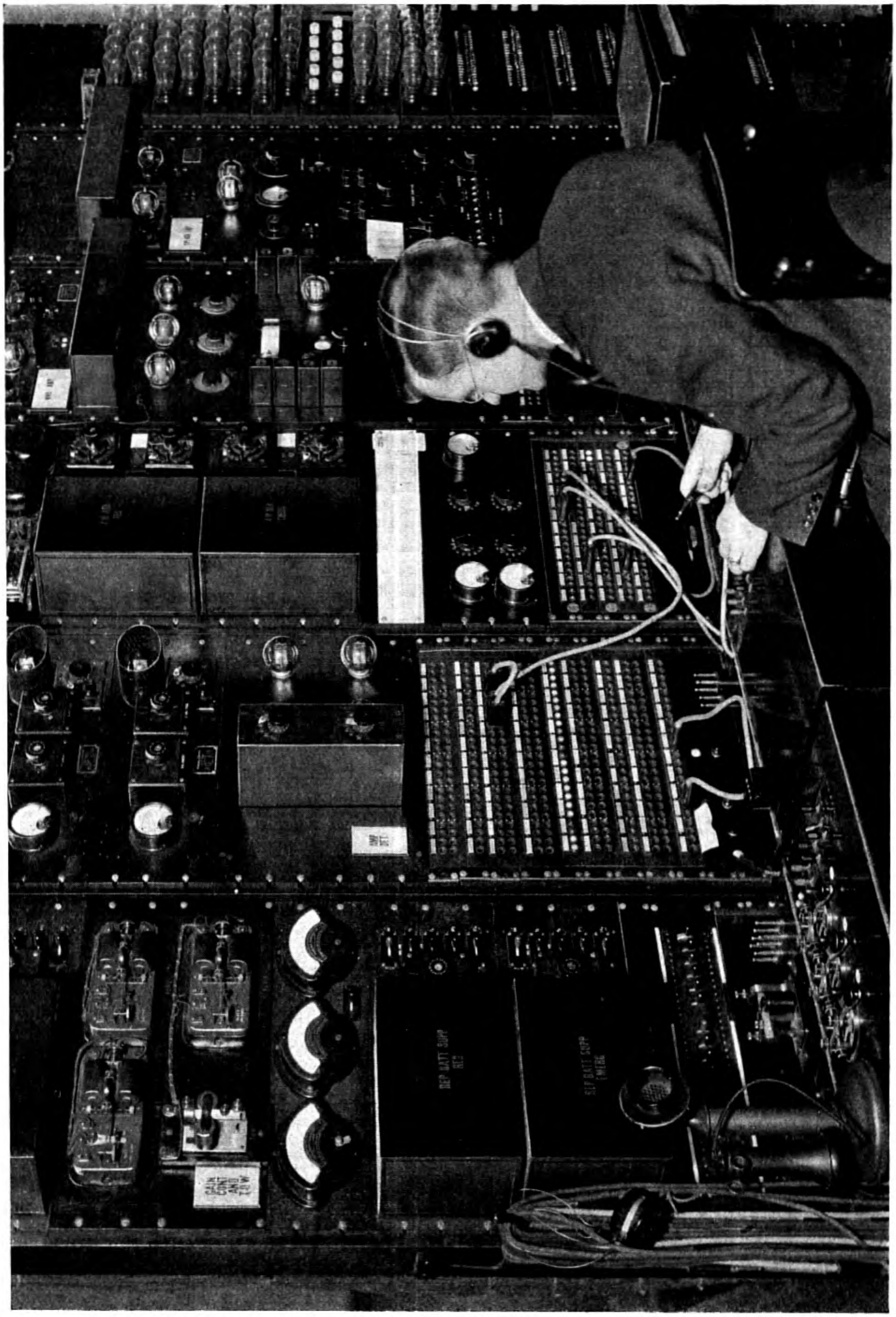


Fig. 3.

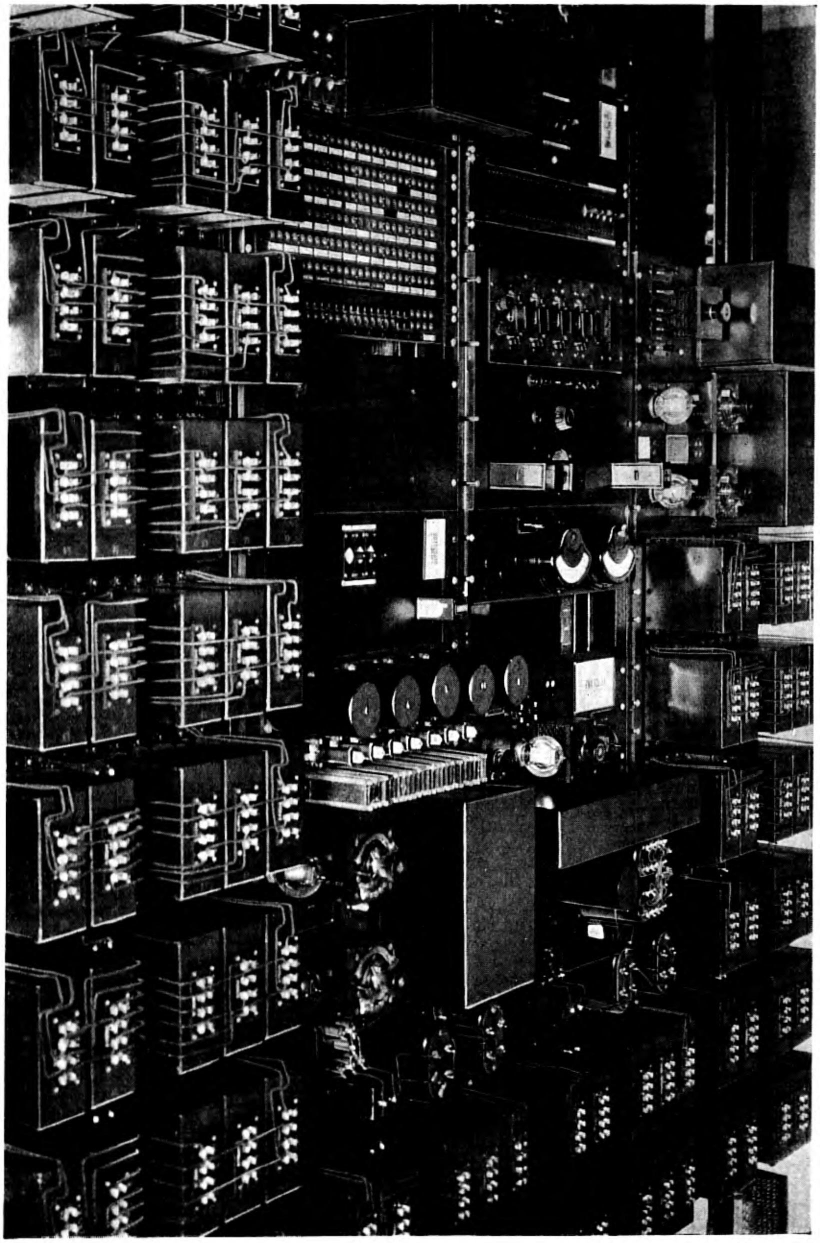


Fig. 4.

England.[1] Were it not for the use of the voice-operated devices, this would result in a very strong echo being returned to the local talker with disconcerting effects, and, dependent upon the adjustment of other parts of the circuit, might even result in violent singing of the circuit.

Referring again to Fig. 2 (*A*), it will be seen that there are three paths capable of giving rise to objectionable echoes or singing, one path at each end of the circuit through the wire lines, radio transmitter and local receiver, and a third path from end to end of the circuit and back again. The first two paths are introduced by using the same frequency band for transmission in both directions. The third path which depends upon the impedance unbalance between the two-wire lines at the terminals and their respective networks is similar to the one which gives rise to echoes in long four-wire land telephone circuits. All three paths are affected by the amplification adjustments.

Suppression of echoes and singing in the circuit requires that all three of these echo paths be kept blocked at all times against unwanted transmission. Furthermore, since there is no single point common to all the echo paths, the system for suppressing echoes comprises two separate installations—one of which is located in New York and the other in London. The devices used to control the echo paths are operated by the voice currents of the two telephone subscribers, in such a manner as to allow transmission to pass first in one direction when one subscriber is talking, and then in the other direction when the second subscriber replies. Transmission in the opposite direction to that in which the waves are traveling is blocked. When no one is talking, the outgoing transmission paths at both ends of the circuit are blocked. The necessary functions at the New York end of the circuit are performed by a combination of electro-magnetic relays, vacuum tube detectors and delay circuits. A photograph of the installation is shown in Fig. 4. At London a device performing similar functions has been developed by engineers of the British General Post Office.[2]

A schematic diagram of the device employed at the New York end is shown in Fig. 5. By tracing the action of the relays it will be seen that for all conditions of the relays, the echo paths shown are blocked at the proper times. Thus, Fig. 5, which shows the conditions when no one is talking, indicates that the circuit from the radio receiver to the terminal is normally in a receiving condition but the transmitting branch of the circuit is kept inoperative by relays *SS* and *CS*.

[1] Directive antenna systems with a blind spot might be used to overcome this effect, but their directive properties would not then be available for use against static and other interference. The general directivity of the receiving systems used, however, reduces the unwanted transmissions about 100-fold.

[2] C. A. Beer and G. T. Evans, "The Post Office Differential Voice-Operated Anti-Singing Equipment," *P. O. E. E. Jnl.*, April, 1927.

APPLICATION OF VOICE OPERATED DEVICES TO NEW YORK-LONDON TELEPHONE CIRCUIT

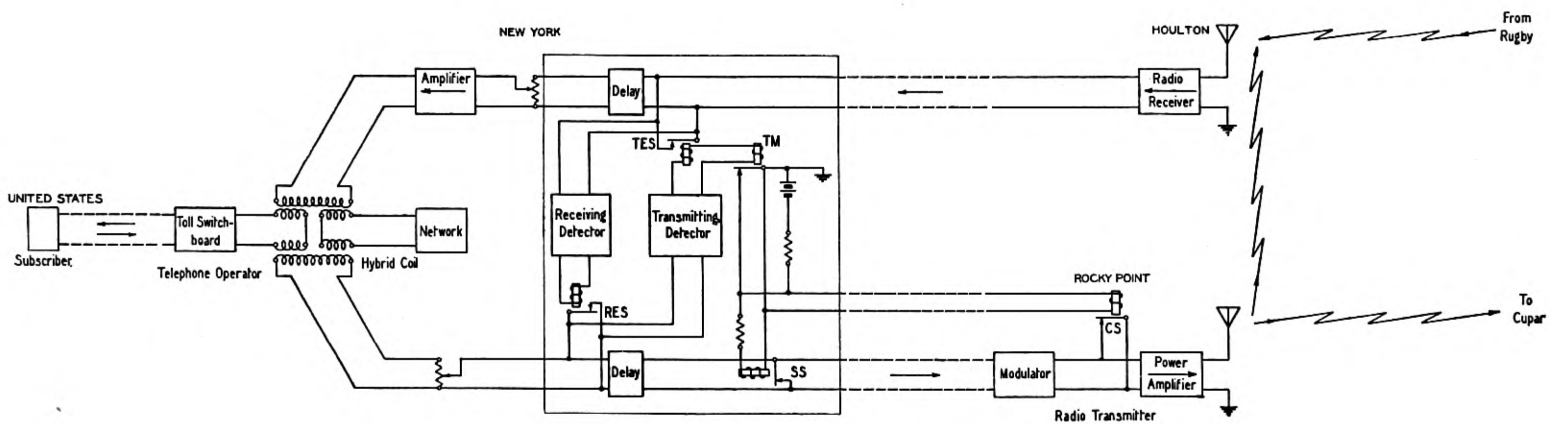
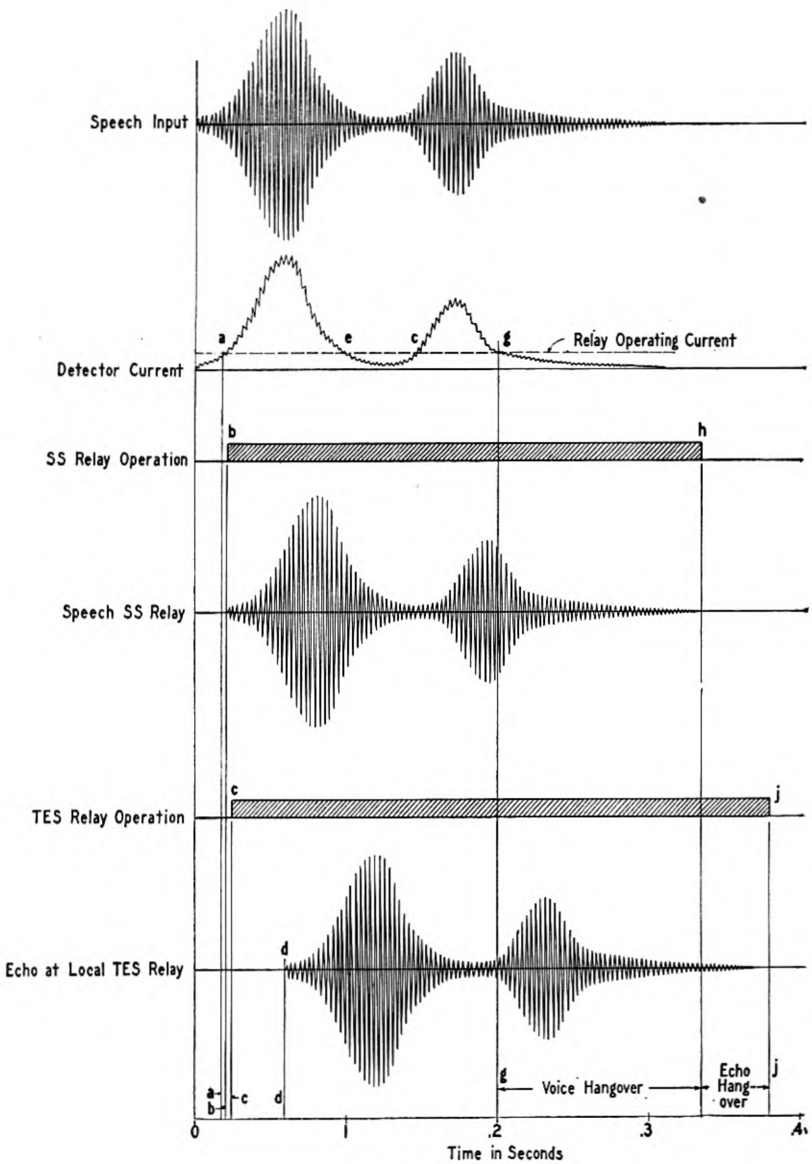


Fig. 5.

When the United States subscriber speaks, a small portion of his voice currents enters a detector, operating relays *TM* and *TES*. The action of relay *TM* causes the operation of relays *SS* and *CS*, thus clearing the outgoing line. The operation of relay *TES* short-circuits the receiving line. The main part of the outgoing voice currents passes on through the delay circuit, wire line and radio transmitter to the distant subscriber. Any transmission picked up by the radio receiver is blocked by relay *TES*. When the subscriber stops talking, the relays are restored to the normal condition.

While relays *SS* and *TES* are sufficient to switch the speech paths back and forth and prevent singing, Fig. 5 shows that there are two other relays which also interrupt undesired transmission. One of these is relay *RES*, which operates when a subscriber in England is speaking and short-circuits the United States transmitting line. This short-circuit prevents the transmitting relays from being operated by the echo of received currents returned from the local subscriber's line. The other relay, shown in Fig. 5 as *CS*, is not used at present but was needed when the circuit was first opened due to the fact that the radio transmitter and receiver in the United States were much closer together than they now are. The action of relay *CS* is similar to that of *SS*, but it was located at the radio transmitting station for an interesting reason. Although the radio transmitter is of a type which should radiate energy only during the actual transmission of speech, it would, were it not for relay *CS*, put a certain amount of noise energy into the air. While this noise, which originates partly in the radio transmitter and partly in the wire lines connecting it to the terminal, is too weak to be heard at the distant terminal, it was strong enough when picked up by the radio receiver at Riverhead, Long Island, to interfere with reception of the distant transmitter. Relay *CS* suppressed any such noise when the transmitter was idle, that is, when no one was speaking from New York. There are a number of ways of operating relay *CS*; either by voice currents rectified at the radio transmitter or via a wire circuit from the action set up by voice currents at the terminal. This latter method is shown in Fig. 5. When the United States radio receiver was moved to Houlton, Maine, the use of relay *CS* was discontinued, as the noise currents picked up there from Rocky Point were negligible.

A graphical representation of the time functions of two of the relays on the transmitting side of the voice-operated device is shown in Fig. 6. This illustrates the action during a representative spoken syllable. It will be noted that relay *SS* does not operate at the exact beginning of the speech. As shown at *a* in Fig. 6, it is necessary for the speech wave



Speech Input
Detector Current
Relay Operating Current
a
e
c
g
SS Relay Operation
b
h
Speech SS Relay
TES Relay Operation
c
j
Echo at Local TES Relay
d
a
b
c
d
g
Voice Hangover
Echo Hang-over
j
0
.1
.2
.3
.4
Time in Seconds

Fig. 6.

to build up to a certain definite amplitude before operation can begin. After this a finite time a–b is required for the relay to operate. During the interval from O to b, the voice currents are passing through the delay circuit. Thus the relay clears the path in time to transmit fully all necessary parts of this speech wave even though they have a very weak beginning.

The final trailing weak part of the speech wave is allowed to pass, by making the *SS* and *CS* relays release slowly by means of suitable circuits not shown on the drawing. This effect is shown from g to h in Fig. 6, as the "voice hangover" action. This action also functions to hold relay *SS* operated during any momentary weak lapses of speech between parts of the syllable as shown at e–f in the speech wave in Fig. 6.

As previously mentioned, a strong echo of the outgoing transmission is picked up at the radio receiver and suppressed by relay *TES*, the operation of which is indicated at c in Fig. 6. This echo is delayed an amount represented by O to d in being transmitted over the wire circuits. Relay *TES* has a releasing time slower than that of relay *SS* by the amount h–j, which is sufficient to care for the time that the echo is delayed.

In the operation of this system it is necessary for relays *SS* and *CS* and the devices which operate them to be sufficiently sensitive to operate on all parts of the outgoing speech sounds in order that none may be lost. On the other hand, relay *RES* need operate only on impulses which, if allowed to be transmitted across the multi-winding transformer ("hybrid" coil) as echoes, would be strong enough to falsely operate the relays associated with the transmitting side. Use of a relay on the receiving side which normally blocks reception would be possible, but this would require very much greater sensitivity. Due to the noise (static) introduced by the radio links, the use of such a sensitive relay is undesirable. Therefore, the device has been made to have a transmission path normally free in the direction in which the noise is high and a normally blocked path in the direction in which only the noise on the two-wire line need be combated.

If it were not for the delay circuit on the transmitting side, it would be necessary to increase the sensitivity of the voice-operated relay device so that the relays would obtain enough current to cause their operation at the very beginning of the speech wave, rather than allow the wave to build up to an appreciable amplitude before operation occurs. This delay circuit, therefore, permits appreciable reduction in the sensitivity of the transmitting side of the device, reducing the probability of operation by noise from the connected subscriber's line.

It may be said in passing that by an increase in sensitivity it is perfectly possible, with the extremely fast relays used, to omit this delay circuit and obtain satisfactory operation. This would, however, make the device more subject to noise effects.

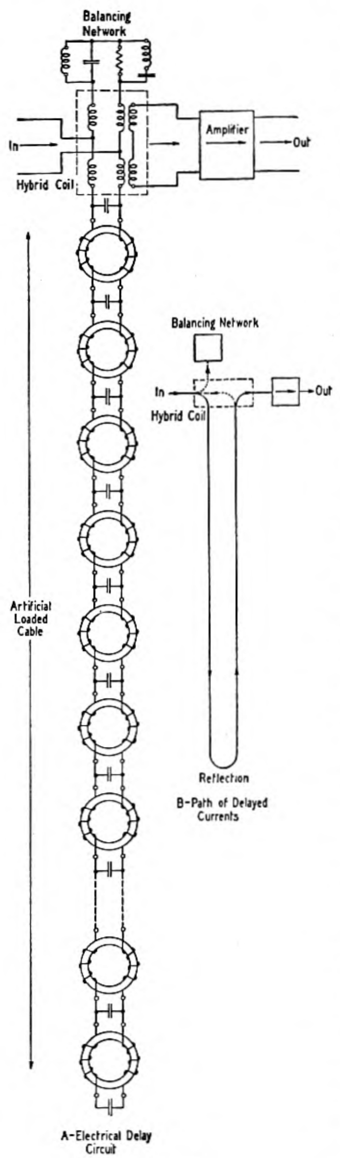


Fig. 7.

Referring to Fig. 5, a delay circuit is also included in the receiving branch of the circuit which, however, performs a somewhat different function from that in the transmitting branch. This delay circuit serves to delay transmission across the hybrid coil, thereby permitting the relay *RES* to operate and apply its protecting short-circuit before the echo from the connected circuit arrives at the input to the transmitting detector. In suppressing the echo from the radio receiver by relay *TES* a similar action is performed by the delay in transmission over the wire lines to the radio stations.

The type of delay circuit used in the voice-operated device just described is shown in Fig. 7 (*A*). This consists of an electrical network by means of which transmission sent into it is received at its output after a finite time interval. To obtain this delay action a loaded artificial line having low attenuation is employed, in conjunction with a network which balances its "surge" impedance, and a hybrid coil. An interesting feature of the arrangement is that the principle of reflection, which tends to cause objectionable echoes in telephone circuits, is here usefully employed to pass the voice currents through the artificial line twice. This results in a considerable saving of apparatus. Fig. 7 (*B*) shows the path of transmission through the delay circuit. Alternating current entering the hybrid coil divides equally between the delay circuit and the balancing network. That part which enters the balancing network is dissipated; that part which enters the delay circuit is transmitted through it, with small attenuation, to the end. Here it meets with a reflecting termination and is transmitted back whence it came. Reaching the hybrid coil, half of it goes back toward the input and half of it is transmitted on in the desired direction. The half which goes to the input meets the output side of a one-way amplifier and is dissipated. The remaining half passes through an amplifier which makes up for the transmission loss of the delay circuit and the loss due to the two divisions of energy at the hybrid coil.

The desirability of maintaining the proper relationships between the time actions of the relays and the delays in the other parts of the system will be apparent from the foregoing. A circuit for measuring the time of operation of the relays is provided which in combination with a detector and a relay may also be used for measuring the time required for alternating currents to travel through a delay network or other telephone circuit. This device is capable of measuring directly time intervals as short as .0001 second and up to 1 second in length. The measuring device is conveniently located along with the apparatus comprising the voice-operated device, as shown in Fig. 4.

In conclusion, it should be pointed out that the method of operation that has been described is expensive and has disadvantages which make it undesirable on any but a very special and necessarily complex telephone circuit, such as the transatlantic. It has, however, proved satisfactory in this service. The more interesting effects which this method of operation accomplishes may be restated as follows:

Given the condition of an anti-singing circuit such as the New York-London radio circuit, it is possible to make the amount of power radiated from the radio transmitting stations independent of the strength of the voice currents arriving over the land lines. For example, a subscriber speaking from a suburb of Chicago is heard just as loudly in London as another person speaking from the terminal of the circuit at New York.

As a result, voices of all talkers, strong or weak, are heard with the same freedom from static.

Both of the above effects result from the adjustment of the strength of outgoing speech so as to load the radio transmitter to maximum output for all messages. If the circuit were operated with amplification fixed at a value required by the strongest talkers, then the voices of weak speakers would often be lost unless the power of the radio transmitter were increased several hundred fold.

Abstracts of Bell System Technical Papers Not Appearing in this Journal

Thermal Agitation in Conductors.[1] H. NYQUIST. At the December, 1926, meeting of the American Physical Society, J. B. Johnson reported the discovery and measurement of an e.m.f. due to the thermal agitation in conductors. The present paper outlines a theoretical derivation of this effect. A non-dissipative transmission line is brought into thermodynamic equilibrium with conductors of a definite temperature. The line is then isolated and its energy investigated statistically. The resultant formula is $E_\nu^2 d\nu = 4kTRd\nu$ for the r.m.s. e.m.f. E_ν contributed in a frequency range one cycle wide by a network whose resistance component at the frequency ν is R. T and k are the absolute temperature and the Boltzmann constant. Experimental data are available for the audible range and there the agreement between the formula and the data is good. It will be observed that neither the charge nor mass nor any other property of the carrier of electricity enters the formula explicitly. They enter indirectly through R. The formula above is based on the equipartition law. If the quantum distribution law is used, the expression becomes

$$E_\nu^2 d\nu = [4h\nu R/(e^{h\nu/kT} - 1)]d\nu.$$

The two expressions are indistinguishable in the range of the measurements.

Light Waves in Metals.[2] THORNTON C. FRY. When a wave of plane polarized light falls obliquely upon a conducting surface, it gives rise to a disturbance inside the conductor which has, among others, the following peculiarities:

(*a*) It is neither plane nor elliptically polarized, but belongs to a third distinct category;

(*b*) It does not travel with what is customarily called "the speed of light";

(*c*) Its velocity varies with the angle of incidence.

There are similar light waves in dielectrics and in free space.

Transatlantic Telephony.[3] F. B. JEWETT. This paper discusses in rather popular terms some of the outstanding problems which

[1] *Phys. Rev.*, Vol. 29, p. 614, April, 1927.

[2] *Opt. Soc. Amer. Jl.*, Vol. 14, p. 473, June, 1927.

[3] *Scientific Monthly*, August, 1927.

were met and solved in the course of the development of commercial transatlantic telephony. The discussion covers the use of single side band transmission, directive receiving antennæ and voice-operated relays which permit of two-way operation upon a single wave-length. The possibilities brought to light by the extended study of receiving conditions are also described.

Some Possibilities and Limitations in Common Frequency Broadcasting.[4] DeLoss K. Martin, Glenn D. Gillett, Isabel S. Bemis. Radio broadcast stations assigned to transmit on the same carrier frequency may cause audible beat notes to be produced when their signals are received simultaneously, due to the inaccuracies in the frequency adjustments of the transmitters. The radio broadcast transmission results that might be obtained from two or more stations transmitting on the same frequency with sufficient accuracy in frequency adjustment to eliminate audio-frequency beat notes are presented briefly in this paper.

Two cases are considered, the first case where there is a difference in frequency of a few cycles and the second case where the frequency of the carrier signal for all stations transmitting on the same frequency is determined by a common oscillator.

The results of preliminary experimental tests with the signals from a station in New York City and a station in Washington, D. C., are given.

[4] *Proceeding Institute of Radio Engineers*, Vol. 15, Number 3, p. 213, March, 1927

Contributors to this Issue

HERBERT E. IVES, B.S., University of Pennsylvania, 1905; Ph.D., Johns Hopkins, 1908; assistant and assistant physicist, Bureau of Standards, 1908–09; physicist, Nela Research Laboratory, Cleveland, 1909–12; physicist, United Gas Improvement Company, Philadelphia, 1912–18; U. S. Army Air Service, 1918–19; research engineer, Western Electric Company and Bell Telephone Laboratories, 1919 to date. Dr. Ives' work has had to do principally with the production, measurement and utilization of light.

FRANK GRAY, B.S., Purdue, 1911; instructor and graduate student in physics at the University of Wisconsin, Ph.D., 1916; member of the Naval Experimental Station during the war. Mr. Gray entered the Bell Telephone Laboratories—then the Engineering Department of the Western Electric Company—in 1919 and has been closely associated with Dr. Ives in his studies on light.

J. W. HORTON, B.S., Massachusetts Institute of Technology, 1914; instructor in physics, 1914–16; Engineering Department of the Western Electric Company and Bell Telephone Laboratories, 1916–. Mr. Horton has been closely connected with the development of apparatus for carrier current communication.

R. C. MATHES, B.Sc., University of Minnesota, 1912; E.E., 1913; Engineering Department of the Western Electric Company and Bell Telephone Laboratories, 1913–. Mr. Mathes has played an important part in the repeat development program and in the design of vacuum tube amplifiers for a wide variety of uses.

H. M. STOLLER, degree of E.E. from Union College, 1913; M.S. in electrical engineering, 1915; Engineering Department of the Western Electric Company and later Bell Telephone Laboratories, 1914 and 1916–. Most of Mr. Stoller's work has dealt with special problems connected with electrical power machinery, particularly voltage and speed regulators. He designed a multi-frequency generator which is now employed in the voice frequency carrier telegraph system.

E. R. MORTON, M.E., Stevens, 1917; Western Electric Company and Electric Test Lab., 1917–18; Air Service, U. S. Army, 1918; development and manufacturing control of color motion pictures, 1919; Harvard, 1920, electro-chemical investigations of nickel;

Fairchild Aerial Camera Corp., 1921–22; Western Electric Company and Bell Telephone Laboratories, 1923–. Since entering the Laboratories, Mr. Morton's work has related principally to vacuum tube regulators.

D. K. Gannett, B.S. in engineering, 1916, University of Minnesota; E.E., 1917; American Telephone and Telegraph Company, Engineering Department, 1917–19, and Department of Development and Research, 1919–. During the war, Mr. Gannett was engaged in development work on amplifiers for submarine cable telegraphy. Since then, his work has been in connection with the transmission development problems of toll line signaling, vacuum tubes, telephotography and television.

E. I. Green, A.B., Westminster College (Fulton, Mo.), 1915; University of Chicago, 1915–16; professor of Greek, Westminster College, 1916–17; Captain, U. S. Army, 1917–19; B.S. in electrical engineering, Harvard University, 1921; Department of Development and Research, American Telephone and Telegraph Company, 1921–. Mr. Green has been engaged largely on work in connection with carrier transmission problems and development.

Edward L. Nelson, B.S. in electrical engineering, Armour Institute of Technology, 1914; Engineering Department of Western Electric Company, 1917; technical expert, Bureau of Engineering, Navy Department, 1917–18; Lieut. U. S. N. R. F., 1918–19; Engineering Department of Western Electric Company and Bell Telephone Laboratories, 1919–. Mr. Nelson has been closely connected with the development of radio apparatus, particularly transmitting equipment for ship-to-shore telephony and for broadcasting.

Karl K. Darrow, S.B., University of Chicago, 1911; University of Paris, 1911–12; University of Berlin, 1912; Ph.D. in physics and mathematics, University of Chicago, 1917; Engineering Department, Western Electric Company and Bell Telephone Laboratories, 1917–. Dr. Darrow has been engaged largely in preparing studies and analyses of published research in various fields of physics.

R. L. Young, B.S. in electrical engineering, University of Pennsylvania, 1907; Westinghouse Electric and Manufacturing Company, 1907–09; Carnegie Institute of Technology, instructor of mathematics, 1909–10; McGraw Publishing Company, district manager Pittsburgh and assistant manager New York of *Metallurgical and Chemical Engineering*, and editorial board *Electrical World*, 1909–11; Engineering Department, American Telephone and Telegraph Company, 1911, and Department of Development and Research, 1911–. Mr. Young has been in charge of Power Development work since 1911.

S. B. Wright, M.E. in E.E., Cornell University, 1919; Department of Development and Research, American Telephone and Telegraph Company, 1919–. Mr. Wright has been engaged in transmission development work on long distance voice frequency telephone circuits.

H. C. Silent, E.E., Cornell University, 1921; Department of Development and Research, American Telephone and Telegraph Company, 1921–. Mr. Silent has been associated with the development work on voice-operated switching devices, and spent some time in England in connection with their application to the New York-London telephone circuit.

Walter A. Shewhart, A.B., University of Illinois, 1913; A.M., 1914; Ph.D., University of California, 1917; Engineering Department, Western Electric Company and Bell Telephone Laboratories, 1918–. Mr. Shewhart is making a special study of the application of probability theories to inspection engineering.