The Bell System Technical Journal

Vol. XIX *April, 1940* *No. 2*

Advances in Carrier Telegraph Transmission

By A. L. MATTE

Introduction

IN the comparatively short period which has elapsed since its commercial introduction in a practical form, the voice-frequency carrier method of operating telegraph has risen to a position of preeminence and is becoming the outstanding means for providing telegraph facilities over main toll routes.

Since the original installation, improvements have been made in the carrier supply, level-compensating devices, maintenance facilities, and in numerous other specific physical parts of the system. Operating speeds have also gone up, and the number of telegraph channels per telephone circuit has been increased. Furthermore, this system, originally designed for cable circuits operating at voice frequencies, has been applied to open-wire lines and adapted by remodulation to other frequency ranges, in particular to those occupied by existing carrier-telephone systems. Some of the chief advances, however, have been of a more intangible nature, not the least of these being the clearer insight which experience and extended tests have given into the possibilities and limitations of carrier-telegraph systems with respect to interference and other causes of signal distortion.

As a result of the success attained by the carrier-telegraph system for open-wire lines,[1, 2] which had been in commercial service in the Bell System since 1918, this company's engineers turned their attention to the adaptation to cable circuits of the carrier method of transmission for telegraph purposes. Following this work and extensive field trials, the voice-frequency carrier-telegraph system went into commercial use in the Bell System in 1923.[3] The initial installation consisted of ten two-way or duplex channels between New York and Pittsburgh, giving an aggregate channel mileage of 3800 miles (6120 km.) including both directions of transmission. Since then, the application of this mode of transmission has spread rapidly so that in spite of the intervening period of retarded business activity it now provides about $1\frac{1}{2}$ million miles (2.4×10^6 km.) of high-grade circuits throughout the Bell System. Its use with variations has extended to other countries so that it bids fair to become an outstanding means for providing overland telegraph facilities, particularly for long distances, where the service is exacting. As indicating the general trend, it may be stated that in England alone about 1700 voice-frequency telegraph channels were reported as available for operation at the end of 1938.[4]

It is interesting to note the role played by carrier telegraph in the evolution of the art of telegraphy. The three major telegraph systems up to about 1890 were those of Hughes on the Continent, Wheatstone in England, and the manual Morse system in the United States. As electrical communication reached out to greater and greater distances, the desire to utilize costly lines more effectively led inventors to concentrate their efforts in two different directions; namely, the development of high-speed systems and of multiplex systems. In high-speed systems, the object, as the name implies, is to secure increased line output by speeding transmission well beyond the ability of a single operator. These devices are characterized by automatic transmitters which can be supplied with perforated tape prepared in advance by a number of individuals. Typical high-speed systems are the Wheatstone automatic, the Murray automatic, the Siemens and Halske high-speed, and the Creed high-speed.[5]

The first efforts at multiplexing circuits were based upon the suggestions of Gintl and Highton, who proposed to take advantage of directional and magnitude effects respectively, and whose ideas were brought together by Edison in his invention of the quadruplex. The multiplex system as we know it, however, was the invention of Baudot, who, putting into practical form a suggestion made by Moses G. Farmer as far back as 1853, produced a system whereby the line was assigned successively to a number of operators. This process had the great advantage that while maintaining the line speed which economy made imperative, it permitted a number of messages to be transmitted simultaneously without delay and with each operator working at his normal pace. The chief examples of the multiplex are the Baudot, Murray, and American.[5, 6]

Owing to the advantages of these higher output systems, the older methods of operation were gradually supplanted for the longer commercial message circuits. This was particularly true in Europe, since certain conditions operating in America tended to favor the survival of the simple Morse arrangement; the chief of these being the availability of large numbers of composited and simplexed circuits, most of which were used in private line service,[7] and the low traffic density on many long multi-section circuits, making it desirable to provide intermediate operating points. By about 1920, the weight of evidence definitely favored the multiplex method of exploitation over the use of the high-speed printer.[6] It was at about this point in telegraph history that the carrier telegraph method of subdividing the line capacity made its appearance and, through its superior flexibility and lesser intricacy of operation, began gradually to supersede the distributor methods of multiplexing circuits for many types of services.

While voice-frequency telegraphy was foreshadowed by Elisha Gray's harmonic telegraph,[8] which was exhibited at the Third French International Exposition in 1878 and at the Electrical Exhibition in Paris in 1881,[9] its practical embodiment had to await the invention of the electrical filter by Campbell, that of the audion by DeForest, and the production of effective means for generating alternating currents of acoustic frequencies.

The success of this system rests mainly upon its adaptability to economical operation over telephone circuits by making effective use of the whole frequency band usually allocated to the voice, and in requiring similar transmission characteristics. Henceforth, every advance in telephony directed to an improvement of the transmitting medium contributes as well towards the improvement of telegraphy; the economies of wide-band carrier telephony, the improved equalization and regulation of circuits, the reduction of interference and the elimination of crosstalk, all tend to make the telegraph a more dependable and efficient tool for modern industry and modern living. Thus telegraphy, one of the oldest of the electrical arts, having fathered the telephone, now finds, within the great technical structure which the latter has created, a fertile medium for the development of its usefulness, not as a competitive but as a complementary service. Thanks to voice-frequency telegraphy, wherever the telephone reaches, a high-speed, reliable, record-form of telegraph may follow. This has brought about a great simplification of the problem of interconnection in such large communication networks as the international postal area in Europe and the Bell System in our own country.

Furthermore, carrier telegraphy has doubtless been a means of advancing the fortunes of the start-stop teletypewriter,[10] by subdividing the frequency band to such an extent that one channel may be economically assigned to a single operator working at normal speed. It has also been a factor in simplifying the switching problems presented by the extensive introduction of teletypewriter exchange (TWX) service.[11]

The purpose of this paper is to describe the principal transmission developments which have taken place in the voice-frequency system since the first commercial installation, to present some of its operating characteristics, and to outline advances in maintenance methods which have developed during this period.

There has been marked and steady improvement in the quality of the service rendered by the voice-frequency telegraph circuits during the last decade within the Bell organization, and while it would be unfair to overlook the part which imaginative management, employee cooperation, and similar factors have played in securing this desirable result, it appears certain that a good deal of it is to be credited to those physical improvements and advances in testing and operating procedures which we are about to recite.

Extensions in Utilized Frequency Range

While the greater part of past experience has been had with the application of voice-frequency systems to extra-light-loaded four-wire cable circuits,* a considerable mileage now utilizes other types of telephone facilities, particularly high-frequency carrier open-wire lines [13] through the less densely populated regions. These latter applications are interesting to the transmission engineer because the association of telephone and telegraph in the same repeaters brings into view new problems which will doubtless grow in importance as the use of broad-band carrier systems becomes more extensive. A typical voice-frequency carrier-telegraph circuit is shown in Fig. 1. It consists of a section of four-wire cable connected in tandem with a three-channel type "C" carrier-telephone circuit,[14, 15] without mechanical repetition at the junction point. The telegraph power is appropriately modified to suit the requirements of the two media by means of pads and amplifiers *P* at the point where they join. In addition to this, the remaining telephone circuits operating through the same carrier repeaters are equipped with volume limiters [16] to prevent voice-energy peaks, which contribute little if any to telephone quality, from overloading the amplifiers, thereby causing excessive distortion to the telegraph.

Shortly after the initial voice-frequency telegraph installation in cables the number of channels was extended from 10 to 12 by the addition of two channels at the upper end of the frequency range, corresponding to carrier frequencies of 2125 and 2295 cycles.

* These circuits are designated as H44. They consist of 19 AWG conductors, loaded with 44-millihenry coils 6000 ft. (1830 m.) apart, with four-wire repeaters spaced at approximately 50-mile (80.5 km.) intervals.[12]

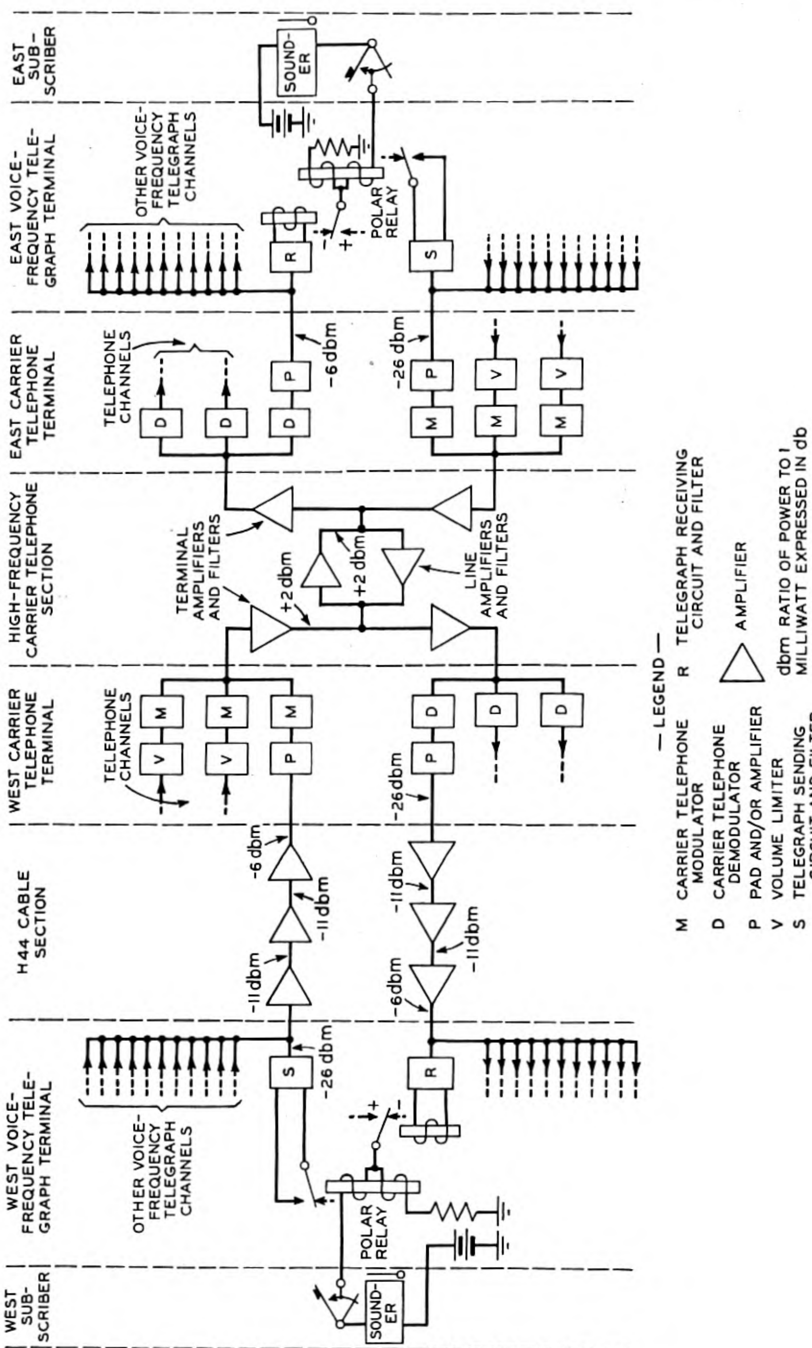


Fig. 1—Typical carrier-telegraph circuit.

At a later date the extensive introduction of H44 circuits, with their relatively high cut-off, was thought to make it desirable to consider a further extension of the frequency band utilized by the telegraph. However, in view of the fact that the then existing state of the art of filter design made it impractical to produce economical filters of the required narrow band-width but having the desired high mid-band frequency, it was decided to develop a carrier-telegraph system suitable for use between dense traffic centers by superposing two standard 12-channel systems over the same cable pair. This was accomplished by causing the various signaling frequencies of one system to modulate a single secondary carrier, thereby transposing all the frequencies of this voice-frequency system to a frequency range above that of a normal voice-frequency system operating over the same cable pair.

The line circuit used with this double system was required to transmit a range of frequencies from about 350 cycles to 4400 cycles, and the stability within this range had to be such as not to cause excessive bias in any channel with the regulating methods available at that time. In order to secure this result, it was necessary to change the transmission characteristic of all repeaters from that used for ordinary four-wire telephone or voice-frequency telegraph transmission and, furthermore, to modify somewhat the regulating repeaters in order to maintain the desired transmission characteristics with changes in temperature.

No changes of any kind were required in the voice-frequency telegraph terminals. The channel frequencies on the line were arranged to extend uninterruptedly at 170-cycle intervals from 425 cycles to 4335 cycles.

This arrangement was called the "double-modulation" system, because operation of two voice-frequency carrier telegraph systems over the same circuit was realized by causing all the channel frequencies of one of these two systems to pass through a common modulator, where a second modulation took place; the individual frequencies of each channel being already considered as having been modulated by the sending relays. A single secondary carrier-frequency was used which was common to all the channels. The allocation of frequencies, which was identical for both directions of transmission, will be readily understood by referring to Fig. 2*B*. The general principle of operation is illustrated in Fig. 2*A*, which shows transmission from west to east, it being understood that the arrangement from east to west is identical. The voice-frequency system denoted as No. 2 will be seen not to differ in any way from the earlier arrangement except that signals from all twelve channels traverse low-pass grouping filters at the sending and receiving terminals. In the case of voice-frequency system No. 1, however, all the frequencies are passed through a modulator, where they are transposed as a group to a position above those of system No. 2. The lower sideband of the secondary carrier is used, so that the order of the channels is reversed on the line. After modulation, this

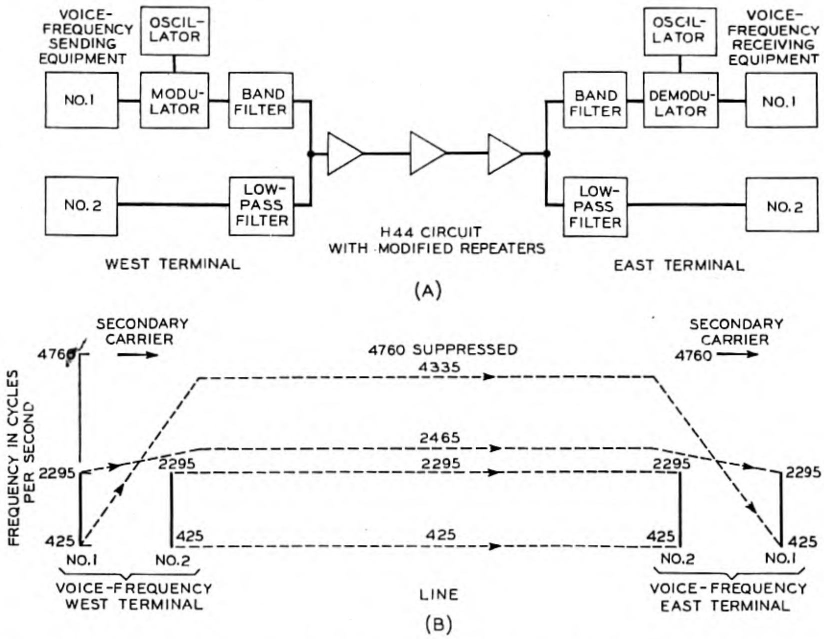


Fig. 2—Double-modulation telegraph system. *A*. Block diagram for one direction of transmission. *B*. Frequency relations at terminals and on the line.

group of frequencies passes through a sending band filter, which elminates all the unwanted frequencies, thereby preventing useless overloading of the repeaters and the creation of undesired modulation products therein. The two groups of frequencies pass through common repeaters over the modified H44 circuit and are then separated at the receiving terminal by a combination of filters similar to the one at the transmitting end of the circuit. The signals pertaining to voice-frequency system No. 1 are next demodulated by a secondary carrier having the same frequency as that used at the sending end and thereby reduced to a frequency range adaptable to the standard terminal equipment. The modulators and demodulators were provided with separate oscillators at both ends of the circuit.

Both modulators and demodulators were of the push-pull type and were arranged as grid-current modulators [17] instead of as plate-current modulators such as were then current in telephone practice. In the case of grid-current modulators, the necessary non-linear characteristic is obtained by so constructing the input circuit that the voltage between the grid and filament of the modulating tubes does not vary directly with the voltage impressed upon the modulator input, while plate modulation is suppressed; in the case of plate-current modulators, on the other hand, there is a linear relation between the grid-to-filament voltage and the voltage impressed upon the modulator input, but advantage is taken of the fact that the plate current does not vary directly with the grid voltage to secure the desired modulation effect. The reason for using grid-current modulators of this type was that the increased power output secured thereby made it possible to produce the required output levels without auxiliary amplifiers.

The trial double-modulation system performed satisfactorily under commercial conditions although the upper group or remodulated system was somewhat less satisfactory than the standard. This wide range system has not been used, however, partly because of reduced demand due to economic conditions and partly because advances in the art of filter design now make possible a considerable extension on a single modulation basis; furthermore, the increasingly wide use of carrier telephone circuits makes it desirable to restrict the band width used by a voice-frequency telegraph system to the frequency range normally assigned to a telephone channel.

Signal Distortion

Improvements in transmission amount essentially to reductions in signal distortion. The principal sources of such distortion [18, 19] are:

Type of Distortion	Source
Characteristic	Filter characteristics.
	Wave shaping characteristic of detectors.
Bias	Variations in circuit net-loss.
	Battery variations at terminals.
	Variations in carrier-current generator voltage.
	Gradual frequency changes.
	High-resistance sending-relay contacts.
	Asymmetrical relay adjustments.
Fortuitous	Noise.
	Lightning.
	Functional switching operations.
	Change in repeater gain with load.
	Intermodulation of several channels in repeaters.
	Infiltration from adjacent telegraph channels.
	Relay contact troubles.
	Irregularities in relay operation.
	Rapid variations in carrier frequencies.

Characteristic distortion attributable to the telegraph channel filters is not an important limitation at the speeds now generally used and need not detain us as it has been discussed at length elsewhere.[20, 21, 22] It might be stated, however, that while the frequency band used at the present time, which provides an effective width of about 110 cycles, allows some margin of transmission with present speeds, the proposition of reducing the spacing of carriers is not very attractive for a number of reasons, among which may be mentioned the reduction in cost per cycle of band width due to the development of carrier-telephone systems, the possibility that higher speed requirements may ultimately make the present spacing desirable, and the greater degree of maintenance demanded by a system designed with less liberal operating margins where a number of sections are operated in tandem. With respect to the speed factor, it may be observed that service is already being rendered commercially in several cases at 75 words per minute, and still higher speeds have been used.

Bias

Variations in circuit loss consequent upon changes in temperature, battery voltages, etc., are a major factor in determining signal distortion, as will be seen by reference to Fig. 3 which shows graphically the

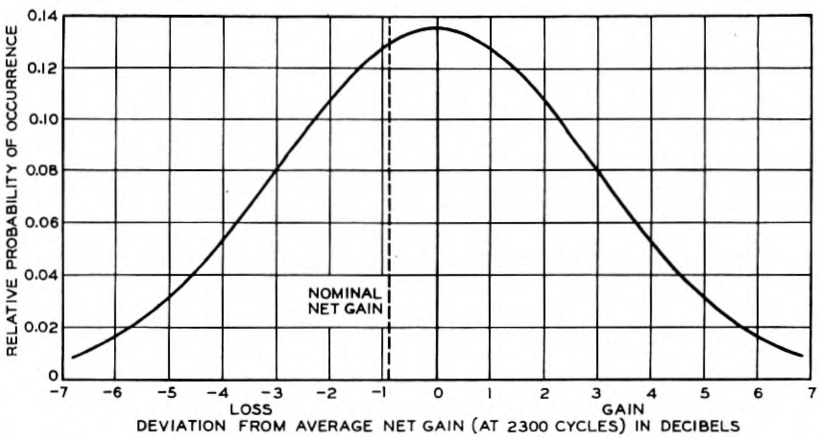


Fig. 3—Probable distribution of net-gain variations at 2300 cycles. H44 circuits about 1000 miles (1600 km.) long. Includes both differences between circuits and variations with time.

approximate manner in which a particular group of 19 gauge H44 circuits about 1000 miles (1600 km.) in length varied through a one-year cycle at the frequency of channel 12 (2295 cycles). The effective range of variations with respect to the optimum input level to the detector depends on the method by which the sensitivity of the latter is adjusted. Two methods have particular advantages: In one of these the detectors are adjusted for a nominal received level which is made the same for all the channels. This is the method usually employed for cable circuits. It permits adjusting the detectors at any time without reference to the particular line with which they are to be used and without the assistance of an attendant at the distant station. It will be evident that the departures from optimum line gain must then be reckoned from this nominal net circuit-gain, which in the illustration is shown as being .9 db below the actual mean value. This discrepancy is principally due to imperfect equalization of the line. In general it is least in the neighborhood of 1000 cycles and increases progressively as one goes away from this frequency. Furthermore, the standard deviation of the distribution curve increases generally in the same manner, so that for a 12-channel system the condition illustrated is perhaps the most unfavorable one.

A second method of lining up is to adjust the detector sensitivity so as to give unbiased operation with signals transmitted from the distant station and with the line loss whatever it happens to be at the moment. On the average, and in the long run, the effect of this procedure is to restore the symmetry of the variations, but the standard deviation is multiplied by a factor equal to the square root of 2. This is because the occurrence of a given departure from the optimum level is then further conditioned by the particular net gain which happens to exist at the time the detector is adjusted, and the chance of a gain of this particular value is given, of course, by the same distribution which has just been discussed. Most of this increased latitude in variations can be eliminated, however, by seasonal adjustments; a procedure which is evidently of no help when the first method is followed. This second method has been found useful on open-wire circuits because the average net loss of telephone channels over these facilities depends somewhat on their frequency allocation and varies more widely than is the case with cable circuits.

If no provision were made to compensate for these line-variation effects the result would be a rapid change in bias as the level at the input of the detector departs from its optimum value. This is shown for a typical telegraph channel by the dotted line in Fig. 4. By the use of a *level compensator* associated with each individual detector a great improvement may be obtained, however, the bias variations being reduced to those illustrated typically by the full line in the same drawing. The resulting changes in teletypewriter orientation

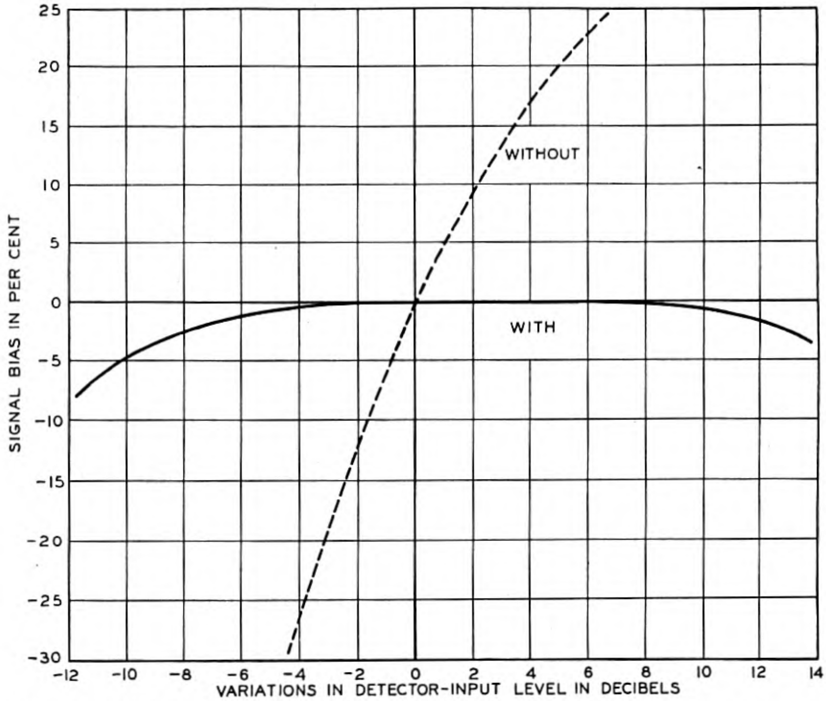


Fig. 4—Effect of level compensator. Signal bias vs. variations in detector input level. Signaling speed 23 dots per second.

range when the level compensator is used are of the order shown in Fig. 5, indicating satisfactory operation over extensive changes in circuit loss. It will be seen that in the absence of a level compensator a single-section telegraph circuit operating at a speed of 23 d.p.s. (46

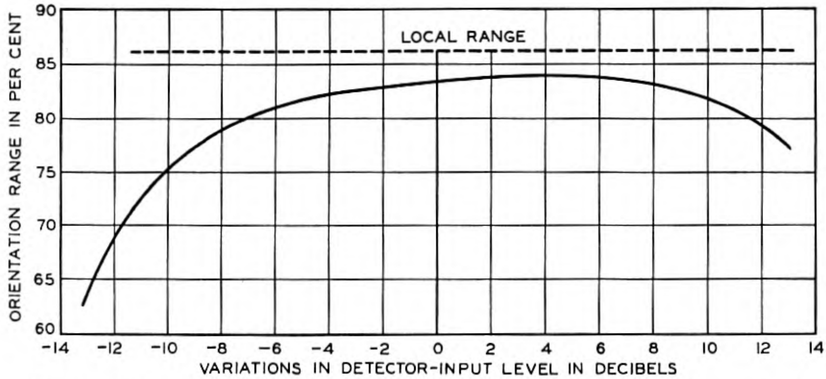


Fig. 5—Teletypewriter orientation-range vs. variations in detector input-level.

bauds) may exhibit a change in signal bias of 4 per cent or more for each db change in input level. With a level compensator of the type described herein, the bias may be kept within a range of $\pm$ 2 per cent for a variation in input level of $\pm$ 8 db.

The elimination of any considerable bias variations in individual telegraph sections due to level changes which usually occur in practice is particularly important in the case of multi-section * circuits. As a result of this improvement it has been found feasible under test conditions to operate satisfactorily as many as 10 telegraph sections in tandem at 60 words per minute for long periods without objectionable bias variations due to level changes and without the use of regenerative telegraph repeaters. In practice, however, other considerations usually make it desirable to use a regenerative repeater when the number of sections in tandem exceeds 4.

The greater part of the effective changes in received level in a given circuit is due to temperature changes which are imperfectly compensated for by the regulators, aggravated by the fact that the conditions prevailing when the circuits are adjusted within the limits specified by the maintenance routines may depart considerably from the average. In addition to this, there are variations of considerable magnitude between individual circuits due to structural differences. In view of the fact that the variations over the whole frequency range are not the same, there is a material advantage in a compensator which adjusts the gain of each detector independently, a feature which could not be secured with a pilot-channel regulator.

Level Compensator

The level compensator,[23] shown diagrammatically in heavy lines in Fig. 6, may be considered as functionally divided into two parts, one of which is in series with the grid of the detector tube and the other of which is connected to the armature of the receiving relay. The first of these will be referred to as the *grid-bias circuit* and the second as the *compensator-relay circuit.*

The grid-bias circuit consists essentially of a condenser C shunted by a high resistance R_c, in series with a biasing battery E_0 of fixed voltage, the secondary of the interstage transformer T, and the grid-filament terminals of the detector tube V. This arrangement functions to keep the effective grid-filament voltage due to the signals nearly constant, irrespective of their magnitude, by setting up a voltage on the condenser which adds algebraically to the grid-bias battery and whose magnitude

* By a multi-section telegraph circuit is meant a connection made up of 2 or more telegraph lines in tandem with mechanical repetition between them.

is automatically adjusted to be proportional to that of the incoming signals.

At any instant the actual voltage between filament and grid is therefore equal to the algebraic sum of (1) the fixed bias voltage E_0;

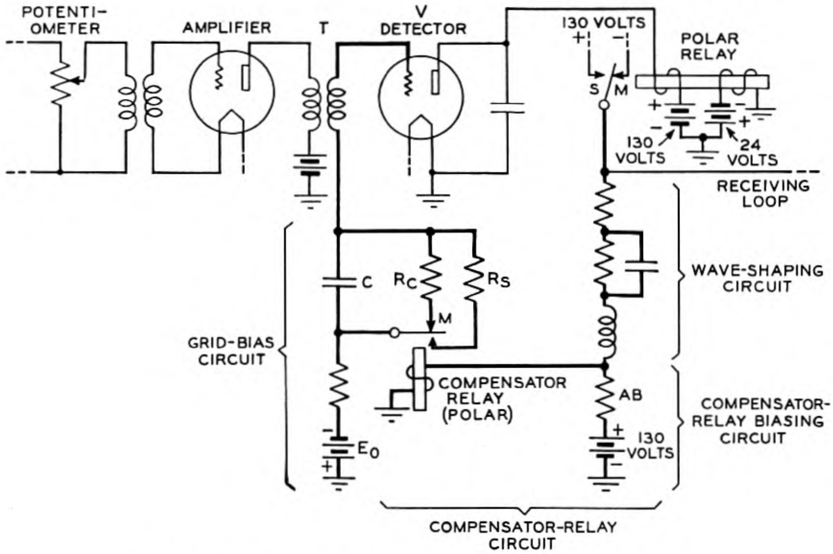


Fig. 6—Schematic diagram of level compensator.

(2) the voltage due to the charge on the condenser C; (3) the signal voltage across the secondary of the interstage transformer T; and (4) the drop in voltage across the transformer. By making the detector sensitivity sufficiently great, the signal voltage is caused to overcome the opposing negative bias during a portion of each positive half of the carrier cycles composing a marking pulse, so that a net positive voltage is periodically impressed on the grid causing a flow of current between it and the filament and consequently through the resistance R_c and the condenser C in parallel.

The resulting voltage across condenser C is in the same direction as that of the fixed grid battery E_0 and adds thereto. The condenser voltage is determined by the amplitude of the received carrier current, increasing with increased input level and decreasing with decreased input level. By a proper selection of the constants of the circuit, the desired compensation action may be obtained. This action will be such that the change in voltage across the condenser will always, within the effective range of compensation, produce the proper negative grid voltage for unbiased reception of telegraph signals by the receiving relay. A quantitative discussion of the operation of the compensator is given in the Appendix.

In order that the voltage across the condenser may not decrease during spacing signals, an auxiliary polar relay, called the compensator relay, is provided which derives its operating current from the armature of the receiving relay and serves to disconnect the resistance R_c during spacing signals. As discussed in the appendix, the unbiased operation of the compensator relay would cause a noticeable decrease in the condenser voltage during the rapid transmission of signals, because the wave shape of the signals impressed on the grid circuit of the detector tube is not square but considerably rounded. In other words, there is a portion of a marking signal during which the receiving relay is operated to marking but the grid is non-conducting. Hence more charge would leak from the condenser during the time the resistance is connected across the condenser than would be replaced by rectification. To prevent this, the compensator relay bridges the discharge resistance R_c around condenser C for a period of time which is shorter than that during which the receiving relay is on its marking contact. The amount by which the compensator condenser must be biased depends on the signaling speed and other factors. It is determined by observing the "drift" in bias suffered by reversals when these are suddenly switched on after a long marking interval.

The operation of the level compensator will be more readily understood by referring to Fig. 7, which shows diagrammatically the manner in which received impulses of different magnitudes are made to operate the receiving relay for equal time intervals. In this diagram, the positive halves of the envelopes of three received marking impulses * of different amplitudes are shown in relation to the grid-voltage plate-current characteristic of the detector tube. For normal input level the sensitivity of the detector is made sufficiently great so that the amplitude of the impulse which is impressed on the interstage transformer is of the magnitude shown at N. The envelope of the received carrier current is symmetrical about the line E_N, which is located at the net value of grid biasing voltage due to the battery voltage E_0 and the grid condenser voltage e_c. The latter voltage, as previously noted, is produced in the grid circuit by rectification of that part of the received carrier current which lies on the positive side of the zero grid-voltage axis OO. By properly adjusting the bias of the receiving relay the latter may be made to operate at a value AA which is one-half the crest value of the envelope. Signals having amplitude N will thus be repeated unbiased by the receiving relay, since the ascending and de-

* In practice such pulses would usually reach the steady state.

scending curves bounding the envelope represent identical time sequences of events.

If the level increases so that the amplitude of the received signals has a value indicated by curve H, a much larger part of the signal current momentarily flows through the grid-filament space and is rectified, thereby increasing the charge on the condenser in such a

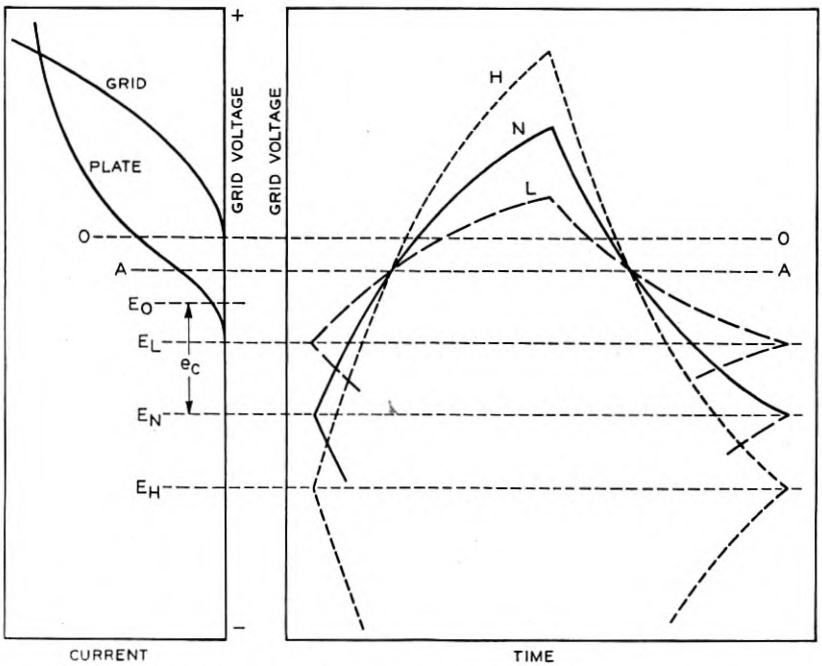


Fig. 7—Principle of level-compensator action.

way as to increase the negative bias of the grid and shift the signal bodily towards the negative side of OO. This shift continues until the total effective grid voltage has reached a point at which the rectified current is just sufficient to compensate for the increased discharge through resistance R_c due to the higher voltage across the condenser. The system is then in equilibrium at the new input level. If the constants of the circuit have been properly chosen, the new grid voltage E_H will be such that the middle of the positive envelope of the received signal again passes through the line AA, thereby giving unbiased signal reception.

The corresponding condition for an input level below normal is shown by curve L.

The condenser is the essential element of the compensator, and it is, of course, able to accumulate a charge in the absence of the resistance R_c; the only function of the latter being to dissipate this charge quickly when the input level drops. If there were only a resistance and no condenser there would be a simple rounding off of that part of the positive crests of the carrier waves which cause grid current to flow but no transfer of the operating axes as a whole to new bias values, such as E_L, E_N and E_H in Fig. 7.

Before leaving Fig. 7, it may be interesting to note parenthetically that a signal with large bias, such as H is much more immune to distortion due to changes in its own magnitude, or variations in relay bias, than a signal such as L. This can readily be seen if we imagine the line AA, which corresponds to the operating point of the receiving relay, to be moved up or down and note the relative lateral displacement of the intersections on these two envelopes. Another advantage of the stronger signal is that the rate of change of energy at the moment when the relay operates is much greater, insuring a more positive operation of its armature. The concave character of the detector characteristic is also favorable to securing a desirably shaped pulse for relay operation. Advantage was taken of these wave shaping possibilities in the design of detectors antedating the use of the level compensator to minimize the effect of circuit and battery variations.

In the absence of resistance R_s, Fig. 6, there would be a tendency during a long spacing interval for leakage in the wiring connected to the grid, to reduce the grid bias to ground potential. Before such discharge had gone very far, however, the receiving relay would close and recharge the condenser. This would give rise to periodic operation or *pulsing* of the relays. The purpose of R_s is to prevent this undesirable effect by making the negative bias voltage approximately equal to E_0 during long spaces. Since this resistance is large compared to R_c, it has a negligible effect during the reception of signals.

Where a circuit is exposed to transient additions of energy from external sources such as lightning, the operation of the level compensator may be stabilized by bridging a large capacitance in series with a resistance around condenser C.

Sending Circuit

In the voice-frequency telegraph system the spacing signals are produced at the transmitting end by short circuiting that portion of the circuit which supplies the sending filter with power, and the marking signals by allowing the current from the generator to flow through freely. This operation is performed by the sending relay.

Experience has shown that owing to the small a-c. voltages involved, there is a tendency for the contacts of the sending relay to increase in resistance, sometimes reaching a value as high as 1,000 ohms or more. In view of the low-impedance of the circuit originally used, this caused considerable residual current to flow during spacing intervals. In order to remedy this condition, the sending relay circuit was modified to the form shown in Fig. 8, in which an auto-transformer is so con-

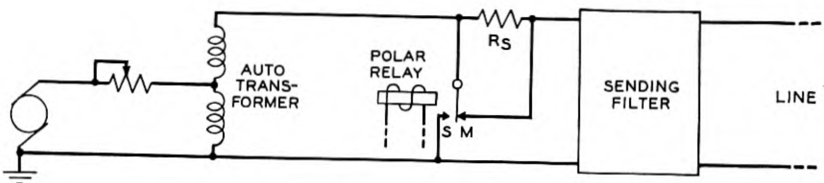


Fig. 8—Relay sending-circuit.

nected as to give a high impedance looking towards the generator, while R_s, which is of the order of 50,000 ohms, provides a correspondingly high resistance towards the output. The sending filter input is suitably padded to insure a satisfactory termination. It will be evident that with this arrangement the contact resistances in both the spacing and marking positions may vary considerably without seriously affecting the transmitting efficiency. Another advantage of R_s is that it eliminates the bias due to the transit time of the sending-relay armature, which may therefore be increased, and need not be kept within such precise limits: a matter of considerable convenience where demountable relays are used.

A trial has also been made of various schemes using varistors (copper-oxide rectifier-elements) to control the flow of carrier current by means of the changes in voltage in the loop circuit, thus dispensing with sending relays of the electromagnetic type. Figure 9*A* shows an arrangement which has been in actual operation for a number of years at several central offices and has given satisfaction. The loop circuit is provided with two equal apex resistances *RR*; hence when the key is closed the point x is positive relative to y regardless of the position of the receiving relay. This follows from the fact that the current through the loop is twice that through the loop-balancing resistance. On the other hand, if the receiving relay is on its marking contact, opening the loop key reverses the relative voltage between points x and y so that x becomes negative relative to y. In other words, polar signals are impressed between points x and y as a result of the transmission of signals in the loop. The a-c. part of the circuit contains a bridge-like arrangement between the carrier source and the sending filter, which is balanced at all times with respect to the d-c. pulses so that these do not tend to be propagated into the line or towards the generator. When x is positive relative to y, rectifier elements a_1 and a_2

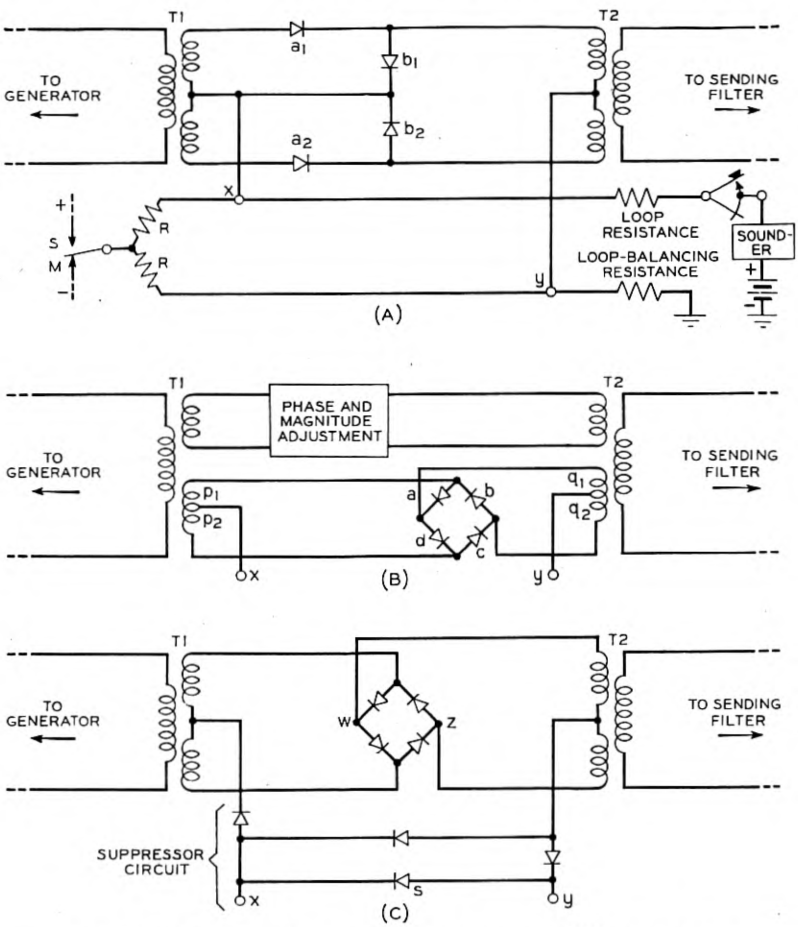


Fig. 9—Varistor sending-circuits. *A*. Series-parallel arrangement. *B*. Phase inverter. *C*. Non-polar arrangement.

are conducting while b_1 and b_2 are non-conducting. This allows a free path for the carrier between transformers T_1 and T_2 and thus from the generator to the sending filter. If, however, point x is negative relative to y, a_1 and a_2 acquire a high resistance, thereby greatly impeding the passage of current between T_1 and T_2, while b_1 and b_2 become conducting and effectively shunt the primary of T_2.

Alternative arrangements have also been tried. Two of these, which were used in actual installations, are shown in Figs. 9*B* and 9*C*. In both cases, the loop arrangement is the same as in Fig. 9*A*. The circuit of Fig. 9*B* consists of two parallel paths between generator and sending filter, one of which impresses a steady a-c. voltage on transformer T_2 while the second path serves to impress a second a-c. voltage of the same magnitude at the same point, but this latter voltage may be either in phase aiding or in phase opposing to the first, depending on the polarity of the d-c. voltage impressed through the varistor bridge. Thus it will readily be seen that if x is positive relative to y, elements a and c are conducting, while b and d are not. Transmission of the carrier then takes place around the path p_1, q_1, q_2, p_2, and the voltages from the two parallel circuits are additive in T_2. If, however, x is negative relative to y, the conducting condition of the varistor elements is reversed so that the carrier path becomes p_1, q_2, q_1, p_2, and the net voltage impressed on the primary of T_2 is zero.

The direct-transmission branch contains a phase and magnitude adjustment network to permit exact neutralization of the carrier voltage for the spacing condition.

Figure 9*C* is much like Fig. 9*B* except that the direct-transmission branch is omitted and a "suppressor circuit" is added, which may be thought of as changing the signals impressed on the varistor bridge from polar to neutral. This is done by inserting element s, which equalizes the voltage between points x and y whenever y is positive with respect to x. This effect is further enhanced by adding other series and shunt elements as shown. The bridge conditions for marking are the same as in Fig. 9*B*, while for spacing, all the elements are normal and alike so that the bridge is balanced for a-c. as well as for d-c.; thus no voltage appears between points w and z, and the carrier is suppressed.

While all these schemes involve balance between groups of varistors, recent advances in design have made it possible to fulfill this requirement to the desired extent and to maintain it over long periods of time.

The limited use to which varistor sending circuits have been put in the telegraph plant of the Bell System is not due to unsatisfactory operation in their present applications, but rather because they have imposed certain operating limitations on recently developed arrangements for interconnecting telegraph circuits.

Grid Bias

The fixed grid bias required by the level compensator exceeds the filament battery voltage, hence the latter cannot be used as a bias source and recourse is had to the negative 130-volt telegraph battery.

It follows that any variations in the latter will cause signal bias. To minimize this effect, several schemes have been used to stabilize the voltage applied to the level compensator. One of these is shown in Fig. 10*A*. In this arrangement the 130-volt battery discharges through

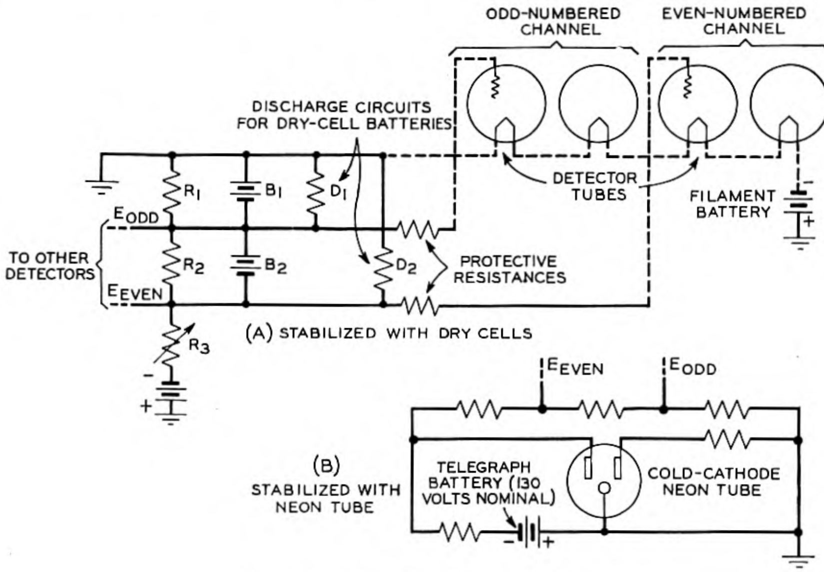


Fig. 10—Grid-bias supply circuits.

a series of resistances R_1, R_2 and R_3 so proportioned as to provide suitable taps giving the required bias when this battery is at its average voltage. Inasmuch as the filament circuits of two detectors are in series, two different voltages to ground are required. Dry cell batteries B_1 and B_2 are bridged between the taps which provide the desired biases and ground, of such values that they would give the proper voltages in the absence of the telegraph battery. These dry cells insure constant bias voltages; they supply no current when the telegraph battery is at its average value; discharge when it is low, and charge when it is high. Resistance R_3 is sufficiently large so that these charging and discharging currents are kept down to very small values and the life of the cells is consequently long. Since the bias batteries are part of a rectifying circuit there is a tendency for the signal current passing from grid to filament to charge the dry cells. To compensate for this, adjustable discharge circuits D_1 and D_2 are bridged respectively across the two grid-bias taps and ground in the manner shown, and their resistances are varied according to the number of detectors deriving their bias from this source.

A second method for securing a stable grid-bias voltage is shown in Fig. 10*B*. Here, advantage is taken of the fact that the voltage required to maintain discharge in a cold-cathode neon-tube is constant, by bridging such a tube across the negative 130-volt telegraph battery in series with a resistance. The desired voltages for the even and odd numbered detectors are then derived by tapping off at suitable points on a second resistance which is connected across the neon tube.

Interference

Interference in a particular channel may manifest itself either by the presence of current when none is intended or by a diminution of the signal current during a marking condition. The former is called *spacing interference*, and tends to change spacing units to marking units; the latter is termed *marking interference*, since it is observed during marking units, tending to change them to spaces.

The principal sources of spacing interference are:

1. Unsuppressed carrier.
2. Noise, lightning, etc.
3. Infiltration from adjacent channels.
4. Modulation products.

For marking interference, these are:

1. Crowding (Saturation effects).
2. Out-of-phase parasitic currents.

Parasitic currents (noise, modulation, etc.) are usually not an important source of marking interference, as their phase relative to that of the carrier forming the signal must fall within a rather narrow range to be effective.

If there exists some unsuppressed carrier during spacing intervals which is due to the design of the sending circuit, it will be fixed in value and may therefore be taken care of in the initial adjustments of the receiving circuit. All the other effects are of a chance character, being for the most part dependent on the transmission circumstances on associated channels or circuits. These effects, therefore, lead to fortuitous distortion.

The effectiveness of all forms of interference is dependent upon the ratio of their magnitudes to that of the signals with which they interfere. On the other hand, the absolute magnitude of the greater part of this interference depends upon the signal level. To establish a balance between these two tendencies, the telegraph power per channel which is transmitted over the circuit is selected so as to minimize the effects of interference on telegraph signals and at the same time cause as little disturbance as possible to associated telegraph or telephone circuits. In the Bell System, the four-wire cable circuits used for telegraph purposes are for the most part devoted to this use exclusively and are organized with terminal repeater gains adapted for this special service. In the case of open-wire carrier-telephone circuits, on the other hand, the overall gain from modulator input to demodulator output is fixed by the telephone requirements, and the telegraph must be adapted thereto. The power per telegraph channel in dbm.* now used on cable and open-wire circuits is shown in Fig. 1 at various points.

The effect of changing the power on the line is illustrated qualitatively in Fig. 11, in which the variations of the received current with increasing transmitted current are sketched diagrammatically for various operating circumstances. By nominal power (*a*), is meant the power which would be received if transmission took place over a linear network having a fixed gain equal to the nominal gain of the circuit. Owing principally to the reduction of repeater gain which takes place with increasing load, the current actually received with all channels marking, is less than this (*c*). If only one channel is marking, some intermediate values (*b*) will be obtained of course, while if, as in the case of regular operation, some of the channels are spacing and others marking, still other values (c') will result. This saturation effect is sometimes called "crowding."

One of the contributions to spacing interference consists of ambient noise due to the combined crosstalk from all the other circuits in the cable and to external induction. This current is represented by curve *e*, which is shown as independent of the power transmitted; this corresponds to the situation existing where telegraph is a small part of the total traffic in the cable under consideration, for evidently if the power were increased on all the circuits the noise power would increase almost proportionally.

A more serious source of spacing interference consists of parasitic currents due to third-order modulation products arising directly or indirectly from the interaction of the several channels of the same system when passing through the non-linear elements of the circuit. Second-order modulation products are taken care of quite effectively by the receiving filters, due to the fact that the carriers are odd harmonics of 85 cycles, while these products being even harmonics thereof fall midway between channel frequencies. Since the attenuation of the receiving filters in the frequency range occupied by neighboring

* The symbol dbm. as used in this paper may be read "db referred to 1 milliwatt." It is intended to denote the ratio expressed in db of the power under consideration to 1 milliwatt; e.g., − 6 dbm. = .2512 milliwatt.

channels is finite, some infiltration of undesired frequencies takes place. Thus if a given channel is spacing and the adjacent ones marking, some small fraction of the flanking carriers will find their way into the idle

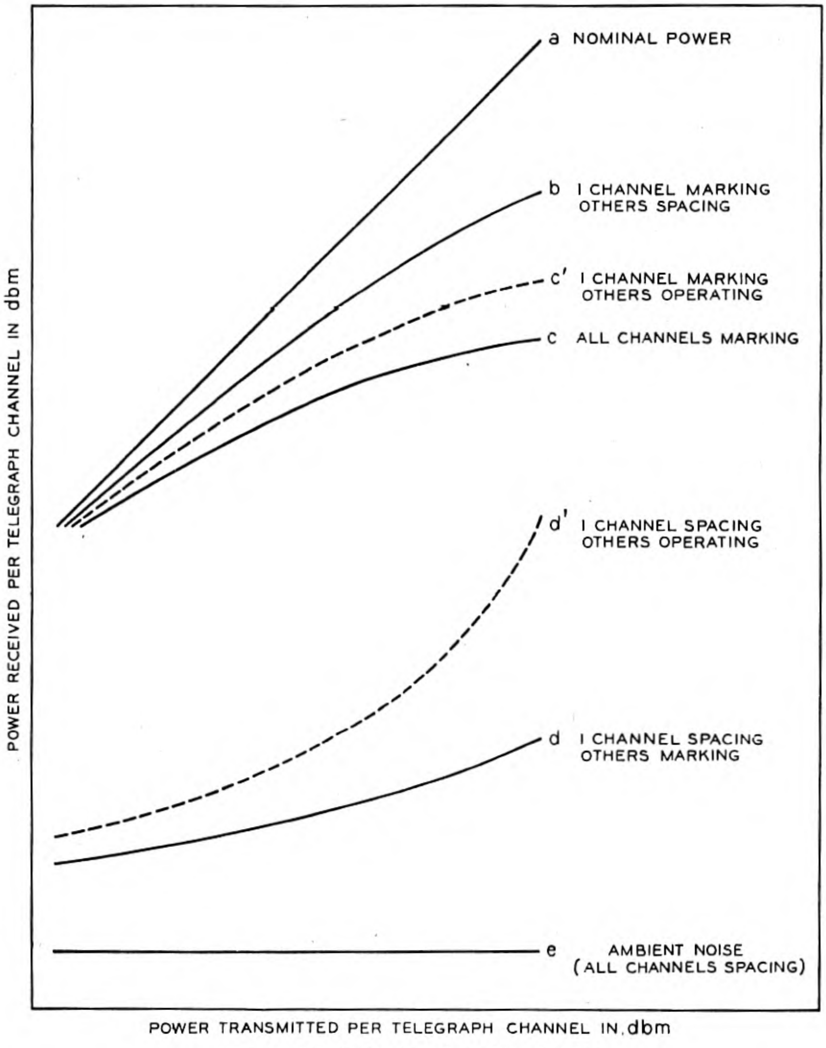


Fig. 11—Interference in cable circuits.

channel. Furthermore, if a channel is in operation, the sending filter does not completely suppress the sideband components lying outside the band assigned to it, and these pass freely through the receiving filter of nearby channels. Unlike modulation, the effectiveness of filter action does not depend, to any large extent, upon the amount of power per channel, since the ratio of the disturbing currents to the signal remains approximately constant. Finally, such residual carrier as exists during spacing intervals may be variable in amount due to changes in the resistance of the sending relay contacts. If the power levels are increased sufficiently the total spacing interference will usually become due preponderantly to the modulation effects and hence a function of total power on the line, as indicated in curves d and d'.

In order to estimate the quality of transmission to be expected from circuits in view of these various interfering factors, it is desirable to establish the following two definitions:

> The signal-to-interference ratio * of a circuit is the ratio, expressed in db, of the normal marking current plus the interference, to the interference alone.
>
> The marking interference is the ratio, expressed in db, of the normal marking current alone, to the marking current plus the interference.

A variety of results may be obtained under these definitions depending upon the methods of observation: it is customary, therefore, to adopt the following practical specifications:

> *Signal-to-interference Ratio:* The change in sensitivity, expressed in db, required in the receiving circuit of a given channel, all other channels being in a marking condition, to just cause the armature of the receiving relay to go to its marking contact; first, with steady marking current transmitted over the channel under test, and second, with the channel under test opened at the sending end. It is understood that the currents are turned on and off by operating the sending relays and moreover that the receiving relay operates on one half the steady marking current. (E.g., no interference $= \infty$ db; complete failure $= 6$ db, approximately.)
>
> *Marking Interference:* The change in sensitivity, expressed in db, required in the receiving circuit of a given channel to just cause the armature of the receiving relay to go to its spacing contact; first, with steady marking current transmitted over the channel under test only, and second, with the interference added thereto. (E.g., no interference $= 0$ db; complete failure $= 6$ db, approximately.)

While the above method for measuring spacing interference is the one used in practice owing to the ease with which it can be applied,

* More precisely the signal to spacing-interference ratio.

a more significant characteristic is the signal-to-interference ratio obtained by observing the aforesaid change in receiving-circuit sensitivity required to operate the receiving relay in a given channel, as we go from a marking to a spacing condition in that channel with all other channels transmitting uncoordinated signals. This is not only a more practical consideration but a more severe condition, as indicated by curve d' in Fig. 11.

Carrier telegraphy, as here considered, is a marginal system of operation in which the current received for a marking condition corresponds on the average to that shown by curve c', while that for a spacing condition is the one represented by d'. The difference between these two characteristics is not all available for operation, however, since the receiving circuit must be made sufficiently sensitive to operate when the marking current has risen to half its final value, thus bringing the threshold of operation 6 db closer to the spacing interference. From these considerations it follows that the actual *operating margin* is an essentially variable quantity whose approximate value is less by about 6 db than the signal-to-interference ratio measured as specified above, since the indicated procedure takes account of marking as well as of spacing interference effects.

The following definitions are also useful in reporting and interpreting test results:

The *interference margins* of a circuit are the ratios, expressed in db, of the actual interference and the amount of this interference which will cause failure. More specifically:

(*a*) The spacing-interference margin is the increased sensitivity, expressed in db, required in the receiving circuit of a given channel adjusted to receive unbiased signals in the absence of interference, to just cause the receiving relay to close when interference alone is present (e.g., no interference = ∞ db; complete failure = 0 db).

(*b*) The marking-interference margin is the decrease in sensitivity, expressed in db, required in the receiving circuit of a given channel adjusted to receive unbiased signals in the absence of interference, to just cause the receiving relay to open when interference is added (e.g., no interference = 6 db, approximately; complete failure = 0 db).

Clearly the various effects which have been defined above, being of a variable and indeterminate character, contribute to the fortuitous distortion of signals. This is illustrated in Fig. 12, which shows what may happen to a single dot. The extreme conditions of operation on the particular channel under consideration are those where all of the channels are marking or spacing. If they are all spacing, there is only noise and unsuppressed carrier, and the change in received current

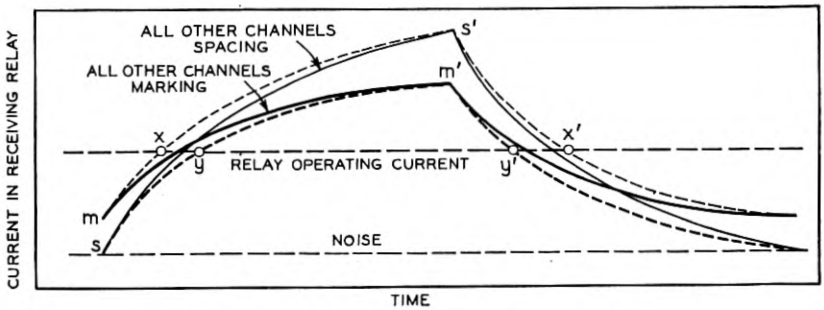


Fig. 12—Effect of interference on signal distortion.

when the one operating channel changes from space to mark is the vertical distance between points s and s', while when all the other channels are marking the change is from m to m'. The ratio of the currents corresponding to either of these pairs of points, when expressed in db, may be conveniently called the *marking-to-spacing ratio*. Practically, of course, the power over the line will be constantly and fortuitously varying, so that the actual arrival curves will lie somewhere between certain limiting values indicated by the dotted lines ms' and sm', giving rise to a range of fortuitous distortions which, if the receiving equipment was adjusted with all the other channels marking, will depend on the length of the extreme intervals xx' and yy'.

The results obtained in the course of experimental observations of some of the above quantities are given below. They were secured on a 700-mile H44 cable circuit of the type described at the beginning of this paper. In order to obtain uniform results, the output of each of the 17 repeaters in tandem in this circuit was adjusted to the same level, so that each output tube contributed about equally to the total modulation effects and a similar uniformity existed relative to the most heavily energized loading coils. In practice, the saturation effects would not be likely to exceed this, and generally would be somewhat less.

Figure 13*A* shows typically the effect on the received current of increasing the total power transmitted over the line when all channels are marking simultaneously. This phenomenon (which may be called crowding) is a measure of the intermodulation which is caused by the circuit during normal operation, since the power transmitted over the circuit then varies fortuitously over a range of 10.8 db for a 12-channel circuit and 13.8 db for a 24-channel system, depending on the number of channels which happen to be marking simultaneously.

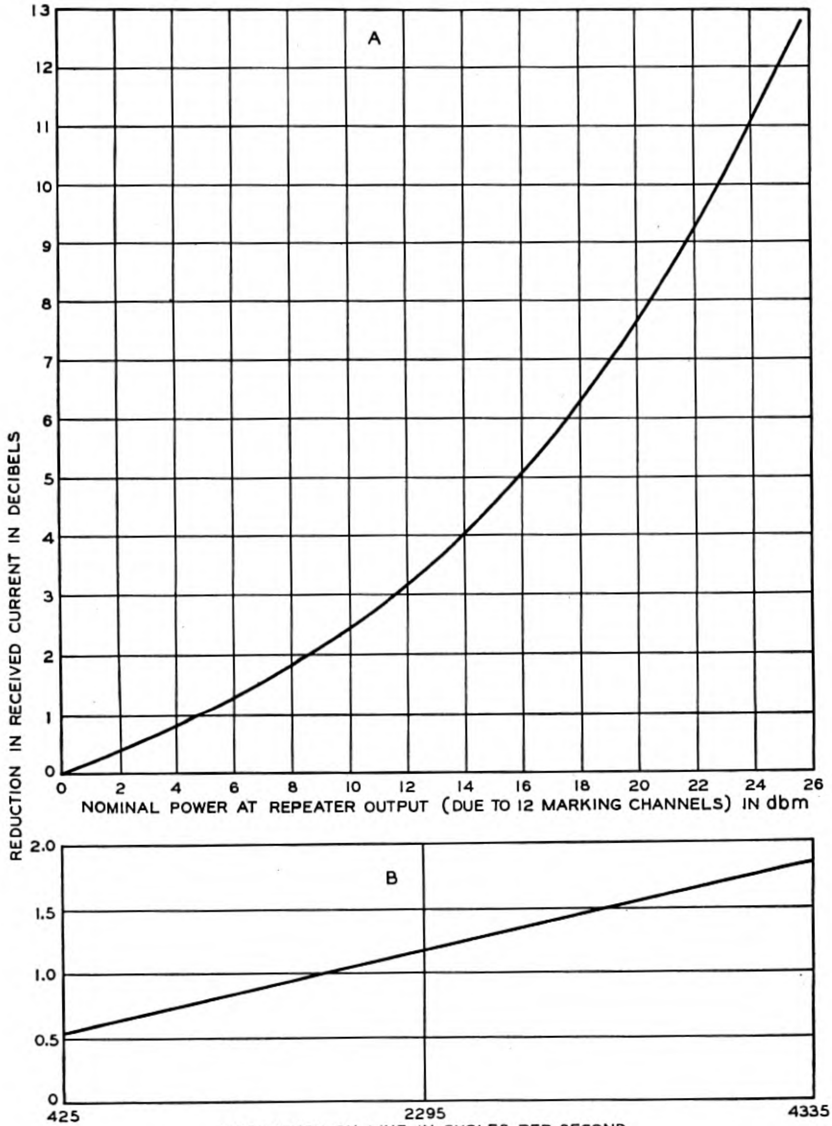


Fig. 13—Crowding. H44 circuits. *A*. Relative crowding vs. power. 12 channel system. Channel No. 5 (1105 cycles per second). *B*. Crowding vs. frequency. 24 channel system. Total power at repeater output increased from 2.2 to 9.2 dbm.

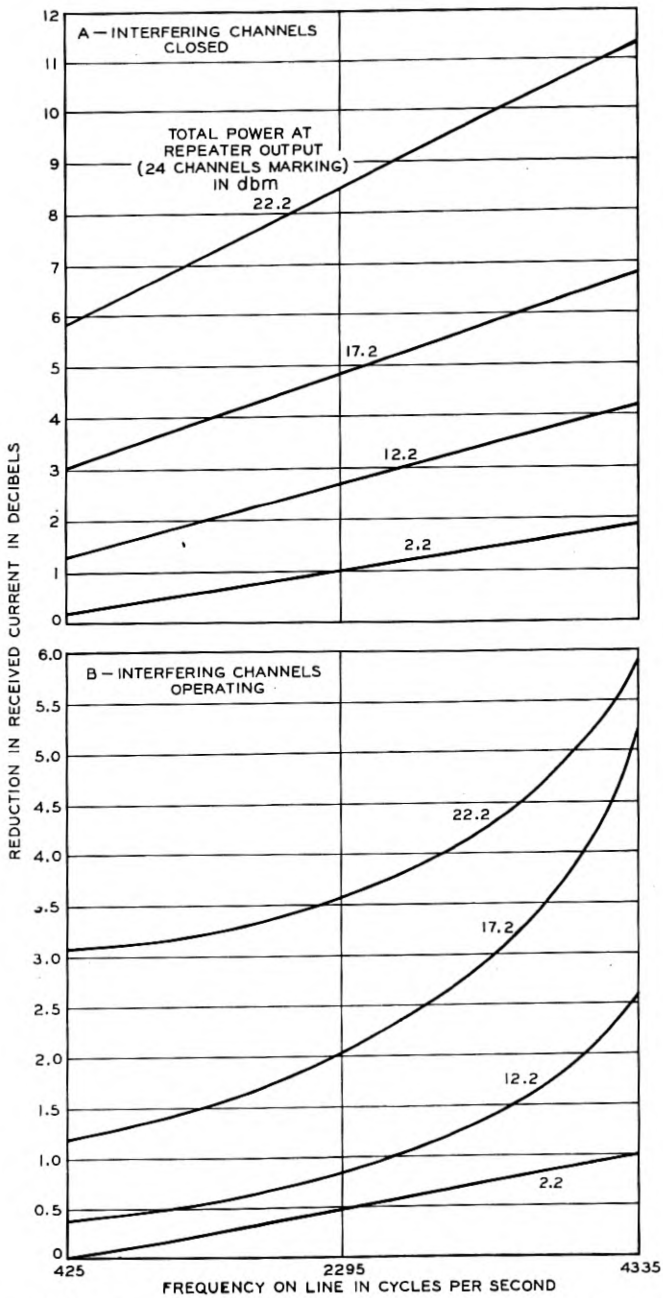


Fig. 14—Marking interference vs. frequency. H44 circuits. 24-channel system.

As will be seen by reference to Fig. 13*B* there is a marked frequency effect; the variations becoming greater for the higher frequency channels. This is particularly noticeable where there is a large number of channels operating simultaneously, as otherwise individual variations between channels tend to obscure the gradual trend.

Figure 14*A* shows the marking interference, expressed in db, due to changing from one channel marking to all channels marking for various changes in repeater output ranging from 2.2 to 22.2 dbm. and a frequency range extending from 425 to 4335 cycles.

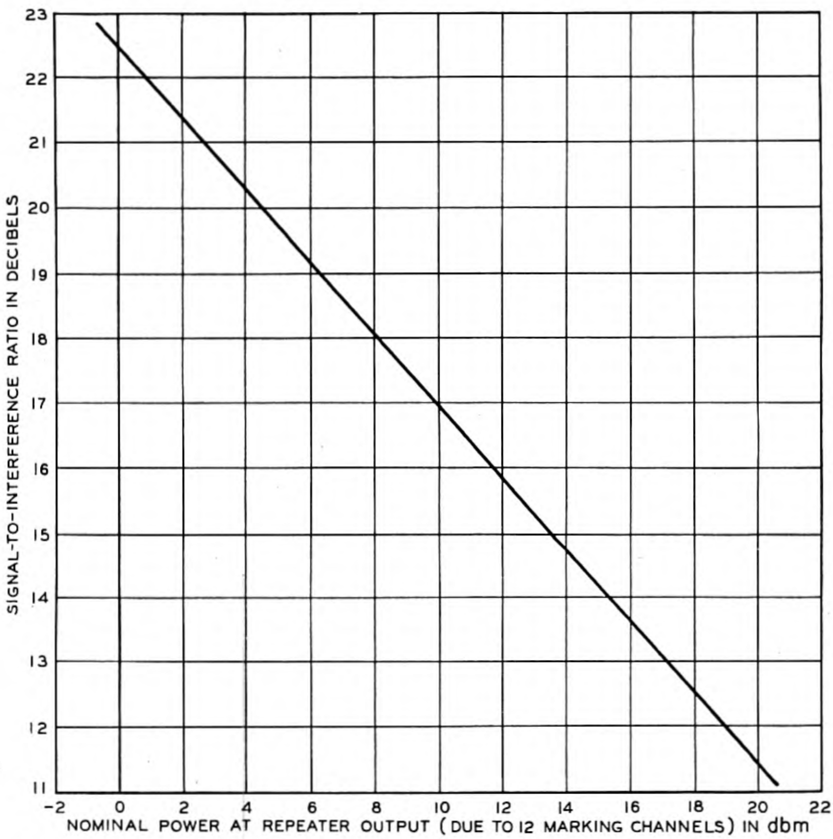


Fig. 15—Spacing interference. H44 circuits. Interfering channels operating. 12 channel system. Channel No. 5 (1105 cycles per second).

Figure 14*B* shows the reduction in received current which results on any given channel over the same range of conditions as Fig. 14*A* except that in this case the associated channels are transmitting uncoordinated reversals.

These interfering effects are subject to considerable variations from channel to channel in an irregular manner depending on repeater spacing, phase relations between the carrier sources, etc., so that the characteristics given in Figs. 13 and 14 must be interpreted as indicating average values and trends rather than specific amounts.

The erratic character of the results due to such fortuitous circumstances is particularly noticeable in spacing interference measurements made as a function of channel frequency, or in those relating to parasitic currents produced by steady currents in the remaining channels of the system. The general effect of repeater load on spacing interference is illustrated for a particular channel in Fig. 15. This refers to the case where the interference is caused by non-synchronized signals on the remaining channels.

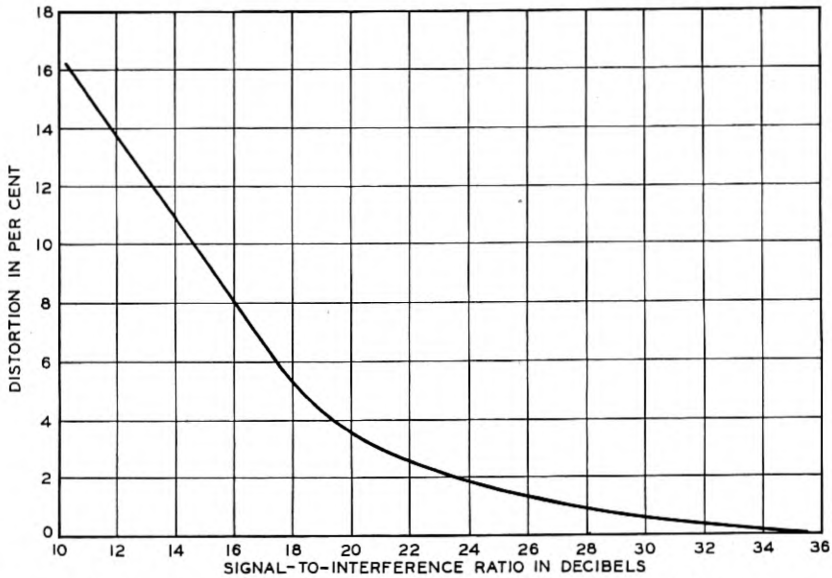


Fig. 16—Increase in signal distortion caused by spacing interference.

As far as the toll line itself is concerned, noise from other circuits is an almost negligible factor in the distortion of signals. It is of a highly fortuitous character and shows no definite trend with frequency except perhaps as it may be somewhat greater in the effective voice range, as modified, however, by variations in cross-talk efficiency with frequency.

The effect of spacing interference on telegraph distortion is given in Fig. 16, which shows that a residual current as small as 30 db below signaling begins to degrade transmission.

Operation Over Carrier Telephone Circuits

Where v-f. telegraph operates through the same repeaters as one or more telephone circuits, as in the case of carrier-telephone systems, a new form of variable interference arises due to the changing load conditions introduced by variations in voice volume. Where there are comparatively few telephone channels involved, as in the type C carrier-telephone system for instance, there is very little averaging and the voice peaks determine the repeater load which is effective in

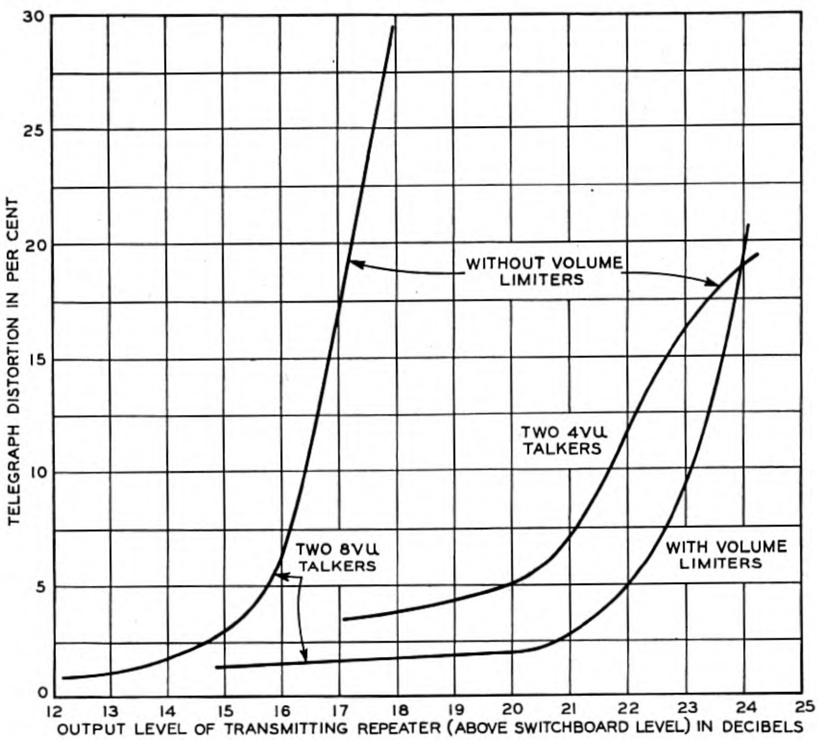


Fig. 17—Voice-frequency telegraph operated over an open-wire carrier-telephone circuit. Effect of associated telephone channels on telegraph distortion.

causing interference to telegraph. This interference is due in part to changes in net-loss of the circuit resulting from the non-linearity of amplification and in part to intermodulation between the various currents passing through the repeaters simultaneously. Figure 17 gives an example of the distortion produced in a telegraph circuit operating over one channel of a 3-channel telephone system when the other two are occupied by two equal-volume talkers. The marked reduction in permissible repeater amplification as we change from two 4-vu to two 8-vu talkers is evident. This figure also shows that by the use of the volume limiters mentioned at the beginning of the paper, the repeater gains with the higher volume talkers may be made at least as great as for the lower volume. This makes it possible to use the present telephone circuits for carrier telegraph purposes without change.

Drainage

The occurrence of atmospheric disturbances, and particularly lightning, constitutes a potential hazard to the operation on open-wire circuits of a service as exacting as carrier telegraph. Where such circuits have been transposed for the operation of carrier telephone, the transverse or metallic-circuit effects due to lightning discharges in the neighborhood of the line are in general not serious, but the voltages generated to ground are very often of sufficient magnitude to cause a breakdown of the protectors. Since it has been found impracticable to devise protectors with such precise limits that they will operate at the same voltage and possess the same discharge characteristics, a transient transverse current is set up in such cases which may cause telegraph errors.

The remedy adopted consists in bridging drainage coils [24] across the line at all points where protectors are required, and in so connecting them that they will either prevent a breakdown of the protectors or assure simultaneous operation with equal discharge currents from either wire to ground.

In Fig. 18*A* the drainage coil is shown bridged directly across each end of the open-wire line between two sections of entrance cable. These coils consist of two carefully balanced windings with the mid-point grounded. They present a high impedance to voice or carrier currents transversely, but offer only a small resistance to ground for longitudinal currents compared with that across the adjacent protector blocks. The chief disadvantage of this method, which is called "direct drainage," is that it prevents the use of grounded telegraph and interferes with the testing of the line by means of direct current. To obviate this, the scheme shown in Fig. 18-*B* has been devised, which is termed "protector drainage." In this case, the drainage coil is connected to the line wires through protectors having a low breakdown-voltage. This combination is backed by high-voltage protectors to insure unimpeded discharge in case of large disturbances. With this arrangement the drainage coil does not come into operation unless there is a severe disturbance, and furthermore owing to the mutual inductance between the two halves of its windings it tends to cause such discharge to occur at the same moment and be of equal magnitude from both wires of the pair to ground whenever it does operate.

Extensive tests and practical experience have shown that the above arrangements are quite effective, affording a reduction of about 95

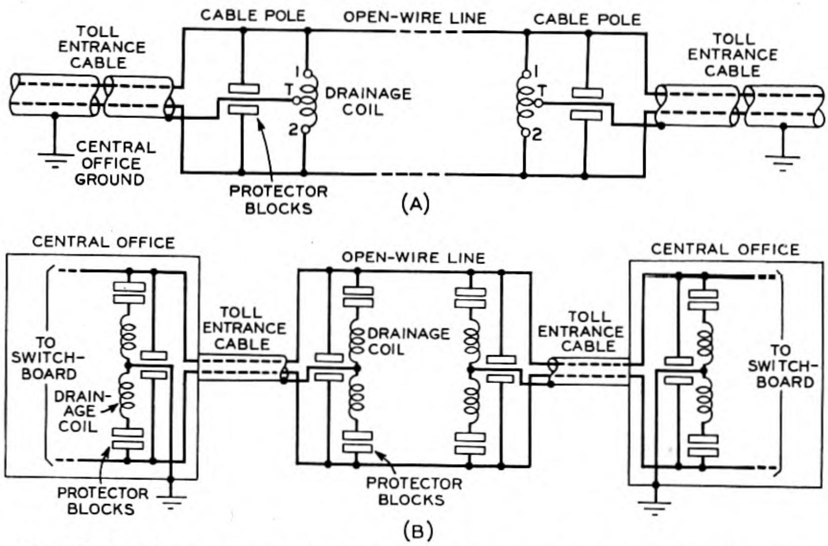


Fig. 18—Drainage arrangements. *A*. Direct drainage. *B*. Protector drainage.

per cent in the number of disturbances occurring on the line as well as in the number of errors in transmission resulting therefrom.

Carrier Supply

The inductor-alternator source for carrier frequencies used in the original installation has been retained except for the addition of two channel frequencies, and the substitution of improved means of speed (frequency) control. The first mechanical governor was soon superseded by a center-contact device which is much less erratic in operation.

The carrier frequency can be shifted over quite a few cycles from its correct value without affecting signal distortion appreciably but rapid speed variations have a more serious effect. Two methods have been used to overcome this difficulty. The first consists effectively in making the generator part of a system having great overall inertia, the second in subjecting it to very rigid control.

To secure the first end the VF generator is driven by a synchronous motor operated from the lighting circuit. This arrangement can, of course, be used only where the frequency of the commercial power is regulated within narrow limits. The required generator speed being

1700 r.p.m. a belt drive is used to reduce the motor speed of 1800 r.p.m. to the proper value; small adjustments in speed being provided by means of a *V* pulley having an adjustable diameter.

In order to insure continuity of operation in case of failure of the power supply the generator is also arranged to be driven by a d-c. motor equipped with a mechanical governor. This motor is supplied from the central office battery and is automatically switched thereto in case the voltage of the commercial power supply falls below a predetermined value.

In the second method the VF carrier-supply unit is governed by means of a vacuum-tube circuit whose output is controlled by frequency variations in the highest frequency channel, and the resulting d-c. current is applied to the motor field. The operation of this device, which has been described elsewhere,[25] may be briefly explained as

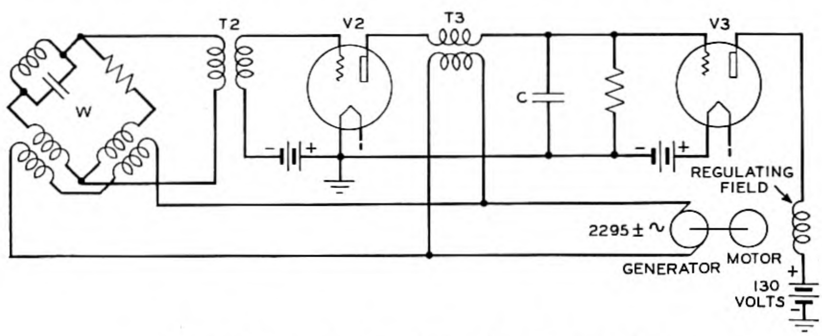


Fig. 19—Principle of tuned-circuit speed-regulator.

follows: It depends for accomplishing its purpose on producing variations in the strength of the current through an auxiliary regulating field associated with the motor which drives the multi-frequency generator. The essential features are shown in Fig. 19. The action is as follows: The voltage of channel 12 (2295 cycles) of the generator whose speed is to be controlled is applied simultaneously through a divided circuit to the input and output of tube V_2 which may be called the phase-detector tube. The plate voltage is applied directly through transformer T_3, there being no *B* battery in the ordinary sense. The grid voltage on the other hand is applied through the bridge *W*, one of the arms of which is an anti-resonant circuit tuned to 2295 cycles. The result is that the magnitude of the grid-filament voltage and its phase relative to the plate voltage are dependent on frequency. When the latter has its correct value the anti-resonant arm exactly balances the bridge, and the a-c. voltage across T_2 is nil. At higher frequencies this circuit acts like a capacitance, while at lower frequencies it acts like an inductance. There is thus a rapid change in both the voltage and in the phase thereof across T_2 whenever there is a variation in frequency on either side of the specified value. The output current from V_2, combining with the current impressed directly through T_3, produces corresponding abrupt changes in the d-c. component of the resultant which in turn varies the bias of tube V_3 and hence the current through the regulating field of the motor.

In order to assure a rapid change in impedance with frequency, the anti-resonant circuit referred to above comprises a carefully shielded air-core coil having a very small resistance relative to its inductance.

The frequency indicator shown in Fig. 20*A* provides means for easily

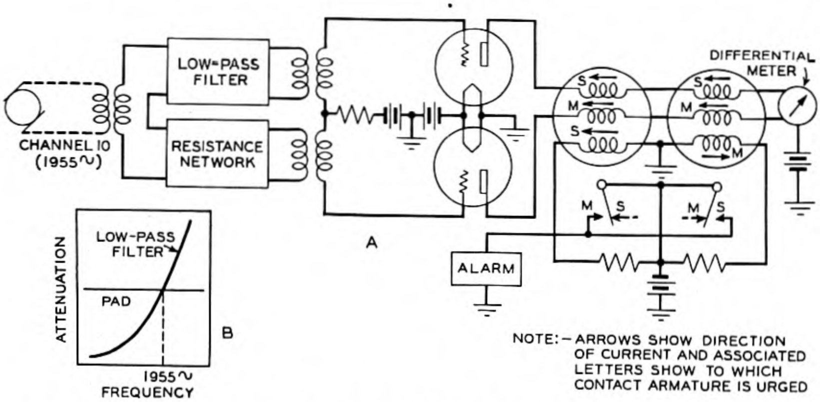


Fig. 20—Frequency indicator. *A*. Schematic diagram. *B*. Attenuation of two input-paths.

observing any departures of the carrier frequency from its nominal value, as well as an automatic maximum-minimum alarm to warn the attendant if these variations become excessive. Since all the carrier currents are derived from generator elements which are mounted on a common shaft, it is sufficient to observe the frequency of a single channel. Current from channel 10 (1955 cycles) is impressed simultaneously on two vacuum tube circuits which are identical except for the fact that there is a low-pass filter in the input of one while there is a simple pad in the input of the other. As indicated in Fig. 20-*B*, the loss through the pad is the same at all frequencies and equal to that of the filter when the generator speed is correct. If the frequency increases, the attenuation of the filter branch goes up; while if it decreases, its attenuation goes down; but in any case the loss through the pad remains constant, of course. The net resulting ampere-turns tend to move one or the other relay armature to the opposite contact according as the frequency is high or low, and thereby ring the alarm. This resulting current also indicates the amount by which the frequency departs from its nominal value.

The frequency indicator may be used with either the mechanical governor or with the synchronous drive, but it is of little use with the vacuum tube tuned circuit control as the latter is too precise to register any indications.

The frequency indicator does not permit a very close adjustment of the mechanical governor nor does it provide a satisfactory check for the correctness of the frequency of the commercial power when a synchronous motor is used, so a stroboscopic method has been adopted as an ultimate standard of comparison. This stroboscope consists of a cylindrical target made up of three distinct peripheral rows of alternate black and white segments mounted on the end of the generator shaft. These segments may be viewed by means of a tuning fork fitted with overlapping metal plates attached to the ends of the tines. Slits cut in these plates lie opposite each other when the fork is at rest. When it is set vibrating, vision through the slits can therefore be established momentarily twice during each complete oscillation. By illuminating the target with a steady source of light the apparent direction of motion of the dots can thus be observed by looking through the slits. The middle row of segments on the target is so proportioned as to appear stationary if the speed is practically correct while the outer rows appear respectively stationary if the speed is approximately 1 per cent above or below the nominal value.

For offices where the frequency of the commercial supply is sufficiently stable, an additional and somewhat more convenient method for checking the speed has been made available. It consists of a special target mounted on the generator shaft, which is illuminated by a neon lamp associated with a wave-shaping circuit which makes the flashing time a very brief portion of each pulse of the 60-cycle current. In other words, the attendant in this case looks at the target constantly under intermittent illumination, while in the former case he views it intermittently under constant illumination.

Testing Facilities

While developments in carrier telegraph equipment have resulted in considerable economies, the ever increasing demand for service which is freer from errors and interruptions and is adaptable to circuits of greater length and complexity has tended to render the maintenance problem increasingly difficult and time consuming. Voice-frequency

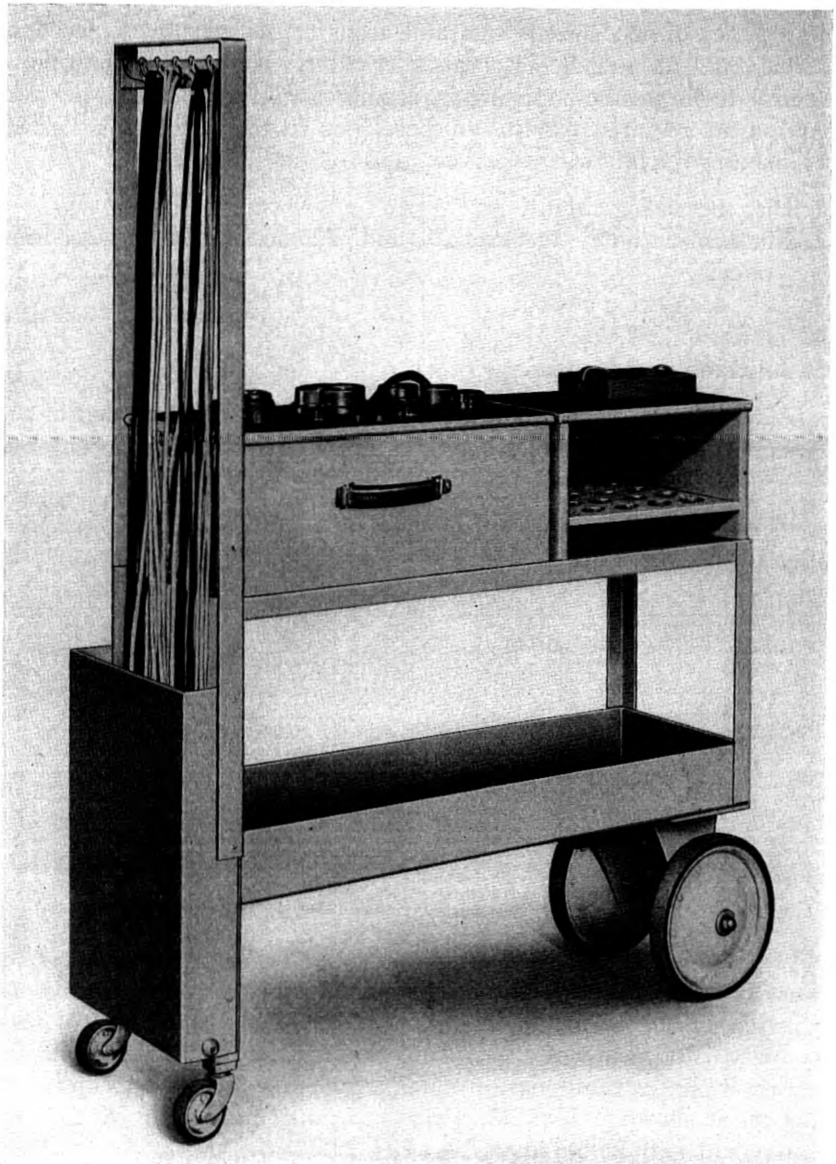


Fig. 21—Carrier-telegraph test set. General view.

telegraph circuits equipped with level compensators must be subjected to a series of specialized tests and adjustments whenever placed in service and at periodic intervals thereafter. To provide for this, a special testing set comprising in readily available form all the test equipment required for this purpose, has recently been introduced. It includes the following features:

1. Bias measuring circuit.
2. Filament-activity test-circuit and filament-current measuring circuit.
3. Drift measuring circuit.
4. Test amplifier.
5. Adjustable attenuator.

With this set all terminal tests may be made from one end of the circuit without the use of a line or external line-simulating repeater as

Fig. 22—Carrier-telegraph test set. Instrument panel.

was done in the past. It is mounted on a small wagon which may be wheeled into position adjacent to the terminals to be tested, then connected by means of a long cord and multiple contact plug to the necessary battery supplies, grounds, etc. The general appearance of the set is shown in Fig. 21, while a more detailed view of the face equipment may be obtained from Fig. 22.

The bias measuring circuit is shown in simplified form in Fig. 23. It uses a 215 type polar relay,[26] or its equivalent, with a meter connected in the armature circuit. This meter, which permits measuring to an accuracy better than 1 per cent, is specially designed to have the proper ballistic and damping characteristics to permit measuring dot signals having a rate of 11 d.p.s. The reason for providing 11-cycle reversals is that experience has shown that these more nearly simulate miscellaneous teletypewriter signals, both with respect to bias and drift, than the higher speed reversals used in the past. An end scale

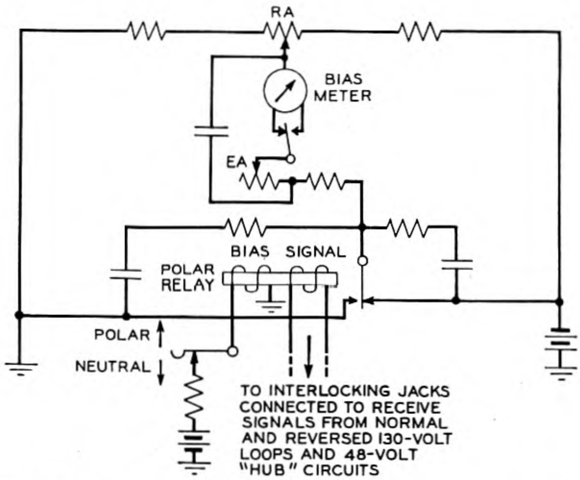


Fig. 23—Carrier-telegraph test set. Bias measuring circuit.

adjustment *EA* permits correction for battery variations, while a second adjustment *RA* allows for correcting any residual bias which may be present in the 215 type relay or in the dot signals used for the tests.

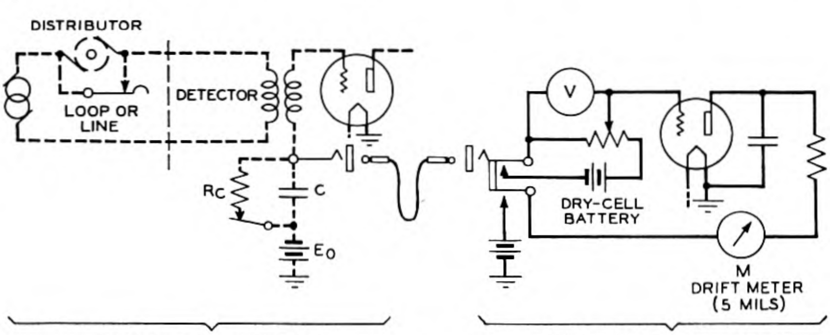


Fig. 24—Carrier-telegraph test set. Drift measuring circuit.

In order to observe and correct for the drift effect discussed previously, a special circuit is provided which is illustrated in Fig. 24. The skeletonized diagram of the carrier-terminal circuit to which it is applied, shown to the left in dotted lines, will help to understand the function of this measuring device. The object is to observe the change in voltage experienced by the upper plate of the level compensator condenser when incoming signals are changed from steady marking to dots. The circuit provided for this purpose is essentially a vacuum-tube electrostatic voltmeter of the conventional type.

Acknowledgment

The advances in the art which have been described in the foregoing pages have taken place over a number of years and cover a considerable variety of subjects. Much of the material is to be found only in test reports and unpublished memoranda. Since such work is, of necessity, the product of the cooperation of many minds, it is impracticable in most cases to apportion credit equitably and the author must therefore confine himself to a general expression of indebtedness to his associates.

APPENDIX

LEVEL COMPENSATOR THEORY

The theory of the level compensator may best be considered in two stages, first taking the steady marking condition, and second the condition where signals are being received.

Steady Marking Condition

Figure 25*A* shows the essential elements involved when the steady current I_0 flows through the receiving relay. Figure 25*B* is the

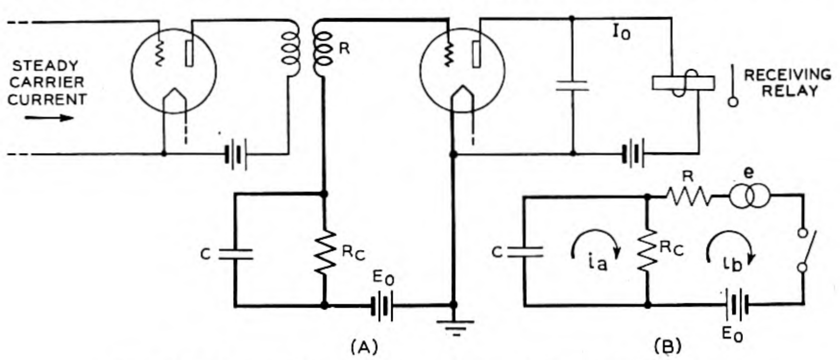


Fig. 25—Theory of level compensator. *A*. Simplified circuit. *B*. Equivalent circuit.

equivalent of the grid circuit shown in heavy lines on Fig. 25*A*, on the assumption that the grid-filament space is effectively a switch which turns the grid current on or off when the grid goes positive or negative, respectively, and that the interstage transformer is perfect. The tube resistance which is effective during the conducting period may be included in R and considered constant, as it is relatively small compared to the other resistances involved.

In describing the principle of action of the level compensator, it has been pointed out that for proper compensation the point of relay operation AA (Fig. 7) should bisect the crest value of the signal pulse in each case. In this figure, the three envelopes shown correspond to very short signal pulses; if the signals were sufficiently long so that a steady marking condition were reached, as is usually the case, the crest of the signal would coincide with the peaks of the steady carrier wave, and AA would bisect its positive loops.

For long signals the condition for proper compensation is therefore that the relay will just operate at one-half the steady-state carrier voltage. This is true for all carrier voltages E which equal or exceed the value required to make the grid positive. In the particular case where E just fails to cause the condenser to charge, $E = -E_0$ and the relay operates when a value $-E_0/2$ is reached. For any other greater value of E it is necessary in addition to overcome the voltage due to the charge on the condenser before the relay will operate. The criterion for perfect compensation where signals are of sufficient duration so that the steady state is reached is therefore:

$$\frac{E}{2} + \frac{E_0}{2} - e_c = 0, \tag{1}$$

where E is the maximum value of the instantaneous steady state carrier voltage, and E, E_0 are arbitrarily taken with such polarities as to urge the mesh current i_a in the direction indicated in Fig. 25*B*, while e_c is negative because it opposes the current i_a which gives rise to it.

Substituting $e_c = Q/C$ in (1) we have

$$\frac{E}{2} = -\left(\frac{E_0}{2} - \frac{Q}{C}\right), \tag{1a}$$

where Q is the average charge existing on the condenser during a long mark for any particular value of E. The problem of compensation then resolves itself in adjusting the constants C, R, R_c, and E_0 so that relation (1a) will hold; in other words an expression for Q must be found in terms of these parameters. The problem is not susceptible of explicit solution, but expressions will be derived which are believed to clarify the operation of the compensator and permit computation.

Referring to Fig. 26 let A represent the positive halves of the open circuit voltage across the secondary of the interstage transformer due to the steadily impressed carrier; and B represent the instantaneous

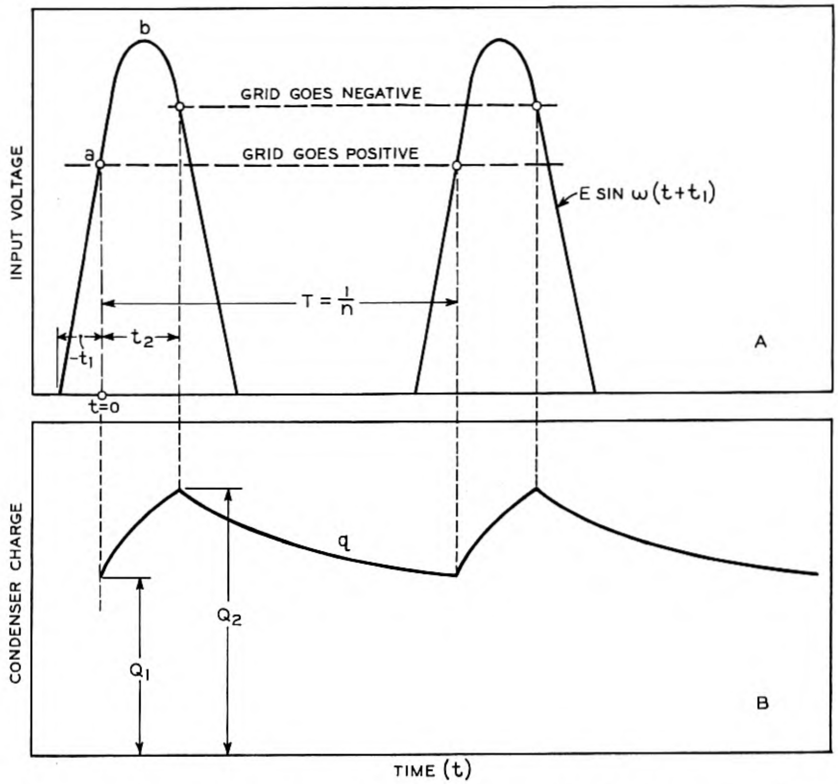


Fig. 26—Theory of level compensator. *A*. Carrier voltage. *B*. Condenser charge.

charge on the condenser. Reckoning time from the instant a when the grid goes positive, we have:

$$e = E \sin \omega(t + t_1), \tag{2}$$

where $\omega/2\pi$ is the carrier frequency.

Consider the instant at which $t = -t_1$: e is increasing through its zero value, the switch representing the grid-filament space (Fig. 25B) is open and the condenser is discharging through R_c. This state of affairs continues until e reaches a value which nullifies the condenser voltage plus the steady grid bias so that the voltage across the switch is zero. At $t = 0$, the switch closes and current i_b starts to flow. If Q_1 is the charge on the condenser at $t = 0$,

$$\frac{Q_1}{C} = E \sin \omega t_1 + E_0. \tag{3}$$

The switch remains closed while e increases to its crest value at b and then decreases until, at time $t = t_2$, the grid goes negative. If Q_2 is the charge on the condenser at $t = t_2$

$$\frac{Q_2}{C} = E \sin \omega(t_1 + t_2) + E_0. \tag{4}$$

The switch now opens, and the condenser discharges through R_c until the switch closes once more at $t = T$ to begin another cycle, T being the carrier period $2\pi/\omega$. During this interval the charge on the condenser is given by

$$q = Q_2 \epsilon^{-(t-t_2)/CR_c}. \tag{5}$$

In order that steady-state conditions may prevail, that is, in order that every cycle be the same as its predecessor, q must again equal Q_1 when $t = T$. Hence

$$Q_1 = Q_2 \epsilon^{-(T-t_2)/CR_c}$$

or, setting

$$D = \epsilon^{-(T-t_2)/CR_c}$$

$$Q_1 = Q_2 D. \tag{6}$$

In connection with equation (1a) it has been pointed out that what is sought is an expression for the average charge Q in terms of the parameters. Referring to Fig. 26B, it will be seen that this average lies somewhere between Q_1 and Q_2, but since Q_1 is very large compared with $Q_2 - Q_1$, we may take either Q_1 or Q_2 as equal to the average charge Q in (1a), when considering the effect of bias on the grid of the detector tube in producing compensator action. For the sake of definiteness, let:

$$\frac{E}{2} = -\left(\frac{E_0}{2} - \frac{Q_1}{C}\right). \tag{1b}$$

Equations 3, 4 and 6 contain the five unknowns E, Q_1, Q_2, t_1, t_2; hence one more relation must be established before we can get the desired formulation between E and Q_1. The circuit equations for Fig. 25B when the switch is closed (charging period) furnish the required relation. Thus:

$$\frac{q}{C} + R_c(i_a - i_b) = 0$$

$$(R + R_c)i_b - R_c i_a = E \sin \omega(t + t_1) + E_0.$$

Eliminating i_b, setting $i_a = \frac{dq}{dt}$ and $\alpha = (R + R_c)/RR_cC$ we have

$$\frac{dq}{dt} + \alpha q = \frac{E}{R}\sin\omega(t + t_1) + \frac{E_0}{R},$$

whence

$$q = EPF(t + t_1) + \frac{E_0}{\alpha R} + A\epsilon^{-\alpha t}, \tag{7}$$

where A is the constant of integration and

$$P = 1/R(\alpha^2 + \omega^2), \qquad F(x) = \alpha\sin\omega x - \omega\cos\omega x.$$

Since q obeys (7) from $t = 0$ until $t = t_2$, we have for the initial and final charges during the charging portion of the cycle:

$$Q_1 = EPF(t_1) + \frac{E_0}{\alpha R} + A$$

$$Q_2 = EPF(t_1 + t_2) + \frac{E_0}{\alpha R} + A\epsilon^{-\alpha t_2}.$$

Eliminating A

$$Q_1 - Q_2\epsilon^{\alpha t_2} = EP\left[F(t_1) - \epsilon^{\alpha t_2}F(t_1 + t_2)\right] + \frac{E_0}{\alpha R}\left[1 - \epsilon^{\alpha t_2}\right], \tag{8}$$

which is the additional equation required.

It now merely remains to eliminate some of the unknowns. To this end we first eliminate Q_1 and Q_2 from (3) and (4) by means of (6) and, solving for E, obtain

$$E = -E_0\frac{D - 1}{D\sin\omega(t_1 + t_2) - \sin\omega t_1}. \tag{9}$$

A second expression for E involving the same variables is next obtained by replacing Q_1 and Q_2 in (8) by their values as given by (3) and (4). Equating these two expressions for E leads to:

$$\tan\omega t_1 = \frac{G_2 + G_3\sin\omega t_2 - BG_2\cos\omega t_2}{G_1 - G_3\cos\omega t_2 - BG_2\sin\omega t_2}, \tag{10}$$

where

$$G_1 = (D - 1)S + H \qquad H = \left(\frac{1}{\alpha R} - C\right)\left(1 - \epsilon^{\alpha t_2}\right)$$

$$G_2 = (D - 1)P\omega \qquad S = P\alpha - C$$

$$G_3 = (D - 1)SB + HD \qquad B = \epsilon^{\alpha t_2}.$$

Referring to Fig. 26A it will be seen that t_2 is 0 if the input voltage is just sufficient to make the grid positive, i.e., if $E = -E_0$; while on the other hand t_2 will not exceed π/ω as E approaches infinity. Hence, if from equation (10) we obtain corresponding pairs of t_1 and t_2 by substituting values of t_2 ranging from $\omega t_2 = 0$ to $\omega t_2 = \pi$ and solving for t_1, we may substitute the pairs of values so obtained in (9), thereby arriving at a relation between E and t_1. Corresponding pairs of these two latter quantities can finally be substituted in (3) thus obtaining a relation between E and Q_1/C.

From (1b) we may express the departure from perfect compensation as a voltage ΔV which would have to be added to the condenser voltage to bring about this ideal condition. Viz:

$$\Delta V = \frac{E}{2} + \frac{E_0}{2} - \frac{Q_1}{C}. \tag{11}$$

By inserting corresponding values of E and Q_1/C, as obtained in the preceding paragraph, the precision of compensation obtained with any given set of parameters may be calculated.

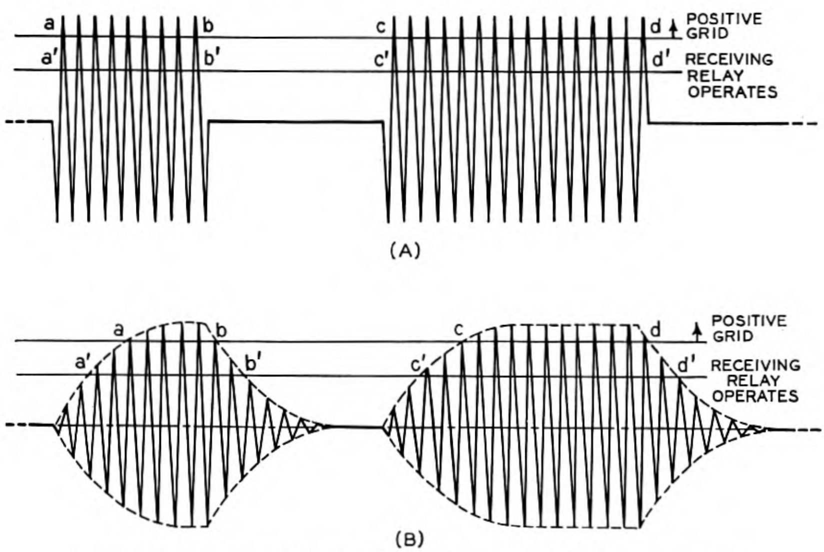


Fig. 27—Function of level-compensator relay. *A*. Square signal. *B*. Rounded signal.

Signaling Condition

Figure 27*A* shows the form which the received signals would have if the transfer admittance of the circuit were independent of frequency. If the level compensator consisted only of the simple condenser-resistance circuit shown in Fig. 25*A*, a large part of the charge accumulated during a mark would be dissipated during the following space, and since both marks and spaces continually change in relative lengths, this would give rise to characteristic distortion. This difficulty could be taken care of by having a relay operating in unison with the receiving relay and serving to disconnect the leakage resistance R_C during spaces in the manner shown in Fig. 6. Actually, however, the presence of the channel filters causes the received signals to resemble more nearly Fig. 27*B*. This leads to a further difficulty due to the fact that the charging intervals *ab*, *cd*, etc., are shorter than the periods $a'b'$, $c'd'$, etc., during which the receiving relay is closed. To remedy this, the compensator relay is biased towards spacing by means of the resistance *AB* (Fig. 6) and associated battery; the operating impulses being first rounded off by the resistances, condenser, and inductance in the *wave shaping circuit* to make them susceptible of such time bias. A necessary condition to be fulfilled by this circuit is that it should supply enough energy to the compensator relay to operate it even under conditions of extreme bias.

Drift

Returning to Fig. 27*B*, we may consider the bias of the compensator relay as equal to $a'b' - ab$ for the shorter signal or $c'd' - cd$ for the longer one. These biases are, of course, equal time intervals which we will denote by δ. From this it can readily be shown that in a given period—one second for instance—the grid will be conducting during a longer time for a group of long signals than for a group of short ones. Thus, let it be assumed for example, that $c'd' = 2a'b'$ and let n be the number of marking conditions of length $a'b'$ in a given interval; we have for the cumulative charging time in the two cases:

$$T_A = n(a'b' - \delta)$$
$$T_B = \frac{n}{2}(c'd' - \delta) = n\left(a'b' - \frac{\delta}{2}\right),$$

whence

$$T_B = T_A + \frac{n\delta}{2}.$$

It follows that the charge on condenser C is greater for the longer signals. If, therefore, the receiving circuit is adjusted to give unbiased signals when dot signals at the rate of 11 d.p.s. are received, similar signals at 23 d.p.s. will be biased positively. Experience shows that an adjustment which gives zero bias with 11 d.p.s. dots will give substantially unbiased signals with the standard test-sentence; hence circuits adjusted with 23 d.p.s. dots should be given a small initial positive bias.

It will be noted that the effective voltage available for charging the condenser is not constant throughout the interval *ab*. The maximum value which it can attain, and does attain if the signal is sufficiently long, equals the effective charging voltage during steady marking; on the other hand if the signal is very short, this value may never be reached. It follows that the charge accumulated by the condenser for any kind of intermittent signals is always less than that for steady marking; the amount of the discrepancy depending on the wave shape of the envelope.

As a result of these effects, and other similar conditions which tend to modify the average charge on the condenser depending on the character of the received signals, there is a perceptible amount of characteristic distortion manifesting itself in a fortuitous manner during the reception of ordinary text.

The change in the mean condenser charge can easily be observed if after a steady marking condition has been maintained for a few seconds, dots are suddenly impressed on the circuit and their bias observed. The latter will be found to *drift* as the charge assumes a new mean value. In practice, an adjustment is made for this by observing the change in voltage across the condenser under these two conditions and adjusting the compensator-relay bias to maintain the drift within limits which experience has shown to give minimum distortion with ordinary text.

Bibliography

1. Colpitts, E. H. and Blackwell, O. B., "Carrier Current Telephony and Telegraphy," *A. I. E. E., Trans.*, Vol. 40, pp. 205–300, 1921. (Has extensive bibliography.)
2. Bell, J. H., "Carrier-Current Telegraphy," *Bell Labs. Record*, Vol. 1, pp. 187–192, Jan., 1926.
3. Hamilton, B. P., Nyquist, H., Long, M. B., and Phelps, W. A., "Voice Frequency Carrier Telegraph System for Cables," *A. I. E. E., Trans.*, Vol. 44, pp. 327–332, 1925.
4. Nancarrow, F. E., "Line Telegraphy: More Voice Frequency Channels," *Electrician*, Vol. 122, p. 108, Jan. 27, 1939.
5. Harrison, H. H., "Printing Telegraph Systems and Mechanisms," Lond., Longmans, 1923.
6. Bell, J. H., "Printing Telegraph Systems," *A. I. E. E., Trans.*, Vol. 39, pp. 167–230, 1920.
7. Pierce, R. E., "Modern Practices in Private Wire Telegraph Service," *A. I. E. E. Trans.*, Vol. 50, pp. 426–435, 1931.
8. Gray, Elisha, "Nature's Miracles," Vol. 3, *Electricity and Magnetism*, pp. 126–128. N. Y., Fords, Howard and Hulbert. 1900.
9. "Il y a Cinquante ans: Essai du telegraphe harmonique de E. Gray," *Rev. Gen. de l'Elec.*, Vol. 32, pp. 361–362, 1932.
10. Watson, E. F., "Fundamentals of Teletypewriters Used in the Bell System," *Bell Sys. Tech. Jour.*, Vol. 17, pp. 620–639, 1938.
11. Pierce, R. E. and Bemis, E. W., "Transmission System for Teletypewriter Exchange Service," *Bell Sys. Tech. Jour.*, Vol. 15, pp. 529–548, 1936.

12. Clark, A. B., "Telephone Transmission over Long Cable Circuits," *A. I. E. E., Trans.*, Vol. 42, pp. 86–96, 1923.
13. Green, E. I., "Transmission Characteristics of Open-Wire Telephone Lines," *A. I. E. E., Trans.*, Vol. 49, pp. 1524–1535, 1930.
14. Affel, H. A., Demarest, C. S., and Green, C. W., "Carrier Systems on Long Distance Telephone Lines," *Bell Sys. Tech. Jour.*, Vol. 7, pp. 564–629, 1928; *A. I. E. E., Trans.*, Vol. 47, pp. 1360–1387, 1928.
15. O'Leary, J. T., Blessing, E. C., and Beyer, J. W., "Improved Three-Channel Carrier Telephone System," *Bell Sys. Tech. Jour.*, Vol. 18, pp. 49–75, 1939.
16. Cowley, G. W., "Volume Limiter Circuits," *Bell Labs. Record*, Vol. 15, pp. 311–315, June, 1937.
17. Peterson, Eugene and Keith, C. R., "Grid Current Modulation," *Bell Sys. Tech. Jour.*, Vol. 7, pp. 106–139, 1928.
18. Nyquist, H., Shanck, R. B., and Cory, S. I., "Measurement of Telegraph Transmission," *A. I. E. E., Trans.*, Vol. 46, pp. 367–376, 1927.
19. Shanck, R. B., Cowan, F. A., and Cory, S. I., "Recent Developments in the Measurement of Telegraph Transmission," *Bell Sys. Tech. Jour.*, Vol. 18, pp. 143–189, 1939.
20. Kupfmuller, K., "Über Einschwingvorgänge in Wellenfiltern," *E. N. T.*, Vol. 1, pp. 141–152, 1924.
21. Nyquist, H., "Certain Topics in Telegraph Transmission Theory," *A. I. E. E., Trans.*, Vol. 47, pp. 617–644, 1928.
22. Stahl, H., "Versuche über eine gunstige Verteilung der Tragerwellen in der Wechselstromtelegraphie," *T. F. T.*, Vol. 19, pp. 340–347, 1930.
23. Herman, J., "Bias Control in Telegraph Communication," U. S. Patent 1,886,808, Nov. 8, 1932; "Transmission Level Control in Telegraph Signaling Systems," U. S. Patent 1,943,478, Jan. 16, 1934.
24. Cash, C. C., "Protection against Lightning Interference," *Bell Labs. Record*, Vol. 15, pp. 125–128, Dec., 1936.
25. Trucksess, D. E., "High Precision Speed Regulator," *Bell Labs. Record*, Vol. 13, pp. 187–190, Feb., 1935.
26. Fry, J. R. and Gardner, L. A., "Polarized Telegraph Relays," *A. I. E. E., Trans.*, Vol. 44, pp. 333–338, 1925.

Electrical Drying of Telephone Cable

By L. G. WADE

DRYING equipment has recently been installed at the Kearny Works of the Western Electric Company whereby exchange area telephone cable is heated to a temperature of 270 degrees Fahrenheit by passing direct electric current through the copper wire conductors. Before discussing this installation in detail, however, it might be well first to outline the type of material handled and to review briefly the history of telephone cable drying methods and the reasons for the changes that have occurred.

Telephone cable consists of a number of individual paper ribbon or pulp insulated wires grouped together and the whole then covered with a serving of wrapping paper before being sheathed in lead. Cables may vary in length from a few feet up to several thousand feet, and in number of pairs from 6 to 2121. The size of the wire ranges from 26 American Wire Gauge to 10 American Wire Gauge, with some of the product often containing as many as two or three different sizes of wire in the same cable. One or more cable lengths are wound on a core truck which may be readily moved from one place to another by means of an electric truck.

Early cables were textile insulated and then dried by placing the cores in a heated oven, followed by boiling in a tank containing a sealing mixture or impregnant. The impregnant was used to keep the cable relatively dry in the rather imperfect lead sheath developed at that time. However, with the advent of an improved lead sheath extruded directly on the core, which would guarantee the excluding of any water, it was found desirable to go to a dry paper insulated telephone cable. This change in design reduced capacitance to about one-half of the previous values. The paper insulated cable was dried in a brick oven with gas retorts below a grille floor for maintaining the oven temperature between 215 degrees and 250 degrees Fahrenheit.

This method of drying produced satisfactory results for short-haul cables in use at that time. However, with the possibility of longer lead covered cable circuits replacing open telephone lines, it became necessary to obtain a greater degree of dryness in order to meet the new demand for transmission quality. The toll cables, which were for the longer haul, were therefore given an additional drying after the sheathing operation. This was accomplished by passing calcium chloride dried air at a temperature of 270 degrees Fahrenheit through the heated cable for a period of twenty-four hours. About 1917 the brick ovens were replaced by steam heated vacuum tanks which reduced the drying period to about one-third that used for the brick ovens and very greatly improved the dryness of all types of telephone cable.

Calcium chloride drying of toll cable was continued until 1927. It then became necessary to provide a means of keeping the cable from regaining appreciable amounts of moisture between the vacuum drying operation and lead covering. After considerable investigation a dry core storage room was developed where the dry cable could be held at .3 per cent to .5 per cent relative humidity until ready for covering.[1] With the improvement in vacuum tank drying and with cable stored under such a dry condition until the protective sheath could be applied, transmission quality of the shorter lengths of cable approached the level of the calcium chloride drying while there was some improvement in drying the longer lengths. This change in handling cable resulted in a large reduction in drying cost.

Still further improvements in drying methods have been obtained by heating the cable electrically rather than by radiation from steam coils in the drying chamber. Considerable thought had been given to this method over a number of years and the first unit of equipment was installed experimentally at the Baltimore Works in 1931. This unit has been in successful operation since that time for drying a part of the toll cable output and has furnished most of the data used in engineering the Kearny installation. The choice of the Kearny Plant for the first large scale installation was due largely to reduced operating and maintenance costs. The following discussion covers the type of equipment used in this installation and points out in closing some of the advantages gained in reduced cost and better drying of cable.

In preliminary experimental work on electrical drying a low voltage transformer was used to supply alternating current and the cable rendered non-inductive on its core truck by short-circuiting one end and dividing the other end and attaching to the source of power. However, such a set-up places the full voltage between a considerable portion of the conductors near the clamped ends, a condition not suited for telephone cable. Since, in telephone cable, the conductors are insulated with a thin tube of pulp or ribbon paper, the insulation

[1] Drying and Air Conditioning in Cable Manufacture, J. Wells and L. G. Wade, *Chemical and Metallurgical Engineering*, March 1932.

resistance between wires is low in an undried cable, particularly when moisture is leaving the cable during the drying operation. The choice of direct current not only eliminated the danger of breakdown between conductors (all conductors are grouped in parallel) but simplified the preparation and clamping of the cable for drying. Direct current also makes it possible readily to obtain practically any starting voltage between the maximum and minimum range of the equipment, a requirement necessitated by the great variety of cable lengths.

The electric drying installation at Kearny consists of a motor generator set for supplying the direct current for heating, Fig. 1, control

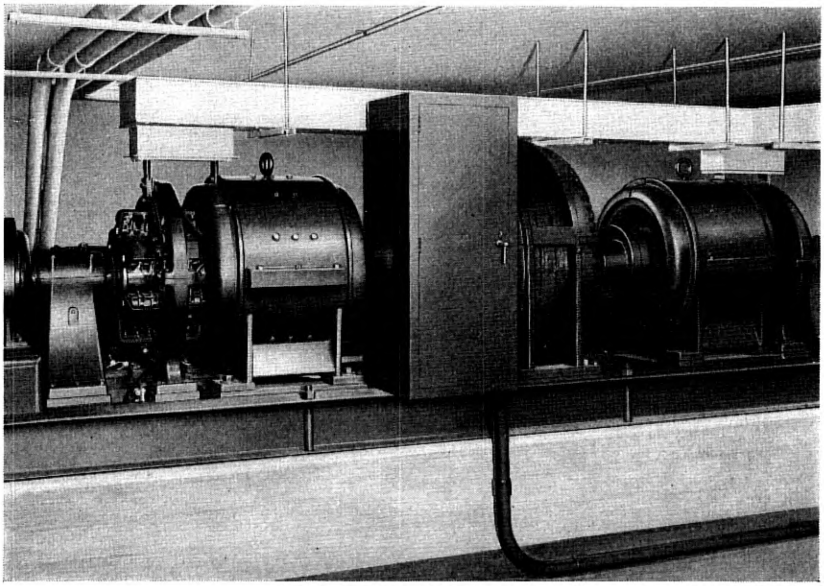


Fig. 1—Power supply set showing two generators for furnishing heating current.

equipment for starting and regulating the motor generator set, Fig. 2, and dome type dryers for holding the cable and its core truck under a vacuum of ½ inch to 1 inch of mercury back pressure during the heating and evacuation period, Fig. 3. A central control panel is located at the dryers on which is mounted apparatus for limiting the heating period to a predetermined amount, as well as visual instruments for indicating information for the operator's use in properly running the drying process.

All cables are dried at the same starting current density per unit of cross-section of the copper wire. With this as a starting point the capacity and type of the motor generator set were determined from the sizes and lengths of cables to be dried, balancing the first cost and efficiency against loss of capacity when handling some cables requiring

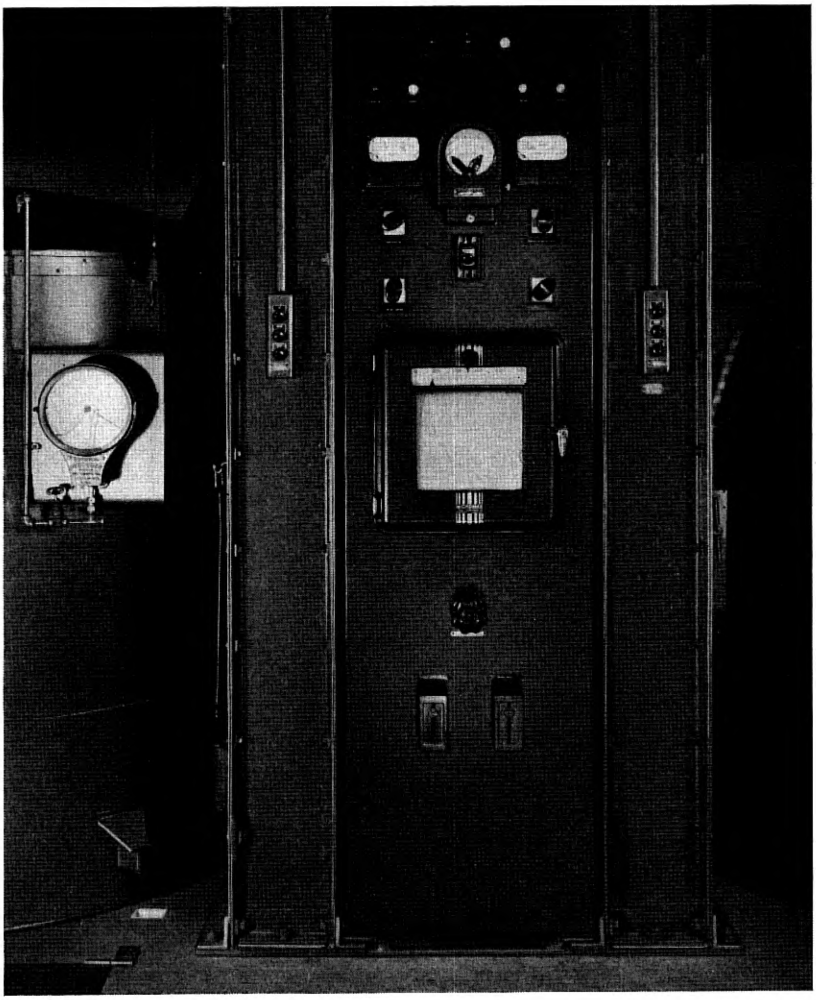


Fig. 2—Control panel for starting motor generator set and controlling the drying operation.

a longer heating period. This latter might be due to too large a cable or too long a length to be heated at the standard rate. In determining the capacity for a fixed starting current density, consideration was given to the cost of equipment, length of heating period and general

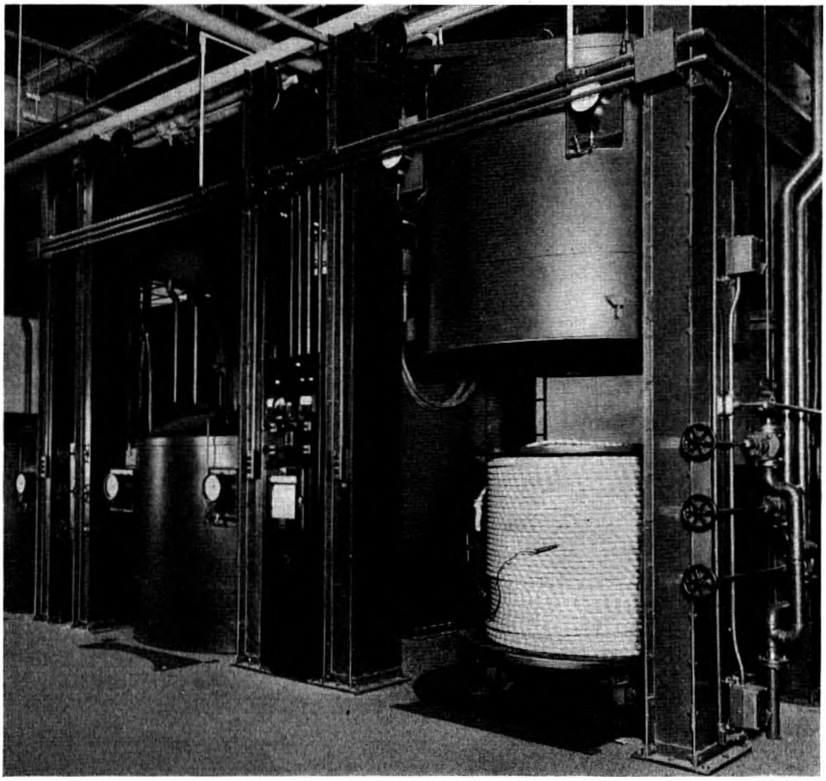


Fig. 3—Dome type dryers for holding cable and its core truck. One dome up, showing core truck in position.

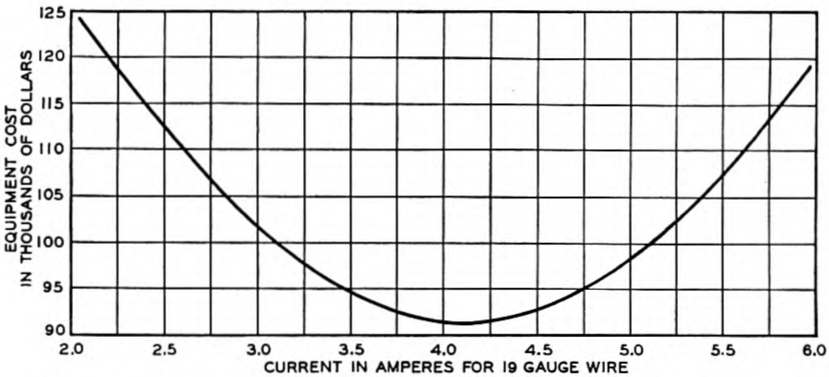


Fig. 4—Curve showing cost of equipment for drying the entire telephone cable output at Kearny Works for various starting current densities for 19-gauge wire or equivalent density for other sizes of wire.

overall efficiency of the drying operation. As shown by the curves in Fig. 4, 4 amperes per 19-gauge wire or equivalent provided about the lowest equipment cost. The heating period under this starting cur-

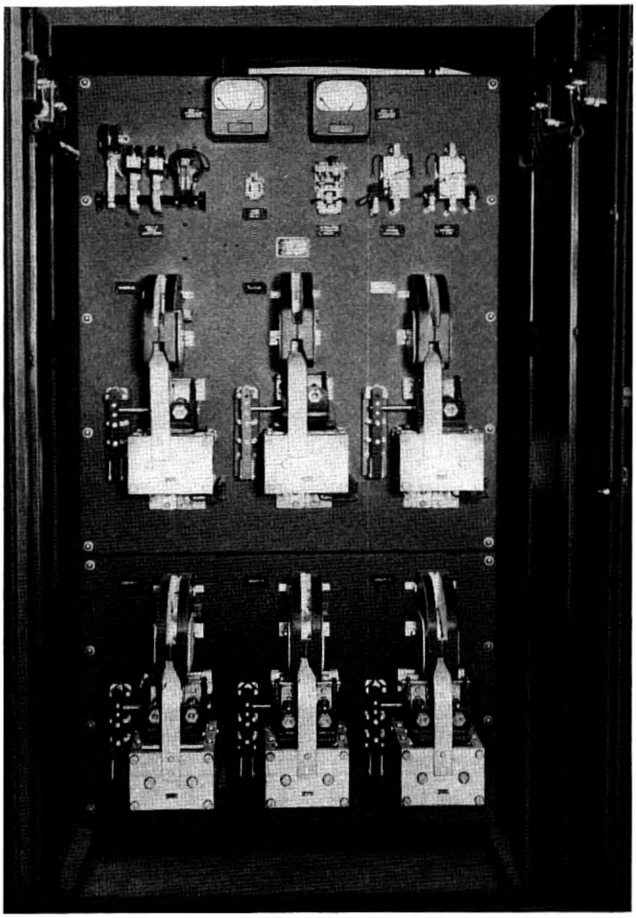


Fig. 5—D-C. power panel. The contactors in the upper row connect the generators either in parallel or in series. The contactors in the lower row connect the power supply to any of three dryers. All contactors are operated by switches on the control panel.

rent density was approximately one-half hour, which fitted very well into the plan for servicing the lead presses with dry cable cores.

In determining the type of equipment it was decided to provide two generators on the set. For the shorter lengths of large cable where low voltage is required, the two generators are operated in parallel. For the long lengths of small cable where higher voltage is required, the generators are operated in series. Suitable switching equipment on the control board provides ready means for changing to either method of operating the generators through large contactors located on the power panel, Fig. 5. Such an arrangement permitted the maximum capacity of the unit to be reduced one-half and the average starting load to be increased to 90 per cent of full load. This not only increased operating efficiency but very greatly reduced installation costs. Three such generating units were provided for handling the entire product at Kearny, one large, one intermediate and one small.

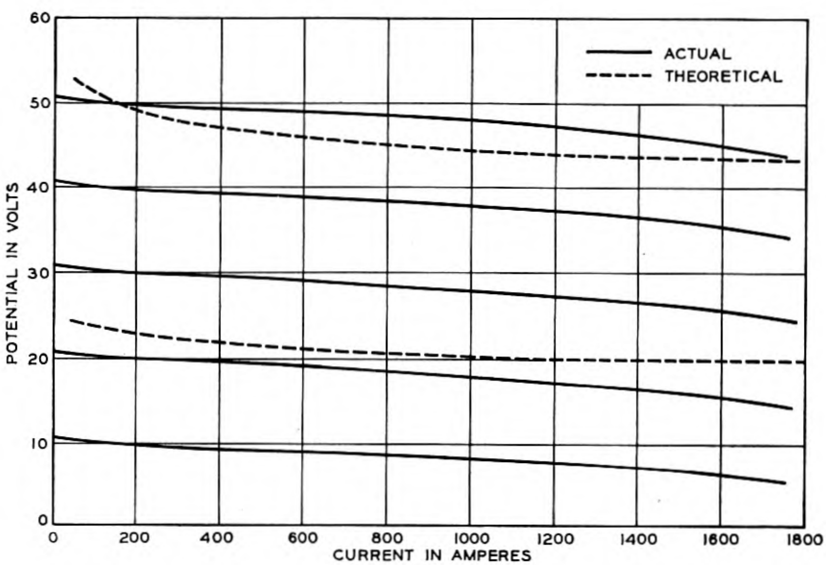


Fig. 6—Curves showing a desired theoretical voltage regulation and actual performance of the d-c. generators.

One of the chief factors contributing to simplification in operating electrical drying is to have the heating period the same for all sizes and lengths of cable. Theoretically, this requires the same voltage regulation at all conditions of the load, from the extreme of high current and low voltage to high voltage and low current. Two such theoretical curves are shown by the dotted lines in Fig. 6. The actual voltage regulation curves for one of the special generators are shown by the heavy lines on the same figure. The maximum amount of variation in the actual voltage regulation for any two load conditions amounted to a change in heating time of approximately one minute. This small difference permits the use of one heating period for all cable as some variation in final temperature is permissible. As will be noted, the generators gave a slightly rising voltage as the current dropped, which contributed to a shortening of the heating period.

The control panel located at the dryers contains switches for starting and stopping the motor generator set, as well as indicating and control equipment for properly conducting the drying operation, Fig. 2. Included in the indicating equipment is a voltmeter and ammeter for showing the potential and current readings for the cable at any time. At the beginning of the heating cycle the operator adjusts the rheostat controlling the field amperage in the generator until the proper voltage is obtained across the cable. This reading is determined from the cable length since the voltage for the specified current density is directly proportional to the cable length. Having established the proper voltage the operator then checks the ammeter for the proper current reading which, in turn, is determined by the size and number of conductors in the cable. To simplify the work, charts are prepared which show the voltage and corresponding starting current for all lengths and sizes of cable to be dried.

As the cable increases in temperature the current falls in direct proportion to the increase in conductor resistance, which in turn is determined by the temperature coefficient of copper (voltage constant). A chart can be used, therefore, to determine the final current for the desired drying temperature. As noted above, the voltage rises slightly as the current falls and the corresponding correction is made in the chart for final current readings for the various sizes of cable.

Such a control by final current readings is all that would be needed if the set were operated manually. However, it is not only more economical to release the operator from watching the ammeter but it has proved advisable for the safety of the product to control the drying cycle automatically. This is accomplished by a time relay and a temperature controller, either of which operates at a predetermined setting to open the field circuit on the direct current generator. The opening of the field circuit operates a signal device to let the operator know that the heating cycle is over so that the next one can be started without delay. The time relay is set from actual experience for a period slightly longer than the required heating time and serves as a protection in case of failure in the heating control circuit. The temperature controller is operated from a thermocouple embedded be-

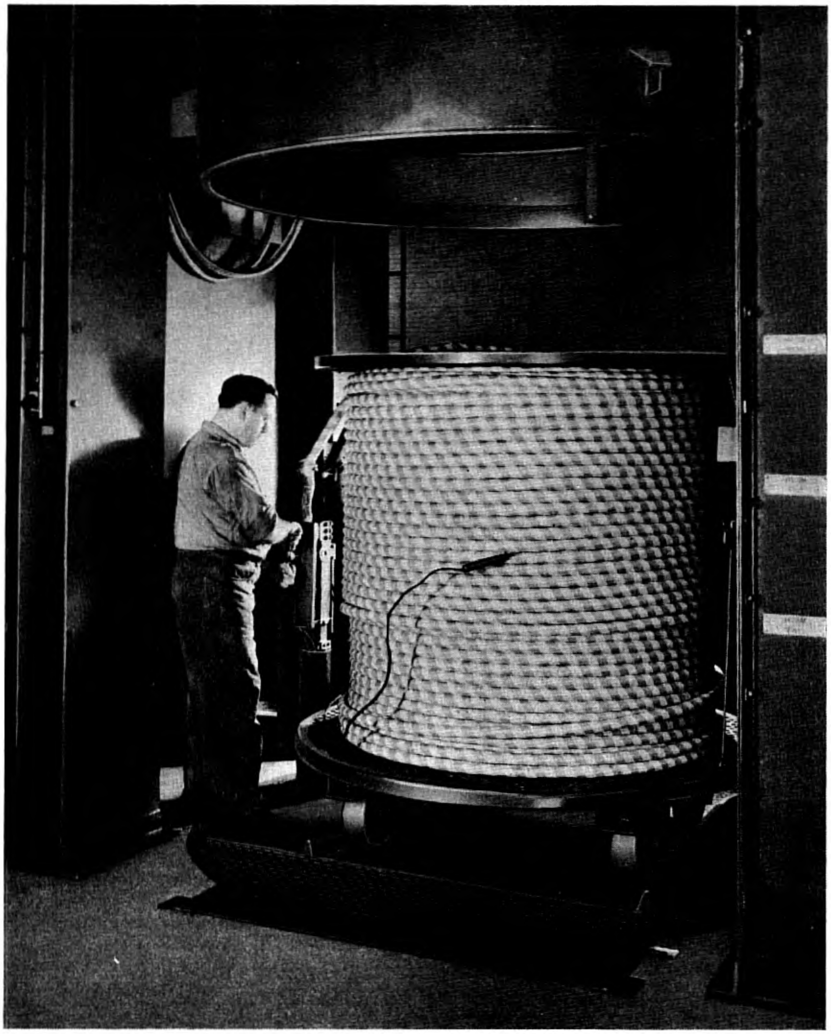


Fig. 7—Close-up of dryer with dome in raised position showing the operator tightening electrode clamp to cable. Thermocouple shown in place between layers of cable.

tween layers of the cable and is set to operate at the final desired temperature.

The drying chamber consists of a dome type shell that can be raised and lowered over a base, Figs. 3 and 7. The base is level with the floor line so that the cable core trucks can be handled readily in and out of the dryer. The dome is fitted with a gasket to provide for sealing with the base during evacuation. Steam coils are also provided inside the dome for keeping the dryer up to temperature during operation and provide a slight amount of heat to the cable during the short electrical heating period. The bus bars, vacuum connections and thermocouple leads enter through the base. The bus bars end in large vise-like clamps for firmly holding the ends of the cable. In the units for drying the larger cables the clamps are water cooled for carrying away the heat due to contact resistance. The core truck and cable are insulated from ground in order to minimize the danger

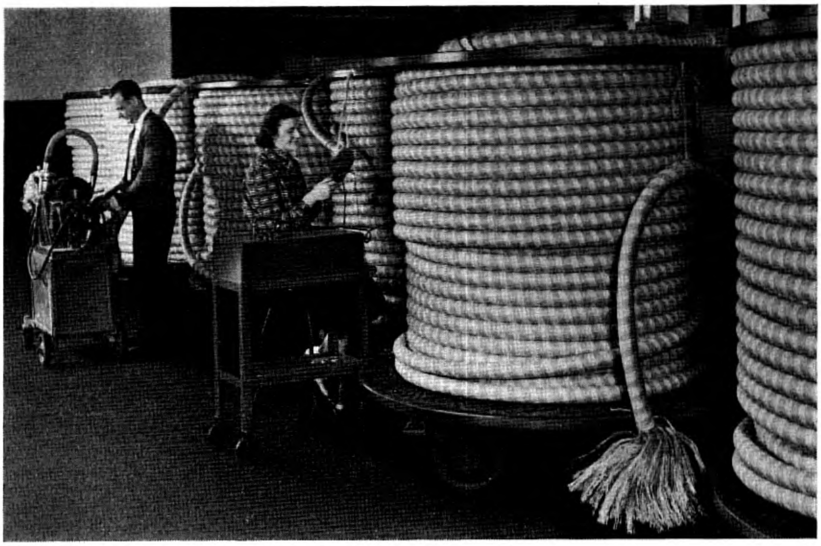


Fig. 8—Removal of paper from cable end and testing of circuits preparatory to drying.

from short circuit should the cable ground on the core truck. The push-button control for raising and lowering the domes is interlocked with the control for the heating cycles so that voltage cannot be applied at the clamp when the domes are up.

For the sizes and lengths of cables to be dried it was found advisable to have three dome dryers to one motor generator set, since all cables need some additional vacuum drying after the heating period is over. This arrangement also permits the removal of the dried cable and the loading of the next cable to be dried while the other two units are in operation. When the current is automatically disconnected at the completion of the heating cycle in progress, the operator turns the rheostat to zero voltage and throws a three-way switch, which connects the set to the next cable to be heated.

In preparing the cable for the regular testing process the insulation is removed from one end for a distance of four or five inches, Fig. 8. After the test shows that the cable is satisfactory for lead covering, the insulation is removed from the other end. At each end the wires are bunched together and tied with cotton string so that the individual wires act together in parallel the same as one large wire. The weight of insulation is sufficiently near enough to the ratio of the wire sizes so that a cable made up of several sizes of individual conductors and their particular insulation will heat satisfactorily by applying heat in proportion to the size of wire.

The electrical drying of cables requires a total of approximately 1½ hours as compared to the process it is replacing, which requires 12 hours or more and necessitates three-shift operation of the vacuum dryers. The short period permits a more rapid turnover of process stock as well as better planning of manufacture. By coordinating the drying and lead sheathing operations cables are lead covered immediately following their removal from the drying tank. The regain of moisture is slight under this condition and therefore the expensive storage oven is unnecessary. As a consequence, the total cost of equipment for the drying operation has been reduced to one-half the former investment value and the floor space from 19,000 square feet to 9500 square feet. Another advantage follows from replacing the continuous three-shift operation by one that operates only in sequence with other operations that may be only on a one- or at most a two-shift basis.

By applying heat to each conductor in proportion to the size of wire, each cable is given an individual treatment which insures a uniformity of drying not possible in the old vacuum tank process. In the replaced system several truckloads of varying amounts of cable were placed in the same vacuum tank and all dried for the same period. It was an averaging process leading to variations in dryness of individual cables and followed of necessity from the fact that a large number of different designs and lengths must be handled each day. To approximate individual handling under such a condition, where the drying period was 12 hours or more, would have involved a large number of dryers and increased floor space and operator-time. Also in the replaced system the layers of cable in the center position on the core truck were not fully up to the temperature of the outer layers, leading to variations in drying throughout the length. Thus, with the tendency in manufacture toward longer and longer lengths on the core truck, the variations in the degree of dryness under the vacuum tank process and the advantages of the electrical drying become more pronounced.

Electrical Wave Filters Employing Crystals with Normal and Divided Electrodes

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

I. Introduction

IN SEVERAL previous papers [1, 2, 3, 4] the application of piezoelectric crystals to electric wave filters has been discussed. The underlying principles and some of the design procedures were given. These filters have received wide application in carrier telephone systems and radio systems both in the United States and abroad.[5] It is the purpose of the present paper to discuss more completely all the standard types of filters with crystals, and methods for determining their constants and attenuation characteristics. In addition some of the newer results for simplifying such filters are given.

The use of a divided plate crystal for filters resulted in cutting the number of crystals in half as was pointed out in three former papers.[2, 3, 4] The theory of the use of such crystals is discussed in this paper and an equivalent circuit is given for a crystal with two sets of plates. The application of this circuit to unbalanced filters allows the results for balanced lattice filters to be realized for unbalanced filters. For one connection of the two plates the resonance of the crystal can be made to appear in one arm of the equivalent lattice, while for the reverse connection the resonance appears in the other arm of the lattice.

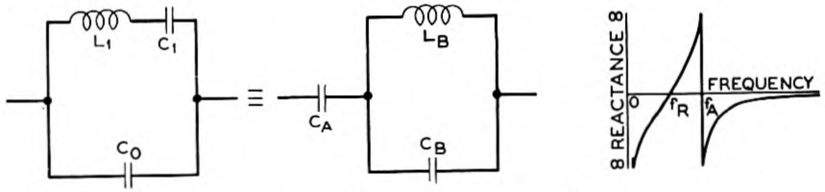
II. Crystal Filter Sections Which Can Be Realized in Lattice Networks

As pointed out in a previous paper[1] the most general filter characteristics for networks employing crystals can be realized in a lattice network, since every known form of a network can be reduced to a

[1] "Electrical Wave Filters Employing Quartz Crystals as Elements," W. P. Mason, *B. S. T. J.*, July 1934, pp. 405–452.
[2] "Resistance Compensated Band Pass Crystal Filters for Unbalanced Circuits," W. P. Mason, *B. S. T. J.*, Oct. 1937, pp. 423–436.
[3] "The Evolution of the Crystal Wave Filter," O. E. Buckley, *Jour. of Applied Physics*, Oct. 1936.
[4] "Crystal Channel Filters for the Cable Carrier Systems," C. E. Lane, *B. S. T. J.*, Vol. XVII, Jan. 1938, p. 125.
[5] "Channel Filters Employing Crystal Resonators," H. Stanesby and E. R. Broad, *P. O. E. E. Jour.*, 31, pp. 254–264, Jan. 1939.
[1] Loc. cit.

lattice network with realizable constants, whereas the converse is not necessarily true.

Let us consider first what types of filter characteristics can be obtained by using a crystal in one arm of a lattice network, and electrical or crystal elements in the other arm. As is well known the equivalent electrical network of a crystal is as shown in Fig. 1. The



$$C_A = C_0 + C_1\ ;\ C_B = \frac{C_0}{C_1}(C_0 + C_1)\ ;\ L_B = \frac{L_1}{\left(\frac{C_0 + C_1}{C_1}\right)^2}$$

$$f_R = \frac{1}{2\pi\sqrt{L_1 C_1}}\ ;\ f_A = \frac{1}{2\pi}\sqrt{\frac{C_0 + C_1}{L_1 C_1 C_0}}$$

Fig. 1—Equivalent electrical circuit and reactance frequency characteristic of piezo-electric crystal.

element values, as calculated in a recent paper, for a plated crystal vibrating longitudinally are [6]

$$C_0 = \frac{K l_w l_y}{4\pi l_t} \times \frac{1}{9 \times 10^{11}} \text{ farads;} \qquad C_2 = \frac{S}{\pi^2}\frac{d'^2_{12}}{S'_{22}}\frac{l_w l_y}{l_t} \times \frac{1}{9 \times 10^{11}} \text{ farads;}$$

$$L_1 = \frac{\rho S_{22}^{2} l_t l_y}{8 d'^2_{12} l_w} \times 9 \times 10^{11} \text{ henries,} \qquad (1)$$

where l_y, l_w, l_t are respectively the length, width, and thickness of the crystal expressed in centimeters, K = specific inductive capacity, S_{22}' = inverse of Young's modulus along the direction of vibration, d_{12}' is the value of the piezo-electric constant along the direction of vibration, and ρ is the density of the crystal. The resistance depends on the clamping resistance, acoustic radiation from the ends of the crystal, internal damping losses, etc. In general the ratio of the reactance of the inductance L_1 to the resistance R at the resonant frequency f_R is from 20,000 to 300,000, depending on how the crystal is mounted, whether it is evacuated, etc. In general this resistance is so small that it can be neglected for design purposes, and only the ideal reactance characteristic need be considered.

[6] "A Dynamic Measurement of the Elastic, Electric and Piezoelectric Constants of Rochelle Salt," W. P. Mason, *Phys. Rev.*, Vol. 55, April 15, 1939, p. 775.

The reactance characteristic of the crystal, as shown in Fig. 1, is a negative reactance at low frequencies up to a resonant frequency f_R. For frequencies greater than f_R, the reactance becomer positive up to the anti-resonant frequency f_A, above which the reactance is again negative. The ratio of the anti-resonant frequency to the resonant frequency is determined directly by the ratio r of C_0 to C_1 existing in the crystal. As shown by Fig. 1,

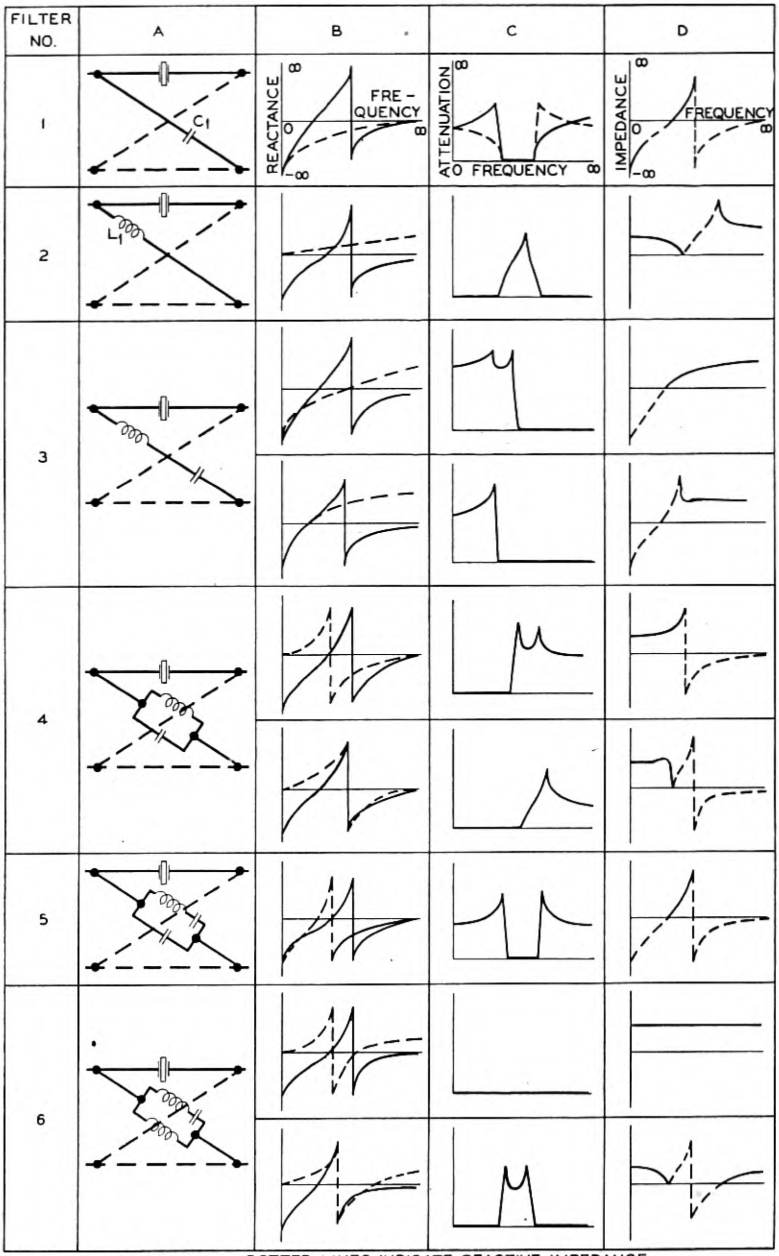
$$\frac{f_A}{f_R} = \sqrt{1 + \frac{1}{r}}. \tag{2}$$

This ratio is usually greater than 125 for a quartz crystal and hence the anti-resonant frequency is less than .4 per cent higher than the resonant frequency.

The previous papers considered mainly band-pass filters and discussed briefly low and high-pass crystal filters. It is also possible to obtain band elimination and all-pass crystal filters by combining electrical elements with the crystals in the proper manner. We consider, first, all the types of filters which can be obtained by using a single crystal in one arm of a lattice filter and electrical elements in the other arms. Figure 2 shows all the possible single-band characteristics which can be obtained by using a crystal in one arm and an electrical impedance, or crystal impedance, in the other lattice arm. For example, the first filter of the table shows a filter with a crystal in one arm and a capacitance in the other arm. Column B shows the reactance characteristic of each arm. A lattice filter will have a pass band when the reactances are of opposite sign and will attenuate when the reactances are the same sign. When the two reactances are equal the filter will have an infinite attenuation. This result follows from the expressions for the propagation constant and characteristic impedance of a balanced lattice network which are

$$\tanh \frac{P}{2} = \sqrt{\frac{Z_1}{Z_2}}; \qquad Z_0 = \sqrt{Z_1 Z_2}, \tag{3}$$

where Z_1 is the impedance of the series arm of the lattice and Z_2 that of the shunt arm. The third column shows the attenuation characteristic of this filter. It is a narrow band filter having a pass band between the resonant and anti-resonant frequencies of the filter. There is a peak of attenuation either above or below the band depending on the value of the capacitance C_1 in the lattice arm. The last column shows the value of the characteristic impedance of the filter as a function of the frequency. The dotted line indicates a



NOTE – IN COLUMN D, DOTTED LINES INDICATE REACTIVE IMPEDANCE
SOLID LINES INDICATE RESISTIVE IMPEDANCE

Fig. 2—Single band lattice filters employing a crystal in one arm.

reactance while a solid line indicates a resistance. In the pass band the filter has a resistive characteristic indicating a transmission of energy, while in the attenuating band the characteristic impedance is reactive indicating a reflection of energy.

Filter No. 2 shows what characteristic will be obtained if an ideal inductance is used in the lattice arm. As can be seen a band elimination filter will result with one attenuation peak. The width of the suppression band will be the separation between the resonant and anti-resonant frequencies.

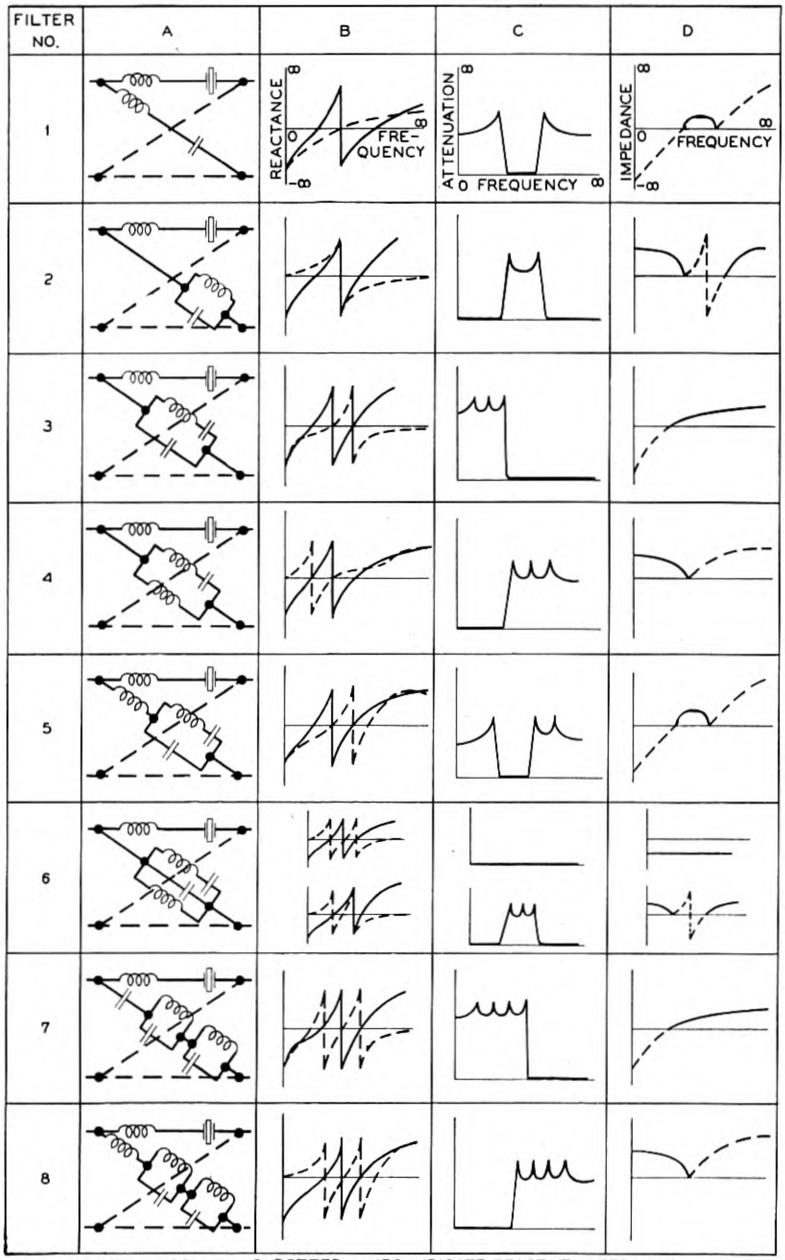
The use of a series resonant circuit results in a high-pass filter as can be seen from filter No. 3. It is possible to obtain two dispositions of the resonant frequencies which will give a single pass band as shown by the two sets of curves. The first set gives a high-pass filter with two attenuation peaks and a simple characteristic impedance. The other arrangement gives one attenuation peak and a more complicated type of characteristic impedance. The theory of this balancing of characteristics obtainable with a lattice filter is well known,[7] and is useful, when it is necessary on account of reflection effects, to make the characteristic impedance constant nearly to the cutoff.

The use of an anti-resonant circuit results in a low-pass filter as shown by filter No. 4. Two characteristics are possible. Filters No. 5 and 6 show the characteristics obtainable by using series resonant circuits shunted by a capacitance or an inductance. In one case a band-pass filter with two peaks results, and in the other either a band suppression filter with two attenuation peaks or an all pass filter. It will be noted that the configurations used in the lattice arm of filter 5 is the equivalent circuit of the crystal and hence a crystal can be used in this arm. In fact the circuit is similar to one discussed in the former paper.[1]

Since the crystal positive reactance region is very narrow ($< .4\%$), all of the band pass and band elimination filters obtained by using a crystal in one arm will of necessity have very narrow band pass or band suppression regions. For high and low-pass filters the attenuation peaks will of necessity come close to the cutoff frequencies. In the all-pass structure the phase shift will be very sharp in the neighborhood of the crystal resonance. It was shown in the first paper,[1] that wider pass bands and more general characteristics can be obtained by employing inductance coils in series or parallel with the crystal. Figures 3 and 4 show the possible types of filters obtained by

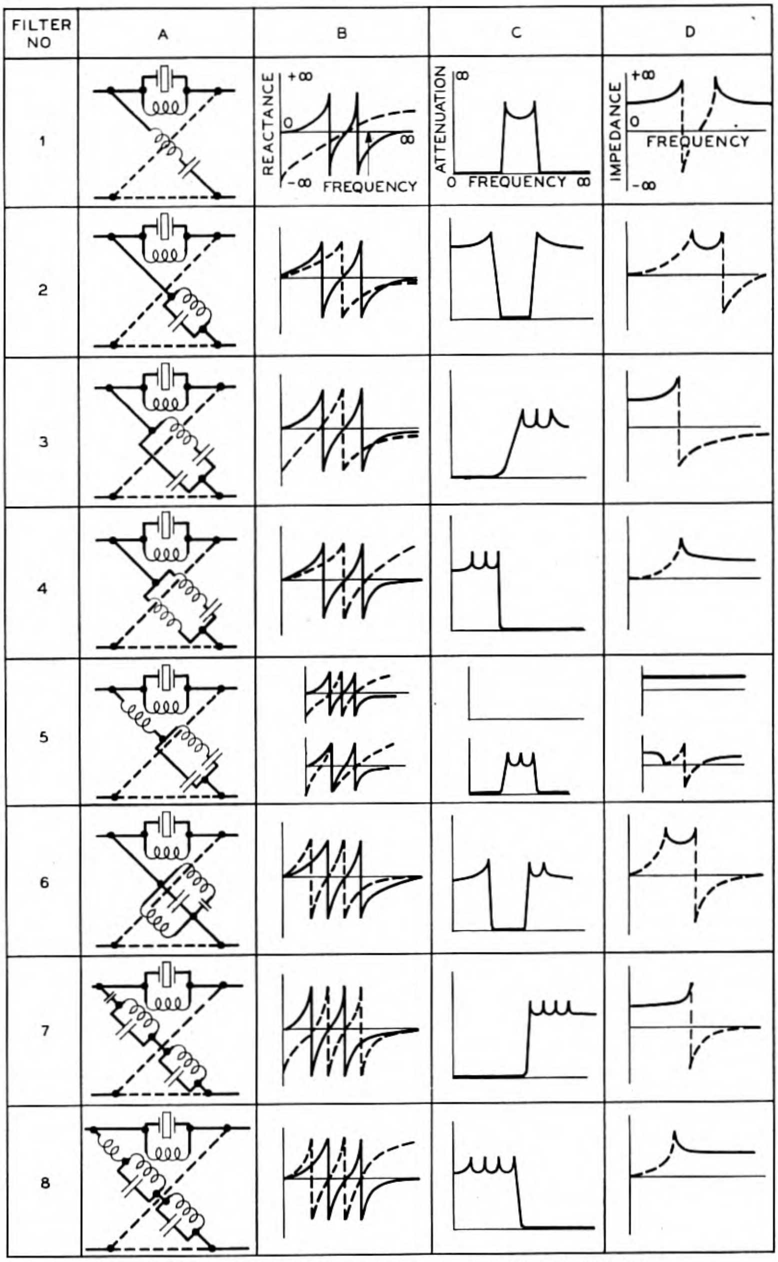
[7] Cauer, Siebschattungen VDI, Verlag Berlin, 1931. H. W. Bode, "A General Theory of Electric Wave Filters," *Jour. of Math. and Physics*, Vol. XIII, pp. 275–362, Nov. 1934.

[1] Loc. cit.



NOTE:—IN COLUMN D, DOTTED LINES INDICATE REACTIVE IMPEDANCE
SOLID LINES INDICATE RESISTIVE IMPEDANCE

Fig. 3—Single band lattice filters employing a crystal and coil in series in one arm.



NOTE:— IN COLUMN D, DOTTED LINES INDICATE REACTIVE IMPEDANCE
SOLID LINES INDICATE RESISTIVE IMPEDANCE

Fig. 4—Single band lattice filters employing a crystal and coil in parallel in one arm.

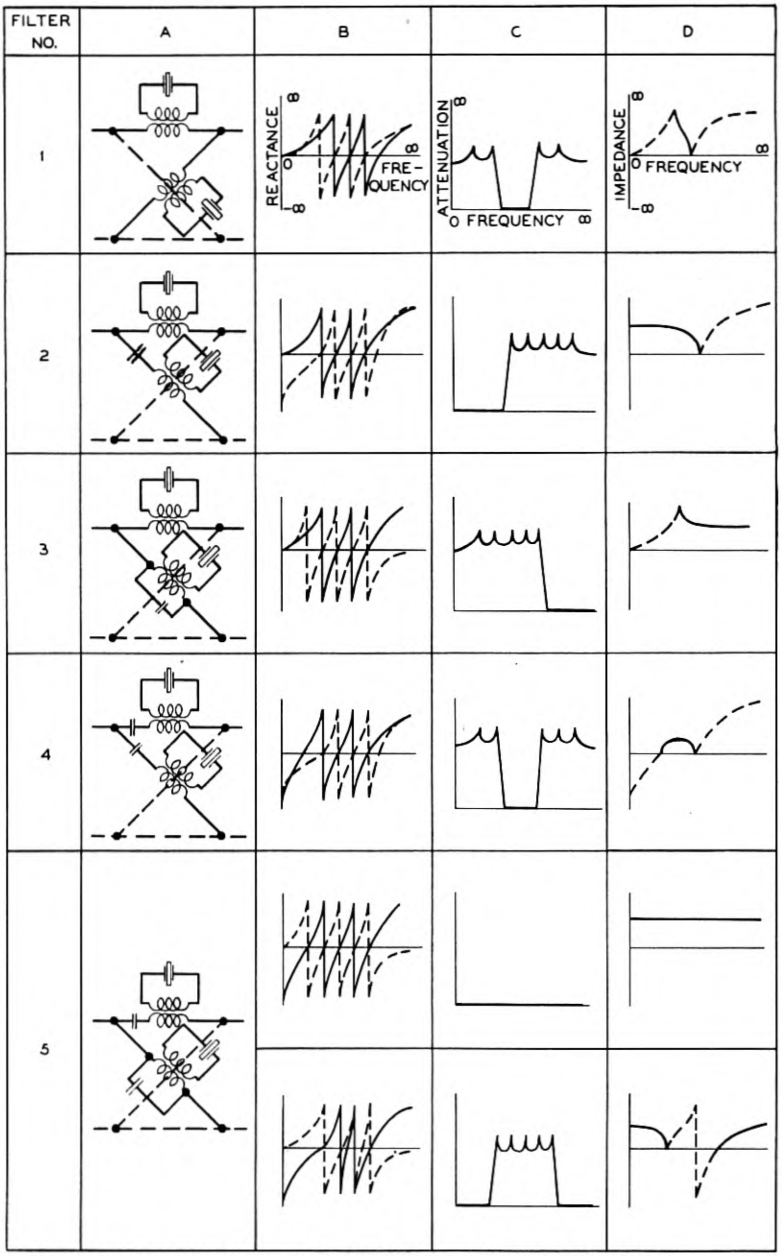
using a crystal and coil in one arm of the lattice and electrical elements in the other. Band-pass, band elimination, high and low-pass, and all-pass filters result. Only the simplest combinations of resonant frequencies giving the highest amount of attenuation are shown. As in filters 4 and 6 of Fig. 2, some of the anti-resonances and resonances of the two arms may be made to coincide giving filters with less attenuation but more flexibility in the impedance characteristics. Condensers can be incorporated in parallel or series with the crystals without affecting the type of characteristic obtained. This procedure is useful in controlling the widths of the pass or attenuation bands and in controlling the position of the peak values. In a number of the filters of Figs. 3 and 4, the equivalent circuit of the crystal occurs in the electrical circuit in the lattice arms. In these filters, crystals can also be used in the lattice arms. Filters 1 and 5 of Fig. 3 and filters 2 and 6 of Fig. 4 are band-pass filters which have been discussed in detail in former papers.[1, 2]

Crystals may also be used in more complicated electrical circuits, for example with transformers as shown in Fig. 5. This figure shows high, low, band-pass, band elimination and all-pass filters which can be constructed by employing transformers and crystals in each arm. More complicated structures still using single crystals can also be constructed but they tend to be of less importance since the dissipation introduced by the electrical elements neutralizes any benefit of using crystals.

It is possible, however, to use more crystals than one in one arm of a lattice and obtain filters having higher insertion losses outside the band without introducing more loss due to the electrical elements in the band. Figure 6 shows a number of such combinations with and without coils. The result of adding an additional crystal in one arm of a lattice is to add another elementary section of the type discussed in Appendix I. An example [8] of the characteristic obtainable with a band-pass filter with two crystals in each arm is shown on Fig. 7.

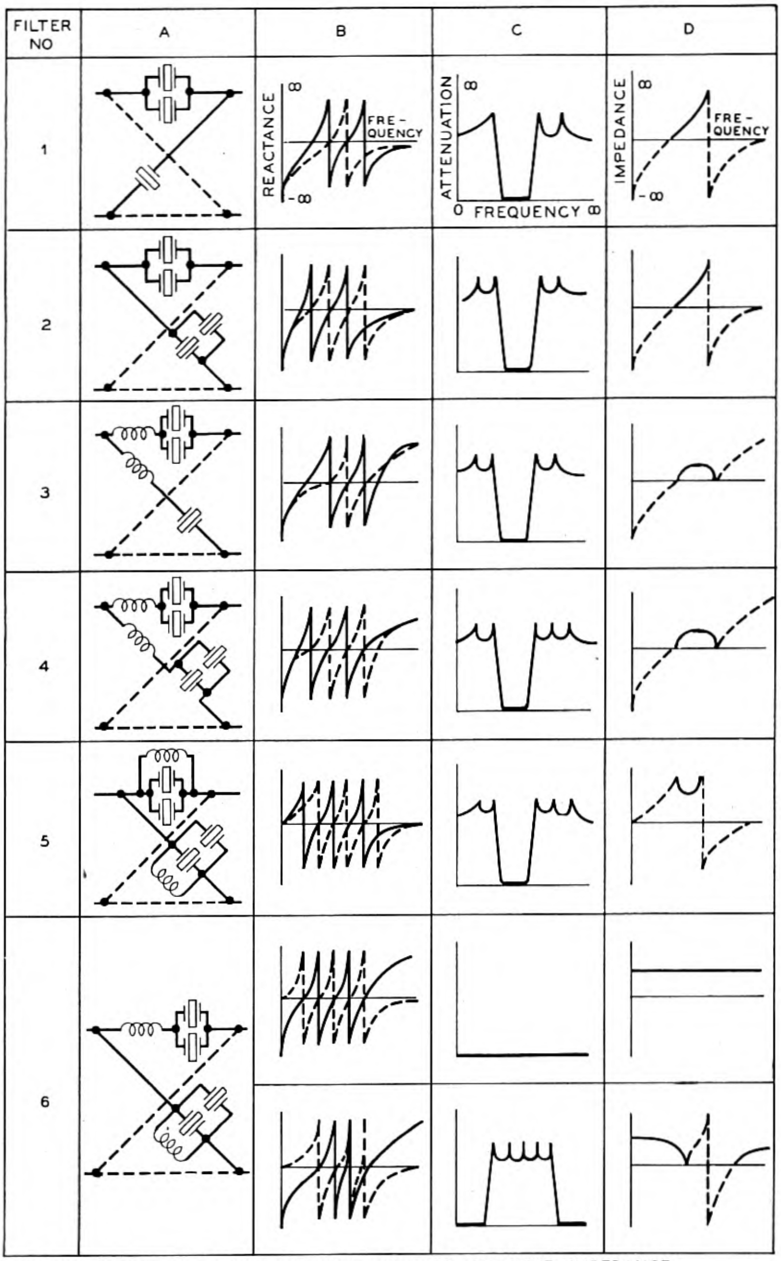
All of the filters discussed above were assumed to be constructed from dissipationless elements. When coils are used, however, a certain amount of resistance is associated with them which may alter the characteristics obtainable. As has been pointed out previously,[1] if the dissipation associated with the coils can be brought out to the ends of the arm either in series or parallel with the complete arm the effect of these resistances will be to add a constant loss independent of

[1, 2] Loc. cit.
[8] This filter was constructed and tested by Mr. H. J. McSkimin.
[1] Loc. cit.



NOTE:- IN COLUMN D, DOTTED LINES INDICATE REACTIVE IMPEDANCE
SOLID LINES INDICATE RESISTIVE IMPEDANCE

Fig. 5—Single band lattice filters employing transformers and crystals.



NOTE:- IN COLUMN D, DOTTED LINES INDICATE REACTIVE IMPEDANCE
SOLID LINES INDICATE RESISTIVE IMPEDANCE

Fig. 6—Single band lattice filters employing more than one crystal in the impedance arms.

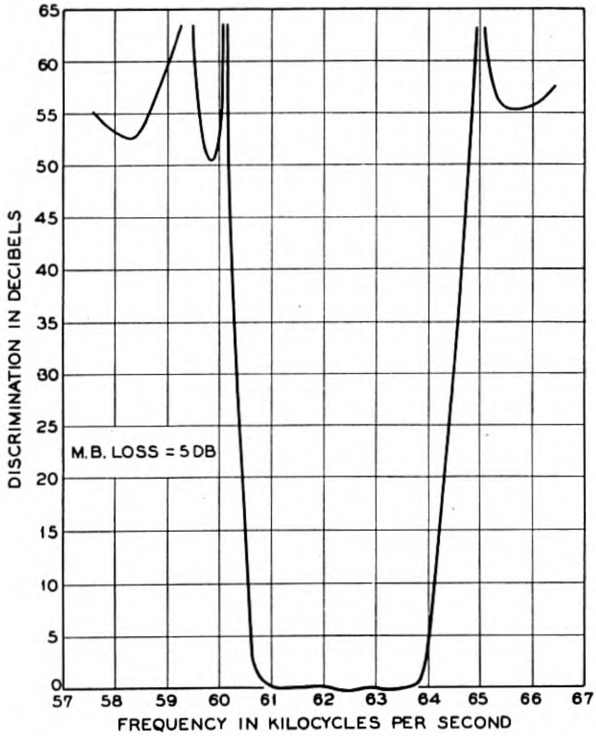


Fig. 7—Insertion loss characteristic of a single section band pass filter employing two crystals in each arm.

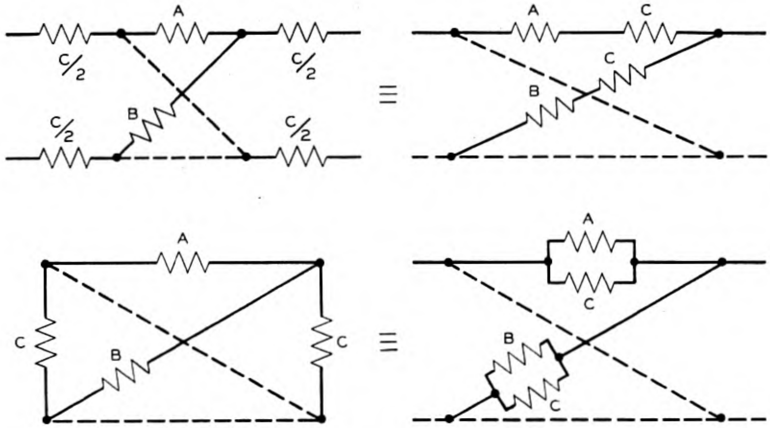


Fig. 8—Equivalences between lattice networks.

the frequency, and hence the discrimination obtainable will not be affected. This follows from the network equivalences shown in Fig. 8 which were first proved in a previous paper.[1] Even if this resistance compensation cannot be completely obtained it can often be obtained over a limited range near the cutoff and the peak frequencies by adding resistances to some of the crystal or electrical elements of such a value that the resistive components of the two arms are nearly equal over a limited frequency range. This results in cutting down the distortion near the cutoff and increasing the loss in the attenuated regions.

The lattice filters of Figs. 2 to 6 can be realized in ladder or bridge T forms in certain cases. If the two arms have two common series elements, then by the first equivalence of Fig. 8 they can be taken outside the lattice. Similarly, if two common shunt elements can be found in the two arms, then, by the second theorem of Fig. 8, the

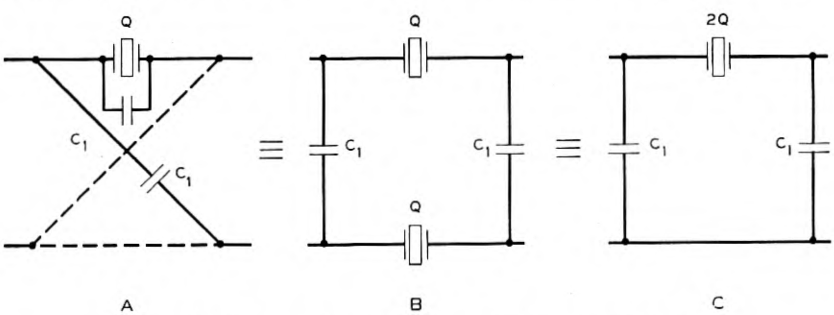


Fig. 9—Method for reducing a lattice filter to a π network filter.

elements can be placed in shunt on the ends of the filter. For example, suppose that we consider filter No. 1 of Fig. 2 and shunt the crystal by a capacitance C_1 which is equal to the capacitance of the lattice arm as shown in Fig. 9A. Then the two capacitances can be taken out in shunt leaving a crystal in the series arm of a π network as shown in Fig. 9C. This has the same type of characteristic as the lattice but considerably greater limitations.

A somewhat more general transformation can be made to a bridge T network of the type shown in Fig. 10A. This network is equivalent to the lattice network shown in Fig. 10B. As is evident, if we have an impedance in parallel with one arm and in series with the other, the resulting lattice can be transformed into a bridge T network. For example, in Fig. 3, filter No. 2, if we reverse the lattice and series arms, which can be done without changing the characteristics except for a 180° phase reversal, the filter can easily be reduced to a bridge

T network as shown in Fig. 11. The shunt coil L_0 is usually considerably smaller than the series coil L_1 so that L_1 can be divided into two coils, L_0 and $L_1 - L_0$. The transformation then becomes as

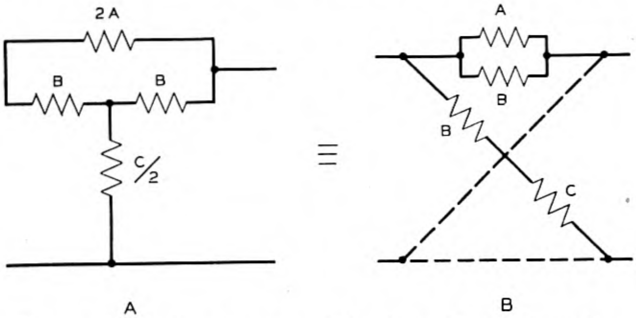


Fig. 10—Equivalence between bridge T and lattice networks.

shown in Fig. 11 with the element values shown. $Q/2$ indicates that the impedance of the crystal in the shunt arm is half that in the lattice arm. This transformation is applicable particularly to low and high-pass filters and band elimination filters.

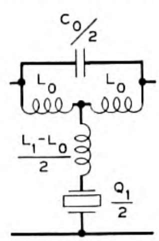


Fig. 11—A bridge T band elimination crystal filter.

Another transformation which can be employed is that for a three-winding transformer, for, as shown in a previous paper,[2] a three-winding transformer connected to two impedance arms as shown in Fig. 12 is equivalent to a transformer and a lattice filter with small

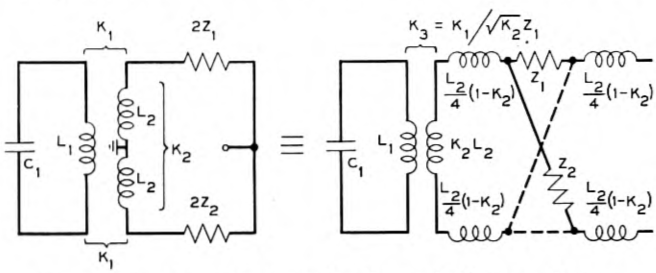


Fig. 12—Lattice equivalent of a three-winding transformer.

[2] Loc. cit.

coils on the ends. By making the coupling in the secondary high these coils can be made very small and can usually be neglected.

Another method for reducing balanced lattice filters to unbalanced circuits is to employ crystals with two sets of plates as described in section IV.

III. Method for Calculating the Element Values of the Filter

The curves in Figs. 2 to 6 give a qualitative picture of what type of characteristics can be obtained by the use of crystals in filter networks. In order to determine what band widths and dispositions of attenuation peaks are realizable with crystals it is necessary to calculate the element values, since a crystal cannot be made with a ratio of capacitances under 125.

The actual process of calculation can be divided into two parts. The first part consists in a determination of the critical frequencies of the arms of the network in terms of the desired attenuation characteristic. The second part consists in calculating the element values from the critical frequencies by means of Foster's theorem.

The attenuation characteristics obtainable with filters are discussed in Appendix I, and it is there shown that the attenuation characteristic of a complicated filter structure can be regarded as the sum of the attenuation characteristics of a number of elementary filters. The critical resonant frequencies of the filter are evaluated in terms of the cutoff frequencies and the position of the attenuation peaks with respect to the cutoff frequencies. The ratios of the impedances of the two arms at zero or infinite frequencies are evaluated in terms of the network parameters. With the aid of these equations, and Foster's theorem discussed below, the element values can be evaluated for any desired attenuation characteristic. Whether the characteristic is realizable or not depends on whether the element values of the equivalent circuit of the crystal calculated have a low enough ratio of capacitances to be realized in practice. The actual value of the series capacitance C_1 of the equivalent circuit of the crystal shown in Fig. 1 may also be too large to be physically realizable.

Having obtained the critical frequencies by the calculations given in the appendix, the element values can be calculated by using Foster's theorem. Foster's theorem [9] deals with impedances in the form of a number of series resonant circuits in parallel as shown on Fig. 13A or a number of antiresonant circuits in series as shown on Fig. 13B.

[9] See "A Reactance Theorem," *B. S. T. J.*, April 1924, page 259.

In either case the impedance of the network can be written in the form

$$Z = -jH \left[\frac{\left(1 - \frac{\omega^2}{\omega_1^2}\right)\left(1 - \frac{\omega^2}{\omega_3^2}\right) x \cdots x \left(1 - \frac{\omega^2}{\omega_{2n-1}^2}\right)}{\omega \left(1 - \frac{\omega^2}{\omega_2^2}\right) x \cdots x \left(1 - \frac{\omega^2}{\omega_{2n-2}^2}\right)} \right], \tag{4}$$

where $H \geqq 0$ and $0 = \omega_0 \leqq \omega_1 \leqq \cdots \leqq \omega_{2n-1} \leqq \omega_{2n} = \infty$. For the series resonant circuits of Fig. 13A, the element values are given by

$$L_i = \frac{1}{C_i \omega_i^2} = \left(\lim_{\omega \to \omega_i} \right) \left(\frac{j\omega Z}{\omega_i^2 - \omega^2} \right); \quad i = 1, 3, \cdots 2n - 1. \tag{5}$$

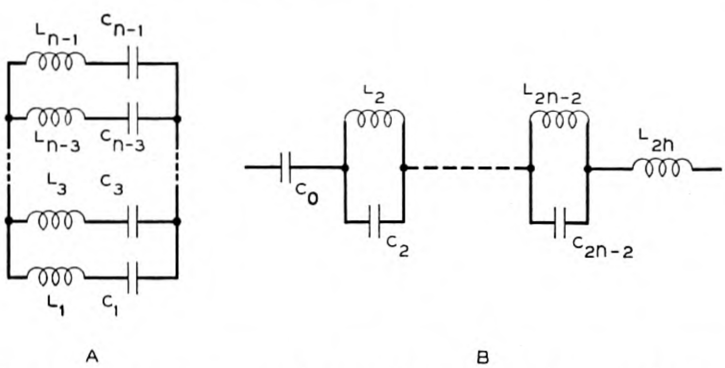


Fig. 13—Impedances arranged in form for application of Foster's Theorem.

For the antiresonant circuits in series the values become

$$C_i = \frac{1}{L_i \omega_i^2} = \left(\lim_{\omega \to \omega_i} \right) \left(\frac{j\omega}{Z(\omega_i^2 - \omega^2)} \right); \quad i = 0, 2, 4, \cdots 2n. \tag{6}$$

These values include the limiting values for the series case of Fig. 13B.

$$C_0 = \frac{1}{H}; \quad L_0 = \infty, \quad C_{2n} = 0; \quad L_{2n} = \frac{H(\omega_2^2 w_4^2 \cdots \omega_{2n-2}^2)}{(\omega_1^2 \omega_3^2 \cdots \omega_{2n-1}^2)}. \tag{7}$$

Hence if the elements of one arm of the lattice are arranged in either of the forms shown on Figs. 13A or B, the element values can be calculated from equations (5) and (6).

If they are not in this form, they can be transformed into one of these two forms by well known network transformations. For example, all the filters of Figs. 2, 3 and 4 are of this form or can be put in this form by employing the simple network transformation of Fig. 1. For the two crystal sections shown on Fig. 6, the series inductance can be evaluated by equation (7) and subtracted from the impedance Z. This leaves an impedance

$$Z' = Z - j\omega L, \tag{8}$$

from which can be evaluated the constants of the two crystals in parallel by employing equation (6). In this way the constants of any filter can be evaluated when the desired attenuation and impedance characteristic are specified. Several of the band-pass filters are discussed in detail in a former paper.[2]

IV. Application of Divided Plate Crystals to Balanced and Unbalanced Filters

The use of a divided plate crystal to cut the number of crystals in half in a balanced lattice filter has been mentioned previously.[2, 3, 4, 10] The theory of this use has not been previously discussed and since it results in further applications it seems worth while to present it here.

In order to use the divided plate crystal in filters it is necessary to find an equivalent circuit for such a crystal which will hold for measurements between any pair of the four terminals. It was shown in a previous paper [6] that an electro-mechanical equivalent of a fully plated crystal free to vibrate on both ends could be represented as shown in Fig. 14A. In this figure the capacitance C_0 is the static capacitance of the crystal, the condenser C_M represents the effective compliance of the crystal at the resonant frequency, and the inductance L_M represents the effective mass. A perfect transformer of impedance ratio 1 to φ^2, where

$$\varphi^2 = \frac{(d_{12}l_w)^2}{(s_{22'})} \tag{9}$$

represents the coupling from electrical to mechanical energy. φ in effect is the ratio of the force exerted by the crystal when it is clamped, to the applied voltage or it is the force factor of the system. If now we use only half the plating on the crystal, for example the plates 1, 2 of Fig. 14B, the same representation will hold. The static capacitance C_0 will be divided by 2, and the force applied by a given voltage will also be divided by 2 or the transformer ratio will be $\varphi/2$. The same compliance and mass will be operative. Hence the equivalent circuit of a crystal with plates covering half the crystal will be as shown on Fig. 14C. For a crystal with two sets of plates, the repre-

[2, 3, 4] Loc. cit.
[10] See patent 2,094,044, W. P. Mason, issued Sept. 28, 1937.
[6] Loc. cit.

sentation shown on Fig. 14D can be used if we are interested only in the transmission from one set of plates to the other. The numbering on the terminals agrees with that shown on Fig. 14B and is necessary in order that a given voltage E will produce the same displacement in the crystal when the voltage is applied between 1 and 2 and 3 and 4.

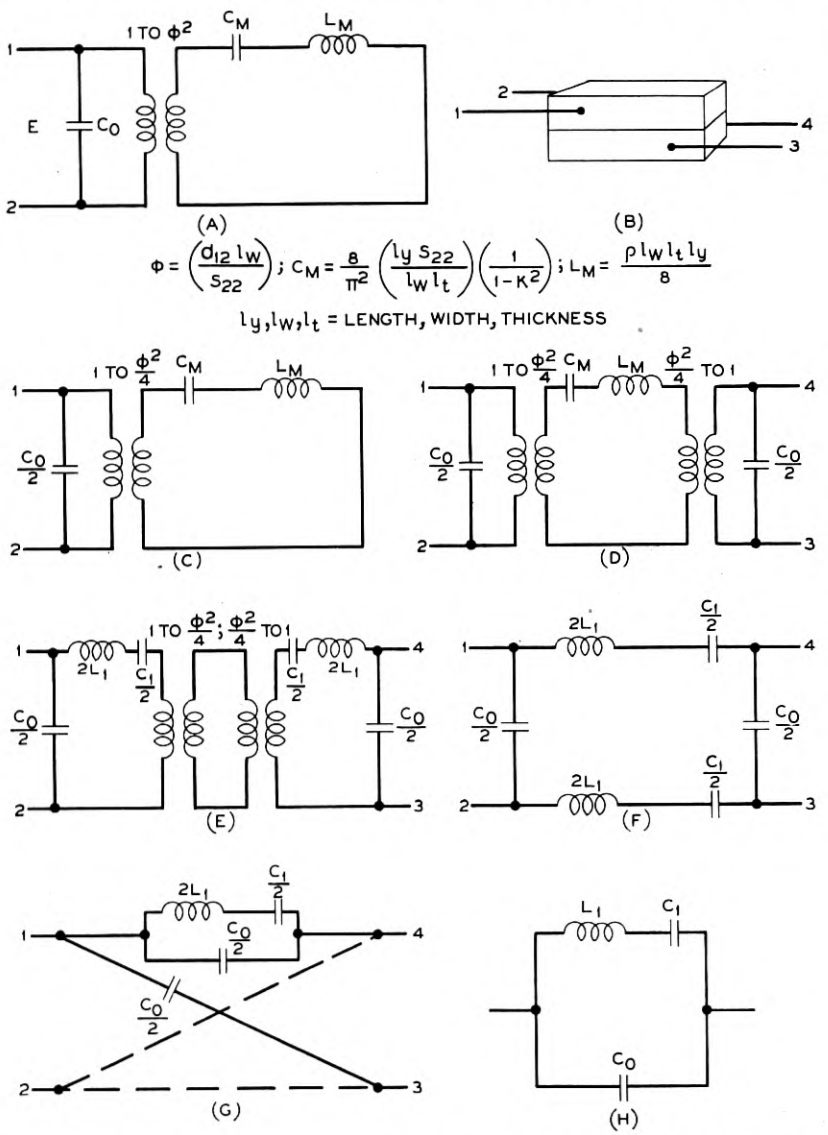


Fig. 14—Equivalent networks to represent transmission through divided plate crystal.

For purely electrical measurement, we can get rid of the two ideal transformers by taking half the impedance of C_M and L_M through each transformer as shown in Fig. 14E. Since we have left two opposing transformers of equal ratio they can be eliminated and the network of Fig. 14F results. This is shown in balanced form. Figure 14G shows the same network expressed in lattice form which is easily done by using the equivalences of Fig. 8. This represents a crystal of twice the impedance of the fully plated crystal in each series arm of the lattice with the static capacitances C_0 in the other arm. If we connect terminal 1 to 3 and 4 to 2, or in other words we use a completely plated crystal, the equivalent circuit reduces to that for a fully plated crystal as shown in Fig. 14H.

The networks of Fig. 14, F and G, represent the two plate crystals for transmission through the crystal, but do not give a four-terminal equivalence. For example, if we measure the crystal between terminals 1 and 3 we should not expect any impedance due to the vibration of the crystal since there is no field applied perpendicular to the thickness. The representation of Fig. 14, F or G, would not indicate this. The same sort of problem arises when it is desirable to obtain a four-terminal representation of a transformer and can be solved by using a lattice network representation with positive and negative inductance elements. The same procedure can be employed for a crystal and the steps are shown in Fig. 15.

We start with the lattice representation of Fig. 14G but employ the series form of the impedance of a crystal shown in Fig. 1. The series capacitance is divided into two parts, $C_0/2$ and a negative capacitance necessary to make the total series capacitance equal to C_0 plus C_1. This negative capacitance and the antiresonant circuit are lumped as one impedance $2Z$ in Fig. 15B. Now by the network equivalence of Fig. 8, we can take the series capacitances $C_0/2$ outside the network. We can also add an impedance $Z/2$ on the ends of the network provided we add a negative Z in series with all arms of the network as shown in Fig. 15C. The network of Fig. 15C is equivalent to that of Fig. 15A as far as transmission through it is concerned, but is different if we measure impedances between any of the four terminals. For example, if we measure the impedance between the terminals 1 and 4, the impedance of the network reduces to that shown in Fig. 15D. The impedance of the parallel circuit reduces to a plus Z shunted by a minus Z which introduces an infinite impedance. Similarly between 1 and 4, 2 and 3, and 2 and 4 the impedance becomes infinite as it should be if we neglect the small static capacitances existing in the crystal. If we take account of these the complete four-terminal representation of a crystal becomes that shown in Fig. 15E. Ordinarily the capacitances C_{13}, C_{14}, C_{23}, C_{24} are small enough to be neglected. Figure 15E then is a complete equivalent circuit for a two-plate piezo-electric crystal which is valid for any kind of impedance or transmission measurements.

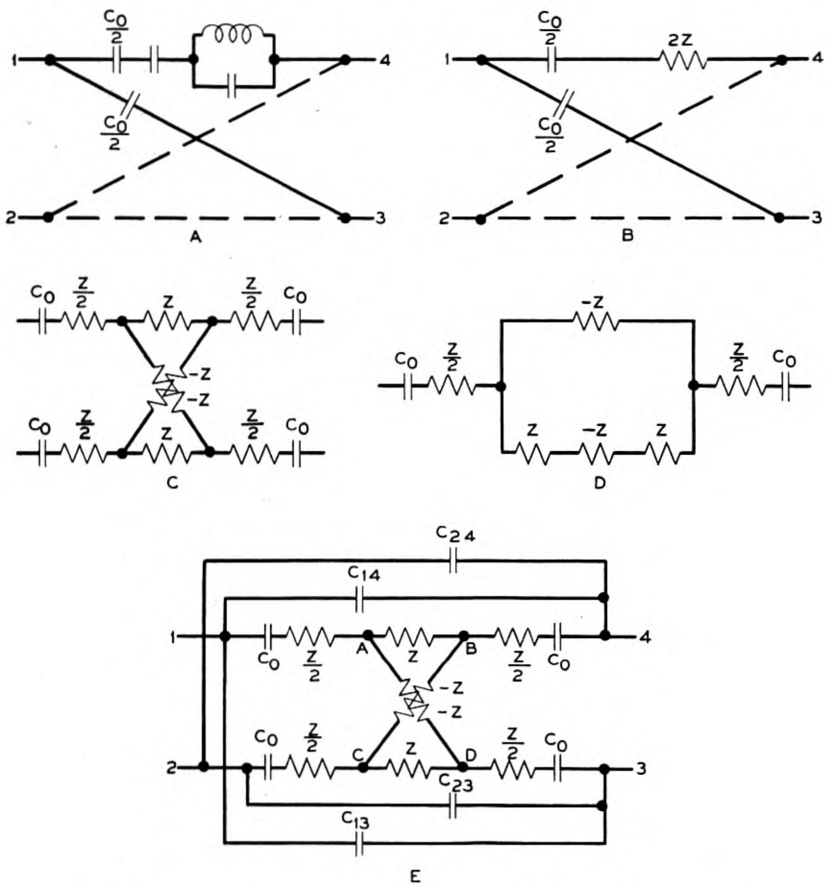


Fig. 15—Four-terminal equivalent network for divided plate crystal.

There are four possible connections for a crystal with two sets of plates used in a balanced filter. These connections and their equivalent circuits for transmission through as used in the filter are shown in Fig. 16. In order to prove these equivalences let us consider the equivalence shown on Fig. 16A. The four-terminal network representation for this case is shown in Fig. 17, which is obtained from

Fig. 15E. The capacitances C_{13} and C_{24} will be equal due to the symmetry in the crystal, while C_{14} will equal C_{23} for the same reason. These capacitances are directly connected to the outside terminals

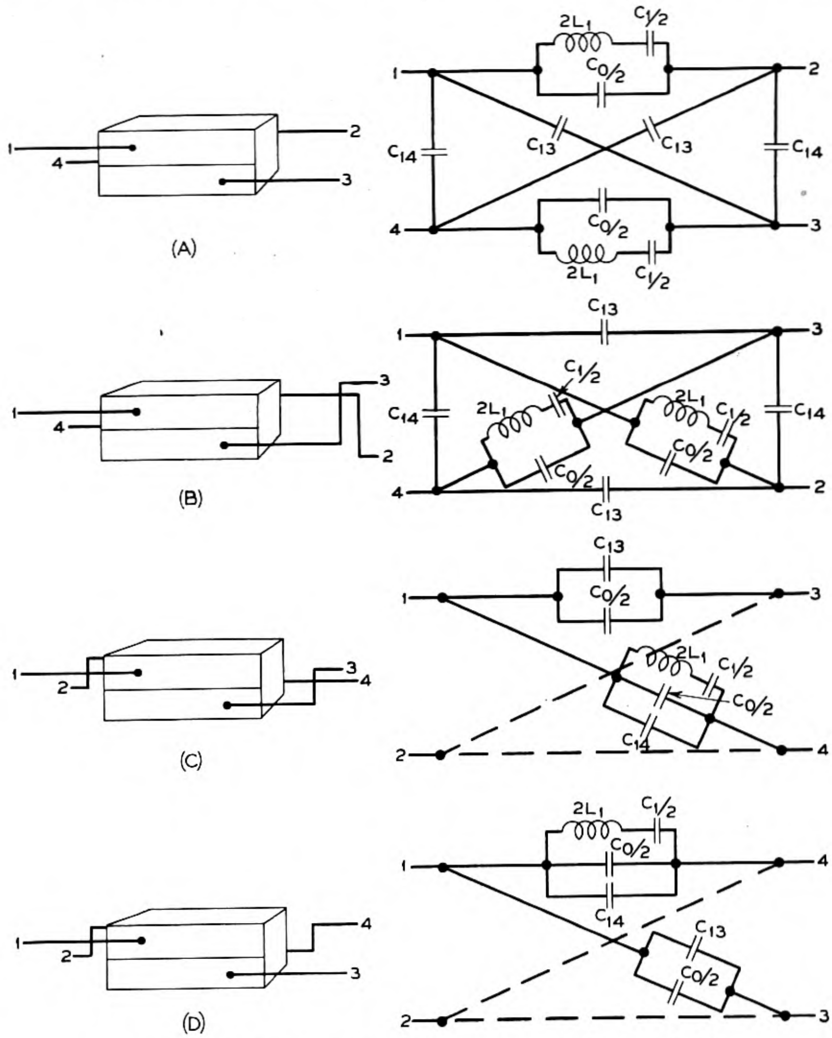


Fig. 16—Balanced divided plate crystal connections and their equivalent lattice electrical circuits.

and hence in obtaining the equivalent lattice they can be connected in directly. The remainder of the circuit can be reduced to its equivalent lattice by employing the equivalence shown in Fig. 8. Taking in the parallel impedance Z and the series impedances $Z/2 + (-j/\omega C_0)$, the network of Fig. 17B results. On account of the paralleling of the $-Z$ and $+Z$ the lattice arm vanishes and the network reduces to that shown in Fig. 16A. In a similar manner the other equivalences result.

The use of divided plating crystals to obtain wide band filters by using series coils to widen the band is obvious. If we connect two

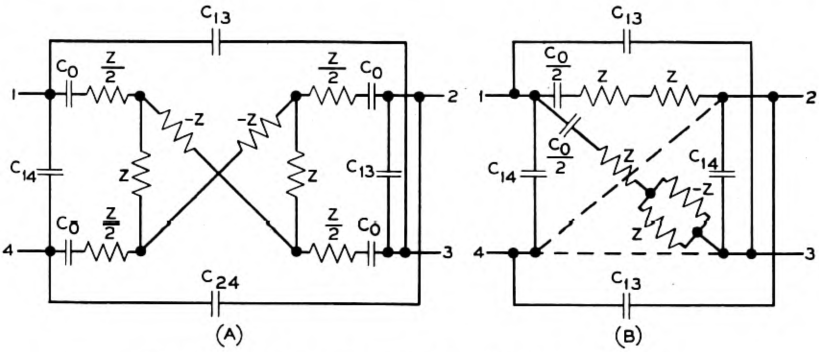


Fig. 17—Method for proving equivalence of Fig. 16A.

crystals as shown in Fig. 18A, one crystal being connected as shown in Fig. 16A and the other in Fig. 16B, a lattice filter equivalent to that is shown in Fig. 18B. In the series arms we have a crystal of twice the impedance of the fully plated crystal Q_1 shunted by the

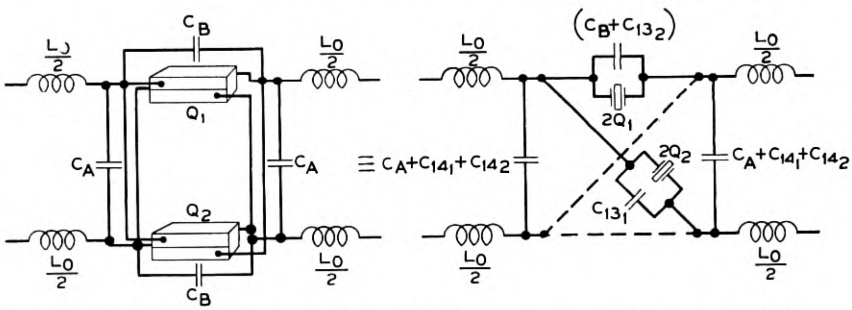


Fig. 18—Band pass crystal filter employing connections of Figs. 16A and B.

capacitance C_B and the capacitance C_{13} of the second crystal Q_2. In the lattice arms we have a crystal of twice the impedance of the fully plated crystal Q_2 in parallel with the capacitance C_{13} of the crystal Q_1. On the ends of the lattice we have capacitances $C_A + C_{14_1} + C_{14_2}$. It is obvious, then, that by using divided plate crystals we can replace two identical crystals in the two arms with crystals having twice the impedance of the fully plated crystals. This result can be utilized in any type of filter where two crystals occur in series or lattice arms of a balanced lattice filter.

The connections C and D of Fig. 16 can also be used to give wide band filters but on account of the extra capacitance shunting each crystal, as wide bands cannot be obtained. This is shown in Fig. 19 which shows two crystals connected as shown in Figs. 16C and D used in a filter. Since each crystal is shunted by half the static capacitance of the other, the ratio of capacitances will be about twice that in the connection shown in Fig. 18 and the band width possible will be about 70 per cent of that shown in Fig. 18. Hence the connections of Fig. 18 are usually desirable.

The connections of 16C and D can be duplicated in unbalanced form as shown in Fig. 20. These equivalents are easily worked out from the network of Fig. 15E by employing Bartlett's theorem. An

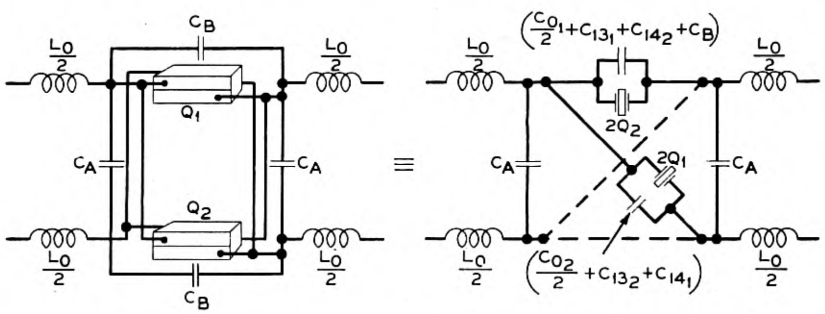


Fig. 19—Band pass crystal filter employing connections of Figs. 16C and D.

unbalanced filter of the type shown in Fig. 19 can be obtained by combining the two connections shown in Fig. 20 as shown in Fig. 21. It will be noted that across the series arm we have a capacitance $C_B + 2C_{14_1} + 2C_{13_2} + C_{0_2}/2$ while the lattice arm has only the capacitance $C_{0_1}/2$. To get attenuation peaks which are separated from the pass band by a large frequency range it is necessary to keep the capacitances C_{14_1} and C_{13_2} small. This can be accomplished by using shielding strips on the plating as shown in Fig. 22 for the two types of connection. In the B connection, the grounding strip is effectively obtained by making the grounded plates 2 and 3 slightly larger than 1 and 4. These grounding strips then act like a guard ring and reduce the stray capacitances.

The same process can be applied to any of the filters of Figs. 2 to 6 to obtain in unbalanced form the characteristics obtained in lattice form. In general the characteristics are somewhat more limited since

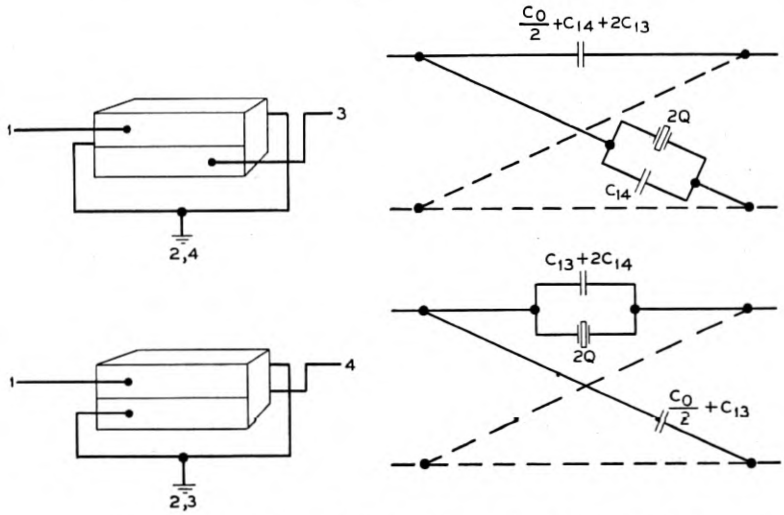


Fig. 20—Unbalanced divided plate crystal connections and their equivalent electrical circuits.

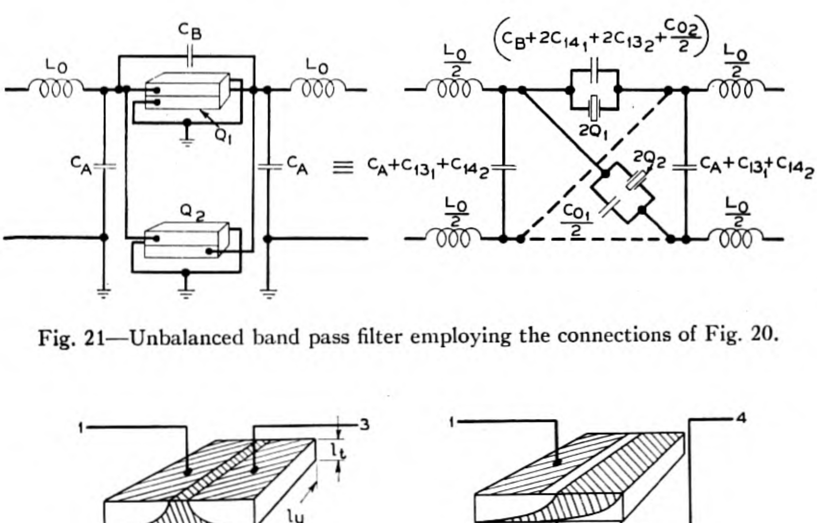


Fig. 21—Unbalanced band pass filter employing the connections of Fig. 20.

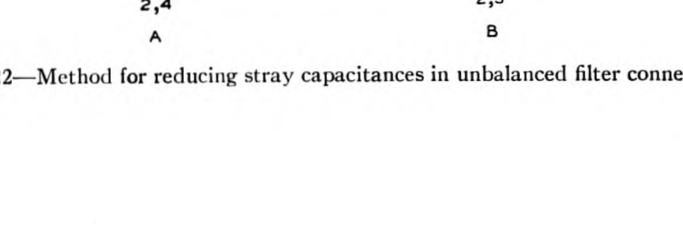


Fig. 22—Method for reducing stray capacitances in unbalanced filter connections.

in effect we have to use crystals with twice the ratio of capacitances than can be used in the balanced case.

APPENDIX

A Determination of the Resonant Frequencies of Lattice Filters

In order to obtain the element values of the filters shown in this paper it is necessary to determine the resonant frequencies of the elements in terms of the desired characteristics of the filter. It is the purpose of this appendix to show how these resonant frequencies may be derived.

The simplest type of band-pass filter section—referred to as the elementary section—is one in which there is one resonance in each arm of a lattice filter as shown in Fig. 23A. The impedance of the

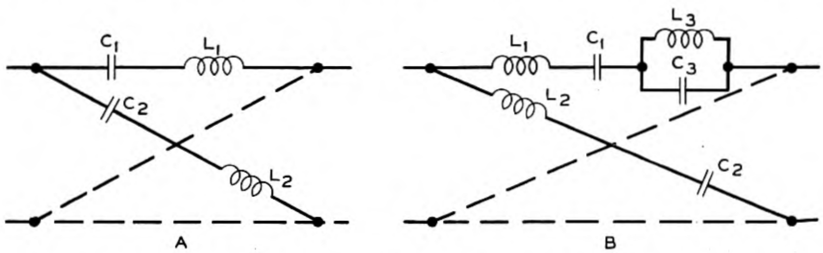


Fig. 23—Lattice filter configuration for elementary band pass sections.

series and lattice arms takes the form

$$Z_1 = \frac{-j}{\omega C_1}\left[1 - \frac{\omega^2}{\omega_A{}^2}\right]; \qquad Z_2 = \frac{-j}{\omega C_2}\left[1 - \frac{\omega^2}{\omega_B{}^2}\right], \tag{10}$$

where ω is 2π times the frequency f, f_A the resonant frequency of the series arm which also is the lower cutoff of the filter, and f_B the resonant frequency of the lattice arm which is also the upper cutoff.

The characteristic impedance and propagation constant are from equation (3)

$$Z_0 = \sqrt{Z_1 Z_2} = \sqrt{-\frac{1}{\omega^2 C_1 C_2}\left[1 - \frac{\omega^2}{\omega_A{}^2}\right]\left(1 - \frac{\omega^2}{\omega_B{}^2}\right)},$$

$$\tanh\frac{P}{2} = \sqrt{\frac{Z_1}{Z_2}} = \sqrt{\frac{C_2}{C_1}\left(\frac{1 - \omega^2/\omega_A{}^2}{1 - \omega^2/\omega_B{}^2}\right)} = m\sqrt{\frac{1 - \omega^2/\omega_A{}^2}{1 - \omega^2/\omega_B{}^2}}. \tag{11}$$

It is desirable to correlate the value of m with the frequency of infinite attenuation in the filter. Since the filter will have an infinite attenuation when $\tanh P/2 = 1$, we have

$$m = \sqrt{\frac{1 - \omega_\infty^2/\omega_B^2}{1 - \omega_\infty^2/\omega_A^2}}, \tag{12}$$

where ω_∞ is $2\pi f_\infty$, where f_∞ is the frequency of infinite attenuation. For a single section since $m = \sqrt{C_2/C_1}$, m must be real and lie between 0 and infinity. The possible attenuation characteristics obtainable with the simple section can be calculated from equation (11). It will be noted that when $\omega = 0$

$$m = \tanh \frac{P_0}{2}. \tag{13}$$

Similar equations for low-pass, high-pass and all-pass filters can be derived from these equations by letting f_A go to zero or f_B to ∞, or both. These equations are:

For Low-Pass Filters

$$\begin{aligned}\tanh \frac{P}{2} &= \lim \omega_A \to 0 \left(\sqrt{\frac{1 - \omega_\infty^2/\omega_B^2}{1 - \omega_\infty^2/\omega_A^2}}\sqrt{\frac{1 - \omega^2/\omega_A^2}{1 - \omega^2/\omega_B^2}}\right)\\ &= \sqrt{1 - \omega_B^2/\omega_\infty^2}\sqrt{\frac{-\omega^2/\omega_B^2}{1 - \omega^2/\omega_B^2}} = m_l\sqrt{\frac{-\omega^2/\omega_B^2}{1 - \omega^2/\omega_B^2}}. \end{aligned}\tag{14}$$

For High-Pass Filters

$$\begin{aligned}\tanh \frac{P}{2} &= \lim \omega_B \to \infty \left(\sqrt{\frac{1 - \omega_\infty^2/\omega_B^2}{1 - \omega_\infty^2/\omega_A^2}}\sqrt{\frac{1 - \omega^2/\omega_A^2}{1 - \omega^2/\omega_B^2}}\right)\\ &= \sqrt{\frac{1}{1 - \omega_\infty^2/\omega_A^2}}\sqrt{1 - \omega^2/\omega_A^2} = m_h\sqrt{1 - \omega^2/\omega_A^2}. \end{aligned}\tag{15}$$

For All-Pass Filters

$$\begin{aligned}\tanh \frac{P}{2} &= \begin{matrix}\lim \omega_A \to 0\\ \lim \omega_B \to \infty\end{matrix} \left(\sqrt{\frac{1 - \omega_\infty^2/\omega_B^2}{1 - \omega_\infty^2/\omega_A^2}}\sqrt{\frac{1 - \omega^2/\omega_A^2}{1 - \omega^2/\omega_B^2}}\right)\\ &= \sqrt{\frac{1}{-\omega_\infty^2}}\sqrt{-\omega^2}. \end{aligned}\tag{16}$$

For this case, since there is no peak in the real frequency range, we must let ω_∞ be imaginary or $i\omega_\alpha$. Then

$$\tanh \frac{P}{2} = \frac{1}{\omega_\alpha}\sqrt{-\omega^2} = m_a\sqrt{-\omega^2}. \tag{17}$$

The band elimination filter cannot be obtained from the band-pass filter by a limiting process. For the simplest band elimination filter with a single peak as shown on Fig. 2, filter 2, the equations are

$$Z_0 = \sqrt{\frac{L_1}{C_2}\left(\frac{1-\omega^2/\omega_A{}^2}{1-\omega^2/\omega_B{}^2}\right)};$$

$$\tanh\frac{P}{2} = \sqrt{\frac{-\,\omega^2 L_1 C_2(1-\omega^2/\omega_B{}^2)}{(1-\omega^2/\omega_A{}^2)}} = \frac{1}{\omega_\alpha}\sqrt{\frac{-\,\omega^2(1-\omega^2/\omega_B{}^2)}{(1-\omega^2/\omega_A{}^2)}}; \quad (18)$$

$$\omega_\alpha = \frac{1}{\sqrt{L_1C_2}} = \sqrt{\frac{-\,\omega_\infty{}^2(1-\omega_\infty{}^2/\omega_B{}^2)}{(1-\omega_\infty{}^2/\omega_A{}^2)}}.$$

Hence when the position of the peak of infinite attenuation and the characteristic impedance Z_0 at zero frequency are specified L_1 and C_2 can be determined.

We next consider the case of a filter with a total of three resonances rather than two. For a band-pass filter this will be represented by the impedance arms shown on Fig. 23B. The impedance of the series and lattice arms will be

$$Z_1 = \frac{-j}{\omega C_1}\left[\frac{(1-\omega^2/\omega_A{}^2)(1-\omega^2/\omega_B{}^2)}{1-\omega^2/\omega_2{}^2}\right]; \quad Z_2 = \frac{-j}{\omega C_2}\left(1-\frac{\omega^2}{\omega_2{}^2}\right). \quad (19)$$

Combining these to form the propagation constant and characteristic impedance we find

$$Z_0 = \sqrt{\frac{-1}{\omega^2 C_1 C_2}(1-\omega^2/\omega_A{}^2)(1-\omega^2/\omega_B{}^2)},$$

$$\tanh\frac{P}{2} = \sqrt{\frac{C_2}{C_1}\frac{(1-\omega^2/\omega_A{}^2)(1-\omega^2/\omega_B{}^2)}{(1-\omega^2/\omega_2{}^2)^2}} \quad (20)$$

$$= B\sqrt{\frac{(1-\omega^2/\omega_A{}^2)(1-\omega^2/\omega_B{}^2)}{(1-\omega^2/\omega_2{}^2)^2}}.$$

We wish to show now that this type of section has an attenuation characteristic equal to that obtained by two sections of the kind shown in Fig. 23A. To show this we write

$$\tanh\frac{P_1+P_2}{2} = \frac{\tanh\frac{P_1}{2}+\tanh\frac{P_2}{2}}{1+\tanh\frac{P_1}{2}\tanh\frac{P_2}{2}}. \quad (21)$$

Substituting the value of $\tanh P/2$ given by equation (14) in (21) and letting the two cutoffs ω_A and ω_B coincide for the two sections, we have

$$\tanh\frac{P_1+P_2}{2} = \frac{m_1+m_2}{1+m_1m_2}\sqrt{\frac{(1-\omega^2/\omega_A{}^2)(1-\omega^2/\omega_B{}^2)}{(1-\omega^2/\omega_2{}^2)^2}}, \quad (22)$$

where

$$\omega_2{}^2 = \frac{\omega_A{}^2\omega_B{}^2(1+m_1m_2)}{\omega_A{}^2+\omega_B{}^2 m_1 m_2}. \quad (23)$$

Comparing (22) with (20) we see that

$$P = P_1 + P_2 \quad \text{and} \quad B = \frac{m_1 + m_2}{1 + m_1 m_2} = \tanh\left(\frac{P_{0_1} + P_{0_2}}{2}\right). \tag{24}$$

We see then that a section with three resonant frequencies can be made to have the same attenuation characteristic as the sum of two simple sections. It is, however, more general since in equations (23) and (24) real values of ω_2^2 and B can be obtained by taking

$$m_1 = m_{1_r} + i m_{1_i}; \qquad m_2 = m_{1_r} - i m_{1_i}; \tag{25}$$

that is, the parameter m_1 can be made complex if the second parameter m_2 is made its conjugate. Such complex sections can be made to have attenuation peaks which are finite even in the absence of dissipation.[7]

By letting $\omega_A \to 0$ or $\omega_B \to \infty$, the equivalent relations for low-pass, high-pass and all-pass filters can be obtained. These are

Low-Pass Filter

$$\begin{aligned} \tanh\frac{P}{2} &= (m_1 + m_2)\sqrt{\frac{-(\omega^2/\omega_B^2)(1 - \omega^2/\omega_B^2)}{(1 - \omega^2/\omega_2^2)^2}}; \\ \omega_2^2 &= \frac{\omega_B^2}{1 + m_1 m_2}; \qquad m_1 = \sqrt{1 - \frac{\omega_{\infty 1}^2}{\omega_B^2}}. \end{aligned} \tag{26}$$

High-Pass Filter

$$\tanh\frac{P}{2} = \frac{m_1 + m_2}{1 + m_1 m_2}\sqrt{\frac{(1 - \omega^2/\omega_A^2)}{(1 - \omega^2/\omega_2^2)^2}}; \quad \omega_2^2 = \frac{\omega_A^2 \omega_B^2 [1 + m_1 m_2]}{\omega_A^2 + \omega_B^2 m_1 m_2}. \tag{27}$$

All-Pass Filter

$$\tanh\frac{P}{2} = (m_1 + m_2)\sqrt{\frac{-\omega^2}{(1 - \omega^2/\omega_2^2)^2}}; \qquad \omega_2^2 = \frac{1}{m_1 m_2}. \tag{28}$$

Band Elimination Filter

For a two-peak band elimination filter such as shown in Fig. 3, filter 2, the equations are:

$$\begin{aligned} Z_0 &= \sqrt{\frac{L_1}{C_2}\frac{(1 - \omega^2/\omega_A^2)(1 - \omega^2/\omega_B^2)}{(1 - \omega^2/\omega_2^2)^2}}; \\ \tanh\frac{P}{2} &= \frac{1}{\omega_\alpha}\sqrt{\frac{-\omega^2}{(1 - \omega^2/\omega_A^2)(1 - \omega_2^2/\omega_B^2)}}; \\ \omega_\alpha &= \frac{1}{\sqrt{L_1 C_2}} = \sqrt{\frac{-\omega_{\infty 1}^2}{(1 - \omega_{\infty 1}^2/\omega_A^2)(1 - \omega_{\infty 1}^2/\omega_B^2)}}, \end{aligned} \tag{29}$$

$$\omega_{\infty 1}\omega_{\infty 2} = \omega_A \omega_B.$$

In a similar manner more sections can be added and the resonant frequencies determined in terms of the cutoff frequencies and the position of the attenuation peaks. The most general section considered in this paper has a maximum of five equivalent sections. For this case by applying the process described above the propagation constant and critical frequencies are given by the equations

$$\tanh\frac{P}{2} = \frac{A+C+E}{1+B+D}\sqrt{\frac{(1-\omega^2/\omega_A{}^2)(1-\omega^2/\omega_3{}^2)^2(1-\omega^2/\omega_5{}^2)^2}{(1-\omega^2/\omega_2{}^2)^2(1-\omega^2/\omega_4{}^2)^2(1-\omega^2/\omega_B{}^2)}}, \tag{30}$$

where

$$A = \sum_1^5 m = m_1 + m_2 + m_3 + m_4 + m_5;$$

$$B = \sum_{m=1}^{5}\sum_{n=1}^{5} m_m m_n; \qquad n \neq m;$$

$$C = \sum_{m=1}^{5}\sum_{n=1}^{5}\sum_{o=1}^{5} m_m m_n m_o; \qquad n \neq m; \qquad n \neq o; \qquad m \neq o;$$

$$D = \sum_{m=1}^{5}\sum_{n=1}^{5}\sum_{o=1}^{5}\sum_{p=1}^{5} m_m m_n m_o m_p; \qquad m \neq n; \qquad m \neq o;$$
$$m \neq p; \qquad n \neq p; \qquad n \neq o; \qquad o \neq p;$$

$$E = m_1 m_2 m_3 m_4 m_5. \tag{31}$$

The resonant frequencies are given by the equations

$$f_2{}^2 = \frac{2f_A{}^2 f_B{}^2(1+B+D)}{f_A{}^2(2+B-\sqrt{B^2-4D}) + f_B{}^2(B+2D+\sqrt{B^2-4D})}, \tag{32}$$

$$f_4{}^2 = \frac{2f_A{}^2 f_B{}^2(1+B+D)}{f_A{}^2(2+B+\sqrt{B^2-4D}) + f_B{}^2(B+2D-\sqrt{B^2-4D})}, \tag{33}$$

$$f_3{}^2 = \frac{2f_A{}^2 f_B{}^2(A+C+E)}{f_A{}^2(2A+C-\sqrt{C^2-4AE}) + f_B{}^2(C+2E+\sqrt{C^2-4AE})}, \tag{34}$$

$$f_5{}^2 = \frac{2f_A{}^2 f_B{}^2(A+C+E)}{f_A{}^2(2A+C+\sqrt{C^2-4AE}) + f_B{}^2(C+2E-\sqrt{C^2-4AE})}. \tag{35}$$

For any smaller number of sections the values can be obtained by letting some of the m's go to zero. For example, for a three section filter $m_4 = m_5 = 0$. For low, high, and all-pass networks the values can be obtained by letting $f_A{}^2 \to 0$, $f_B{}^2 \to \infty$ or $f_A{}^2 \to 0$; $f_B{}^2 \to \infty$.

The Coronaviser, an Instrument for Observing the Solar Corona in Full Sunlight *

By A. M. SKELLETT

Introduction

BECAUSE of the rarity of solar eclipses, their short duration, and their occurrence usually at inconvenient places on the earth's surface, the problem of observing the solar corona in full sunlight is an important one for astronomers. It is also of considerable interest to those telephone engineers who are concerned with radio transmission over long distances. The major disturbances of such transmission have their origin in the sun and studies to date have indicated that a day-to-day knowledge of the activity of the corona might prove useful in predicting the transmission conditions.

The first attempt to solve this problem was made by Huggins in 1878 and since that time every conceivable optical means to accomplish the desired result has been tried. The problem is to observe the corona, not in itself a faint object, through the blinding glare of the sky in the region around the sun. If one holds his hand at arm's length so that it blots out the sun, he will find the glare in the sky around it so intense as to be painful. It is generally at least a thousand times brighter than the corona. The trials have usually been made at very high altitudes where the atmospheric glare is greatly reduced but since the scattered light from the telescope itself, particularly the objective, is some hundreds of times brighter than the corona no success was obtained until M. Lyot[1] invented his *coronographe*, a telescope in which this latter kind of glare is greatly reduced. With this instrument at the top of Mt. Pic du Midi in the Pyrenees mountains he has obtained several photographs which show some of the features of the inner corona. At best he has to work with a glare that is nearly as bright as the brightest parts of the corona. There are only a few days in the year when the intensity of the glare is low enough to enable him to observe coronal features through it.

It is obvious that a method of greater discrimination is needed if day-to-day observations are to be made. Such a method was proposed several years ago.[2] It is based on the use of television technique;

* Presented at mtg. of Nat. Acad. Sciences, Providence, R. I., October 1939, and before Amer. Philos. Soc. in Philadelphia, Pa., November 1939.

[1] Lyot, B., *M. N. R. A. S.*, **99**, 8, 580, 1939.

[2] Skellett, A. M., *Proc. Nat'l. Acad. Sci.*, **20**, 461, 1934.

the corona is separated from the glare by electrical filters while the image of the sky around the sun is temporarily represented by an electric current.

Apparatus and Technique

Dr. G. W. Cook, Director of the Cook Observatory, kindly offered the use of his 15-inch horizontal telescope for a trial of the method. Special television apparatus was developed at the Bell Telephone Laboratories for use in conjunction with this telescope. This apparatus has been called the coronaviser. Figure 1 shows a layout of the complete apparatus. Starting at the left, the plane mirror M is coupled to driving mechanism to form a siderostat so that sunlight may be continuously held on the axis of the telescope. The solar disc, about 2″ in diameter, is focussed by the objective of approximately 18 feet focal length to fall on the tilted mirror which reflects the direct sunlight out through a hole in the side of the telescope and into a light trap consisting of a black walled tube with a black velvet end. Figure 2 shows the scanning mechanism in greater detail. Immediately in back of the mirror R, which is the one just referred to for throwing sunlight out of the telescope, a black masking disc D further prevents sunlight from getting into the scanning apparatus. Several different sizes of this disc are used to take care of the different diameters of the solar image which occur throughout the year and to shield the scanner from sunlight which spills over with bad seeing. This mirror and disc are supported by means of the plate glass P so that there is no obstruction in the field around the sun.

To the right of the plate glass lies the scanning apparatus. This is a mechanical device which scans the region of the sky around the sun in a spiral path. The scanning motion thus consists of a circular and a radial motion.

The simple plano-convex lens L which is silvered on the back is equivalent to a concave mirror and forms an image of a portion of the sky image that lies in the plane of D on the scanning hole H which is on the axis of the telescope and scanner. The light that enters the scanning hole passes through the lens, the prism, and the light tunnel U into the photo-cell E. When the lens L is rotated about the axis by the motor, the effect is the same as moving the scanning hole around in the image plane at D. This takes care of the circular component of the scanning motion.

The radial component is obtained by changing the angle of tilt of the lens L while it is rotating. A worm W is mounted on the motor shaft but held stationary so that as the gear G revolves as a whole

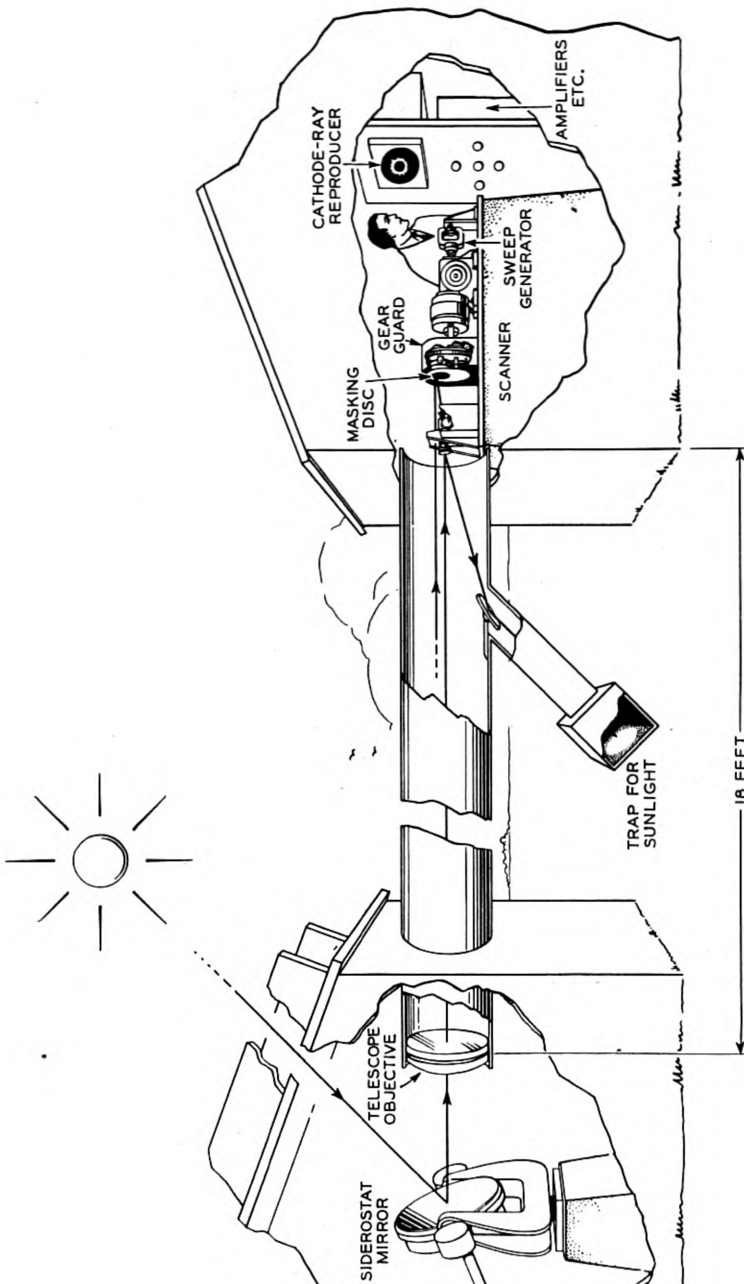


Fig. 1—Layout of apparatus at the Cook Observatory.

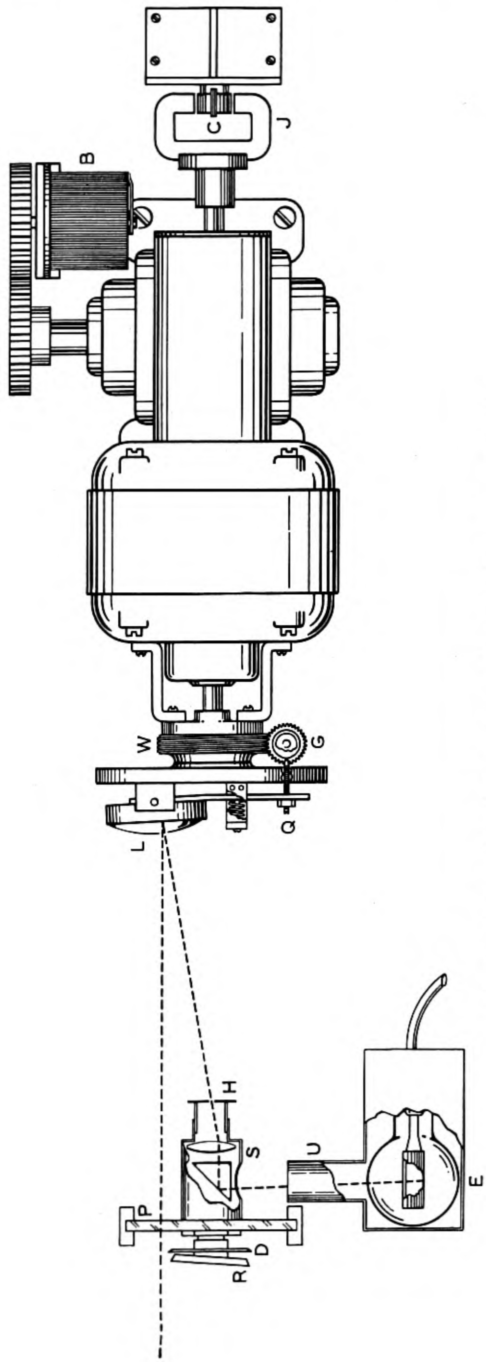


Fig. 2—The mechanical scanner and sweep frequency generator.

about the axis of the scanner it is turned more slowly about its own axis. A hardened pin Q attached to the arm of the lens mounting rides on a cam which is fixed to this gear and thus imparts a cyclic tilting motion to the lens. There are another similar gear and cam (not shown) mounted opposite this one which serve only to compensate for the weight of the working gear so that the rotating mechanism may be balanced to a high degree of precision.

The lens behind the scanning hole forms a stationary enlarged image of the scanning hole on the photosensitive surface of the photoelectric cell.

The reproduction of the image is done entirely by electrical means. At the opposite end of the motor a C-shaped permanent magnet rotates about a cylindrical iron core around which there is wound a coil of wire. The electric current thus generated in the coil is smoothed into a true sine wave by tuning the circuit of the coil so that it resonates at the frequency of circular scan which is approximately 30 cycles per second. The resulting wave is split into two components 90° apart in phase and these, after amplification, are impressed on the deflector plates of the reproducing cathode ray tube so that the spot may move in a circular path.

At the generator end of the motor there is also a reduction gear box, the slow speed shaft of which turns once for a complete cycle of the radial scanning motion, i.e. once a second. This shaft is geared to that of the potentiometer unit B so that its sliding contact also revolves at this rate. The potentiometer winding is continuous around the circle and connections are made at the opposite ends of a diameter. The scanning voltage from the generator circuit is fed into this unit and as the arm revolves the amplitude of the sine wave is made to vary uniformly between its minimum and maximum values. Thus the scanning spot spirals outward, scanning a complete image of the field, and then inward, giving another complete image, so that there are two images per cycle of the radial scanning motion. The spot does not follow the identical path on the two scans; the outward scan crosses over the lines of the inward scan along one diameter and interlaces along the diameter at 90 degrees to this. The radial resolution along the latter diameter is therefore double that along the former. Since the frequency of the radial component of the scanning motion is approximately one cycle per second, these two resolutions are 15 and 30 lines respectively.

The glare of a clear sky is uniform around the solar image and therefore as the scanning spot travels around the field it gives rise mainly to a direct current in the photo-cell. The coronal features, however, that is the streamers, arches, etc., give rise to alternating components and only these components are amplified, the direct current being eliminated by resistance capacity coupling of the cell to the amplifier. Inaccuracies of alignment and any non-uniformity of the intensity of the glare across the field give rise to strong low-frequency components and because of the high glare levels that were occasioned by the location and by the siderostat mirror it was found necessary to filter these components out of the amplified currents. A high-pass filter is inserted between the first and second stages of the amplifier for this purpose. A low-pass filter with a cut-off at 3750 cycles is also included to cut out the noise at frequencies above the desired band. The top frequency is 3600 cycles for the 0.04 inch diameter scanning hole that was generally used. After amplification the signal current is made to modulate the intensity of the cathode-ray beam.

The electrical frequency spectrum of the television image consists of the fundamental scan frequency (about 30 cycles) plus a large number of its harmonics. By varying the characteristics of the high-pass filter it was possible to eliminate the fundamental and several of the lower harmonics as desired. This became advantageous in studying the prominences and smaller coronal details which give rise almost entirely to higher harmonics and are reproducible therefrom with a good degree of accuracy. In addition, for these smaller details, the coupling capacity between the first and second stages of the amplifier was greatly reduced so that the gain at the upper end of the band was considerably enhanced in relation to that at the lower end.

The light of the corona is practically identical with that of the sun in its spectral characteristics and a caesium sulphide photo-cell which has a maximum sensitivity in the green was used. See Fig. 3. It was found that by using gas amplification in the photo-cell adequate sensitivity was obtained. The inner corona has a surface brightness of about the same magnitude as the full moon and the sensitivity of the apparatus was checked by obtaining images of the moon in its various phases.

It is convenient to measure light intensity levels in millionths of the brightness of the sun's surface. The brightness of the full moon is about 2 millionths and it is known that the inner corona falls off fairly rapidly with distance from a brightness of a little more than one millionth measured within one minute of the solar limb. The level of the glare was measured from time to time by means of a photronic cell behind an aperture placed at the focus of the objective. On days when the haze in the sky was very noticeable but yet not in the form of clouds its brightness at 2 minutes from the limb was as high as 6000 millionths or approximately 6000 times as bright as the corona. On very clear days the brightness was as low as 1250 millionths and this limit may have been set by the scattered light from the telescope parts, particularly the siderostat mirror, rather than the sky. The

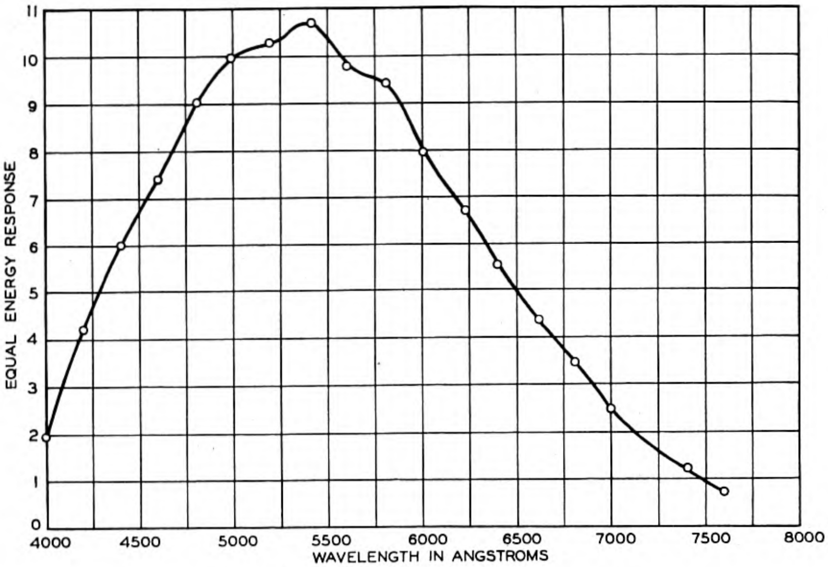


Fig. 3—Spectral characteristic of photo-cell.

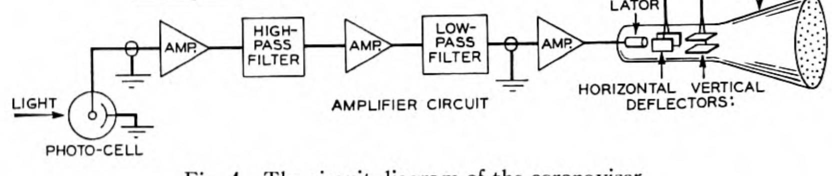


Fig. 4—The circuit diagram of the coronaviser.

siderostat mirror had a surface of evaporated aluminum but by the time the development of the equipment had progressed to the point where useful images could be obtained, its surface had begun to show blemishes.

The scanning rate was so slow (2 images per second) that even with a cathode ray screen having an appreciable time lag there was difficulty in studying the complete image visually. The expedient was therefore adopted of photographing the images and this proved advantageous in comparing images as well as in greatly reducing the effects of noise. The latter advantage was realized by taking exposures of from 20 to 30 seconds during which there were reproduced some 40 to 60 images and although the noise patterns might be very noticeable in a single scan the fluctuations balanced out in a statistical manner, leaving a uniform field.

Theoretically the limiting amount of glare through which it is possible to work with this method is determined by the shot noise in the photoelectric current but there are practical limits, set mainly by the cleanness of certain of the optical parts, which were the important factors in this case. Although the whole scanner is designed to reduce, as much as possible, the scanning of the optical parts of the instrument by the beams of light, certain parts are of necessity scanned, particularly the plate glass used to support the scanning hole unit. This plate is also near the focal plane and the slightest smudge or speck of dust on its surfaces gives rise to an overloaded image on the cathode ray screen. The glass itself was specially selected to be free from bubbles or blemishes of any kind and in addition it was carefully washed at frequent intervals. Great pains had also to be taken in keeping the other glass surfaces in the optical train clean, for the essence of the method is the amplification of minute variations in the intensity of the illumination from point to point in the field.

Occasionally tiny specks of brilliant light would float across the screen, the sources of which were very puzzling at first. They were finally traced to insects or wind-borne seeds which drifted across the sky in the path of the shaft of light. Being illuminated by direct sunlight, they scattered enough light in the direction of the telescope to give a bright diffraction pattern. They ruined quite a few plates.

Since the glare decreased in the direction outward from the sun (though not so rapidly as the coronal light), patterns that were caused by instrumental defects took on at times appearances which might easily have been confused with that of a coronal image. It was necessary, therefore, to have an absolute criterion by which one could distinguish between these spurious images and those which were associated with the sun. The siderostat mounting of the telescope furnished such a test. With this type of mounting the celestial field rotates about the optic axis of the telescope with time. This rate varies with the declination of the object and other factors and in our case it ranged within a degree or two of 7° per hour. Thus by taking a series of photographs over a period of several hours it was possible to determine definitely whether or not the image in question was associated with the sun or with the apparatus. In addition to this test, for the prominences, there was the spectrohelioscope at hand by which a direct comparison could be made. Another test applied to the prominence images was furnished by their color. A red glass filter, such as the Schott RG 2 which has a cut-off just below the H_α line, reduced the general glare level by about 30 times whereas its reduction of the light of the prominences which is a maximum at this wave-length was not nearly so great.

Results

The prominences shown in Fig. 5 are among the first of which good images were obtained. Seven photographs were taken of them be-

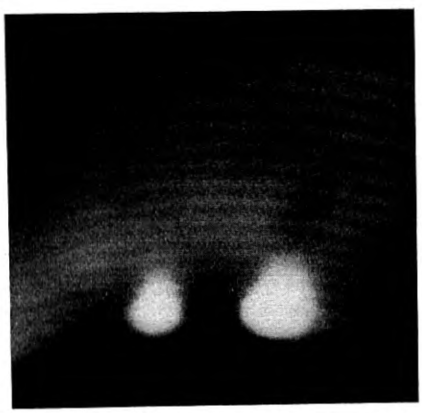


Fig. 5—Prominences taken with red filter on Feb. 21, 1938.

tween $16^h\ 58^m$ and $19^h\ 11^m$ G.C.T. on February 21, 1938, some in white light and the others with the red filter in front of the photo-cell. This particular photograph was taken in red light; those taken in white light were of considerably less contrast.

Figure 6 is another one of the many prominence photographs that have been taken with the apparatus. These are the prominences that were present around the sun at $18^h\ 30^m$ G.C.T. on October 31, 1938. This was also taken with the red filter.

Figure 7 shows a pair of bright prominences photographed in white light on October 3, 1938.

Figure 8 shows a jet or flare in the corona that was photographed on October 18, 1938. It is one of 11 photographs that were taken

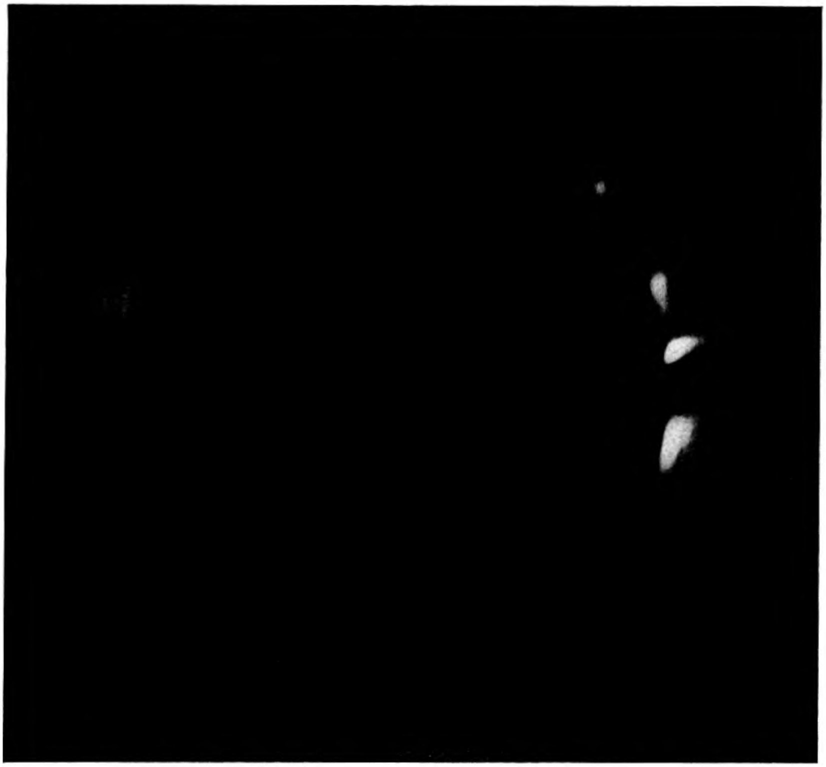


Fig. 6—Prominences around the sun on Oct. 31, 1938. This image was obtained with a scanning hole 0.013 inch in diameter which had 1/10 the area of that used for the other photographs.

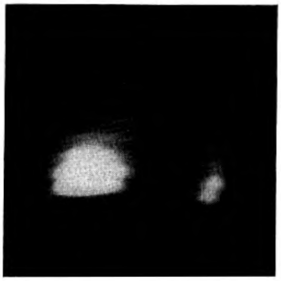


Fig. 7—A pair of prominences taken without optical filter (i.e., in white light) on Oct. 3, 1938.

Fig. 8—A jet or flare in the corona photographed on October 18, 1938.

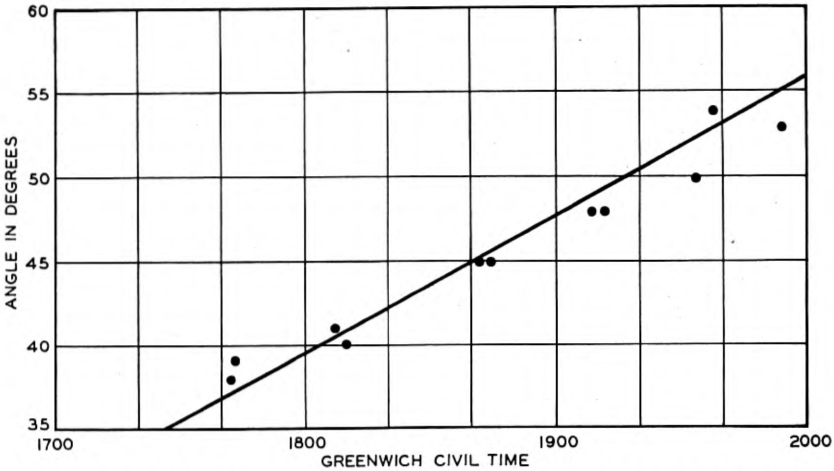


Fig. 9—Positions of the flare of Fig. 8 at the times shown. The turning of the image is due to the horizontal mounting of the telescope.

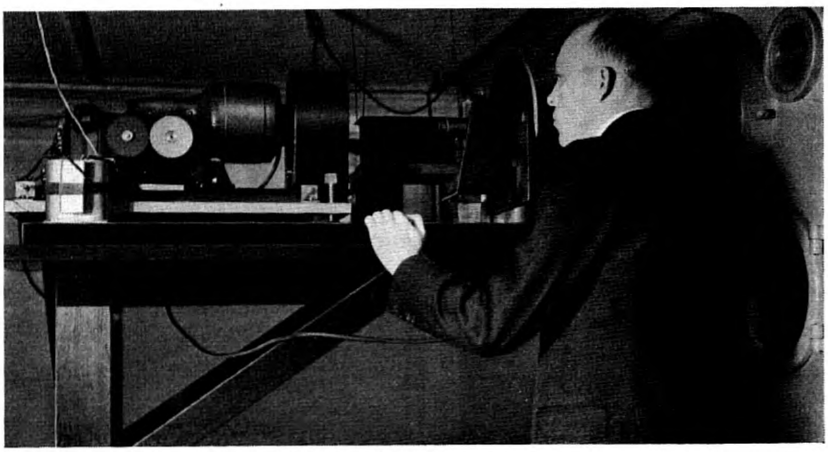


Fig. 10—The input scanner.

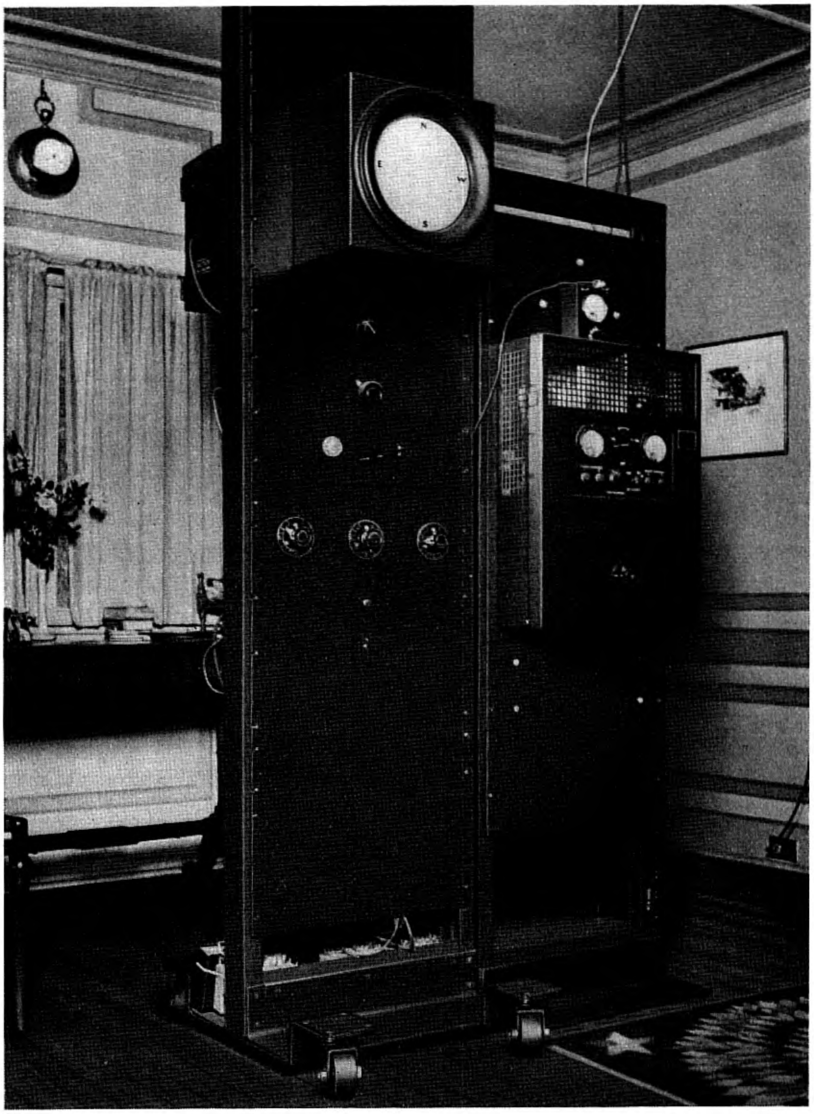


Fig. 11—The amplifying and sweep circuits and reproducing tube.

over a period of more than 2 hours. All these show this feature, and its position on each plate is plotted against the time of exposure on Fig. 9. The slope of the line is the correct rate for the turning of the celestial field in the neighborhood of the sun at that time. It was not brighter on the plates that were taken with the red filter than on those taken with white light, and it is concluded, therefore, that it was white in color.

There was no prominence at its position on the limb of the sun, but the next day the observatory at Huancayo reported that an eruptive prominence blew off from the limb at this position (28° N. Lat.). It seems quite probable that the unusually bright jet in the corona was lying over and probably associated with this active region.

A number of other plates have shown details which appear to have their origin in the corona, but generally they have been partly obscured by other patterns of instrumental or other origin.

The major objectives of this phase of the work which were the development of an adequate instrument and the proving in of the method have been achieved. The next phase of the investigation should be carried out under the most favorable conditions possible and this means a location on a mountain top with a telescope preferably pointing directly at the sun.

I wish to acknowledge the helpful cooperation of Dr. Cook and his associates at the Cook Observatory and of many of my colleagues in the Bell Telephone Laboratories. In particular, Dr. J. B. Johnson has greatly contributed to the investigation by his counsel and aid.

Lead-Tin-Arsenic Wiping Solder *

By EARLE E. SCHUMACHER and G. S. PHIPPS

SOME fourteen or more wiped joints occur in every mile of lead-sheathed telephone cable, and in making these joints from one to two million pounds of solder are used per year. To join cables a lead sleeve of sufficient diameter to accommodate the bundle of spliced wires is slid in place at the junction, the ends of the sleeve are beaten to conform to the circumference of the cable, and an air-tight and mechanically strong joint formed at each end of the sleeve by molding a solidifying mass of solder into the desired shape. This last step is called the wiping operation.

The making of a successful wiped joint depends upon a satisfactory composition in the solder and considerable skill on the part of the splicer. The two factors are inter-related in that the more dextrous operators can produce satisfactory joints with solder compositions which could not be shaped by the average operator. The most satisfactory composition for a wiping solder from practical tests has been found to be about 38 per cent tin, 62 per cent lead. A solder containing 40 per cent tin also possesses satisfactory handling qualities and is used to some extent. If the tin content is much above 40 per cent the workable temperature range in which the solder is plastic becomes too limited for practical handling. The plastic range can be increased by increasing the lead content above 62 per cent but then it is found that the joint becomes coarse-grained and porous. The highest practicable lead content is of course advantageous from an economic standpoint.

The impurities allowable in a wiping solder are also closely controlled since in general small percentages of most impurities have been found to have a harmful effect upon the handling character of the solder or the properties of the joints. One exception, which has hitherto not been recognized, is arsenic whose beneficial effects in small quantities are discussed in this paper.

An engineer would prefer to interpret the handling of a wiping solder in terms of basic properties which can be measured in the laboratory. Such attempts [1] have been made but with only limited success since

* *Metals and Alloys*, Vol. 11, pp. 75–76, March 1940.

[1] "Some Physical Properties of Wiping Solders," D. A. McLean, R. L. Peek, Jr., and E. E. Schumacher, *Journal of Rheology*, Vol. 3, January 1932, p. 53.

not only does the wiping process itself not admit of scientific measurement but also the basic properties related to the process are difficult to determine and are of restricted practical bearing. These difficulties arise because a complex solid-liquid system is involved, and during most of the time of wiping the joint the system is far from being in an equilibrium condition.

Experience has shown, however, that a wiping solder should possess certain general characteristics which are enumerated below, although in many instances the characterization cannot be extended beyond a qualitative statement.

1. The temperature at which the solder begins solidification should be lower than the temperature of beginning melting of the cable sheath and sleeve. The temperature of beginning solidification for the 38–62 tin-lead solder is 240° C. while the lead alloy sheath begins to melt at approximately 310° C. Since with this solder no trouble is encountered with melting sheath it appears that a 70° differential is satisfactory.

2. The solidification range of the solder should be such as to provide, during cooling, an ample forming period between the time when enough primary lead has precipitated to give sufficient body to permit forming to begin until the mass is too solid to manipulate. In the 38–62 wiping solders the solidification range is approximately 60° C. while the forming range is about 40° C.

3. The tendency for the joint to drain or slip and break apart during wiping should be a minimum. These properties are associated with surface tension and plasticity. The desired condition is sometimes referred to as a "buttery" texture.

4. The solder should readily wet and alloy with the parts to be joined. This implies a freedom from non-reducible oxides in the solder and a minimum in the tendency to form reducible oxides. Suitable non-corrosive fluxes are used to clean the surfaces being joined and reduce the oxides which cannot be entirely excluded from the melted solder. To facilitate handling, the solder should not adhere to the splicers' wiping cloths.

5. The solder should be such that the strength of the joint formed should be equal to or greater than that of the parts being joined. The joint must also be gas-tight. This property is secured through a fine texture in the solder and freedom from draining of the lower melting constituents. The test for porosity of a joint is simple. A positive gas pressure of from six to nine pounds per square inch is applied inside the sleeve and soap-suds are then painted on the joint. Observation of the soap-suds will then show whether or not the joint is porous.

6. The health hazard under normal conditions of use should be negligible.

An investigation was made of modifications of the lead-tin type wiping solder to produce a solder which more fully meets the practical requirements than does the present compositions. Through the addition of 0.1 per cent of arsenic substantial improvements have been

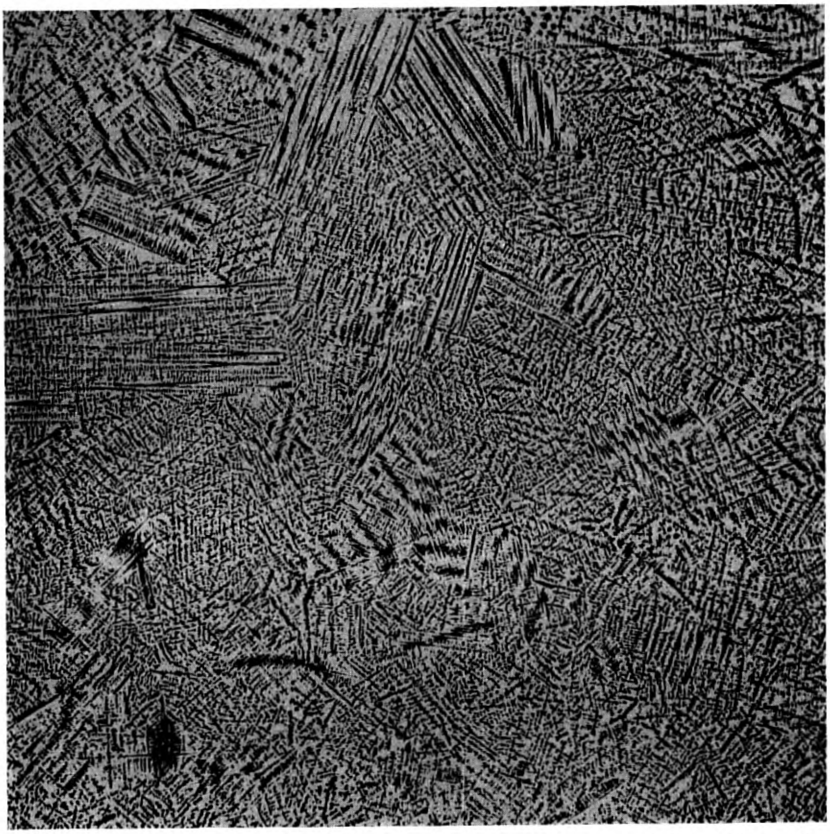


Fig. 1—Photomacrograph of a nominal 38% tin–62% lead wiping solder. The sample was slowly cooled from 300° C. Specimen etched with a mixture of 4 parts of glacial acetic acid and 1 part of 30% hydrogen peroxide.

achieved in certain of the characteristics. The nominal composition of an alloy which shows this improvement is tin—37.25 per cent, arsenic—0.1 per cent, and balance lead. The effect of arsenic manifests itself at percentages as low as 0.04 per cent. Amounts much over 0.1 per cent should be avoided because of a tendency to segregation at the higher percentages. As the tin is reduced below the recommended percentage the difficulties in producing a satisfactory joint increase. There is the possibility, however, that a lower tin content can be tolerated but definite conclusions await further substantiation in subsequent investigation.

Although the percentage of arsenic in the new solder is relatively small it has two important beneficial effects. The amount of dross formed on the arsenic-bearing solder is but a fraction of that ex-

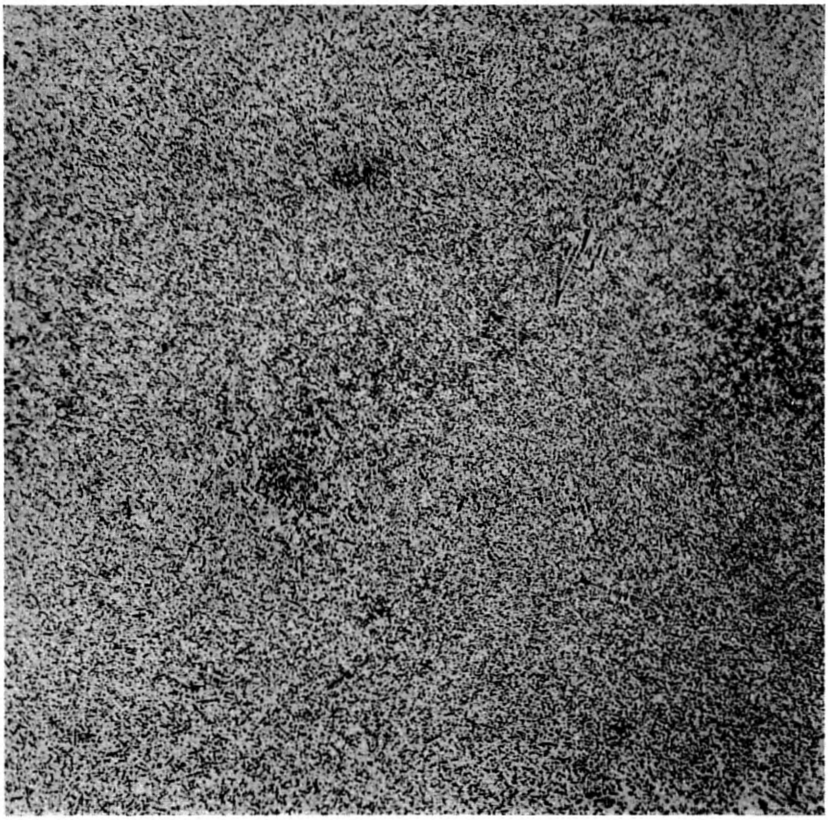


Fig. 2—Same as Fig. 1, except for an addition of 0.1% arsenic. Note the finer grain.

perienced in the ordinary lead-tin solders. The practical advantage is that less time need be spent in skimming the molten solder in the melting pot, and there is less possibility for the inclusion of dross in the finished joints. The presence of dross in the wiped joints is to be avoided because of its possible contribution to porosity.

The second beneficial effect of arsenic is related to the grain size in the solidified alloy. This is illustrated by the photomacrographs shown in Figs. 1 and 2 which compare the grain structure of a lead-tin solder and an arsenic-modified solder which have undergone similar handling and cooling treatments. The arsenic-bearing solder exhibits a finer and more uniform texture. The finer texture is associated with improved handling characteristics and freedom from porosity in the finished joint. The texture of the solder is more buttery in the wiping process and there is less tendency for the lower freezing constituents to drain from the partially finished joints than with the lead-tin compositions. Practically, this provides for the splicer a longer forming range, although the solidification range is materially the same as for the corresponding unmodified alloy.

Although the mechanism by which arsenic reduces the size of the dendrites and refines the grain of wiping solder has not been definitely established, it seems probable that it either provides new and more numerous nuclei of crystallization around which the primary lead precipitates, or imposes barriers against the growth of the primary lead crystals or both. Arsenic forms a compound, Sn_3As_2, with tin and it is probable that this constituent is responsible for the mechanism postulated. In slowly cooled solders this compound is discernible beginning at approximately 0.1 per cent arsenic. With a greater number of precipitated crystals present a greater surface is made available to which the molten eutectic may cling, producing a readily formable mass.

Laboratory observations on the solder have been verified by field tests in the Bell System. The arsenic-bearing solder is handled in the same way as the ordinary lead-tin compositions. Joints have been wiped on large and small aerial and underground cables and in difficult situations involving large branched joints. The consensus of splicers from several different localities is that the solder possesses handling properties superior to those of the lead-tin compositions. These joints in all cases were pressure tested after wiping and found to be sound.

Regarding the possible health hazard involved in using the new solder, tests have been made to determine whether arsenic or arsenic compounds would be volatilized from this alloy under the conditions encountered in practice. These tests gave entirely negative results and showed that no additional hazards would be introduced by substituting arsenic-bearing solder for standard solder.

Nuclear Fission

By KARL K. DARROW

This article pertains to the most newly-discovered and most sensational mode of transmutation, in which the entry of a neutron into a massive atom-nucleus brings about an internal explosion in which the nucleus is "fissured" or divided into two fragments which share the total mass and charge between them in nearly equal proportions. (In all other modes of transmutation except those affecting the very lightest elements, the division is into fragments of very unequal mass and charge.) The conversion of rest-mass into kinetic energy, or (as is more commonly said) the release of energy, is unprecedented in scale. A multitude of radioactive bodies, many hitherto unknown, is formed; and there is spontaneous emission of fresh neutrons in great quantities, possibly sufficient to convert the process once initiated into a self-perpetuating one under realizable conditions.

EVERY now and then a physicist is liable to receive a letter from some yearbook or other, in which he is invited to write x thousand words on the "most important developments in physics during the year just ending." The only safe reply is of course that for ten years at least and perhaps for fifty it will be impossible to tell which is the most important development in physics during the year just ending. This year, however, it looks as though one need not be so cautious; for ever since the first few weeks of the year many have felt pretty sure that one particular discovery would long be recognized as the most important to be made, or at any rate to be revealed, in 1939. It came early—the first publication was on the sixth of January, and there was a rain or perhaps I should say a deluge of others before the end of February. Inasmuch as these others proceeded from laboratories sprinkled all of the way from Copenhagen to Berkeley, it is literally true for once that a discovery commanded immediate attention. Nor is attention even yet diverted, though the pace of publication has grown less.

The phenomena of fission are as yet confined to the last three elements of the periodic table: thorium, protactinium, uranium. I show their chemical symbols, their atomic numbers or nuclear charges, and the mass-numbers—to wit, the nearest integers to the actual values of the masses—of their several isotopes (charges expressed of course as multiples of e, masses as multiples of one-sixteenth the mass of the commonest kind of oxygen atom); charge appears as a subscript before the symbol, mass as a superscript after it:

$$_{90}Th^{232}; \qquad _{91}Pa^{231}; \qquad _{92}U^{234}; \qquad _{92}U^{235}; \qquad _{92}U^{238}$$

From this list I omit several very unstable isotopes of which we shall probably never be able to assemble enough to observe their fission. Protactinium 231 is itself so rare that only one man in the world (he is von Grosse, of Chicago) ever got enough of it together for this experiment. He brought his precious sample—less than 9 mg.—to Dunning at Columbia for the test, and the three of them found fission. I make this allusion at the start, because there will be little further occasion to refer to protactinium, and yet it should not be forgotten. There is danger of forgetting even thorium, since so disproportionately great an amount of study has been lavished on uranium. Neither thorium nor uranium is a very rare element, but more than 99 per cent of any sample of uranium consists of the isotope 238, so that the two other isotopes must also be classed as rare; yet it is believed at present that 235 is responsible for some of the most remarkable of the phenomena.

Now let me indicate two qualities shared by all five of these nucleus-types. First: all are radioactive, that is to say, they are unstable. I must not be too emphatic with this word; the average lifetime of nuclei of either Th^{232} or U^{238} is hundreds of millions of years, and there are not many organizations which would be considered unstable if they could bank on a probable lifetime of that scale. Still they are, in the physicist's sense, unstable; and this suggests that it might be relatively easy to disorganize, to disrupt, to explode them by a fitting agency coming from without.

Now to any physicist the term "fitting agency coming from without" suggests at once the bombarding particles by which transmutation was first effected: alpha-particles, protons, deuterons—the positively-charged nuclei of the elements helium and hydrogen at the other end of the periodic table from uranium. Should one not project these nuclei against uranium nuclei, and see what happens? Well, it has often been done, and nothing has happened; [1] and an adequate reason is supplied by the second important quality of these five nucleus-types, their greatness of atomic number. All of them are so highly charged with positive electricity, that the proton, the deuteron and the alpha-particle, however fast they are when they start, cannot approach them closely enough to do them any harm. (What with the current progress in cyclotrons that statement may soon be out of date!) Even with our

[1] Until in October of this year Gant reported strong indications of fission of uranium produced by very energetic (8-Mev) deuterons.

present resources of energy we could not tamper with any of these nuclei, had we not at our disposal those chargeless particles the "neutrons" with which to assail them.

One might of course foretell that mighty powers of transmutation would be possessed by a particle which is *not* repelled as it approaches a nucleus. Actually the transmuting powers of the neutron are greater than, I should think, anyone can have expected; nor can many people, if any, have foreseen that the *slow* neutron—the neutron having no more speed and kinetic energy than a molecule of air at room-temperature—would prove to be more potent than the fast one. Yet so it is. When the other agents of transmutation were first applied—alpha-particle, proton, deuteron, photon, fast neutron—it took years to get proof of the transmutation of even a few elements; but when the slow neutron was first applied, Fermi and his half-a-dozen colleagues at Rome managed to do something to almost every element in a very few months! Let me recall that neutrons mostly are what we call fast—i.e., they have energies of millions of electron-volts—when they start their careers. Slow neutrons are initially-fast ones which have been sent through layers of paraffin or water, and have lost nearly the whole of their initial energy by making elastic impacts with hydrogen nuclei. We shall later have to distinguish between the fast neutrons and the slow as agents of nuclear fission.

Now to supply a fitting historical background to the discovery of fission, I must draw attention to a theorem which until the end of 1938 was believed to govern the whole of transmutation, and which still governs nearly the whole of the field. It is this: with the exceptions presently to be related, *no transmutation ever produces a change in atomic number greater than 2 or a change in mass-number greater than 4.* I am going to illustrate this theorem by writing in symbolic form three of the reactions of transmutation produced by neutrons and recognized before the end of 1938. Here I use E as the general symbol for element; Z and A as the general symbols for atomic number and mass-number; and α, p, d, n, and ϕ for alpha-particle, proton, deuteron, neutron and photon; and I recall that the mass-numbers of these five particles are 4, 1, 2, 1, 0 respectively.

$$ {}_{Z}E^{A}\ (n,\alpha)\ {}_{Z-2}E^{A-3}; \qquad {}_{Z}E^{A}\ (n,p)\ {}_{Z-1}E^{A}; \qquad {}_{Z}E^{A}\ (n,\varphi)\ {}_{Z}E^{A+1}. $$

The first of these (for example) is to be read: a neutron enters a nucleus (Z, A) and an alpha-particle comes out, leaving behind a nucleus $(Z - 2, A - 3)$. There was a similar (not identical) rule setting a limit on the changes of atomic number and mass-number suffered by radioactive bodies. Every radioactive nucleus emits either a positive electron or a negative electron, or an alpha-particle; the corresponding changes in Z are -1, $+1$, -2, those in A are 0, 0, 4.

Now it is high time that we get on to uranium. This, like thorium, was one of the elements that Fermi exposed to slow neutrons, and to which he observed that something was happening. As he continued these researches, and as the great institutes of nuclear physics of Hahn in Berlin and the Joliots in Paris followed suit, it became evident that a great deal was happening. With nearly all of the other elements, there occurred just one of the reactions which I symbolized above, or maybe one reaction with some of the nuclei and another with others. Sometimes the reaction would lead to a stable nucleus-type; in such a case, when the neutron-bombardment ceased all the excitement was instantly over. Sometimes it would lead to a radioactive nucleus-type; in such a case, the radioactivity would continue after the bombardment ceased, but it would steadily die away to nothing, in accordance with the well-known law. But with uranium and thorium there was a swarm of radioactive products, so many that to this day they have not all been separated and identified. Moreover, some of these were descendants of others, for after the bombardment ceased they grew in strength for a while before declining. All of them were emitters of electrons, and the electrons (in every case in which their sign is known) were *negative*.

Owing to the theorem which has just been stated, it was taken for granted that the immediate effect of the neutron entering the nucleus of uranium was to provoke one of the three reactions which I lately listed. Of these the one most commonly assumed was the reaction,

$$_{92}\mathrm{U}^{238}\ (n, \phi)\ _{92}\mathrm{U}^{239}$$

the so-called "reaction of pure neutron-capture"—called pure, because no massive particle goes forth. I mention it here because it is still believed in, and we shall meet it later. Uranium 239 is radioactive, and some of the other radioactive products were believed to be direct or indirect descendants of it. Well, every one of the radioactive substances resulting from the reaction or reactions $U(n)$—for so I will symbolize them in general—is an emitter of *negative* electrons. Therefore each has a greater positive charge on its nucleus than does its predecessor; therefore by this theory, each descendant from U^{239} must have a greater positive charge than the $92e$ of the uranium nucleus. But uranium is the final element of the periodic table; therefore by this theory the radioactive bodies in question had to be isotopes of new elements beyond uranium.

These so-called "trans-uranic elements" were for several years the principal study of Hahn and Meitner and their colleagues at the great institute in Berlin-Dahlem. The groups at Paris and at Rome contributed also—not very much, but enough to signify their full adhesion to these concepts. Other physicists scarcely entered the field, but had the fullest reliance in views sustained by such authorities. Yet now, the trans-uranic elements are gone! This is regrettable, because it was pleasant to think that human artifice had succeeded even in lengthening the list of the elements. It is regrettable for the chemists especially, because they were looking forward to getting information about the chemical properties of elements beyond 92. Whether on balance there is regret among physicists I doubt, because the knowledge that has replaced the trans-uranic elements seems even more spectacular than they did. Let us see how this knowledge was attained.

Some time in 1938, Hahn observed that three of the radioactive substances resulting from the exposure of uranium to neutrons had some of the chemical properties of barium—enough to follow barium in certain of the distinctive precipitations which are known to chemists. Now this is a statement which is true of radium. Hahn assumed that he had three new isotopes of radium, and this was entirely natural, for two reasons. First, radium and its isotopes already known are all radioactive, suggesting that any which remained to be discovered should also be so; and second, the atomic number of radium is 88, so that radium isotopes could conceivably come into being through the reaction $U(n, \alpha)$ followed by the spontaneous emission of an alpha-particle from the residue. Yet (and this is the fact which came out on the 6th of January 1939) these substances were much too much like barium! When Hahn and Strassmann used some of the procedures which separate radium from barium, the novel substances declined to be separated. In a typical experiment, one of them together with some well-known isotope of radium would be introduced into a solution of some salt of barium. Fractional crystallization being performed, it was found as usual that the relative concentration of the radium isotope was greatly changed in the first-to-be-formed of the crystals; but not the relative concentration of the new substance, which entered into the crystals in just the same proportion as the barium itself.[2]

There are people who have revolutionary and false ideas about questions of science, and who irritate the scientists by their overconfident, their often arrogant ways of offering those ideas to the world.

[2] The salts were the bromide and the chromate (perhaps also the chloride and carbonate) of barium; the isotopes of radium were ThX and $MsTh_1$.

Listen now to men of science having a revolutionary and true idea, and expressing it in a befitting way:

"Now we must speak of some more recent investigations, which we publish only with hesitation because of the strange results. . . . We come to the conclusion: our 'radium isotopes' have the properties of barium; as chemists, we really ought to say that these new substances are barium, not radium. . . . As chemists, we ought to use the symbols Ba and La and Ce where we have been using Ra and Ac and Th. But as 'nuclear chemists' closely associated with physics, we cannot yet bring ourselves to make this leap, in contradiction to all previous lessons of nuclear physics. Perhaps, after all, our results have been rendered deceptive by a series of strange accidents. . . ."

Here I ought to mention another famous group of nuclear physicists who at an earlier date might have taken the leap, but recoiled before it so vehemently that they could not even bring themselves to mention in print the possibility of making it. These were the physicists of the Institut du Radium at Paris: Irene Curie and Savitch discovered in early 1938 that one of the products of $U(n)$ was indistinguishable, by all the tests that they applied, from the rare-earth element lanthanum ($Z = 57$). Later on they said that at a certain moment they had envisaged what we now call the fission of the uranium nucleus, but had preferred to believe that they had before them one of the transuranic elements resembling lanthanum more closely than anyone as yet had foreseen.

Now we come on to the middle of January 1939, and I must introduce the grand idea which with the force and suddenness of revelation burst upon several people far apart in the world, as soon as they heard of the experiments of Hahn and Strassmann and of the leap which these two had dared to envisage and publish if not quite to take.

Let us be audacious enough to take the leap, and let us further imagine that after the entry of the neutron the nucleus divides itself into two pieces or "fragments" of which one shall be barium. I must say directly that this assumption is more specific than need be, and that the same conclusions would be reached if we assumed that one of the fragments is some other element close to barium in the Periodic Table. It will be simpler, however, to be definite: let us assume barium, and for still greater definiteness let us suppose that the isotopes concerned are 238 of uranium and 139 of barium. The neutron, then supposedly enters a nucleus $_{92}U^{238}$ and with it forms the transitory "compound nucleus" $_{92}U^{239}$, and from this there splits off a nucleus $_{56}Ba^{139}$. What is left behind must be (if in a single piece) the nucleus of which the charge added to 56*e* makes 92*e*, and of which the mass-number added to 139 makes 239—that is to say, the nucleus $_{36}Kr^{100}$.

This is an example of the type of process which has been named by borrowing the word "fission" from biology. The biologists seem not to have found a specific verb to correspond (I am told that they use "divide") and the physicists have had no better inspiration. The dictionaries, however, authorize the use of "fissure" as a verb both transitive and intransitive, and I will henceforth so use it.[3]

Now a difficulty looms up, or rather what seems to be a difficulty but is really a great advantage, for the grandeur of the idea depends on it. Mass-numbers are only approximations to true masses, and the true mass of the nucleus U^{239} is greater than the sum of the masses of Ba^{139} and Kr^{100}. There is a superfluity of mass, and by classical ideas this superfluity might have to vanish, which would indeed be a stumbling-block. However, that stumbling block does not exist, because of something I have now to introduce. *It is the rest-mass, in the sense of relativity, of* U^{239} *which exceeds the sum of the rest-masses of* Ba^{139} *and* Kr^{100}. Now U^{239} before the explosion is practically at rest, but we are not obliged to make the same assumption about the fragments, and in fact we can assume that *the fragments fly apart at just such speeds that their relativistic increase of mass with speed brings up the sum of their masses to exactly the right value.* If so, their kinetic energies must be 50 to 100 Mev apiece. These on the nuclear scale are immense amounts of kinetic energy, and particles possessing it must be easy to isolate and easy to detect. This is why the idea is a grand one.

As it might occur to some reader to go to the tables of constants and look up the mass-values of U^{239} and Ba^{139} and Kr^{100}, I must say at once that he will not find them. Generally speaking, the mass-spectrograph cannot be used on radioactive and unstable atoms because one cannot get enough of them together for the experiment (exception being made for very long-lived ones like U^{238} and Th^{232}). All those three belong in that category, and therefore we have to estimate their masses by extrapolation from those of stable isotopes. The extrapolations for Ba^{139} and U^{239} are so small that the uncertainty is trivial, but Kr^{100} is no less than fourteen units heavier than the heaviest stable isotope of krypton, and this is serious. However, one is not so much concerned about conceivable defects in grand ideas when the ideas have already done their work by leading with success to grand experiments. I lay emphasis again, for a reason later to appear, on the extent to which Kr^{100} is out of line with the stable krypton isotopes; and now we pass to the experiments.

[3] I am indebted to Dr. Elizabeth Patterson of Bryn Mawr College for this solution.

There are actually two grand experiments, which I tried to distinguish above in a sentence by saying that the energetic particles must be *easy to isolate* and *easy to detect*. "Isolate" is not a very happy word: the fact is, that if so energetic they must be able to fly right out of the bombarded sheet of uranium (unless they start too deep beneath its surface)—thus, if some sort of a collector is placed across from the uranium and not too far away, they must assemble on it and there they should be found together with all their descendants. Joliot published this experiment before the end of January. He found radioactive substances on his collector even when more than two centimetres of air [4] had separated it throughout from the uranium.

The experiment has been performed by many, some introducing new refinements. Meitner and Frisch for instance used a bowl of water for collector, and then could concentrate the radioactive bodies by letting the water evaporate, or by precipitating various salts which in advance they had dissolved in it. This last is the chief technique for finding out the chemical nature of the radioactive products, to wit, the elements of which they are unstable isotopes; but we have not space for entering into the details of the technique, already practiced these five years. Glasoe, McMillan and others modified the method by piling very thin foils of very light substances—aluminium, filter-paper, cigarette-paper—on the uranium. Some of the radioactive matter is found embedded in each of the first few foils, and one may study thus their "distribution-in-range," an almost self-explanatory term. In McMillan's experiment the utmost perceptible range was slightly above 2.2 cm of air.

Already in the first experiment Joliot observed that in respect of its decay in time, the radioactivity on the collector was very like that remaining on the uranium. Later more accurate work has merely strengthened that conclusion, and Segré in particular affirms that out of many radioactive bodies there are only two which are found in the bombarded uranium itself and not on the distant collector also. On the distant collector there are found, in particular, the substances once classed as "trans-uranic elements." This is very important, for in the theory of the trans-uranic elements there occurs no stage in which the fragments of the uranium nucleus (or any other) are thrown apart with so tremendous energies. Were these elements trans-uranic, they should not be able at all to escape from the bombarded uranium

[4] To give the thickness of air (of the density corresponding to 15° C. and one atmosphere of pressure) which can just be traversed by a charged particle is the ordinary way of stating the "penetrating power" of the particle. Often some other substance than air is used in the tests; it is then not the actual thickness of the substance, but the "air-equivalent" thereof, which is ordinarily stated. Joliot appears to have used actual air in the experiments.

target. When in defiance of this the radioactivity crossed over to the collector, the trans-uranic elements were doomed.

In these experiments, then, the fragments of the initial explosion are found *en masse* together with their descendants upon the distant plate. In those of the other grand type, they are detected each by itself *en route*. Being charged particles of great momentum, they cleave through any gas in nearly linear paths, along which quantities of ions stay behind.[5] The Wilson chamber may be used to make these visible, and has indeed already been so used (by Joliot, and by Corson and Thornton); but another device gave the first and as yet most instructive results. This is the ionization-chamber equipped with linear amplifier and oscillograph. In the first, the ions due to the passage of a single particle are drawn to a collector and their charges united; in the second, the united charge is multiplied by a large fixed factor; in the third, the multiplied charge produces a sharp sidewise motion of the oscillograph-beam and the spot which this last produces. On the photographic plate the moving spot produces a line, the length of which is measured. Instances of these lines appear in Fig. 1.

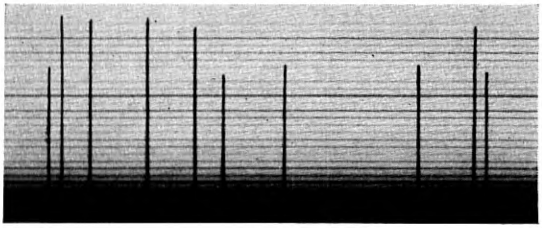


Fig. 1—Records of fission-fragments obtained with ionization chamber, linear amplifier, and oscillograph. The short lines due to alpha-particles are lost in the hazy dark band beneath. (Courtesy of J. R. Dunning)

Uranium is a spontaneous emitter of alpha-rays (this is how the radioactivity of U^{238} and U^{234} is manifest) and so the apparatus will show "kicks" even when neutrons are absent. This is an advantage really, since when the neutrons are admitted and the kicks due to the fragments appear they are so much the larger that there is no danger of confusing them with alpha-particle kicks, while these last may be pressed into service for calibrating the device. The calibration reposes upon a theorem of very great value in physics: viz., the (average) amount of energy expended by a fast charged particle in producing an ion-pair is fixed and constant, whatever the charge and mass and

[5] This is correct whether they travel as isolated nuclei, or are attended by some though not the full quota of orbital electrons which would environ them were they already the nuclei of completed atoms. Capture of electrons along the course is almost certain (it has been proved to occur with alpha-particles).

speed of the particle. So, the ratio of the kicks caused respectively by a fragment and an α-particle is the ratio of their initial energies, provided the chamber is so "deep" that they both run their courses completely to the end in the gas thereof. If on the other hand the chamber is so shallow or "thin" that fragment and α-particle shoot across it and only a small part of the total course of each is comprised within it, then the ratio of the kicks may be the ratio of the densities-of-ionization along the two tracks. Both the initial energy and the density-of-ionization are known for the α-particles, permitting the calibration. Also the constant value of the energy-expended-per-ion-pair is known (it is about 30 ev.) so that if the experimenter can measure the actual amount of charge set free in his ionization-chamber he need not bother with the α-particles.[6] In Fig. 1, by the way, the alpha-particle tracks are quite lost in the black band of the "background."

The second grand experiment, then, consisted in showing that when the neutrons were falling upon the uranium, there instantly appeared among the smallish kicks due to the α-particles others which were much greater—ten- and twenty-fold greater. This was done in four places [7] at least in America in the closing days of January 1939; in Copenhagen, however, a fortnight earlier.

The greatness of the kicks when the ionization-chamber is deep signifies the greatness of the initial energies of the fragments: I shall presently quote the latest data of these. But when the chamber is thin, the kicks due to the fragments again stand out very much over those due to the α-particles; and this signifies that the ionization-density along the fragment-tracks is great. (Take note, by the way, that one and the same chamber may be thin or thick, according as the density of the gas within is low or high—a very convenient fact.) The fragments, then, not only have remarkably great energy to start with, but also spend it at a remarkable rate in ionization along their courses. The course or "range" of a fragment must therefore be much shorter than would be that of an α-particle of the same energy. This is a verifiable fact, the ranges being easily measured by this method. We have just seen how Joliot was able to estimate them earlier, wiping out by this observation the possibility that each of the great kicks may be due to many α-particles starting off together. From the ion-

[6] All this is contingent upon the ions being completely gathered in by the collector of the ionization-chamber before any serious fraction of them is annulled by recombination, or (failing that) upon the loss by recombination being the same in proportion for α-particles and for fragments. Owing to the (unprecedently) great density of ion-pairs along the tracks of the fragments, this is by no means sure.

[7] New York (Columbia), Baltimore (Johns Hopkins), Washington (Carnegie Institution), Berkeley (University of California). In these cases the suggestion originated with Fermi.

ization-density and the range and the energy all taken together, it may be inferred that in both charge and mass these particles much exceed the α-particles; but here, better data and fuller theory are urgently required.[8]

Now we will consider the energies of the particles according to the data of Kanner and Barschall of Princeton.

If the immediate products of the fission are really just a pair of fragments nearly but not quite identical, we may expect a distribution-

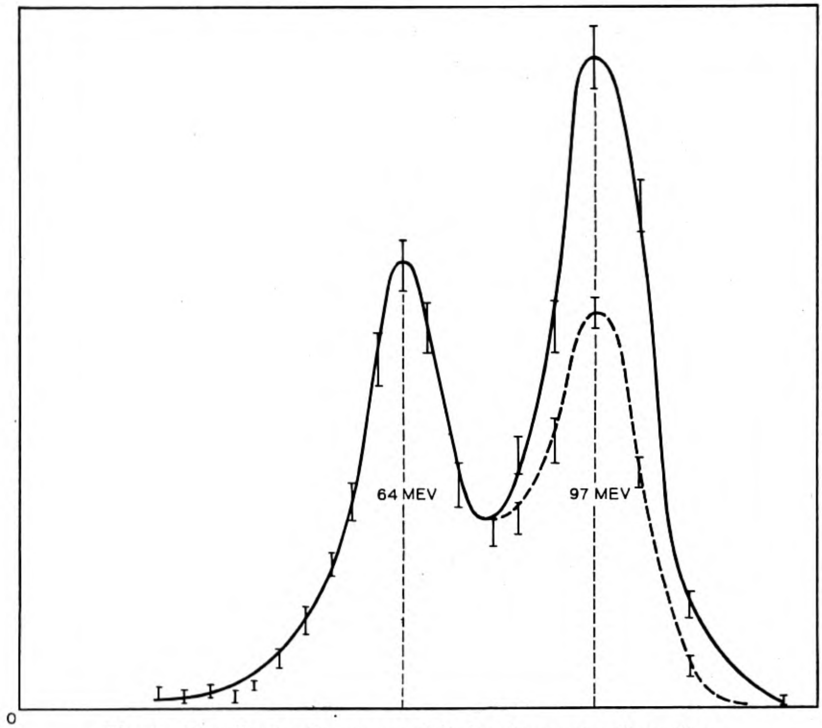


Fig. 2—Distribution-in-energy of fission-fragments of uranium. (Kanner and Barschall; *Physical Review*)

in-energy curve with two sharp peaks. If different fissions result in different fragment-pairs, the peaks must be broadened. If three or more particles are formed at a fission, there should be a broad continuous distribution of energies. This third of the possibilities is well excluded by the curve of Fig. 2; it remains to be seen whether the breadths of the humps speak for the second over the first.

[8] The inferring of charge and mass from energy, range and ionization-density is much practiced in the field of cosmic-ray research, in which, however, the particles usually have charge e and masses between the proton-mass and the electron-mass.

In giving the data for Fig. 2, as my words have implied, only a single fragment from each fission escaped into the gas of the ionization-chamber; this was arranged by laying down a very thin film of uranium upon a thick sheet of another metal. Figure 3 was obtained by laying down the uranium film upon a foil so very thin, that from most of the fissions both of the fragments entered into the gas. The great peak of Fig. 3 therefore indicates the sum of the energies of the fission-fragments, the peaks of Fig. 2 the components of that sum. Making al-

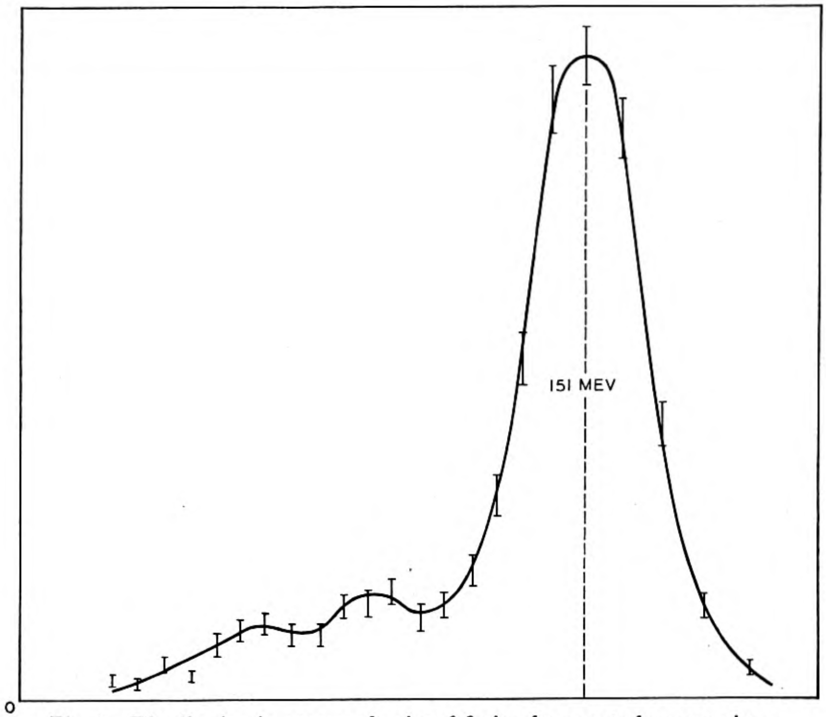


Fig. 3—Distribution-in-energy of pairs of fission-fragments from uranium. (Kanner and Barschall)

lowance for the average energy-loss suffered by the fragments in passing through solid matter before they escape to the gas, Kanner and Barschall decide on 159 Mev. for the sum, 98 and 65 Mev. for the components: the discrepancy between 159 and (98 + 65) lies within the uncertainty of experiment. By the law of conservation of momentum, the ratio of the component energies is the ratio of the fragment-masses; if one of these is about 100, the other is therefore about 150—and the uncertainty implied by "about" is broad enough to permit the hypotheses which we have made and are to make about the nature of the fragments. The Columbia school has done the same experiment, with like results.

Another way of ascertaining the energy released by fission was adopted by Henderson of Princeton; it is the oldest and most unimpeachable of all the methods of measuring energy, for he determined the rate at which heat was being developed in a uranium target and a container surrounding it while the fissuring was going on. His value was 175 Mev. per fission, with an uncertainty of some ten per cent. As some of this energy belonged not to the fragments but to the electrons emitted after the fission, the agreement is better than passable.

Now we come back to the question of the masses of the initial fragment-pair; and I will develop a second consequence of these masses, entirely different from the first. I revert to the use of mass-numbers, since the corrections needed for converting these into actual masses have not the slightest bearing on the point which is now to occupy us.

If the members of the initial fragment-pair are Ba^{139} and Kr^{100}, then the second of these two is fourteen units heavier than the heaviest stable isotope of $_{36}Kr$. It is therefore much too massive for its charge. This suggests that it may be able to shed neutrons, and so bring down its weight to the highest value compatible with its charge. But one may also say that Kr^{100} is too feebly charged for its mass. One has to go no fewer than six steps along the periodic table—to $_{42}Mo$—to find an element with a stable isotope of mass-number 100. Yet there is nothing to prevent us from assuming that the nucleus $_{36}Kr^{100}$ may shoot out six negative electrons, and so increase its charge to the minimum value compatible with its mass. The six might come out *seriatim*, in which case there would be a chain of six radioactive substances comprising all the elements from $_{36}Kr$ to $_{41}Cb$. Again, the nucleus might conceivably eject any number of neutrons under fourteen and some number of negative electrons under six, arriving at a sort of compromise pair of values of mass and charge compatible one with the other. One guesses already a mighty number of possible radioactive bodies resulting from the fission!

But now let us discard the assumption that Ba^{139} and Kr^{100} are the actual fragments of the fission, replacing these with any two nuclei which (*a*) lie in the middle region of the Periodic Table and (*b*) have atomic numbers adding up to 92 and mass-numbers adding up to 239. What, then, will happen to our two inferences from the masses? Essentially, *nothing*. Whichever such pair we take, one at least of its members must be too heavy for its charge and too feebly charged for its mass. (With most conceivable pairs, this will be true of both the members!) This derives from one of the fundamental facts of nuclear physics: the fact that when mass-number is plotted against atomic number, the points representing the stable nuclei cluster about a concave-upward curve. Moreover, whichever pair of fragments we assume, there will be a superfluity of rest-mass which will manifest itself in a high kinetic energy of the two fragments. This derives from another fundamental fact: when the percentage of excess of mass-number over true mass is plotted against the mass-number (or for that matter the atomic number) the points representing the nuclei cluster along an upward-trending curve.

And so, the kinetic energy of the fragments and the facts of the emerging negative electrons tell us neither which is the initial fragment-pair, nor even whether the initial pair is in all cases the same! Can these questions be answered out of the study of radioactive substances? Some of these we can indeed exclude by observing that they grow out of others; but as to these others, we shall never be able to exclude the possibility that they grow out of still others so short-lived, as to be quite unidentifiable. The half-period of a radioactive substance must be appreciable, if the substance is to be detected and its chemical character recognized; and "appreciable" thus far has signified, among the products of fission, "several seconds or more." It is true that certain *tours de force*, whereby much shorter half-periods have been measured among the natural radioactive bodies, have not yet been applied to the fission-products (so far as publications tell); they might prove workable.[10]

Thus it may be necessary for the nonce to lay aside the problem of deciding which is the true initial fragment-pair (or pairs) and be contented with identifying as many as possible among the radioactive substances and tracing their interrelations. Of these—hereafter to be called "the fission-products"—there is indeed a multitude. Among them, chemical elements have been recognized as follows: ${}_{34}Se$, ${}_{35}Br$, ${}_{36}Kr$, ${}_{37}Rb$, ${}_{38}Sr$, ${}_{39}Y$, ${}_{40}Zr$, ${}_{41}Cb$, ${}_{42}Mo$, ${}_{52}Te$, ${}_{53}I$, ${}_{54}Xe$, ${}_{55}Cs$, ${}_{56}Ba$, and ${}_{57}La$. Yet by counting these one does not count all of the distinguishable products; experimenters say that they can tell apart three isotopes of barium, three of strontium, four of iodine and no fewer than seven of tellurium! Some of these agree in their half-periods with radioactive isotopes of those same elements already formed by the older ways of transformation, and frequently we can thus identify their mass-numbers with a fair degree of certainty. (Ba^{139}, which I introduced into the hypothetical reaction of page 272, is such a one; but we shall see

[10] I refer particularly to the use of a rapidly-turning wheel to carry a target swiftly from a place where it is under bombardment (or receiving a deposit of radioactive nuclei) to another place where it is opposite a detector; and the measurement of the radioactivity of a beam of fast-flying nuclei at various points along the beam (the method applied by Jacobsen to RaC′).

that certainly it is not always and possibly it is never an initial fragment.) Others were unknown till 1939.

So numerous are these and the other fission-products still unrecognized, that the "decay-curve" for a piece of bombarded uranium or for the deposit on a nearby collector, due to all of them conjointly, looks like the resultant of contributions practically limitless in number and with a random distribution of half-periods. Not only is this also true when neutrons impinge on thorium, but the curves for the two elements cannot be told apart! Only after chemical separations have been made can individual half-periods be sorted out from among the welter; and if there are some characteristic differences between the results of the fission of uranium and those of the fission of thorium, they have not yet been proved.

Special interest attaches to the fission-products which are gaseous. They can be separated physically from the rest: the fission-products are received or dissolved into water (or indeed the uranium may be exposed to neutrons while in aqueous solution) and through the water a stream of air is bubbled, which takes along these particular ones to distant points in the system of tubing where they and their descendants can be studied. They cannot themselves be identified, but among their descendants are found (radioactive) isotopes of ${}_{37}$Rb and ${}_{55}$Cs; therefore the gases comprise unstable isotopes of krypton (${}_{36}$Kr) and xenon (${}_{54}$Xe). Could these be initial fragments of various types of fission? If so, their mates are ${}_{56}$Ba and ${}_{38}$Sr. Now, barium and strontium are found indeed among the fission-products, which seems to sustain this idea. But barium and strontium may also be the immediate descendants of the caesium and the rubidium aforesaid. This alternative idea is testable; and according to Hahn and Strassmann, two among the three barium isotopes (Ba^{139} being one of the two) are surely descendants of caesium, while the third may be an initial fragment.

Many other such "genetic" relationships have been published, but it would be lengthy and might be premature to quote them. I will mention at least that several sequences have been traced by Abelson in greater or less detail among the many fission-products which are isotopes of the three consecutive elements ${}_{51}$Sb, ${}_{52}$Te, and ${}_{53}$I. A special interest attaches to one of these bodies, the "77-hour tellurium"; for it has been identified as tellurium not only by its chemical properties but also by its X-rays. Let us pause to consider this.

The ordinary way of evoking an X-ray spectrum is to use the element in question as the target, or a constituent of the target, of an X-ray tube. This means that the atoms are excited by projecting electrons against them. They may also be excited by projecting photons against them, and this is sometimes done. Both of these ways are completely out of the question as yet with any artificial radioactive substance, for the greatest amount yet produced of any of these is so small that if its atoms were placed in a target, the hits made upon them by electrons or photons projected in streams of any feasible strength would not be numerous enough to produce detectable X-rays. If, however, the necessary photons proceed from the nuclei of the atoms themselves, then the whole situation is changed, because now the efficiency of excitation is so great. Such is the case with many of the natural radioactive substances, and now also (it appears) with the "77-hour tellurium." Excited presumably by photons proceeding from their nuclei,[11] the atoms emit X-rays, and these have been found (by Abelson in Berkeley, by Feather and Bretscher in England) to be the characteristic rays of the K-series of iodine. "Iodine" here is *not* a misprint for tellurium! When the nucleus is radioactive by virtue of the emission of an electron, the photon (if any) leaves after the electron is gone, by which time the atom is already an atom of the daughter-substance.[12]

Now we take up the yield of the fission-process: how does it depend on the energy of the incident neutrons?

Here uranium sets itself apart from the two other fissurable elements. Thorium and protactinium respond to fast neutrons only, uranium both to slow and to fast (but not to intermediate) neutrons. It is, however, believed that with uranium, one isotope is sensitive only to fast and another both to fast and to slow (or possibly only to slow). There are good theoretical grounds for this belief, and also for choosing the respective isotopes; but as yet there is not the certainty to be expected from some future and probably imminent experiment on separated isotopes.[13] Accepting nevertheless the current belief, we sup-

[11] Another mode of excitation is now known: an electron may fall into a nucleus, and by quitting its place in the orbital electron-family create the condition for the emission of an X-ray photon. Whether this or the other or both be the mode of excitation of the 77-hour tellurium is not yet certainly known.

[12] As this is likely to cause confusion, I emphasize that when the chemical separation is made, the atoms which have not yet emitted nuclear electrons are still tellurium atoms, and when they manifest themselves by that emission it appears among the tellurium; from then on they are iodine atoms among the tellurium, but no longer manifest themselves except through these X-rays. One may wonder whether the "transuranic elements," to which the 77-hour body was formerly thought to belong, would have been discredited if this measurement on the X-rays had been made earlier. Well, the measurement *was* earlier made (though not so precisely) and the rays were interpreted as characteristic X-rays of the L-series of a transuranic element. It is hard to make a guess as to whether further and better measurements would have destroyed this possibility.

[13] As these pages start for the press I am authorized to say that the separation has been achieved by Nier and the experiment performed by Dunning, Booth and von Grosse. The "light fraction," consisting of U^{235} with a small proportion of the very rare isotope U^{234} is definitely sensitive to slow neutrons; U^{238} is definitely *not*.

pose that of four known types of nuclei three (Th^{232}, Pa^{231}, U^{238}) are fissurable by fast neutrons only, one (U^{235}) by slow neutrons and probably also by fast. The mass-numbers just given are those of the nuclei awaiting the invading neutrons. If one prefers (as many do) to think of the transient composite nuclei formed by the neutron-invasions, one must write Th^{233}, Pa^{232}, U^{239} and U^{236}.

To speak of "fast" neutrons is vague, but not much vaguer than the state of knowledge, which as yet is rudimentary. Fission has been detected of thorium at neutron-energy of about 2 Mev, of protactinium at about one Mev, of uranium at about 0.5 Mev. It thus appears that the "threshold," or least contribution of energy demanded for fission, declines as the end of the Periodic Table is approached; and this seems natural. (Remember always that even with the fastest neutrons ever used, the contribution is very small compared with the energy released.) Values given in the literature for the "cross-section for fission by fast neutrons" include: $0.5 \cdot 10^{-24} cm^2$ and $0.1 \cdot 10^{-24}$ for uranium and for thorium bombarded by 2.4-Mev neutrons (Princeton) and $0.1 \cdot 10^{-24}$ for uranium bombarded by the "RnBe" neutrons.[14]

If between the source of neutrons and a target of uranium a screen of paraffin or water is inserted, the fissions become more abundant; but if now between the paraffin or water and the uranium a shield of cadmium is placed, the fissions become very rare. Now, paraffin and water convert fast neutrons into slow or "thermal" ones,[15] and cadmium is a very efficient absorbent for slow neutrons. We recognize, therefore, a specific effect of slow neutrons, peculiar to uranium. "Slow" or "thermal" signify in this usage: having kinetic energies of the very modest magnitudes possessed by molecules of air (or anything else) at ordinary temperatures: fractions of one electron-volt, and rarely more. Clearly then it is not the energy of motion of the neutron which is the insignificant spark setting off the mighty explosion; it is the mere presence of the neutron within the nuclear system.

A beam of slow neutrons falling upon a thin uranium layer produces many more fissions than does a fast-neutron beam of identical strength. Twenty-to-one was the ratio of yields found by the Columbia school, in the same experiment as gave them the value $0.1 \cdot 10^{-24}$ for the cross-section for fission by the fast "RnBe" neutrons. If, however, we put $2 \cdot 10^{-24}$ for the cross-section appropriate to the thermal neutrons, we

[14] Cross-section for fission, σ_f, is so defined that if N neutrons strike a thin layer comprising M nuclei per unit area, $MN\sigma_f$ fissions occur.—The "RnBe" neutrons, *viz.*, those released when α-particles from radon and its descendants impinge on beryllium, have a very broad energy-range extending at least from 14 Mev indefinitely downward (cf. Dunning, *Phys. Rev.* **45**, 586; 1934).

[15] The neutrons lose their great kinetic energies in repeated elastic impacts with hydrogen nuclei.

are in effect assuming that all the nuclei in the layer are equally liable to being fissured by these. To remain faithful to the well-grounded assumption that only the nuclei U^{235} are liable thus, we must multiply by 140, since only one nucleus in one hundred and forty is of this isotope.[16] The resulting value is large-sized for the nuclear scale, though not unprecedented: there are elements which absorb thermal neutrons so voraciously (without however suffering fission) that the cross-section for absorption is found to be hundreds of times more extensive.

Now in conclusion we turn to the particles other than nuclei, which go forth into space when or after the fission occurs. These comprise photons, electrons, and newborn or "secondary" neutrons; and the last are by far the most sensational.

Of the electrons, almost all has been said that should find place in this account. I recall that by virtue of the second argument from the masses (page 279) the nuclei of the fission-products should go from instability over to stability by emitting electrons which are negative. Observation shows that the emitted electrons are negative indeed (and yet there must be many among the products for which the sign has not been ascertained). Unstable nuclei emitting positive electrons are not at all unknown; indeed they are formed in many transmutations; their absence from among the fission-products is therefore significant. Many of the electrons coming forth are of "secondary origin," i.e. released by photons from the electron-families of the atoms. When classified with the many radioactive bodies formed by other modes of transmutation, some of the fission-products are found to be identical with some of those others, and the rest are in no wise peculiar.

Of the photons, some are X-ray photons engendered as I have recently described (page 282). Others are of the gamma-ray type, *i.e.*, they spring from unstable nuclei among the fission-products. Their existence not being in the least surprising, they have in the main been left for future study.

Coming now to the secondary neutrons, I will begin by dividing them into the "delayed" and the "instantaneous." The former come forth and are detected during an appreciable time—a few seconds up to a few minutes—after the fissions cease. Here then are radioactive bodies, of which the radioactivity consists in the emission of neutrons! Nothing of the sort had ever been known, and the discovery (made at the Carnegie Institution of Washington) created a sensation. In number they are much fewer than the "instantaneous" neutrons, define

[16] The figure is from Nier, *Phys. Rev.* **55**, 150 (1939), who gives 139 ± 1 as the abundance-ratio of U^{238} and U^{235}.

(for the present) as those which come out within a few thousandths of a second of the moment of fission. The ratio, according to the Columbia school, is about one to sixty. Delayed electrons and delayed photons have also been observed. Most observations on secondary neutrons are made while the target is being bombarded, and therefore relate to a mixture of the instantaneous with a small proportion of the delayed.

In energy the secondary neutrons differ greatly from the primary, a remarkable contrast! This is shown by several neat and pretty experiments, in which the secondaries manifest themselves by acting on detectors which cannot perceive the primaries at all. Thus if the primaries are thermal neutrons, an expansion-chamber or an ionization-chamber full of gas can be set among them without showing any sign of them,[17] since they cannot strike hard blows against the molecules therein. Let, however, a piece of uranium be set nearby, and the chamber will show dense trains of ions, produced by nuclei struck very hard and driven out of the molecules by neutrons which are fast. In this manner Halban and Joliot and Kowarski in Paris detected secondaries running in energy up to 11 Mev and beyond, while Zinn and Szilard of Columbia mapped the energy-spectrum up to 3.5 Mev. But also there are detectors able to discriminate between fast and faster neutrons: e.g. phosphorus, which becomes radioactive when bombarded by neutrons if, but only if, these have energy greater than 2 Mev. Dode and others in Paris prepared a source producing neutrons of energy one Mev; placed it next to a uranium target; surrounded source and target with a tank of liquid carbon disulphide, in which phosphorus was dissolved; and the liquid grew radioactive.

But how many neutrons are released per fission? This is a question of singular and perhaps of devastating importance, as will presently appear.

The obvious way to answer it seems to be that elected by Zinn and Szilard, who measured the number of fissions and also estimated, from the number of recoiling nuclei observed in their expansion-chamber, that of the secondary neutrons. Most of the trials have been made by a different method, in which all of the secondary neutrons are reduced to thermal energies before they are detected. (Incidentally there is the advantage, that if neutrons are released with low initial energies they will be counted by this way but not by the other.)

In this more customary method, the neutron-source and target are close together in the midst of a great tank of water, as large as can

[17] Except that if nitrogen is contained in the chamber, the thermal neutrons will react with the nitrogen nuclei so as to release protons (Zinn and Szilard).

conveniently be made. Paraffin may surround [18] the target and the source, to slow down the primary neutrons; or the water itself may perform this office. Again, the target may be diffused throughout the entire water-mass, in the form of a soluble salt of uranium. The detector is a substance becoming radioactive when exposed to slow neutrons. It may itself be spread throughout the water in the form of a soluble salt, or it may be in the form of a thin foil which can be moved from place to place in the water. In the former case, the water is thoroughly stirred after the exposure is over, and then a sample is taken, the activity of which is a measure of the average density—and therefore of the total quantity—of thermal neutrons in the entire tank during the exposure. In the latter case, the foil is used for mapping out the density of thermal neutrons in the water as function of the distance r from the target in the middle, and what is usually plotted is the "Ir^2 curve," I standing for the strength of the activity of the foil.

The total quantity of thermal neutrons, existing at any moment dispersed throughout the water, is greater in the presence of the uranium than in the absence thereof [19] (Anderson, Fermi and Hanstein); this is the simplest proof of the fundamental result. When the Ir^2 curves are compared, it is found that the presence of the uranium lowers the curve in the close neighborhood (within 13 or 14 cm) of the neutron-source, but raises it further out. Presumably this is because the uranium swallows up the slow primary neutrons, and those which it gives out in exchange are themselves not slow until they have gone a long way onward in the water. In tanks of sufficient size, the increase farther out more than balances the diminution nearer in, and the total quantity of thermal neutrons is augmented by the presence of uranium (Halban, Joliot and Kowarski); this agrees with the other result. It is therefore established what when the primaries are slow, the fission-process delivers more neutrons than it consumes. The same holds true when the primaries are fast, for when a beam of RnBe neutrons is sent through a plate of uranium oxide the detector beyond reveals a greater quantity of rapid neutrons than when the plate is absent (Haenny and Rosenberg, experimenting with a plate 8 cm thick).

How many neutrons then emerge, for every one which is spent in producing a fission? This is a remarkably difficult question to put to

[18] It is not necessary that the "slowing-down" substance be actually between the target and the source, since slowed-down neutrons come out of it in all directions.

[19] To make the situations strictly comparable, the uranium is replaced in the control experiment by some substance possessing an equal absorbing-power for neutrons, but not liable to fission.

the test of experiment. About all that the several answers have in common is, that *more neutrons emerge than are spent.* Zinn and Szilard say, two or three times as many; Anderson, Fermi and Szilard say, between one and two; the Paris school, three or four. A yet higher value (eight) published from Paris seems to comprise some " tertiary " neutrons produced by the secondaries.

But if every fission produces a fresh neutron to replace the one which caused it, and then some extras in addition, must we not anticipate a self-sustaining, nay even a self-amplifying effect? Must we not fear, in fact, a cataclysmic explosion?

Were anything of the sort to happen, we may take it for granted that the world would know of it, though in all probability the experimenter would not himself survive to report it. Evidently then it has not happened, and there must be a brake or brakes in Nature which impede the slide toward the catastrophe, and have thus far averted it. In other words, there must be ways in which neutrons are made harmless by some innocuous type of capture, before they ever produce a fission.

Some of these other ways are known already. If the uranium is mixed with other elements—as, in Nature, it invariably is—the nuclei of these can take up some of the neutrons. Whether the composite nuclei so formed are stable or radioactive is in this connection not important; they give no neutrons out in exchange for the ones absorbed, and so the chain is broken. But if all other elements are carefully extracted, do any brakes remain?

Two surely do, and one is the fact that the newborn neutrons are rapid, and cannot be efficacious as agents of fission until they are slowed down to thermal energies. In pure uranium the slowing-down can only be extremely gradual, so unfavorable is the huge mass-ratio—238 to 1—for the energy-transfer in the elastic impacts. Yet if the volume of purified metal were great enough, this brake would relax. Thus the durability of small-size pieces of uranium made chemically pure, well attested as it is, is not by itself a proof that much larger pieces would be safe. Those who are trying to approach the catastrophe, while hoping not to provoke it, are engaged in piling up uranium in greater and greater masses.

The other brake is supplied by the "reaction of pure neutron-capture," which I mentioned on an early page (p. 270). Every now and then, when a neutron enters a nucleus of uranium, the composite nucleus finds itself able to live on without fissure. It survives for a time, then emits a negative electron of energy trivial compared with fission-energies, and relapses into permanent stability. This is especially likely to happen when the neutron-energy is about 10 ev. Suppose then a volume of pure uranium so great, that the rapid neutrons released within it can make collisions numerous enough to bring their energies down to the thermal range where they are dangerous. Before they reach this range they must pass successfully through that other where they are liable to be disarmed—or put away in prison, rather. This second brake does not diminish in its strength as the volume of uranium is raised.

Perhaps the second by itself is powerful enough to avert the explosion. In this case there is no danger of incurring the cataclysm by piling up uranium, however pure. There remains however the chance of separating the two isotopes 235 and 238, verifying that the fission by thermal neutrons occurs only in the one and the reaction of pure neutron-capture only in the other, and then accumulating the dangerous one by itself. Enough has just now been separated, as I said in an earlier footnote, for the verification to begin. To separate enough for the dangerous trial will take a good deal longer in the doing. After it is done, there is yet another brake which may avert catastrophe. When the cumulative processes begin, the heating of the metal may and probably will so affect the energies of the neutrons, that their efficiency for fissuring the nuclei will be greatly abated and so the processes find a natural limit. Otherwise it is to be hoped that those who build up great masses of sensitive uranium will recognize preliminary signs that the danger-point is close, before they actually attain it.

Literature

(Year always 1939 unless otherwise stated)

Early Chemical Identifications of Fission-Products: Hahn and Strassmann, *Naturwiss.* **27**, *passim;* also Curie and Savitch, *Jour. de Phys.* **9**, 355 (1938).

Isolation of Fission-Products by their Spontaneous Egress from Bombarded Uranium or Thorium: Joliot, *C.R.* **208**, 341; Meitner and Frisch, *Nature* **143**, 471; McMillan, *Phys. Rev.* **55**, 610; Segre, *Phys. Rev.* **55**, 1104.

Detection of Fission-Products in Ionization-Chamber: Meitner and Frisch, *Nature* **143**, 239, 276; Columbia school, *Phys. Rev.* **55**, 511; Roberts, Meyer and Hafstad, *Phys. Rev.* **55**, 416; Green and Alvarez, *Phys. Rev.* **55**, 417; Fowler and Dodson, *ibid.*

Fission-Products Detected in Expansion-Chamber: Joliot, *C.R.* **208**, 647; Corson and Thornton, *Phys. Rev.* **55**, 509.

Distribution-in-Energy of Fission-Products: Booth, Dunning and Slack, *Phys. Rev.* **55**, 981, 982, 1273; Kanner and Barschall, *Phys. Rev.* **57**, 372 (1940).

Further Chemical Studies of Fission-Products: Hahn and Strassmann, *Naturwiss.* **27**, 451; *Preuss. Akad.* 1939, no. 12; Savitch, *C.R.* **208**, 646; Thibaud and Moussa, *C.R.* **208**, 652; Heyn, Aten and Bakker, *Nature* **143**, 517; Bretscher and Cook, *Nature* **143**, 559; Dodson and Fowler, *Phys. Rev.* **55**, 880; Glasoe and Steigman, *Phys. Rev.* **55**, 982; Abelson, *Phys. Rev.* **56**, 1.

Decay-Curves for Totality of Fission-Products Unseparated: Bjerge, Brostrom and Koch, *Nature* **143**, 794.

Identification of Iodine among Fission-Products by X-rays: Abelson, *Phys. Rev.* **55**, 418; Feather and Bretscher, *Nature* **143**, 516.

Delayed Radiations: Roberts, Meyer and Wang, *Phys. Rev.* **55**, 510; Booth, Dunning and Slack, *Phys. Rev.* **55**, 876; Barschall *et al.*, *Phys. Rev.* **55**, 989; Gibbs and Thomson, *Nature* **144**, 202.

Yield of Fission-Process: Columbia School, *Phys. Rev.* **55**, 511; Ladenburg *et al.*, *Phys. Rev.* **56**, 168.

Energy and Yield of Neutrons Released at Fission: Anderson, Fermi and Hanstein, *Phys. Rev.* **55**, 797; Halban, Joliot and Kowarski, *Nature* **143**, 470, 680, 939; Haenny and Rosenberg, *C.R.* **208**, 898; Dode *et al.*, *C.R.* **208**, 995; Zinn and Szilard, *Phys. Rev.* **56**, 619; Halban *et al.*, *Jour. de Phys.* **10**, 428.

Scattering and Pure Neutron-Capture in Uranium: Whitaker *et al.*, *Phys. Rev.* **55**, 793; Anderson and Fermi, *Phys. Rev.* **55**, 1107; Halban, Kowarski and Savitch, *C.R.* **208**, 1396; Goldstein, Rogozinski and Walen, *Jour. de Phys.* **10**, 477; Anderson, *Bull. Amer. Phys. Soc.* Feb. 5, 1940.

Fission by Deuterons: Gant, *Nature* **144**, 707.

Fission of Protactinium: Grosse, Booth and Dunning, *Phys. Rev.* **56**, 382.

Fission of Separated Isotopes of Uranium: Nier, Booth, Dunning and Grosse, *Phys. Rev.* **57**, 546 (1940).

Theory of Fission: Bohr and Wheeler, *Phys. Rev.* **56**, 426, 1065.

General Review: Turner, *Rev. Mod. Phys.* **12**, 1 (1940).

A Solution for Faults at Two Locations in Three-Phase Power Systems

By E. F. VAAGE

This paper is an outgrowth of studies of double faults to ground in three-phase power systems made by the author in connection with work of the Joint Subcommittee on Development and Research, Edison Electric Institute and Bell Telephone System. The paper provides a systematic solution, based on the method of symmetrical components, by means of which currents and voltages can be determined at times of fault involving any combination of phases at one or two locations on three-phase power systems.

1. Introduction

A KNOWLEDGE of the magnitude and phase relation of power system voltages and currents for various types of faults in three-phase systems is of importance in the study of various problems, among which are relaying studies, the efficacy of current limiting devices and their reaction on the power network, and estimates of induction in paralleling communication circuits.

The method of symmetrical components as developed by Fortescue [1] and others is now extensively used in the solution for currents and voltages in three-phase power systems under fault (short-circuit) conditions. Formulas for special cases of faults, such as single and double line-to-ground faults, can be found in various text books on this subject. The solution for simultaneous faults at two locations has been treated by Miss Clark,[2] in a form particularly adaptable to the use of a calculating board.

The present development provides a complete and systematic solution for currents and voltages at times of fault on any number of phases at one or two locations in a three-phase system, in which generators may be assumed in phase and of the same internal voltage, and where load currents can be neglected. These are the usual assumptions made in computing fault currents, except for certain special problems, such as that of power system stability. The methods employed herein could be extended to cases where generators of different phase angles and voltages of more than two points of fault are involved. Formally such cases can be treated in a manner similar to that given in the paper. The number of impedances to which an n-terminal network

[1] Reference numbers refer to references appearing at the end of the article.

can be reduced is given by the expression $\frac{1}{2}n(n-1)$. For $n = 3$, the case treated in this paper, three impedances are required which necessitate six equations for the general solution of the six fault currents. For $n = 4$, which would be the case for three points of fault (or two points of fault and two generating voltages), six impedances would appear in the reduced network and this would necessitate twelve equations for the general solution of the fault currents. For larger values of n, the necessary number of equations increases rapidly, thus making the solution impractical. Such problems usually, as a practical matter, are more readily solved by the use of a-c. calculating boards.

While no departure from the general methods of symmetrical components is made in the present development, a systematic method of handling the equations is presented and means of determining the coefficients given so that numerical calculations can be directly carried out when the constants of the network are known.

2. General Solution

The equations developed in this paper are based on the sequence impedances looking into a three-phase network from two points of fault.

Consider the network shown in Fig. 1. This system can be reduced

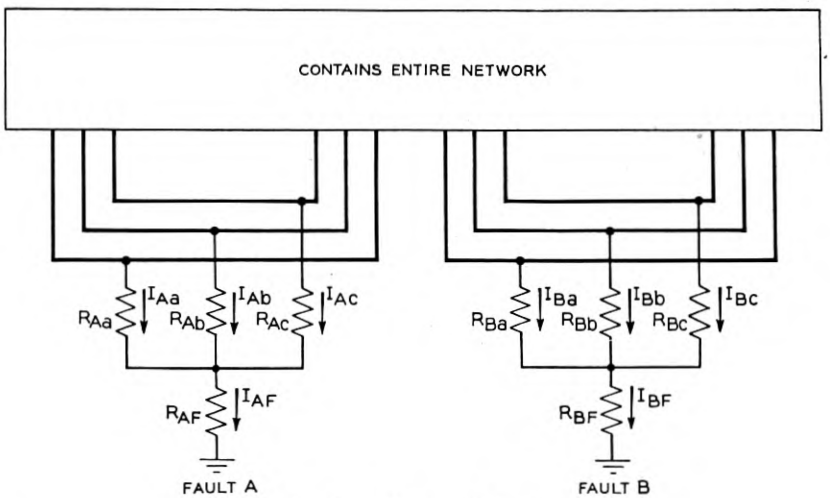


Fig. 1—General network diagram.

to an equivalent star for each of the positive, negative and zero sequence networks, with legs to the points of generation and to the faults at A and B. Figure 2 shows the reduced positive sequence network. Similar diagrams can be made for the negative and zero sequence systems except for the fact that in these cases there are no generated voltages, and the impedances and currents are the negative and zero sequence quantities.

The reduction of a network to an equivalent star is usually a tedious and sometimes a difficult process especially in large interconnected systems. Methods of accomplishing the reduction, such as delta-star transformations, simultaneous equations or direct measurements on calculating boards can be found in the literature.[3, 5]

Having reduced the three sequence networks to equivalent stars, the equations are developed as shown in the Appendix.

The following set of equations (1) is the general solution for fault currents during simultaneous three-phase faults to ground at two different locations in a power network. The set applies directly to the calculation of ground fault currents on a system having finite

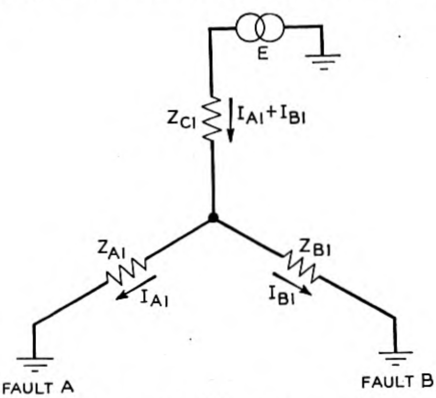


Fig. 2—Reduced positive sequence diagram.

neutral impedances or an isolated system in which the zero sequence capacitance is taken into account. For other types of faults, such as faults to ground in isolated systems in which zero sequence capacitance has been neglected, or for phase to phase faults, set (1) is not directly applicable since some of the constants become infinitely large. However, by certain transformations of set (1), more convenient sets (2) and (3) are obtained, directly applicable for solution of these latter cases.

Neutral Grounded System

I_{Aa}	I_{Ab}	I_{Ac}	I_{Ba}	I_{Bb}	I_{Bc}			
A_{11}	A_{12}	A_{13}	A_{14}	A_{15}	A_{16}	$3E$	(Aa)	
A_{21}	A_{22}	A_{23}	A_{24}	A_{25}	A_{26}	$3a^2E$	(Ab)	
A_{31}	A_{32}	A_{33}	A_{34}	A_{35}	A_{36}	$3aE$	(Ac)	(1)
A_{41}	A_{42}	A_{43}	A_{44}	A_{45}	A_{46}	$3E$	(Ba)	
A_{51}	A_{52}	A_{53}	A_{54}	A_{55}	A_{56}	$3a^2E$	(Bb)	
A_{61}	A_{62}	A_{63}	A_{64}	A_{65}	A_{66}	$3aE$	(Bc)	

The six equations are written in matrix form with the currents and voltages outside the system matrix. For example the first row in (1) is interpreted as:

$$A_{11}I_{Aa} + A_{12}I_{Ab} + A_{13}I_{Ac} + A_{14}I_{Ba} + A_{15}I_{Bb} + A_{16}I_{Bc} = 3E$$

The values of the A's in (1) are given in Table I. It should be noted that of the 36 constants only 13 are distinct. Six of these are in the nature of self-impedances, two are transfer impedances between phases at A and two between phases at B. The remaining three are transfer impedances between the two faults at A and at B.

Considerable reductions in the constants are obtained when the positive and negative sequence impedances are assumed equal. These values are given in Table II.

Faults to ground on less than three phases at one or both locations are accounted for by assuming the corresponding fault resistances infinitely large. The currents to ground in the sound phases are zero. Striking out the columns containing these currents and the corresponding rows, indicated by the index at right in equation (1), a reduced set of equations is obtained from which the desired currents can be found. A few examples are given in subsequent sections.

In power networks with isolated neutral the zero sequence impedance Z_{C0} reduces essentially to the capacitance of the system. In this case equations (1) are still appropriate and will give a rigorous solution for the six currents. However, in many cases it is sufficiently accurate to neglect the capacitance of the system. This results in infinitely large values of all of the A's in Table I (Table II), since each depends on Z_{C0} which is infinitely large. For this condition it is desirable to transform the set of equations in (1) to a more convenient set with finite constants.

The transformation required is obtained by observing that, with Z_{C0} infinitely large, the sum of the zero sequence currents $I_{A0} + I_{B0}$ must be equal to zero. Making use of this relation the difference of the zero sequence voltages at A and B (equation (50) of appendix) reduces to:

$$V_{A0} - V_{B0} = (Z_{A0} + Z_{B0})I_{B0}$$

The last equation shows that subtraction of equations associated with phases at A from those at B removes the infinitely large element Z_{C0}. This can be done in nine ways (ignoring reversals of sign), but three of these result in the single equation:

$$I_{Aa} + I_{Ab} + I_{Ac} + I_{Ba} + I_{Bb} + I_{Bc} = 0$$

This equation with any five of the remaining six constitutes an independent set; for convenience in dealing with special cases the redundant set of seven equations is shown in the following array:

Isolated System—Capacitance Neglected

I_{Aa}	I_{Ab}	I_{Ac}	I_{Ba}	I_{Bb}	I_{Bc}		
B_{11}	B_{12}	B_{13}	B_{14}	B_{15}	B_{16}	$3(1-a^2)E$	$(Aa-Bb)$
B_{21}	B_{22}	B_{23}	B_{24}	B_{25}	B_{26}	$3(1-a)E$	$(Aa-Bc)$
B_{31}	B_{32}	B_{33}	B_{34}	B_{35}	B_{36}	$3(a^2-1)E$	$(Ab-Ba)$
B_{41}	B_{42}	B_{43}	B_{44}	B_{45}	B_{46}	$3(a^2-a)E$	$(Ab-Bc)$
B_{51}	B_{52}	B_{53}	B_{54}	B_{55}	B_{56}	$3(a-1)E$	$(Ac-Ba)$
B_{61}	B_{62}	B_{63}	B_{64}	B_{65}	B_{66}	$3(a-a^2)E$	$(Ac-Bb)$

(2)

$$I_{Aa} + I_{Ab} + I_{Ac} + I_{Ba} + I_{Bb} + I_{Bc} = 0 \tag{2a}$$

The index to the right indicates which of the equations in (1) have been used. The values of the B's are given in Table I and Table II.

In case of faults to ground on less than three phases, as in equations (1), columns and rows associated with sound phase currents are to be deleted; with respect to the rows, however, the index is double and all rows having the index of the sound phase or phases are deleted. If, for example, the sound phase is Aa, rows 1 and 2, each of which contains Aa in its index, as well as column 1, are deleted. This leaves only four equations, which together with (2a) give the necessary five equations for the five currents. For this reason all six equations are given in (2), since any phase might be involved in special cases.

Phase-to-phase faults are obtained from the general case (1) by allowing the resistances R_{AF} and R_{BF} to become infinite. In this case phase-to-phase quantities at the same location remain finite and the appropriate set of equations is obtained by subtracting equations having the corresponding phase indexes; thus $Aa - Ab$, $Aa - Ac$ and $Ab - Ac$ indicate subtractions at A. There are six possible ways of doing this, ignoring reversals of sign. The resulting set is given in (3). The four equations obtained by taking any two of the first three and any two of the last three equations in this set together with the two equations (3a) relating to the sum of the currents at each fault location, which from physical considerations equal zero, constitute an independent set. For convenience in dealing with special cases all eight equations are given below:

Phase-to-Phase Faults

I_{Aa}	I_{Ab}	I_{Ac}	I_{Ba}	I_{Bb}	I_{Bc}		
C_{11}	C_{12}	C_{13}	C_{14}	C_{15}	C_{16}	$3(1-a^2)E$	$(Aa-Ab)$
C_{21}	C_{22}	C_{23}	C_{24}	C_{25}	C_{26}	$3(1-a)E$	$(Aa-Ac)$
C_{31}	C_{32}	C_{33}	C_{34}	C_{35}	C_{36}	$3(a^2-a)E$	$(Ab-Ac)$
C_{41}	C_{42}	C_{43}	C_{44}	C_{45}	C_{46}	$3(1-a^2)E$	$(Ba-Bb)$
C_{51}	C_{52}	C_{53}	C_{54}	C_{55}	C_{56}	$3(1-a)E$	$(Ba-Bc)$
C_{61}	C_{62}	C_{63}	C_{64}	C_{65}	C_{66}	$3(a^2-a)E$	$(Bb-Bc)$

(3)

$$\begin{aligned} I_{Aa} + I_{Ab} + I_{Ac} &= 0 \\ I_{Ba} + I_{Bb} + I_{Bc} &= 0 \end{aligned} \tag{3a}$$

The index to the right indicates which equations of (1) have been used. The values of the C's are given in Table I and Table II.

The total ground fault currents at the two fault locations are:

$$I_{AF} = I_{Aa} + I_{Ab} + I_{Ac} \tag{4}$$

$$I_{BF} = I_{Ba} + I_{Bb} + I_{Bc} \tag{5}$$

and the total residual current in the two faults is:

$$I_R = I_{AF} + I_{BF} \tag{6}$$

In an isolated system in which capacitance has been neglected this current is zero and equation (6) will be identical with (2*a*).

The distribution of these currents in the network can be found as follows. From equations (51) and (52) of the appendix the calculated fault currents are transformed into sequence currents. By working back into the original sequence networks the sequence currents in each branch of the system can be found and later combined from similar expressions as shown in (43) and (44) to obtain the actual branch currents.

A combination of the equations in (1) and (3) can be used for cases involving faults to ground at one location and faults between phases (not involving ground) at the other location.

For faults to ground at A and between phases at B, the three first equations in (1) and the three last in (3) together with the last in (3*a*) constitute the most convenient set of equations for this type of fault.

It should be noted that if all three phases are involved at B any two of the three last equations in (3) together with the last in (3*a*) can be used, while for less than three phases involved, the rules for striking out rows and columns automatically will result in the proper equations to be used.

Vice versa the three first equations in (3) together with the first in (3*a*) and the three last equations in (1) will give the solutions for phase-to-phase faults at A and ground faults at B.

This will be illustrated with an example in a later section.

The voltages to ground at the two locations of faults can be obtained directly from equations (53) and (54) of the Appendix, after the currents have been evaluated. At any other point in the system the voltages to ground are found by adding the voltage drops of the lines in question to these voltages, treating each sequence network separately, then adding the sequence voltages together according to equations (46) or (47).

TABLE I

CONSTANTS FOR EQUATIONS (1), (2) AND (3)

Equation (1)	Equation (2)	Equation (3)	Equalities
$A_{11} = S_a + 3R_{Aa} + 3R_{AF}$	$B_{11} = S_a - T_c + 3R_{Aa} + 3R_{AF}$	$C_{11} = S_a - S_c + 3R_{Aa}$	Equation (1)
$A_{12} = S_b + 3R_{AF}$	$B_{12} = S_b - T_a + 3R_{AF}$	$C_{12} = -S_a + S_b - 3R_{Ab}$	$A_{12} = A_{23} = A_{31}$ $A_{45} = A_{56} = A_{64}$
$A_{13} = S_c + 3R_{AF}$	$B_{13} = S_c - T_b + 3R_{AF}$	$C_{13} = S_c - S_b$	$A_{13} = A_{21} = A_{32}$ $A_{46} = A_{54} = A_{65}$
$A_{14} = T_a$	$B_{14} = -U_c + T_a - 3R_{BF}$	$C_{14} = T_a - T_c$	$A_{14} = A_{25} = A_{36} = A_{41} = A_{52} = A_{63}$
$A_{15} = T_b$	$B_{15} = -U_a + T_b - 3R_{Bb} - 3R_{BF}$	$C_{15} = T_b - T_a$	$A_{15} = A_{26} = A_{34} = A_{42} = A_{53} = A_{61}$
$A_{16} = T_c$	$B_{16} = -U_b + T_c - 3R_{BF}$	$C_{16} = T_c - T_b$	$A_{16} = A_{24} = A_{35} = A_{43} = A_{51} = A_{62}$
$A_{22} = S_a + 3R_{Ab} + 3R_{AF}$	$B_{21} = S_a - T_b + 3R_{Aa} + 3R_{AF}$	$C_{21} = S_a - S_b + 3R_{Aa}$	Equation (2)
$A_{33} = S_a + 3R_{Ac} + 3R_{AF}$	$B_{22} = S_b - T_c + 3R_{AF}$	$C_{23} = -S_a + S_c - 3R_{Ac}$	$B_{12} = B_{43} = B_{51}$ $B_{22} = B_{33} = B_{61}$
$A_{44} = U_a + 3R_{Ba} + 3R_{BF}$	$B_{23} = S_c - T_a + 3R_{AF}$	$C_{32} = S_a - S_c + 3R_{Ab}$	$B_{13} = B_{41} = B_{52}$ $B_{23} = B_{31} = B_{62}$
$A_{45} = U_b + 3R_{BF}$	$B_{24} = -U_b + T_a - 3R_{BF}$	$C_{33} = -S_a + S_b - 3R_{Ac}$	$B_{14} = B_{45} = B_{56}$ $B_{24} = B_{35} = B_{66}$
$A_{46} = U_c + 3R_{BF}$	$B_{25} = -U_c + T_b - 3R_{BF}$	$C_{44} = U_a - U_c + 3R_{Ba}$	$B_{16} = B_{44} = B_{55}$ $B_{25} = B_{36} = B_{64}$
$A_{55} = U_a + 3R_{Bb} + 3R_{BF}$	$B_{26} = -U_a + T_c - 3R_{Bc} - 3R_{BF}$	$C_{45} = -U_a + U_b - 3R_{Bb}$	Equation (3)
$A_{66} = U_a + 3R_{Bc} + 3R_{BF}$	$B_{32} = S_a - T_b + 3R_{Ab} + 3R_{AF}$	$C_{46} = U_c - U_b$	$C_{13} = -C_{22} = C_{31}$
	$B_{34} = -U_a + T_c - 3R_{Ba} - 3R_{BF}$	$C_{54} = U_a - U_b + 3R_{Ba}$	$C_{14} = -C_{26} = C_{35} = C_{41} = -C_{53} = C_{62}$
	$B_{42} = S_a - T_c + 3R_{Ab} + 3R_{AF}$	$C_{56} = -U_a + U_c - 3R_{Bc}$	$C_{15} = -C_{24} = C_{36} = C_{42} = -C_{51} = C_{63}$
	$B_{46} = -U_a + T_b - 3R_{Bc} - 3R_{BF}$	$C_{65} = U_a - U_c + 3R_{Bb}$	$C_{16} = -C_{25} = C_{34} = C_{43} = -C_{52} = C_{61}$
	$B_{53} = S_a - T_c + 3R_{Ac} + 3R_{AF}$	$C_{66} = -U_a + U_b - 3R_{Bc}$	$C_{46} = -C_{55} = C_{64}$
	$B_{54} = -U_a + T_b - 3R_{Ba} - 3R_{BF}$		
	$B_{63} = S_a - T_b + 3R_{Ac} + 3R_{AF}$		
	$B_{65} = -U_a + T_c - 3R_{Bb} - 3R_{BF}$		

where:

$$S_a = (Z_{A1} + Z_{C1}) + (Z_{A2} + Z_{C2}) + (Z_{A0} + Z_{C0}) \quad T_a = Z_{C1} + Z_{C2} + Z_{C0} \quad U_a = (Z_{B1} + Z_{C1}) + (Z_{B2} + Z_{C2}) + (Z_{B0} + Z_{C0})$$
$$S_b = a(Z_{A1} + Z_{C1}) + a^2(Z_{A2} + Z_{C2}) + (Z_{A0} + Z_{C0}) \quad T_b = aZ_{C1} + a^2Z_{C2} + Z_{C0} \quad U_b = a(Z_{B1} + Z_{C1}) + a^2(Z_{B2} + Z_{C2}) + (Z_{B0} + Z_{C0})$$
$$S_c = a^2(Z_{A1} + Z_{C1}) + a(Z_{A2} + Z_{C2}) + (Z_{A0} + Z_{C0}) \quad T_c = a^2Z_{C1} + aZ_{C2} + Z_{C0} \quad U_c = a^2(Z_{B1} + Z_{C1}) + a(Z_{B2} + Z_{C2}) + (Z_{B0} + Z_{C0})$$

TABLE II

CONSTANTS FOR EQUATIONS (1), (2) AND (3) WHEN POSITIVE AND NEGATIVE SEQUENCE IMPEDANCES ARE EQUAL

Equation (1)	Equation (2)	Equation (3)	Equalities
$A_{11} = S_a + 3R_{Aa} + 3R_{AF}$	$B_{11} = S_a - T_b + 3R_{Aa} + 3R_{AF}$	$C_{11} = S_a - S_b + 3R_{Aa}$	Equation (1)
$A_{12} = S_b + 3R_{AF}$	$B_{12} = S_b - T_a + 3R_{AF}$	$C_{12} = -S_a + S_b - 3R_{Ab}$	$A_{12} = A_{13} = A_{21} = A_{23} = A_{31} = A_{32}$
$A_{14} = T_a$	$B_{13} = S_b - T_b + 3R_{AF}$	$C_{13} = 0$	$A_{45} = A_{46} = A_{54} = A_{56} = A_{64} = A_{65}$
$A_{15} = T_b$	$B_{14} = -U_b + T_a - 3R_{AF}$	$C_{14} = T_a - T_b$	$A_{14} = A_{25} = A_{36} = A_{41} = A_{52} = A_{63}$
$A_{22} = S_a + 3R_{Ab} + 3R_{AF}$	$B_{15} = -U_a + T_b - 3R_{Bb} - 3R_{BF}$	$C_{23} = -S_a + S_b - 3R_{Ac}$	$A_{15} = A_{16} = A_{24} = A_{26} = A_{34} = A_{35}$
$A_{33} = S_a + 3R_{Ac} + 3R_{AF}$	$B_{16} = -U_b + T_b - 3R_{AF}$	$C_{44} = U_a - U_b + 3R_{Ba}$	$= A_{42} = A_{43} = A_{51} = A_{53} = A_{61}$
$A_{44} = U_a + 3R_{Ba} + 3R_{BF}$	$B_{26} = -U_a + T_b - 3R_{Bc} - 3R_{BF}$	$C_{45} = -U_a + U_b - 3R_{Bb}$	$= A_{62}$
$A_{45} = U_b + 3R_{BF}$	$B_{32} = S_a - T_b + 3R_{Ab} + 3R_{AF}$	$C_{56} = -U_a + U_b - 3R_{Bc}$	Equation (2)
$A_{55} = U_a + 3R_{Bb} + 3R_{BF}$	$B_{34} = -U_a + T_b - 3R_{Ba} - 3R_{BF}$		$B_{11} = B_{21}$ $\quad B_{15} = B_{65}$
$A_{66} = U_a + 3R_{Bc} + 3R_{BF}$	$B_{53} = S_a - T_b + 3R_{Ac} + 3R_{AF}$		$B_{12} = B_{23} = B_{31} = B_{43} = B_{51} = B_{62}$
			$B_{13} = B_{22} = B_{33} = B_{41} = B_{52} = B_{61}$
			$B_{14} = B_{24} = B_{35} = B_{45} = B_{56} = B_{66}$
			$B_{16} = B_{25} = B_{36} = B_{44} = B_{55} = B_{64}$
			$B_{32} = B_{42}$ $\quad B_{34} = B_{54}$
			$B_{26} = B_{46}$ $\quad B_{53} = B_{63}$
			Equation (3)
			$C_{13} = C_{16} = C_{22} = C_{25} = C_{31} = C_{34}$
			$= C_{43} = C_{46} = C_{52} = C_{55} = C_{61}$
			$= C_{64} = 0$
			$C_{14} = -C_{15} = C_{24} = -C_{26} = C_{35}$
			$= -C_{36} = C_{41} = -C_{42} = C_{51}$
			$= -C_{53} = C_{62} = -C_{63}$
			$C_{11} = C_{21}$ $\quad C_{12} = -C_{32}$ $\quad C_{23} = C_{33}$
			$C_{44} = C_{54}$ $\quad C_{45} = -C_{65}$ $\quad C_{56} = C_{66}$

where:

$$
\begin{aligned}
S_a &= 2(Z_{A1} + Z_{C1}) + (Z_{A0} + Z_{C0}) \\
S_b &= -(Z_{A1} + Z_{C1}) + (Z_{A0} + Z_{C0}) \\
T_a &= 2Z_{C1} + Z_{C0} \\
T_b &= -Z_{C1} + Z_{C0} \\
U_a &= 2(Z_{B1} + Z_{C1}) + (Z_{A0} + Z_{C0}) \\
U_b &= -(Z_{B1} + Z_{C1}) + (Z_{A0} + Z_{C0})
\end{aligned}
$$

3. Special Cases

The application of the three sets of equations (1), (2) and (3), will be illustrated with a few examples. For simple cases, such as a single or double line-to-ground fault at one location, the equations reduce to formulas frequently found in the literature on this subject.

From set (1) equations for faults to ground at one or two locations can be obtained directly when the zero sequence impedance is finite. Set (2), obtained from (1), is the most convenient set for solutions of faults to ground in isolated systems in which capacitance has been neglected. The phase-to-phase fault currents are best obtained from set (3).

3.1 Single Line-to-Ground Fault at A

Consider a fault to ground on phase "b" at A. The solution can be obtained from (1) by letting:

$$R_{Aa} = R_{Ac} = R_{Ba} = R_{Bb} = R_{Bc} = \infty \tag{7}$$

This results in:

$$I_{Aa} = I_{Ac} = I_{Ba} = I_{Bb} = I_{Bc} = 0 \tag{8}$$

Striking out all columns in (1) containing the currents in (8) and the corresponding rows indexed by Aa, Ac, Ba, Bb and Bc only one equation is left:

$$A_{22}I_{Ab} = 3a^2E \tag{9}$$

The numerical value of A_{22} can be calculated directly from Table I, or on substituting the symbolic value of A_{22} in equation (9) the result will be:

$$I_{Ab} = \frac{3a^2E}{Z_1 + Z_2 + Z_0 + 3R_F} \tag{10}$$

where

$$\begin{aligned} Z_1 &= Z_{A1} + Z_{C1} \\ Z_2 &= Z_{A2} + Z_{C2} \\ Z_0 &= Z_{A0} + Z_{C0} \\ R_F &= R_{Ab} + R_{AF} \end{aligned} \tag{11}$$

Equation (10) is the well-known formula for a single line-to-ground fault at one location in a three-phase system.

3.2 Double Line-to-Ground Fault at A

Consider a double line-to-ground fault on phases "a" and "b" at A. Then:

$$R_{Ac} = R_{Ba} = R_{Bb} = R_{Bc} = \infty \tag{12}$$

and

$$I_{Ac} = I_{Ba} = I_{Bb} = I_{Bc} = 0 \tag{13}$$

Striking out the columns of (1) containing the currents in (13) and the corresponding rows (Ac, Ba, Bb and Bc) the following two equations remain:

$$\begin{aligned} A_{11}I_{Aa} + A_{12}I_{Ab} &= 3E \\ A_{21}I_{Aa} + A_{22}I_{Ab} &= 3a^2E \end{aligned} \tag{14}$$

from which on substituting the numerical values for the A's from Table I the two currents I_{Aa} and I_{Ab} can be found. The total fault current to ground at A is:

$$I_{AF} = I_{Aa} + I_{Ab} \tag{15}$$

In the special case where R_{Aa}, R_{Ab} and R_{AF} are zero the expression for I_{AF} can be reduced to the following expression after a direct substitution for the A's in (14) is made:

$$I_{AF} = \frac{-\,3aZ_2E}{Z_0Z_1 + Z_0Z_2 + Z_1Z_2} \tag{16}$$

where:

$$\begin{aligned} Z_1 &= Z_{A1} + Z_{C1} \\ Z_2 &= Z_{A2} + Z_{C2} \\ Z_0 &= Z_{A0} + Z_{C0} \end{aligned} \tag{17}$$

3.3 Simultaneous Double Line-to-Ground Fault at A and Double Line-to-Ground Fault at B

Consider a fault-to-ground on phases "a" and "b" at A and phases "a" and "c" at B. Then:

$$R_{Ac} = R_{Bb} = \infty \tag{18}$$

Hence:

$$I_{Ac} = I_{Bb} = 0 \tag{19}$$

Striking out the two columns containing I_{Ac} and I_{Bb} and the two corresponding rows (Ac and Bb), the four following equations remain:

$$\begin{aligned} A_{11}I_{Aa} + A_{12}I_{Ab} + A_{14}I_{Ba} + A_{16}I_{Bc} &= 3E \\ A_{21}I_{Aa} + A_{22}I_{Ab} + A_{24}I_{Ba} + A_{26}I_{Bc} &= 3a^2E \\ A_{41}I_{Aa} + A_{42}I_{Ab} + A_{44}I_{Ba} + A_{46}I_{Bc} &= 3E \\ A_{61}I_{Aa} + A_{62}I_{Ab} + A_{64}I_{Ba} + A_{66}I_{Bc} &= 3aE \end{aligned} \tag{20}$$

A symbolic solution in terms of the sequence impedances for these currents becomes quite involved and it is advisable to substitute numerical values of the constants before solving for the four currents. The total fault currents at A and B, respectively, are (from (4) and (5) in connection with (19)):

$$I_{AF} = I_{Aa} + I_{Ab} \tag{21}$$

$$I_{BF} = I_{Ba} + I_{Bc} \tag{22}$$

In a similar manner faults to ground for any other combination of faulted phases can be found.

3.4 Single Line-to-Ground Faults at A and B in an Isolated System

Consider a fault-to-ground on phases "a" at A and "b" at B in an isolated system in which capacity can be neglected. Then:

$$R_{Ab} = R_{Ac} = R_{Ba} = R_{Bc} = \infty \tag{23}$$

and

$$I_{Ab} = I_{Ac} = I_{Ba} = I_{Bc} = 0 \tag{24}$$

Striking out the columns of (2) containing the currents in (24) and the corresponding rows $Aa - Bc$, $Ab - Ba$, $Ab - Bc$, $Ac - Ba$ and $Ac - Bb$ (all rows containing Ab, Ac, Ba and Bc), leaves only one equation in (2), which together with (2a) gives:

$$\begin{aligned} B_{11}I_{Aa} + B_{15}I_{Bb} &= 3(1 - a^2)E \\ I_{Aa} + I_{Bb} &= 0 \end{aligned} \tag{25}$$

Solving for these currents the result is:

$$I_{Aa} = -\,I_{Bb} = \frac{3(1 - a^2)E}{B_{11} - B_{15}} \tag{26}$$

Inserting the values of the B's from Table I this reduces to:

$$I_{Aa} = -\,I_{Bb} = \frac{3(1 - a^2)E}{Z_{1i} + Z_{2i} + Z_{0i} + 3(R_A + R_B)} \tag{27}$$

$$\begin{aligned} Z_{1i} &= Z_{A1} + Z_{B1} + 3Z_{C1} \\ Z_{2i} &= Z_{A2} + Z_{B2} + 3Z_{C2} \\ Z_{0i} &= Z_{A0} + Z_{B0} \\ R_A &= R_{Aa} + R_{AF} \\ R_B &= R_{Bb} + R_{BF} \end{aligned} \tag{28}$$

The subscript i (isolated) is used to distinguish these impedances for the isolated system from those used in (11), (17) and (38).

3.5 Phase-to-Phase Fault at A

Consider a fault between phases "a" and "b" at A. Then let:

$$I_{Ac} = I_{Ba} = I_{Bb} = I_{Bc} = 0 \tag{29}$$

Striking out the columns in (3) containing the currents in (29) and the corresponding rows (all rows containing Ac, Ba, Bb and Bc) only one equation is left:

$$C_{11}I_{Aa} + C_{12}I_{Ab} = 3(1 - a^2)E \tag{30}$$

It is further known from (3a) that:

$$I_{Ab} = -I_{Aa} \tag{31}$$

Substituting (31) and the constants C_{11} and C_{12} from Table I in (30) the result is:

$$I_{Aa} = -I_{Ab} = \frac{(1 - a^2)E}{Z_1 + Z_2 + R_{Aa} + R_{Ab}} \tag{32}$$

$$\begin{aligned} Z_1 &= Z_{A1} + Z_{C1} \\ Z_2 &= Z_{A2} + Z_{C2} \end{aligned} \tag{33}$$

which is a well-known expression for a phase-to-phase fault.

3.6 Three-Phase Fault at A

For this case:

$$I_{Ba} = I_{Bb} = I_{Bc} = 0 \tag{34}$$

Striking out the columns in (3) containing the currents in (34) and the corresponding rows, three equations remain, any two of which together with the equation from (3a) relating to the currents at A give:

$$\begin{aligned} C_{11}I_{Aa} + C_{12}I_{Ab} + C_{13}I_{Ac} &= 3(1 - a^2)E \\ C_{21}I_{Aa} + C_{22}I_{Ab} + C_{23}I_{Ac} &= 3(1 - a)E \\ I_{Aa} + I_{Ab} + I_{Ac} &= 0 \end{aligned} \tag{35}$$

from which the currents can be found.

In the special case where the fault resistances are all zero, the three currents are equal in magnitude and related as follows:

$$I_{Aa} = aI_{Ab} = a^2I_{Ac} \tag{36}$$

The rank of the system determinant in (35) is therefore 1. Using any of the first two equations in (35) in connection with (36) and the constants in Table I, the result is:

$$I_{Aa} = aI_{Ab} = a^2I_{Ac} = \frac{E}{Z_1} \tag{37}$$

$$Z_1 = Z_{A1} + Z_{C1} \tag{38}$$

3.7 Phase-to-Phase Fault at A and Phase-to-Ground Fault at B

Consider a fault between phases "a" and "b" at A and a fault to ground on phase "c" at B. Then:

$$R_{Ac} = R_{Ba} = R_{Bb} = \infty \tag{39}$$

and

$$I_{Ac} = I_{Ba} = I_{Bb} = 0 \tag{40}$$

As explained in a preceding section the three first equations in (3) together with the first in (3a) and the three last equations in (1) may be used for this case.

Striking out the columns I_{Ac}, I_{Ba} and I_{Bb} and the corresponding rows $Aa - Ac$, and $Ab - Ac$ in the three first equations in (3) leaves only the first equation. Similarly by striking out the columns I_{Ba}, I_{Bb}, and the corresponding rows Ba and Bb in the three last equations in (1) leaves only the last equation. Hence:

$$\begin{aligned} C_{11}I_{Aa} + C_{12}I_{Ab} + C_{16}I_{Bc} &= 3(1 - a^2)E \\ A_{61}I_{Aa} + A_{62}I_{Ab} + A_{66}I_{Bc} &= 3aE \end{aligned} \tag{41}$$

and finally from (3a):

$$I_{Aa} + I_{Ab} = 0 \tag{42}$$

from which the three currents can be found. The A's and C's are given in Table I and Table II.

4. Conclusion

While the probability of all phases being faulted at both locations simultaneously is very remote, the three sets of equations (1), (2) and (3) have been given in such a form that they conveniently will provide a solution for any combination of phases faulted from a single line-to-ground fault at one location to the most involved fault condition.

In Section (3) of this paper, in which special cases have been treated, only simple types of fault conditions have been shown in order to illustrate the method to be used and to prove that the general equations reduce to well-known formulas.

The constants given in Table I consist of the nine quantities S_a, S_b, S_c, T_a, T_b, T_c, U_a, U_b and U_c arranged as shown for each set of equations. Table II gives somewhat simpler values for the constants in cases where the positive and negative sequence impedances are assumed equal.

The voltages to ground at the two fault locations are given by (46) and (47) in the Appendix.

It is hoped that this development will provide a more unified presentation of fault current calculations in power networks.

ACKNOWLEDGMENT

The author wishes to express his appreciation to P. A. Jeanne and J. Riordan for valuable suggestions and checking of the mathematical relations in preparation of the paper.

APPENDIX

The standard notation for phase and sequence quantities is usually indicated by a subscript. Thus I_a, I_b, etc., means the current at the point of fault of phase "a" and "b," respectively. I_1, a^2I_1 and aI_1 are the positive sequence currents in phase "a," "b" and "c," respectively. In this treatment, however, complication arises from the fact that two points of faults are involved and it will be necessary to distinguish between the quantities at these two locations. This is most conveniently done by a double subscript, the first referring to the point of fault and the second to the phase or sequence in question. Thus I_{Aa}, I_{Ba}, etc., are the currents in the fault at A and B of phase "a" and I_{A1}, I_{B1} the positive sequence current at the two points of fault, respectively. Making use of this notation the fundamental equations for the sequence currents at fault A are:

$$\begin{aligned} I_{A0} + I_{A1} + I_{A2} &= I_{Aa} \\ I_{A0} + a^2 I_{A1} + a I_{A2} &= I_{Ab} \\ I_{A0} + a I_{A1} + a^2 I_{A2} &= I_{Ac} \end{aligned} \tag{43}$$

And at fault B:

$$\begin{aligned} I_{B0} + I_{B1} + I_{B2} &= I_{Ba} \\ I_{B0} + a^2 I_{B1} + a I_{B2} &= I_{Bb} \\ I_{B0} + a I_{B1} + a^2 I_{B2} &= I_{Bc} \end{aligned} \tag{44}$$

where the coefficient "a" is the sequence operator, having the value:

$$\begin{aligned} a &= -\frac{1}{2} + j\frac{\sqrt{3}}{2} \\ a^2 &= -\frac{1}{2} - j\frac{\sqrt{3}}{2} \end{aligned} \tag{45}$$

The voltages to ground at the two fault locations are given by:

$$\begin{aligned} V_{Aa} &= V_{A0} + V_{A1} + V_{A2} \\ &= (R_{Aa} + R_{AF})I_{Aa} + R_{AF}I_{Ab} + R_{AF}I_{Ac} \\ V_{Ab} &= V_{A0} + a^2 V_{A1} + a V_{A2} \\ &= R_{AF}I_{Aa} + (R_{Ab} + R_{AF})I_{Ab} + R_{AF}I_{Ac} \\ V_{Ac} &= V_{A0} + a V_{A1} + a^2 V_{A2} \\ &= R_{AF}I_{Aa} + R_{AF}I_{Ab} + (R_{Ac} + R_{AF})I_{Ac} \end{aligned} \tag{46}$$

$$
\begin{aligned}
V_{Ba} &= V_{B0} + V_{B1} + V_{B2} \\
&\qquad = (R_{Ba} + R_{BF})I_{Ba} + R_{BF}I_{Bb} + R_{BF}I_{Bc} \\
V_{Bb} &= V_{B0} + a^2V_{B1} + aV_{Ba} \\
&\qquad = R_{BF}I_{Ba} + (R_{Bb} + R_{BF})I_{Bb} + R_{BF}I_{Bc} \\
V_{Bc} &= V_{B0} + aV_{B1} + a^2V_{B2} \\
&\qquad = R_{BF}I_{Ba} + R_{BF}I_{Bb} + (R_{Bc} + R_{BF})I_{Bc}
\end{aligned}
\tag{47}
$$

Consider the positive sequence diagram in Fig. 2. Evidently:

$$
\begin{aligned}
V_{A1} &= E - (Z_{A1} + Z_{C1})I_{A1} - Z_{C1}I_{B1} \\
V_{B1} &= E - Z_{C1}I_{A1} - (Z_{B1} + Z_{C1})I_{B1}.
\end{aligned}
\tag{48}
$$

where V_{A1} and V_{B1} are the positive sequence voltages to ground at the two fault locations. Similar expressions can be obtained for V_{A2}, V_{B2}, V_{A0} and V_{B0}, except for the fact the E is zero in these cases and the impedances and currents are the negative and zero sequence quantities. They are:

$$
\begin{aligned}
V_{A2} &= -(Z_{A2} + Z_{C2})I_{A2} - Z_{C2}I_{B2} \\
V_{B2} &= -Z_{C2}I_{A2} - (Z_{B2} + Z_{C2})I_{B2}
\end{aligned}
\tag{49}
$$

$$
\begin{aligned}
V_{A0} &= -(Z_{A0} + Z_{C0})I_{A0} - Z_{C0}I_{B0} \\
V_{B0} &= -Z_{C0}I_{A0} - (Z_{B0} + Z_{C0})I_{B0}
\end{aligned}
\tag{50}
$$

Solving (43) and (44) for the sequence currents the result is:

$$
\begin{aligned}
I_{A0} &= \tfrac{1}{3}(I_{Aa} + I_{Ab} + I_{Ac}) \\
I_{A1} &= \tfrac{1}{3}(I_{Aa} + aI_{Ab} + a^2I_{Ac}) \\
I_{A2} &= \tfrac{1}{3}(I_{Aa} + a^2I_{Ab} + aI_{Ac})
\end{aligned}
\tag{51}
$$

$$
\begin{aligned}
I_{B0} &= \tfrac{1}{3}(I_{Ba} + I_{Bb} + I_{Bc}) \\
I_{B1} &= \tfrac{1}{3}(I_{Ba} + aI_{Bb} + a^2I_{Bc}) \\
I_{B2} &= \tfrac{1}{3}(I_{Ba} + a^2I_{Bb} + aI_{Bc})
\end{aligned}
\tag{52}
$$

Substituting the expressions for the sequence currents in (48), (49) and (50) the result is:

$$
\begin{aligned}
V_{A1} &= E - \tfrac{1}{3}(Z_{A1} + Z_{C1})(I_{Aa} + aI_{Ab} + a^2I_{Ac}) \\
&\qquad - \tfrac{1}{3}Z_{C1}(I_{Ba} + aI_{Bb} + a^2I_{Bc}) \\
V_{A2} &= -\tfrac{1}{3}(Z_{A2} + Z_{C2})(I_{Aa} + a^2I_{Ab} + aI_{Ac}) \\
&\qquad - \tfrac{1}{3}Z_{C2}(I_{Ba} + a^2I_{Bb} + aI_{Bc}) \\
V_{A0} &= -\tfrac{1}{3}(Z_{A0} + Z_{C0})(I_{Aa} + I_{Ab} + I_{Ac}) \\
&\qquad - \tfrac{1}{3}Z_{C0}(I_{Ba} + I_{Bb} + I_{Bc})
\end{aligned}
\tag{53}
$$

$$\begin{aligned} V_{B1} &= E - \tfrac{1}{3}Z_{C1}(I_{Aa} + aI_{Ab} + a^2I_{Ac}) \\ &\qquad - \tfrac{1}{3}(Z_{B1} + Z_{C1})(I_{Ba} + aI_{Bb} + a^2I_{Bc}) \\ V_{B2} &= -\tfrac{1}{3}Z_{C2}(I_{Aa} + a^2I_{Ab} + aI_{Ac}) \\ &\qquad - \tfrac{1}{3}(Z_{B2} + Z_{C2})(I_{Ba} + a^2I_{Bb} + aI_{Bc}) \\ V_{B0} &= -\tfrac{1}{3}Z_{C0}(I_{Aa} + I_{Ab} + I_{Ac}) \\ &\qquad - \tfrac{1}{3}(Z_{B0} + Z_{C0})(I_{Ba} + I_{Bb} + I_{Bc}) \end{aligned} \tag{54}$$

Substituting (53) and (54) in (46) and (47) the six original equations in (1) are obtained.

REFERENCES

1. C. L. Fortescue: "Method of Symmetrical Coordinates Applied to the Solution of Polyphase Networks," *Trans. A. I. E. E.*, 1918, Vol. 37, pages 1027–1114.
2. E. Clarke: "Simultaneous Faults on Three-Phase Systems," *Trans. A. I. E. E.*, 1931, Vol. 50, pages 919–941.
3. Wagner and Evans: "Symmetrical Components" (McGraw-Hill, 1933).
4. Monseth and Robinson: "Relay Systems—Theory and Application" (McGraw-Hill, 1935).
5. Lyon: "Applications of the Method of Symmetrical Components" (McGraw-Hill, 1937).
6. G. Oberdorfer: "Der Erdschluss und seine Bekämpfung" (Springer, 1930).
7. M. J. Fallou: "Propagation des Courants de haute Frequence Polyphasés le long des Lignes Aériennes de Transport d'Energie Affectées de Courts—Circuits ou de Defauts d'Isolement," *Bulletin de la Société Française des Électriciens*, August, 1932, pages 787–864.

A Single Sideband Musa Receiving System for Commercial Operation on Transatlantic Radio Telephone Circuits*

By F. A. POLKINGHORN

In the operation of short-wave radio telephone circuits selective fading is observed which is a result of the combination at the receiving antenna of waves which have arrived from the transmitter over paths of different lengths. The poor quality resulting from this fading may be mitigated by increasing the directivity of the receiving antenna in the vertical plane so as to favor the waves arriving at one angle to the exclusion of others. Friis and Feldman have described an experimental system designed to accomplish this end which they call a "Musa" receiving system. This system was found under certain transmission conditions to give an improvement in the grade of circuit which could be obtained. A commercial installation of this type has now been constructed for use on the single sideband circuits of the American Telephone and Telegraph Company from England. Two receivers have been provided for the operation of four radio telephone circuits.

The antenna system consists of a row of sixteen rhombic antennas two miles long, each antenna connected by a separate transmission line to receivers located near the center of the row of antennas. In each receiver the signals from the antennas are combined in the proper phase to permit simultaneous reception from three adjustable vertical angles. The three signals are then added through delay equalizing circuits or discretely selected on the basis of amplitude to obtain diversity reception. A fourth branch of the receiver has its vertical angle of reception continuously varying and is used to set automatically the angles of reception of the three diversity branches. The delay equalization is also automatically adjusted. A recorder is provided which continuously registers the relative carrier field strength with variation of vertical angle of reception, and the amount of delay equalization.

Introduction

In the operation of short-wave radio telephone circuits fading is observed which is caused by the combination at the receiving antenna of waves which arrive at different vertical angles and which have traveled from the transmitter over paths of different lengths. This fading may be mitigated by increasing the directivity of the receiving antenna in the vertical plane so as to favor the waves arriving over one path to the exclusion of the others.[1, 2] It is not

* *Proc. I.R.E.*, April 1940.

[1] E. Bruce, "Developments in Short Wave Directive Antennas," *Proc. I.R.E.*, Vol. 19, pp. 1406–1433, August 1931.

[2] E. Bruce and A. C. Beck, "Experiments with Directivity Steering for Fading Reduction," *Proc. I.R.E.*, Vol. 23, pp. 357–371, April 1935.

possible, however, to increase this directivity to any great extent with an ordinary antenna system before it is found that the signal arrives outside the angular range of the antenna an appreciable part of the time. To overcome this difficulty Friis and Feldman experimented with a receiving system consisting of a number of antennas, each having moderate directivity and each connected by a separate transmission line to a receiver where the outputs are phased by a variable phase shifting system in such a manner as to give a system of high, variable directivity. A system of this kind, which they called a "musa" system from the initial letters of "multiple unit steerable antenna," was built and found to give under most transmission conditions an improvement in the grade of circuit which could be obtained.[3] Accordingly it was decided that a commercial system should be built for use on the circuits of the American Telephone and Telegraph Company from England. A corresponding system of modified design has been built by the British General Post Office.[4] The purpose of this paper is to review a few of the principles upon which a musa receiver operates, to describe the equipment which has been built in this country, and to discuss some of its operating characteristics.

The transmissions which are to be received are of the so-called twin single-sideband reduced-carrier type described by Oswald [5] and consist of two sidebands, representing two distinct speech channels, on opposite sides of a carrier which is 16 to 26 db below the maximum sideband amplitude. Under normal conditions one of the sidebands is adjacent to the carrier while the second is spaced by the width of one sideband from the carrier. A single sideband receiver for this type of transmission has been described by Roetken [6] and many of the features discussed by him were developed for use also in the musa receivers. These features include highly stable oscillators, crystal filters and automatic tuning circuits.

Outline Description of Receivers

A block schematic of one channel of the commercial musa system is shown in Fig. 1. The sixteen rhombic antennas are placed in a line two miles long in the direction of the English transmitting station.

[3] H. T. Friis and C. B. Feldman, "A Multiple Unit Steerable Antenna for Short-Wave Reception," *Proc. I.R.E.*, Vol. 25, pp. 841–917, July 1937; *B.S.T.J.*, Vol. XVI, No. 3, pp. 337–419, July 1937.

[4] A. J. Gill, Wireless Section, Chairman's Address, *Jour. I.E.E.*, Vol. 84, No. 506, pp. 248–260, February 1939.

[5] A. A. Oswald, "A Short-Wave Single Sideband Radio Telephone System," *Proc I.R.E.*, Vol. 26, No. 12, pp. 1431–54, December 1938.

[6] A. A. Roetken, "A Single Sideband Receiver for Short-Wave Telephone Service," *Proc. I.R.E.*, Vol. 26, No. 12, pp. 1455–65, December 1938.

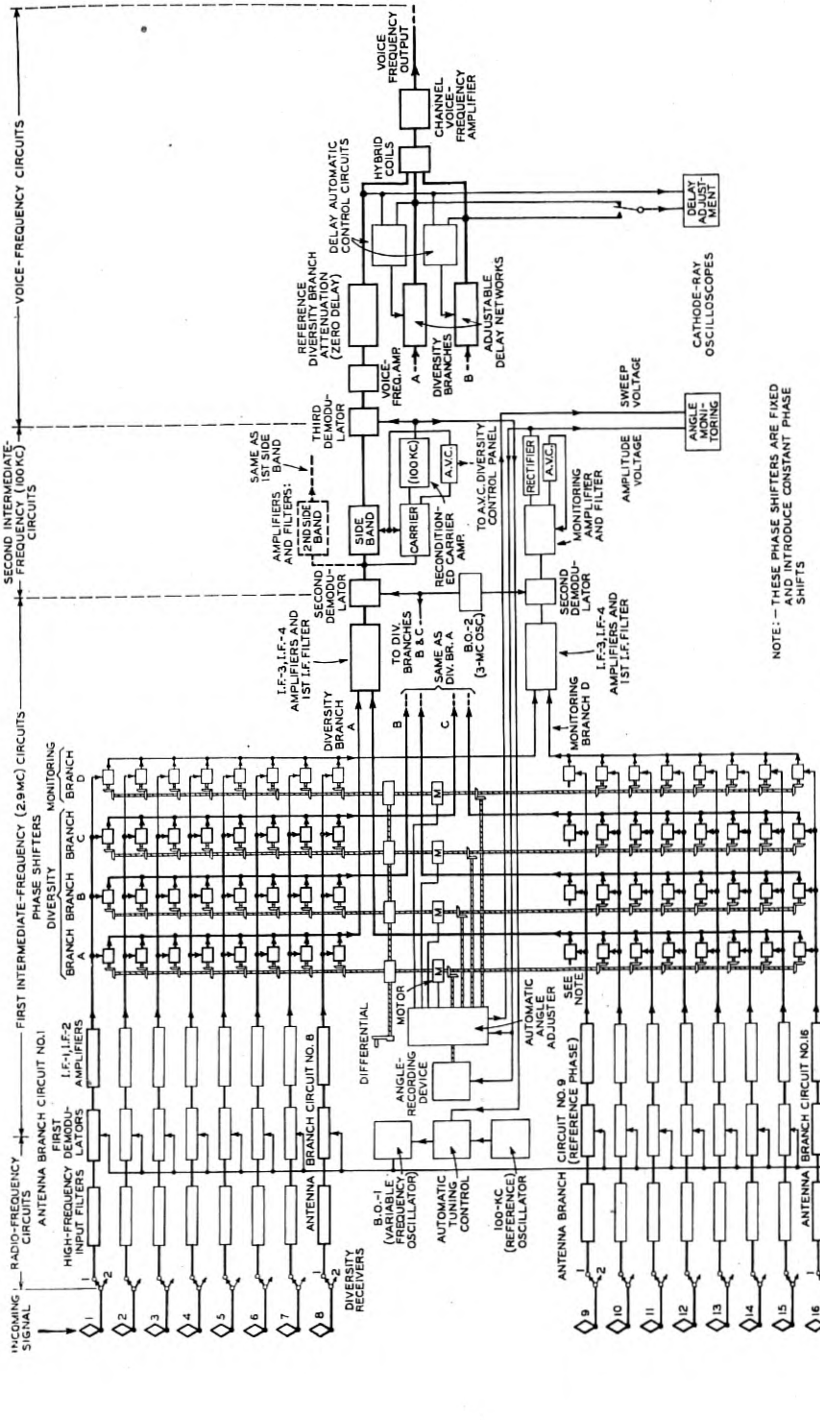


Fig. 1—Block schematic diagram of one musa receiver.

Separate transmission lines lead from each antenna to a building placed a little to one side of the rear of the ninth antenna. Two receivers, only one of which is shown in the figure, are connected in parallel to each transmission line. Each receiver is designed to receive five specific frequencies, ranging from 4,810 kc. to 18,620 kc., assigned to the corresponding transmitter in England.

After passing through selective input circuits the signals from each antenna are demodulated by a common oscillator to a band adjacent to a carrier frequency of 2,900 kc. The signals, after going through two stages of intermediate frequency amplification are then applied to the inputs of four phase shifter systems in parallel. In each of these phase shifter systems the signals from the sixteen antennas are combined so as to give reception from a particular vertical angle. This angle can be varied by a mechanical movement of a phase shifter drive shaft. Three of the phase shifter system outputs are used for a three-branch angular-diversity system in which the signals arriving over three separate paths are separately received and then either combined or individually selected for connection to the line, while the fourth branch is used for monitoring to determine where the phase shifters of the three diversity branches should be set in order to receive the best signals. From each phase shifter group the circuit continues through the first intermediate frequency filter and two further stages of amplification to the second demodulator where the 2,900 kc. carrier frequency is shifted to 100 kc. The carriers and sidebands of each diversity branch are then amplified separately and again combined in the final demodulators to give three distinct voice frequency outputs for each sideband. These three outputs are either combined after inserting variable delay in two of the branches or, optionally, the branch having the greatest signal at any instant is connected to the line. Both of these operations are performed automatically. The output of the monitoring phase shifter group is also heterodyned to 100 kc. and after amplifying the carrier only it is rectified and applied to an automatic system for adjusting the phase shifters of the three diversity branches.

A general view of the receivers is shown in Fig. 2. The principal parts of the two musa receivers occupy three rows of bays each about 25 feet long and $11\frac{1}{2}$ feet high. The row shown on the right contains the input circuits and first demodulators for both receivers. The middle row contains the remaining equipment for one receiver and the left row that for the second receiver. In addition there are five bays of rectifiers and power control equipment located in a fourth row which is not shown.

Fig. 2—View of musa receivers.

General

When waves arriving over several paths from the same transmitter are demodulated in a simple receiver the severity of the resultant selective fading is dependent upon the relative amplitudes at the demodulator of the several path contributions, the differences in the times of transmission over the several paths, and the rates at which the path lengths are varying.[7] When the difference in the time of

[7] R. K. Potter, "Transmission Characteristics of a Short-Wave Telephone Circuit," *Proc. I.R.E.*, Vol. 18, No. 4, pp. 581–648, April 1930.

transmission over two paths is t there are alternate maxima and minima in the frequency spectrum caused by these two components which are separated by $1/2t$. Continuous small changes in the lengths of the paths cause these maxima and minima to wander back and forth through the spectrum. By separating the waves arriving at distinctly different angles the musa receiver succeeds, for the most part, in separating those waves which have greatly different transmission times and thus widens the frequency interval between a maximum and an adjacent minimum. As the interval increases the fading appears less selective. The signal appearing to arrive at any one angle, however, is in reality composed of a bundle of waves, the components of which have traveled over slightly different paths and which might be expected to be nearly alike in amplitude and transmission time but not in phase. As a consequence it is to be expected that the general fading on a single-angle musa receiver will be greater than on an ordinary receiver and it is essential that some form of diversity be used to insure a satisfactory output amplitude at all times. Sudden shifts in the received angle of signals will also give general fading which will be greater the greater the angular discrimination of the musa system.

A musa receiver differs from an ordinary receiver in that there are a number of separate antenna branches, the outputs of which must be added in the proper phase over an appreciable band of frequencies.

When delay equalization is used between the various diversity branches, these branches must also have equal phase shifts if the audio-frequency outputs are to add properly. In both antenna and diversity branches the problem of keeping equal phase shifts is complicated by the action of the automatic volume control system which changes the operating condition of the vacuum tubes over a wide range. In designing these receivers a nominal overall value of non-uniformity of $\pm 10°$ was taken as acceptable and an effort made to keep the phase uniformity of individual elements to within one or two degrees wherever possible.

Within the receiving station all radio-frequency wiring is made with a flexible coaxial cable having rubber insulation. The various panels composing the receivers are placed on the racks with a view to operation and maintenance rather than ease of wiring and consequently long leads between panels are frequently necessary. For this purpose the circuit impedance is dropped to 70 ohms and at a number of points brought out to jack panels to facilitate testing.

Coaxial jacks are used which fit into the usual jack strips. Normalling jacks are not available and consequently it is necessary to have plugs in jacks during operation. In order to avoid cords, which would be in the way, the jacks to be connected regularly are mounted adjacent to each other and connected together by two plugs mounted in a shell similar to that commonly used for terminating resistances. Alternating current for cathode heating is supplied to conduit outlets near each panel. Flexible cords with plugs complete the circuit to the panels. All audio frequency, bias, and signal wiring is made into cables in the usual telephone manner. Wires having a potential of over 150 volts to ground are placed in conduit and safety switches are provided to remove the voltage from a panel when the panel cover is removed.

Antenna System

The degree of vertical resolution and the signal-to-noise improvement of a musa antenna system are functions of the overall length and number of unit rhombic antennas used. The decision to build a sixteen-antenna system was based on experience with the six-antenna system and took into consideration the land necessary, the cost of antennas and transmission lines, and the complexity of the receiving equipment, as well as the resolution which it would be practical to use and still have it possible for the operator or automatic equipment to follow changes in the direction of signal arrival.

When the spacing between unit antennas of a musa system is several wave-lengths there will be more than one vertical angle at which the phase shifters will simultaneously phase the antenna outputs. The spacing between antennas is so chosen that the range traversed by the lowest of these angles will be the range covered by useful signals. Fortunately the angle of useful signals varies with frequency in such a manner as to permit a variation in frequency over the range desired with a fixed spacing. The unit rhombic antenna is designed to have a null at the position of the second phasing maximum and reception is thus confined essentially to the lowest phasing maximum.

Extensive study did not disclose a better unit rhombic antenna than the one used in the experimental system and consequently an antenna 590 feet (180 meters) long, 60 feet high, and having each side angle equal to 140 degrees was used. The spacing is 656 feet (200 meters) between corresponding parts of adjacent antennas.

A representative directive diagram of a unit rhombic and the 16-unit array in a vertical plane in the line of the antennas is shown plotted in rectangular coordinates in Fig. 3. A polar diagram of the major lobe in three different positions, corresponding to three possible angles of diversity reception, is shown on Fig. 4*a*. In this figure the

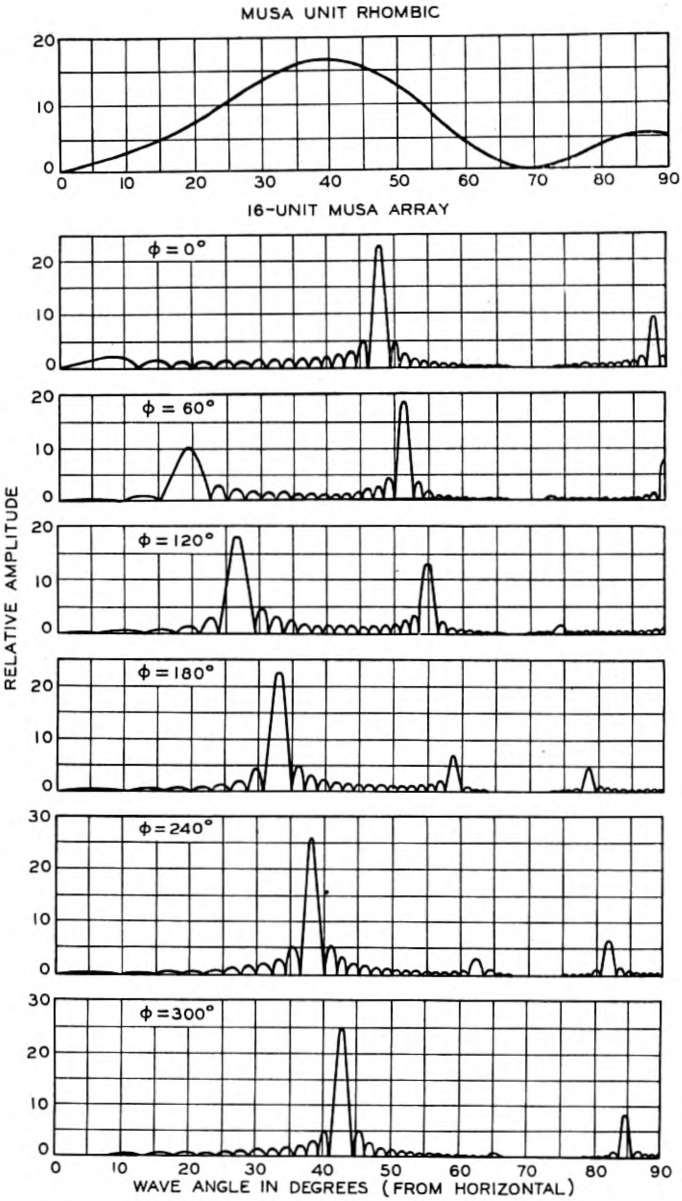


Fig. 3—Directive diagram in the vertical plane of a unit rhombic antenna and the 16 unit musa array. Frequency 4700 Kc. ϕ = phase shifter setting.

dotted outline is the diagram of the unit rhombic antenna characteristic enlarged 16 times. The complete diagram is of course a solid. Figure 4*b* is an attempt to show how the middle lobe of Fig. 4*a* looks when viewed from the ground plane at a horizontal angle of 45° from the line of the antennas. The contour lines on this leaf-shaped figure are lines of equal reception. All of these diagrams are for a frequency of 4,700 kc., near the low end of the range of received signals. At higher frequencies the lobes will be more slender and the angle of maximum reception will be lower. Fortunately the latter corresponds to the trend of the received signals.

When the outputs of several antennas are connected to a receiver and added in the proper phase the total signal can be made equal to

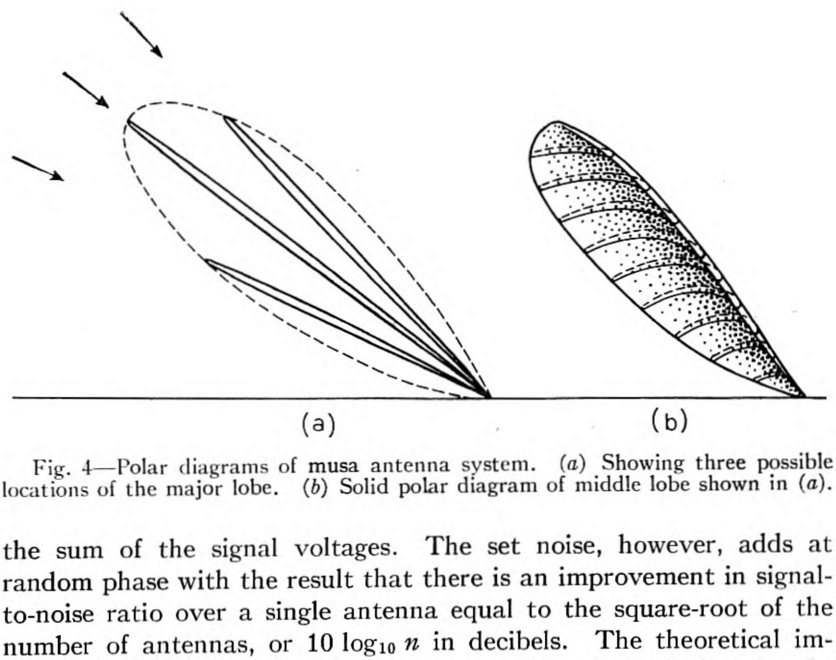


Fig. 4—Polar diagrams of musa antenna system. (*a*) Showing three possible locations of the major lobe. (*b*) Solid polar diagram of middle lobe shown in (*a*).

the sum of the signal voltages. The set noise, however, adds at random phase with the result that there is an improvement in signal-to-noise ratio over a single antenna equal to the square-root of the number of antennas, or $10 \log_{10} n$ in decibels. The theoretical improvement of a 16-antenna system is, therefore, 12 decibels. On the assumption that received noise comes from a random direction with respect to the signal a similar result is obtained. At a particular time, however, the principal noise may be arriving from such a direction as to allow of no discrimination by the antenna system, or at another time to allow much more than 12 db discrimination.

The received signal in an antenna is a result of both the direct wave, arriving from some angle above the horizon, and the same wave front reflected from the ground at a point ahead of the antenna. If the phase of the induced voltage is to progress uniformly from antenna to antenna it is essential that the ground be homogeneous and flat. For this and other reasons a flat marsh near Manahawkin, N. J. was chosen for the station site. The ground is level to within less than one foot, except where there are inlets, for the entire length of the antenna and for a considerable distance ahead.

As shown in Fig. 3, in addition to the major lobe caused by the phasing factor there are fourteen minor lobes. The amplitude of the first three starting from a major lobe are approximately $2/3\pi$, $2/5\pi$, $2/7\pi$ of the major lobe. However, when the major lobe is at a very low or very high angle it is greatly reduced in amplitude by the unit antenna directional characteristic while the adjacent minor lobes on one side may not be reduced to any such extent. Consequently the ratio of the amplitude of the major to minor lobes may be much less than the values given and signals from two or more angles might be received simultaneously with comparable amplitude on the same diversity branch and so defeat the purpose of the system.

It has been shown by John Stone Stone and others that if the amplitudes of the currents contributed by the various units of the antenna system are tapered in such a manner that the central units contribute more than the end units a reduction in the amplitude of the minor lobes can be obtained. However, this is accompanied by a widening of the major lobe and a reduction in the signal-to-noise improvement obtained. For antenna systems having only a few units there appears to be a net advantage in tapering but for the sixteen-unit system under discussion a large amount of tapering broadens the major lobe so that it extends over the normal first and possibly second minor lobes. Since the remaining lobes are already of a low enough amplitude to be comparable with those which might be produced by inescapable errors in phase and amplitude of the various unit contributions, there appears to be no particular advantage in much tapering in this system. Provision has, however, been made to obtain tapering should it ever be found desirable. Under normal conditions all antenna branch amplifiers are operated at the same gain so as to use only the small tapering caused by the losses in transmission lines.

The antennas are coupled to the transmission lines through metallic core transformers which pass a band from 4,000 kc. to over 20,000 kc. with a loss of less than 1 db. The transformers are equipped with lightning arrestors and arranged so that the total d.c. loop-resistance of the transmission line, antenna, transformer, and antenna terminating resistance can be checked from the station.

The transmission lines from the antennas to the receiver are of the coaxial type made of copper tubing $1\frac{1}{4}''$ in inside diameter surrounding a central conductor of $\frac{3}{8}''$ outside diameter. Ceramic insulators $\frac{1}{8}''$ thick are spaced every 16″ and a locking insulator is placed every 250 feet to prevent creeping of the inside pipe. The velocity of propagation of the line is 0.980 of that in space. The attenuation amounts to about 1 db per 1,000 feet at 20 megacycles.

The lines are buried to protect them from mechanical injury and to prevent phase errors due to differences in expansion. Bacterial growth in the marsh makes the soil extremely corrosive and it was necessary to protect the lines by coatings of tar and asbestos tape. The lines are kept under gas pressure at all times.

In a musa receiving system a saving of nearly one half of the transmission line can be made if the receiving equipment is located at the center of the antenna system rather than at one end. Furthermore, since the average length of transmission line will be cut in half, the diameter of the coaxial transmission line also could be cut in half and still maintain the same signal loss. The economic, and to a lesser extent the locational, advantages of the center position were so great that the equipment was so placed in spite of certain technical disadvantages.

Theory of Phase Shifter System with Center Location of Receivers

At this point some of the theory of the musa phase shifting system will be reviewed and extension made to cover the situation of the receiver being located near the middle of the antenna system.

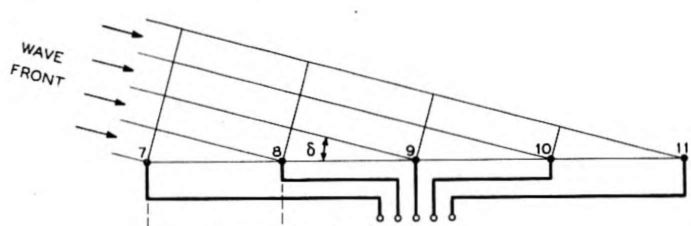


Fig. 5—Diagram of wave front approaching antenna.

Assume a plane wave front progressing toward the earth from the Kennelly-Heaviside layer at an angle δ with respect to the horizon and impinging upon the receiving antennas indicated as points 7, 8, 9, 10, 11, etc. (Fig. 5). Let the receiving station be located at antenna 9. Let the time of arrival at the receiving station of the voltages induced in the various antennas by this wave front be computed with respect to that of antenna 9. The wave front will arrive at antenna 9 at a time $\frac{1}{c}(d \cos \delta)$ later than at antenna 8. The voltage from antenna 8 will require a time $\frac{d}{v}$ to reach the receiving station over the transmission line, making the net time delay between the two outputs equal to

$$t_{9-8} = d\left(\frac{1}{v} - \frac{1}{c}\cos \delta\right), \tag{1}$$

where d is the horizontal distance between antennas
c is the velocity of transmission in space
v is the velocity of transmission in the transmission line

Similarly the difference in time of arrival of the voltages from antennas 9 and 7 will be:

$$t_{9-7} = 2d\left(\frac{1}{v} - \frac{1}{c}\cos \delta\right). \tag{2}$$

The voltages from all three antennas could be made to add in phase if a delay of

$$2d\left(\frac{1}{v} - \frac{1}{c}\cos \delta\right) \tag{3}$$

were added to the output from antenna 9, and half as much to the output of antenna 8. For a greater number of antennas these delays would have to be correspondingly increased.

Now consider the outputs of antenna 9, and those antennas which lie behind it. It will immediately be seen that the time between the arrival of the voltage caused by the wave front at antenna 9 and antenna 10 will be the sum of the times of transmission of the wave front from 9 to 10 in space and on the transmission line back from 10 to 9, which is similar to equation (1) except that the sign between the space and transmission line components is reversed. Similarly the equation for the delay between the outputs of antennas 9 and 11 is similar to that between 7 and 9 with the same sign reversed, and likewise with the other antennas.

Considering either the group of antennas ahead of the station or the group of antennas behind the station, and considering antenna 9 as a part of whichever group is being considered, it will be seen that the delay compensation which it is necessary to insert in the various antenna outputs to make the signal voltages add up in phase at the receiving station is always an integral multiple of either:

$$d\left(\frac{1}{v} - \frac{1}{c}\cos\delta\right) \tag{4}$$

or

$$d\left(\frac{1}{v} + \frac{1}{c}\cos\delta\right), \tag{5}$$

so that for either the front or the rear group the delay compensation could be adjusted by means of a single shaft geared to various individual delay compensators through gears having integral ratios.

Practically, the difficulty in building continuously variable delay circuits resulted in the use of continuously variable phase shifters. For any particular delay t there is a corresponding phase shift $\frac{2\pi ct}{\lambda}$ radians. Since this phase shift is a function of frequency, for a given phase shifter setting the frequencies in a wide band transmission will not all be in phase. The substitution of phase shifters for delay compensating circuits therefore results in a restriction of the band of simultaneous reception from a given angle. For the group of antennas ahead of the station the band restriction is small, since the space and transmission line paths give a partial delay compensation. For the antennas behind the station the net delay difference between antennas is large and the band is restricted to a much greater extent.

When using phase shifters it is possible to get minus as well as plus values of phase shift. This makes it possible to reverse the order in which delay compensation was assumed to be added in the above discussion so that no phase compensation is added to antenna 9, one negative unit is added to antenna 8, two to antenna 7, etc. In the equipment herein described the phase shifters were so designed.

Assume that the phase shifters of the front group are all set alike and then changed so as to receive from an angle δ. There will have been introduced a change of phase of

$$\phi = \frac{2\pi cd}{\lambda}\left(\frac{1}{v} - \frac{1}{c}\cos\delta\right) \tag{6}$$

between each of the antenna outputs.

Similarly if the phase shifters of the rear group were initially set alike, and similar to the initial condition of the front group, and then changed to receive from an angle δ the change of phase which would be required would be

$$\phi = -\frac{2\pi cd}{\lambda}\left(\frac{1}{v} + \frac{1}{c}\cos\delta\right). \tag{7}$$

The negative sign ahead of this equation takes account of the fact that an increase in the angle δ requires that more phase will have to be taken from line 8 while less phase will have to be subtracted from 10.

It will be noted that for a given change in δ, however, the amount of change of phase is the same absolute amount for the two groups. This indicates that it should be possible to connect the front and rear groups of phase shifters together and drive them as a single unit. To do this it will be necessary to connect the shafts together after they have been moved from their initial position by the amounts given in the above equations. The difference in phase is

$$\phi_{8-10} = \frac{4\pi d}{\lambda} \frac{c}{v}. \tag{8}$$

It will be noted that for a given installation this value is dependent upon the received frequency and will have to be changed when the received frequency is changed.

In the receivers herein described the phase shifters connected to antenna 9 are not driven. The phase shifters connected to antennas 1 to 8 are driven at ratios of 8 : 1, 7 : 1, 6 : 1, 5 : 1, 4 : 1, 3 : 1, 2 : 1 and 1 : 1 respectively. The phase shifters of antennas 10 to 16 are driven in the opposite direction at ratios of 1 : 1, 2 : 1, 3 : 1, 4 : 1, 5 : 1, 6 : 1, and 7 : 1 respectively. A differential gear mechanism is inserted between the two groups. Changing the position of the ring gear of the differential permits a mechanical shift equivalent to the phase shift given in equation (8) to be inserted between the two groups. This change may be made while the drive shafts are in motion. The ring gear drives of the differentials of the four groups of phase shifters are all connected by a common adjustment shaft so as to insure that the monitoring branch does not give a false indication of receiver performance by being set differently than the diversity branches.

A front view of the phase shifting system is shown on the left of Fig. 6. The vertical shafts drive the individual phase shifting condensers through spiral gears, of which there are eight inside each of the lower cast boxes and seven inside each of the upper cast boxes. The horizontal shafts connect to the cam switches used with the automatic adjustment feature. A rear view showing a few of the individual phase shifting condensers is given in Fig. 7. Each phase shifting condenser consists of four quadrantal stators each consisting of two plates between which revolves an eccentric circular plate. The rotor plates of the condensers in a horizontal row (see Figs. 1 and 7) are connected in parallel through the demountable brushes, and a wire which is behind the brush support, to the output of one of the antenna branch amplifiers. The corresponding stator plates of the condensers in a vertical row are connected together with coaxial wiring and to the four terminals of an artificial quarter-wave transmission line, which

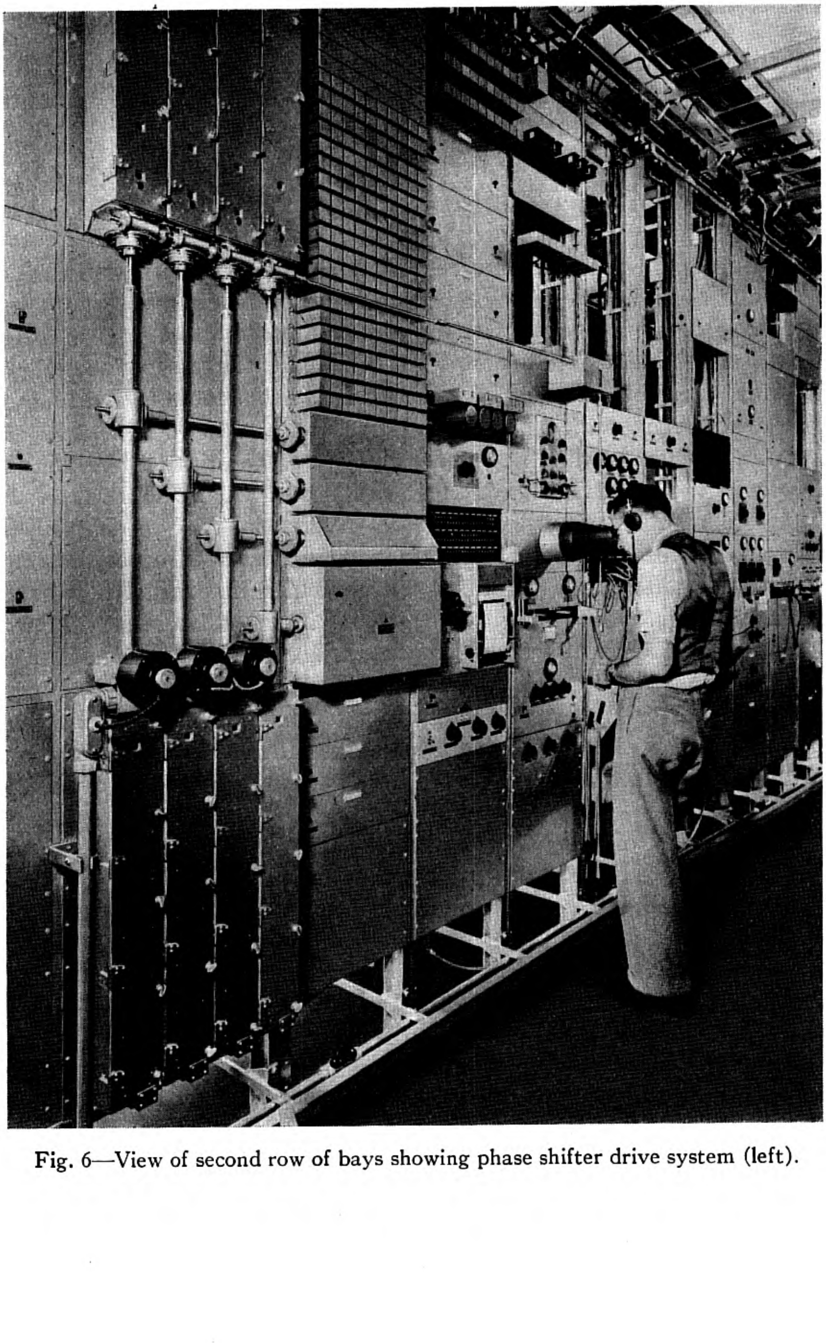


Fig. 6—View of second row of bays showing phase shifter drive system (left).

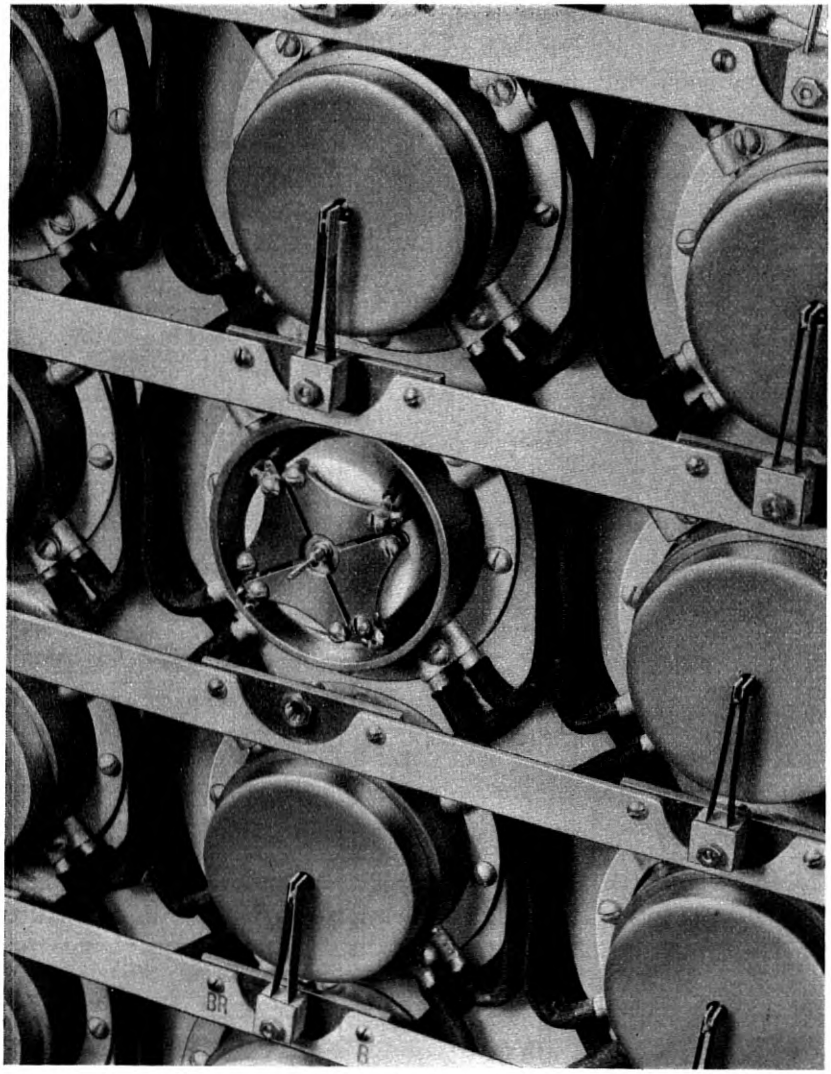


Fig. 7—Portion of a group of phase shifting condensers. Brush and cover removed from center condenser to show construction.

in turn is connected to the output amplifier. The quarter-wave line forms a combining network for the contributions of the various phase shifters. The two groups of eight phase shifters forming one diversity branch are each treated as units so as to facilitate isolating one group in case of trouble or the necessity of reducing the resolution of the antenna system.

There are two opportunities for undesirable interchanges of power in the parallel phase shifter system; there can be an interchange between antenna branches through condensers in the same diversity group, and an interchange between diversity branches through condensers in the same antenna branch. Both of these can be reduced to the required values by a proper proportioning of input and output impedances with respect to the condenser reactance. This results in a large loss, about 40 db, through the phase shifting system.

Input Circuits and First Demodulators

Obtaining uniformity of phase shift in the circuits preceding the first detectors is facilitated by reducing the number of circuit elements and selectivity to a minimum. No high-frequency amplification is used for this reason. The selectivity required to avoid image reception is materially lowered by the choice of a high operating frequency for the first intermediate frequency amplifiers. In addition to the usual requirement of high selectivity and high voltage transformation ratio, the input circuits must have uniformity in gain and phase, and must properly terminate the transmission lines so that multiple reflections will not shift the effective input amplitude or phase.

These factors seemed to preclude use of the usual variable tuned circuits on account of the time required to change from one frequency to another, and consequently fixed tuned circuits were used. Five input circuits are mounted on a panel and the switches for changing the input and output circuits of the two panels of each receiver which are mounted on a single bay are ganged. Each input circuit consists of two anti-resonant circuits capacity-coupled to each other and to the transmission line. It operates into the grids of the two first demodulator tubes which are paralleled. The first beating oscillator input is applied in push-pull between the cathodes of the tubes.

Oscillators and Automatic Tuning System

Three highly stable oscillators are used in the receivers, one for beating the signal frequency down to 2,900 kc., the second for beating it down further to 100 kc., and the third for use as a reference frequency for the automatic tuning system, or it may be used in the final demodulator when the automatic branch selector is used.

The first beating oscillator is of the coil-and-condenser type and covers the range from 7,000 kc. to 17,000 kc. Automatic tuning is used to compensate for long-time variations in its frequency, as well as any variations in the transmitter frequency, but every effort has been made to keep short-time variations to a minimum. The oscillator is contained in a cast box mounted on rubber supports. The inductance coils have an extremely low temperature coefficient and are mounted on cast supports for rigidity. The condenser is also of rigid construction. A variation in plate voltage of one volt gives a frequency variation of about 6 cycles at 18,000 kc.

A buffer amplifier connects between this oscillator and a push-pull power amplifier which delivers ten watts of output. This output is delivered to a transformer located on the center bay of the row of first demodulator bays. Coaxial cables of equal length distribute it to the first demodulators on the adjacent bays. A vacuum tube voltmeter connected across this transformer gives an alarm if the voltage fails.

In the automatic tuning system the incoming carrier at approximately 100 kc. is beaten with the local 100 kc. oscillator. The phase of the beat frequency is then split and the resultant two-phase output applied to a motor which drives a condenser in the first beating oscillator circuit until the beat frequency is reduced to zero. In order to avoid interruption of control due to fading, all three diversity branch carriers are simultaneously connected to the circuit.

The second beating oscillator operates at 3,000 kc. and is a standard broadcast oscillator which has been slightly modified.

The 100 kc. oscillator is required to have the same frequency as the center of the pass band of the carrier filters. Since these are only 40 cycles wide both filters and oscillators are made with low temperature coefficient crystals. The oscillator is of the bridge type described by Meacham [8] but without temperature control.

Automatic Volume Control and Final Demodulators

Only the carrier is rectified for automatic volume control purposes. In the 100 kc. amplifiers where the carrier and sideband are amplified separately, separate automatic volume control circuits are used. The time-constant of the carrier amplifier control is made 0.1 second, which is as fast as is practical with the narrow carrier filter used, and the time-constant of the sideband volume control is made variable but is generally set at a value of 1 second.

With the common automatic volume control used with diversity receivers the rectifiers are so connected as to give outputs which vary according to the square or first power of the branch input. The sum of the rectifier outputs is held substantially constant. If the combined signal output is also to be held constant, the final demodulators of the

[8] L. A. Meacham, "The Bridge-Stabilized Oscillator," *Proc. I.R.E.*, Vol. 26, No. 10, p. 1278, October 1938; *B.S.T.J.*, Vol. XVII, No. 4, p. 574, October 1938.

branch circuits must follow the same law, i.e., if linear rectifiers are used, the final demodulators should be linear, and if square-law rectifiers are used, the final demodulators should be square-law. When linear demodulators are used the output noise is independent of the strength of the incoming carrier and since the gains of all branch amplifiers are the same the noise output of each branch will be the same, assuming that received noise does not vary with the vertical angle of reception. As a consequence the total noise will be equal to the product of a single branch noise and the number of branches regardless of the signal contributions of each branch. If square-law detectors are used, however, the noise output of a branch will go down when the carrier in the final demodulator of that branch goes down and consequently the total noise will be proportional to the total signal. In a three-branch diversity system a theoretical improvement in signal-to-noise ratio varying up to 4.77 db can be had by using square-law demodulators rather than linear. For this and other reasons square-law final demodulators have been used in this equipment.

When delay equalization is used between the various diversity branches it is essential that the received carrier be used in the final demodulation process. Small changes in the lengths of the paths in space traversed by the sidebands being received by the various diversity branches make the phases at random and if all branches were demodulated by a common carrier the audio outputs would not add in phase. By using the carrier arriving over each path for the demodulation of the accompanying sideband the random relation disappears and the audio outputs can be added in phase.

When using an automatic branch selector which discretely chooses one branch at a time for connection to the line it is no longer necessary to consider phases in the diversity branches and a local carrier is used because it reduces output amplitude variations. With only one diversity branch connected to the output, only the corresponding volume control rectifier should be contributing to the automatic volume control voltage if the output volume is to be held as constant as possible. This result is obtained by putting a rectifier in the d.c. output lead of each branch volume control rectifier so that only the volume control rectifier having the highest amplitude will supply current to the load resistance.

Delay Circuits and Switches

On account of the fact that waves arriving at different vertical angles have taken different times in transit it is necessary to insert delay compensation in two of the three diversity branches when they are all connected to the output at once so that the audio outputs will add in phase. The branch receiving the waves at the highest angle, which are always the waves which have traversed the greatest distance, does not contain a delay circuit. The other two branches contain variable delay circuits having a maximum of 2,768 microseconds delay in steps of 31 microseconds. The band covered by the delay circuits is 6,000 cycles, the same as the width of the filter in the last intermediate frequency amplifier. The delay steps were made small enough so that it would be technically possible to phase properly over the entire band within less than a quarter cycle provided that the phase distortions of the transmission path, and the other parts of the receiver made it practical to do so.

The delay circuits each consist of 8 units having a delay of 31 microseconds and 9 units having a delay of 280 microseconds. Hand operated switches are provided to vary the delay in the usual decade switch manner. A motor driven switch is also provided which is arranged with two shafts connected by an intermittent movement so that when the eight small units of delay have been added a continued movement of the shaft removes these units from the circuit and simultaneously connects one of the large units, which is equivalent to nine small units. Further movement in the same direction successively adds in the smaller units again.

Automatic Delay Adjusting Circuits

A block schematic of the automatic delay adjusting apparatus as well as the automatic angle adjusting and recording equipment is shown in Fig. 8.

The proper delay compensation to phase the output of two diversity branches can be determined by connecting the outputs of the two branches to the two pairs of plates of a cathode ray oscilloscope and varying the delay in the lower angle path until a straight line pattern is obtained on the oscilloscope screen. This adjustment is facilitated by the restriction of the band in each branch at the oscilloscope to about an octave. If only a single frequency were used for this adjustment several positions of the delay adjustment might be found to give a straight line on the oscilloscope, the number of positions depending upon the frequency. In order that there may be only one position when the maximum delay is 2,768 microseconds the phase shift caused by this delay must be less than 180°, and consequently the frequency of operation must be less than 180 cycles. Where a band of frequencies a few hundred cycles wide is used, however, this difficulty is not encountered and there is only one adjustment of the delay which gives a straight line on the oscilloscope.

With the cathode ray oscilloscope there is no indication as to whether the delay in the circuit is too small or too great, but only that it is either correct or incorrect. A direct indication of whether the delay should be increased or decreased can be obtained by connecting one diversity branch to the push-pull input of a balanced modulator and another diversity branch to the parallel input of the same modulator, after having shifted the phase between the two branches by 90°. When the two receiver branch outputs are in the same phase the two modulator input voltages will add in quadrature and the currents in the plate circuits of the two modulators will be the same, but when the phases are not the same the current in one plate circuit will be greater and the other less than in the in-phase case, the sense of the unbalance depending upon whether too little or too much delay is in the circuit. A center-zero meter in a bridge connection in the plate circuits therefore can be made to indicate the sense of the necessary correction. By substituting a voltmeter relay for the indicating meter a motor drive of the variable delay circuit can be operated in such a manner as to adjust the delay to the correct value.

This automatic equipment must operate satisfactorily with circuits having types of privacy in which the energy bearing components of speech are shifted from their normal position in the frequency spectrum. The equipment is made, therefore, to operate on a band of frequencies from 250 to 750 cycles. Volume limiters are provided which keep the input to the automatic equipment from this band substantially constant. If the delay is never incorrect by more than 1,000 microseconds, which has been found to be true under all conditions of normal operation, the automatic equipment will bring it to the correct value. For greater errors the equipment tries to set the delay at values about 2,000 microseconds higher or lower than the correct value.

Since the delay adjustment operates on speech, the relay operation will be intermittent at a syllabic rate and the motor drive relay system must incorporate a suitable hangover circuit to keep the motor operation as constant as possible when the delay is far from the proper value, and still not cause an over-run when the proper adjustment is reached. Freedom from over-run caused by motor inertia is obtained by using a motor having a multiple-pole permanent magnet armature with a speed of 75 r.p.m., which under no load will brake itself in about twenty degrees of armature travel.

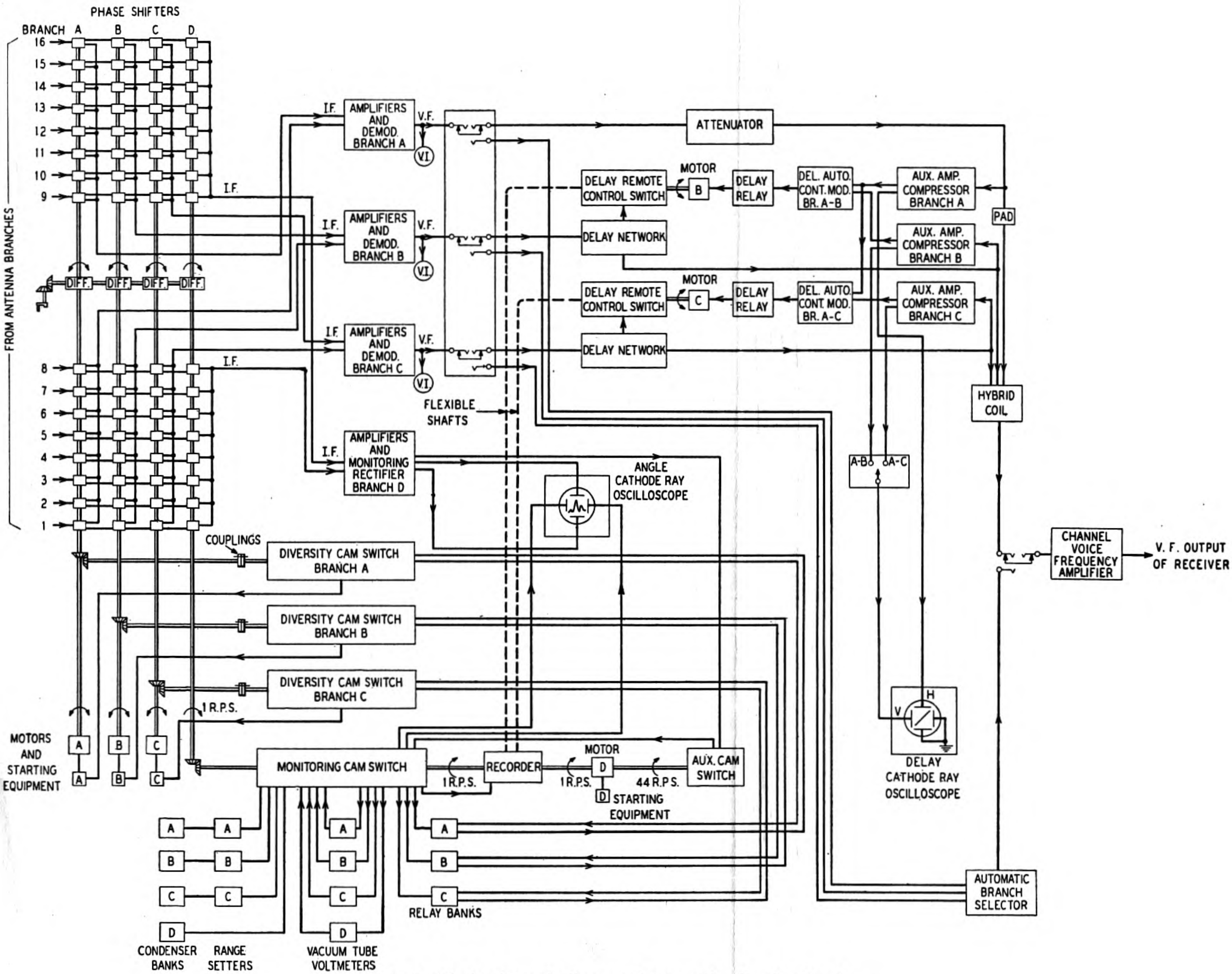


Fig. 8—Block schematic diagram of automatic adjusting circuits.

Automatic Branch Selector

As an alternative to the combination of the outputs of the diversity branches after delay equalization it is possible to use a system which discretely chooses one branch for connection to the line. Equipment to do this has been provided.

The common automatic volume control which is used with both systems operates on the carrier and if the carrier and sideband fade alike the output volume is held substantially constant. To the extent that the received noise on the various branches is the same, the branch having the highest output volume will be the most desirable to use. In the equipment used part of the audio output of each of the three diversity branches is rectified and applied differentially to three polar relays in such a manner that the relay corresponding to the branch having the highest amplitude is operated. This relay reduces the bias of an amplifier which connects between that branch and the output line from a high to a normal value, and thus permits voice frequency signals to flow through that branch to the output. The noticeable effects of switching are eliminated by several expedients. Push-pull amplifiers with feedback are used so as to balance out the low-frequency thump. The variation of biasing takes place with a time-constant of 0.01 seconds in order to aid in this matter as well as to render unnoticeable the instantaneous differences in the two channels. A biasing winding on each relay insures that once a contact is broken the relay moves to the opposite contact in a fixed time, which permits the selection of time-constants for the suppression and build-up which are as nearly complementary as possible, and so keeps the volume constant.

A difference in volume of 1 to 2 db is required to cause a switch. During the switching period, which lasts about 20 milliseconds, the output varies about $\pm$ 1 db. When no speech is being transmitted the relays remain inoperative and consequently the line may not be connected to the branch which at the next instant may deliver the highest volume. It might be expected that clipping of the succeeding initial syllable would be intolerable. To reduce selective fading to an unnoticeable amount it is only necessary to suppress the unwanted branches by 12 to 15 db. With the equipment adjusted to give this suppression it is found that there is always sufficient signal transmitted through one or more of the branches to practically eliminate noticeable clipping.

The use of the automatic branch selector has the disadvantage that the effect of having more than one diversity branch contribute to the output at any one time is lost. This is estimated to be equivalent to not more than an increase of one decibel of transmitted power. To offset this there must be considered the possibility of the delay being out of adjustment for brief periods when changes in angle require sudden changes in delay equalization. The necessity for close phase uniformity between the various carrier and sideband amplifiers over a wide range of automatic volume control voltage is also eliminated when the automatic branch selector is used. Further, it is no longer necessary to use the received carriers for demodulation of the various branch outputs and a locally generated carrier of uniform amplitude can be used with a resultant increased stability of output volume. The practicability of trying to phase branch outputs by delay equalization over a range of more than 3,000 cycles from the carrier has not been demonstrated and consequently the use of automatic branch selection with the channel whose sideband is spaced by one sideband width from the carrier is to be recommended. The use of the branch selector also permits simpler operation of the automatic angle adjusting equipment as will be explained later. For all of these reasons it is to be expected that the automatic branch selector may eventually be used to the exclusion of the delay compensating equipment.

Automatic Angle Adjusting Equipment

In the experimental musa receiver the rectified carrier output of the monitoring branch was connected to one set of deflection plates of a cathode ray oscilloscope and a sweep circuit mounted on the monitoring phase shifter shaft connected to the other set of deflection plates. The oscilloscope screen displayed a graph of the amplitude of signal received for each phase shifter setting. The pattern frequently changed rapidly from moment to moment so that only by constantly observing the screen was it possible to determine at what phase shifter settings the best signals were being received and to set the diversity phase shifters accordingly.

The attention necessary to operate satisfactorily the equipment in this manner was believed to be too great for commercial operation, particularly since it might vary widely from hour to hour. With a mind to the fact that improper adjustment might give poorer reception than would be obtained with ordinary receivers it was decided to make the settings of the phase shifters of the diversity branches automatic.

In Fig. 8 the motors A, B, C and D drive the phase shifters of the corresponding branches through the vertical shafts. Motor D operates continuously so as to vary the phase shifting system through its complete range once a second, while motors A, B and C operate only when a change in the angle of reception is required. Connected to the vertical drive shafts by the horizontal shafts are the three diversity cam switches and the monitoring cam switch.

The incoming carrier in the monitoring branch D is amplified in such a manner as to keep the average peak amplitude constant. It is then rectified, and applied to the vertical deflection plates of a cathode ray oscilloscope in the same manner as in the experimental equipment, the monitoring cam switch being provided with a set of contacts and resistances which act as a sweep circuit for the horizontal plates of the oscilloscope.

The rectified signal from the monitoring rectifier is also connected through the auxiliary cam switch, the monitoring cam switch, a high resistance, and the range setters, to three separate banks of condensers, each consisting of 44 four-microfarad condensers. The condensers in each bank are connected successively to the rectifier circuit once each second for a short period by the cam switch so that each is charged at a rate depending upon the amplitude of the received signal for a particular phase shifter position. A vacuum tube voltmeter is connected successively across each condenser of a bank. When one condenser becomes charged to one-half volt or more during the preceding second the vacuum tube voltmeter operates a relay. With the Branch A, B, and C voltmeters this results in a relay corresponding to that particular condenser being locked up and all the condensers in that particular bank being discharged. The operation of the second relay causes the motor of the corresponding diversity branch to start and turn in the right direction so that the branch phase shifters are adjusted with the least movement to the position corresponding to the relay, at which point a contact in the diversity cam switch trips the relay and stops the motion.

If no further control were provided all the diversity branch circuits would be set in the same position and consequently no diversity action would be obtained. To prevent this the range setters, A, B, and C, have been provided which are operated manually to limit the angular range of operation of each diversity branch. These switches merely short-circuit the condensers of a particular branch in the range which it is not desired to use. Since the short-circuited condensers do not acquire a charge the automatic adjusting equipment will never move the phase shifters to a position corresponding to a short-circuited condenser in that particular branch.

In setting the range switches when using delay equalization it is necessary to know what the condenser position is which corresponds to the highest angle which it is possible to receive at the particular frequency being used. The short circuit will then be removed from this condenser and from the condensers representing successively lower angles in Branch A until it seems probable from the recorder or cathode ray oscilloscope pattern that a good signal will be received in that branch. The remaining condensers in that branch are left short-circuited. The short-circuits are then removed in a like manner from part of the remaining range for Branch B and in the remainder of the range for Branch C, the best division line being determined from the cathode ray oscilloscope or recorder pattern. This procedure is necessary inasmuch as branch A is used for a reference in adjusting the audio frequency delay compensation and must always have a satisfactory signal if that equipment is to operate.

When using the automatic branch selector other settings of the range switches are possible. One arrangement is to allow one branch to stop on even numbered contacts for a part of the range and another branch to stop on the remainder of the even numbered contacts. The third branch may then stop on any odd numbered contact. This permits two diversity branches to be set in the range of maximum signal. A difference of one contact has been found sufficient to give satisfactory diversity action in most cases and the recorder pattern always shows that the signal is more than one contact wide. The third branch is free to follow a signal in another part of the range, which may grow to be the strongest at any moment.

In order to improve the accuracy of this equipment and reduce the maintenance of the monitoring cam switch an auxiliary high speed cam switch is used which operates 44 times faster than the main switch, closes just after each contact of the monitoring switch, and opens just before each contact opens. The charging time for all condensers in a given bank is thus determined by the same cam and set of contacts.

To prevent over-running on the diversity branches from mechanical inertia a special motor is used. This motor is similar to the one used for automatic delay adjusting equipment.

At times there may be only one angle at which a satisfactory signal may be arriving. It is possible to get diversity action at these times by setting the diversity branches on opposite sides of the average angle of best reception. Provision is made for doing this with the automatic equipment by allowing one bank of condensers and one voltmeter relay to control all three diversity branches and then mechanically off-setting the phase shifters of two branches by means of adjustable couplings in the diversity cam switch drives.

It will be seen that to obtain accurate operation of the automatic angle adjusting equipment it is necessary that the charging voltage should be large as compared with the final voltage on any condenser and that the final voltage must be the sum of a number of charges. The time necessary for a condenser to reach the final voltage can be varied from 8 to 45 seconds or longer and successive movements of a diversity branch phase shifter drive will not be oftener than this. Once the motor has started, however, it will move the phase shifter shaft through an angle of 180°, the maximum which would ever be necessary, in 6 seconds.

Recorder

In order properly to set the phase shifters manually or to set the range adjusters of the automatic angle adjusting equipment it is necessary to know the phase shifter positions corresponding to the angles at which signals are arriving. The angle monitoring cathode ray oscilloscope shows how the signal amplitude *vs.* phase shifter position varies from second to second. By using a retentive screen on the oscilloscope it is possible to see the traces for the previous few seconds at the same time as the most recent trace. The traces, however, normally vary appreciably in position of maximum amplitude and it is somewhat difficult to form an opinion from looking at the oscilloscope as to just where to set the diversity branches. By integrating the value of received signal over a number of seconds a better conception can be obtained.

In addition to the cathode ray oscilloscope it also seemed desirable to have a record available to the operator of the variation of signal intensity with phase shifter position as it changes from minute to minute so that he would not continuously have to observe the oscilloscope to determine whether the range adjusting switches were set properly. This required one more variable to be considered than the ordinary recorder is designed to register and it was consequently necessary to devise a new type of device.

The scheme of recording operates in a somewhat similar manner to the automatic angle adjusting equipment. A set of 44 condensers is charged by the incoming signal through the monitoring switch. Each condenser corresponds to a particular position of the phase shifters and consequently with a particular vertical angle of arrival at a particular frequency. A vacuum tube voltmeter is successively connected to the condensers until one is found which has acquired a predetermined potential in the order of two volts. A relay in the plate circuit of the voltmeter then operates, causing only that particular condenser to be discharged and making a record on a paper strip.

The recorder consists of a mechanism for driving a paper strip 5″ wide at a constant speed over a drum having a spiral wire on its periphery. Above the drum and paper are a typewriter ribbon and a thin bar which may be made to come down on the ribbon, paper tape, and spiral wire, by the action of an electromagnet. The action of this striker bar is to cause a dot to be made on the paper strip at the position where the striker bar and spiral wire intersect. The drum carrying the spiral wire revolves in synchronism with the phase shifters and there is consequently a lateral position on the paper corresponding to each one of the 44 condensers previously mentioned. When each condenser is discharged by the action of the vacuum tube voltmeter a dot is made in a particular lateral position on the paper strip and successive dots caused by the discharge of the same condenser fall in the same longitudinal line on the strip. The frequency of dots in a particular longitudinal line is, therefore, proportional to the relative field strength at a vertical angle corresponding to that line. As a result of the action of the automatic volume control on the monitoring branch amplifier, the maximum frequency of dots along a longitudinal line is kept approximately constant regardless of the absolute value of signal received so that the device does not record the variation of signal at a fixed angle from minute to minute.

A sample of a section of a record is shown on Fig. 9, together with a scale showing the angles corresponding to the rows of dots for a particular received frequency. The angle record is contained in the section above the "Phase Shifter Position" scale.

In order to have a check on, and a record of, the operation of the automatic angle adjusting equipment, provision has been made so that three longitudinal lines are drawn on the paper corresponding to the three angular positions of the diversity branches.

A record of the operation of the automatic delay adjusting device is also made by the recorder. This was done by inserting a mechanism which uses the margins of the paper on either side of the main record. The delay recording device consists of two drums mounted concentrically with the main recorder drive shaft which are similar in nature to the tens and units drums of an ordinary counter. Flexible shafts extend from the delay adjusting switches to the recorder where they drive the drums on each end of the shaft. The two drums on one end are connected together with an intermittent movement so that one revolution of the small units drum causes the large units drum to move forward one step.

With the paper tape normally running at only $\frac{1}{4}''$ per minute it is impractical to stamp numbers on the paper since the delay adjustment varies several times in a minute and thus would cause the numbers to record on top of one another. Recourse was accordingly taken to a mark in a definite lateral position to indicate the magnitude of the delay. The drums have segmental ridges on their periphery which are displaced in various lateral positions. Cam operated hammers descend on the typewriter ribbon above the drums and paper once each second leaving a mark in a lateral position corresponding to the

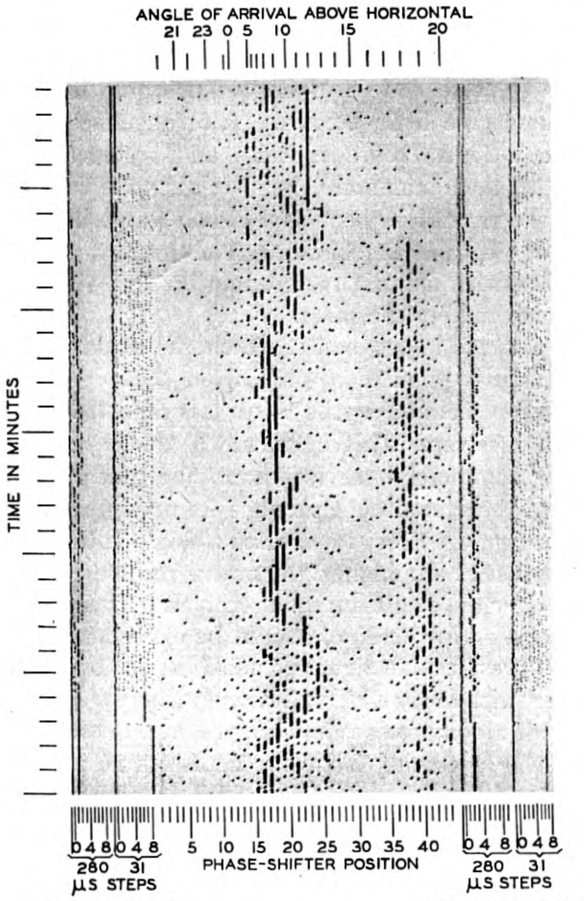


Fig. 9—Sample of musa recorder chart. Station GAW. Freq. 18,200 Kc.

segmental ridge which is beneath and consequently to the delay setting. Two reference lines are used to facilitate the reading of the delay values.

Operation and Performance

The musa system can be expected to give an improvement in signal-to-noise ratio and in selective fading over a receiver using only a single antenna. The improvement in signal-to-noise ratio caused by the use of 16 antennas should average 12 db, but instantaneous improvements might vary from large negative values, if the equipment were not kept properly adjusted, to values of 25 or 30 db, which might be expected when the noise came from the direction of a null in the musa antenna directive diagram.

In the operation of radio telephone circuits there is a minimum signal-to-noise ratio below which commercial service cannot be given. As the signal-to-noise ratio is increased a value is reached where further increases give little benefit. The range between these two values is about 25 db. Transmitters and receivers generally are designed so that their maximum signal-to-set-noise ratio is somewhat greater than the maximum beneficial circuit value in order that set-noise shall not degrade the circuit. The maximum signal-to-noise ratio obtainable with a musa receiver and a single-antenna receiver should be approximately the same.

The 12 db average improvement which the musa receivers should give should make it possible to obtain, on the average, commercial circuits with signal field strengths 12 db less than those usable with a single-antenna receiver. This in turn will decrease the amount of time in which commercial service cannot be given.[9] On the other hand the musa receiver should produce its maximum signal-to-noise ratio for field strengths 12 db lower than a single-antenna receiver and at field strengths 12 db higher the musa receiver would show no improvement over the single-antenna receiver. The net improvement in signal-to-noise ratio therefore should be expected to average from 12 db at the lowest usable signal-to-noise ratios to 0 db at fields 25 or 30 db higher, with fairly wide variations with time from the average.

The results of a comparison between the musa system and a single sideband receiver operating from one of the same antennas confirm the theoretical expectations to a fair degree. The fraction of the time that given improvements in decibels are obtained follows approximately a normal probability curve.

The reduction in selective fading effected by the musa receivers is difficult to state numerically. Most of the objectionable selective fading is removed. There are times when waves that have traveled over distinctly different paths arrive at so nearly the same angle that they cannot be resolved. Fortunately these times are fairly rare. When waves of closely adjacent angle are present the monitoring system does not give a true indication of reception angles, as can be

[9] R. K. Potter and A. C. Peterson, Jr., "Reliability of Short-Wave Radio Telephone Circuits," *Bell Sys. Tech. Jour.*, Vol. XV, pp. 181–196, April 1936.

shown by theory. It is a fairly common occurrence for a wave group which has apparently traveled over a single general path to have components which vary in transmission times by 100 or 200 microseconds from others of the same group. The fading caused by such small delay differences is not distinctly selective in effect and its chief detriment is in causing volume variations which must be overcome by the use of special devices.

During some severe magnetic storms successive traces on the angle monitoring cathode ray oscilloscope show little relation to each other. It has been reasoned that a reduction in resolution might be beneficial at such times but sufficient experience to prove this has not been obtained. A reduction in resolution can be obtained easily by switching off the amplifiers associated with one group of eight antennas. It has been found that fading on the front group of antennas is generally at random to that on the rear group so that the two groups can be used in space diversity if desired. No particular advantage has been found to this arrangement.

The use of delay compensation between the diversity branches does not seem to have any advantage over the use of the automatic branch selector. The output volume variations are slightly greater with delay compensation because of the use of reconditioned carrier for demodulation. On the other hand the use of the automatic branch selector promises materially to reduce maintenance by eliminating the necessity for keeping the phases of the various carrier and sideband amplifiers alike.

It has been amply demonstrated that the automatic adjusting features provided are essential to the efficient operation of the equipment.

Abstracts of Technical Articles by Bell System Authors

Thermionic Emission, Migration, and Evaporation of Barium on Tungsten.[1] J. A. Becker and G. E. Moore. When barium is deposited on tungsten, the thermionic activity of the tungsten increases, comes to a maximum, and then decreases. It has frequently been found by emission measurements that this optimum corresponds to about a monomolecular layer. However, data obtained in this work show that some regions of the filament require more than five times as much barium as others for optimum emission.

Photographs are presented which show that the rates of both migration and evaporation depend on the crystal surface, the temperature, and amount of barium on the surface. Barium migration on tungsten can be observed at temperatures as low as 970° K., is readily observed at 1025° K., and is rapid at 1070° K. Evaporation is observed on some crystals at temperatures as low as 1025° K., while on others it is slow even at 1260° K. At 1300° K. it is rapid for all crystals. These temperatures probably vary with the oxygen contamination which comes over to the filament with the barium. For barium concentrations near the optimum there exists a range of temperature over which migration is readily observed, but where evaporation is not noticeable.

Measurements of electron emission after all the barium is evaporated show that the filament was contaminated by an electronegative material, probably oxygen.

Barium tends to migrate toward the negative end of the filament, thus indicating ionization of adatoms.

A mechanism for migration is suggested.

The Vocoder—Electrical Re-creation of Speech.[2] Homer Dudley. In the Bell Telephone Laboratories have been developed electrical circuits for the artificial production of speech. One form of the device is itself voice-controlled, thus differing fundamentally from the Voder of the World's Fair which is controlled by keys and pedals. It has been christened the "Vocoder" or "voice coder."

Many startling effects are possible when the code is varied, for the Vocoder then re-creates sounds quite different from those used by the person speaking. Cadences may become monotones, rising inflections may be turned to falling inflections, a vigorous voice may become a

[1] *Philosophical Magazine*, February 1940.
[2] *Jour. S. M. P. E.*, March 1940.

quaver, or a single voice may accompany itself at any desired musical interval—thus converting a solo into a duet, etc. Also non-speech sounds may be coded into intelligible speech and instrumental music into vocal music.

Statistical Measurements on Conversational Speech.[3] H. K. DUNN and S. D. WHITE. Using apparatus designed to collect a large number of data in a short time, the following measurements have been made: peak and r.m.s. pressures in one-eighth-second intervals, and in various bands of frequencies up to 12,000 cycles per second, from the voices of six men and five women; comparison of r.m.s. pressures in one-eighth- and one-fourth-second intervals, from a single male voice; and distribution of the instantaneous pressures in whole speech, from a single voice. Derived from these data are peak factors in one-eighth-second intervals, and frequency distribution of speech energy in long intervals. Both the absolute value and the distribution of energy are found somewhat different from previously published results.

Auditory Patterns.[4] HARVEY FLETCHER. During the last two decades considerable progress has been made in understanding the hearing processes taking place when we sense a sound. The application of the same instrumentalities that have brought such a wonderful development in the radio and sound pictures to this problem is largely responsible for this progress. Such instrumentalities have made it possible to make accurate measurements which are the basis for understanding any physical process.

To understand this problem then we need to know first how to describe and measure the sound reaching the ears; then we need to know how to describe and measure the sensations of hearing produced by such a sound upon a listener. To do this quantitatively we must also know the degree and kind of hearing ability possessed by the listener. It is with these three phases of the problem that this paper deals.

[3] *Jour. Acous. Soc. of America*, January 1940.
[4] *Reviews of Modern Physics*, January 1940.

Contributors to this Issue

Karl K. Darrow, B.S., University of Chicago, 1911; University of Paris, 1911–12; University of Berlin, 1912; Ph.D., University of Chicago, 1917. Western Electric Company, 1917–25; Bell Telephone Laboratories, 1925–. Dr. Darrow has been engaged largely in writing on various fields of physics and the allied sciences.

W. P. Mason, B.S. in Electrical Engineering, University of Kansas, 1921; M.A., Columbia University, 1924; Ph.D., 1928. Bell Telephone Laboratories, 1921–. Dr. Mason has been engaged in investigations on carrier systems and in work on wave transmission networks both electrical and mechanical. He is now head of the department investigating piezoelectric crystals.

A. L. Matte, B.S. in Electrical Engineering, Massachusetts Institute of Technology, 1909; Graduate Studies, M.I.T., 1912–13. New England Investment and Securities Company, 1910–12; Detroit United Railways, 1913–18. American Telephone and Telegraph Company, Department of Development and Research, 1918–34; Bell Telephone Laboratories, 1934–. Mr. Matte has been engaged principally in transmission studies relating to carrier telegraphy.

G. S. Phipps, B.S. in Electrochemical Engineering, Pennsylvania State College, 1930; M.S. in Metallurgy, Columbia University, 1939. Bell Telephone Laboratories, 1930–. Mr. Phipps has been engaged principally in the metallurgical investigation of solders and lead base alloys.

F. A. Polkinghorn, B.S., University of California, 1922; U. S. Naval Radio Laboratory at Mare Island Navy Yard, California, 1922–24; A-P Radio Laboratories, San Francisco, 1924–25. Pacific Telephone and Telegraph Company, San Francisco, 1925–27; Bell Telephone Laboratories, 1927–. Mr. Polkinghorn's work has been primarily in connection with the design of radio receiving and test equipment for use at high and ultra-high frequencies.

Earle E. Schumacher, B.S., University of Michigan; Research Assistant in Chemistry, 1916–18. Engineering Department, Western Electric Company, 1918–25; Bell Telephone Laboratories, 1925–. As

Associate Research Metallurgist, Mr. Schumacher is in charge of a group whose work relates largely to research studies on metals and alloys.

A. M. SKELLETT, A.B., 1924, M.S., 1927, Washington University; Ph.D., Princeton University, 1933; Instructor, 1927–28, Assistant Professor of Physics, 1928–29, University of Florida. Bell Telephone Laboratories, 1929–. Dr. Skellett, formerly engaged in investigations pertaining to the transatlantic radio telephone, is concerned with applications of electronic and ionic phenomena.

R. A. SYKES, Massachusetts Institute of Technology, B.S. 1929; M.S. 1930. Columbia University, 1931–33. Bell Telephone Laboratories, Research Department, 1930–. Mr. Sykes has been engaged in the application of piezoelectric crystals to selective networks, and more recently in the use of coaxial lines as filter elements.

EMIL FRIDSTEIN VAAGE, E.E., Technical University of Darmstadt, Germany, 1921–26; M.S., Brooklyn Polytechnic Institute, 1932. Elektrisk Bureau, Oslo, Norway, 1926–27. American Telephone and Telegraph Company, 1927–34; Bell Telephone Laboratories, 1934–. Mr. Vaage has been engaged in inductive coordination studies, 1927–39. Since 1939 his work has been concerned with open-wire television transmission.

L. G. WADE, B.S.E.E., University of Idaho, 1918; M.S., Cornell University, 1924. Engineering Department, Western Electric Company, 1924. Mr. Wade has been engaged in developing equipment and processes for the manufacture of lead covered telephone cable, particularly on drying and dry storage of cable cores preceding the lead sheathing operation.