The Bell System Technical Journal

October, 1924

"The Stethophone," An Electrical Stethoscope

By H. A. FREDERICK and H. F. DODGE

1. Acoustic Stethoscopes

AUSCULTATION is commonly practiced by means of the ordinary stethoscope, a device with which the physician is able to study sounds produced within the heart, lungs, or other portions of the body and to determine whether such abnormal conditions exist as are evidenced by abnormal sounds. Of particular importance are the characteristics of the normal heart sounds, heart murmurs, breathing sounds and râles.[1] It is well known that the intensity of certain of these sounds is not in itself of fundamental significance, that, for example, certain very faint murmurs may represent serious organic lesions; hence it is of pathological importance that these sounds be heard and understood.

Most acoustic and mechanical vibratory systems introduce distortion by discriminating in favor of certain frequency bands. Extreme distortion may alter a sound beyond recognition. If a moderate amount of distortion is unavoidable, it may be possible to control it judiciously so as to give most accurate reproduction in the frequency region of major importance.

From this standpoint it is of interest to consider the frequency characteristics of the two common types of stethoscopes shown in Figs. 1 and 2. The stethoscopes used in these tests were equipped with thick-walled soft rubber tubing such that the distance from the chest piece to the ear pieces was approximately 55 cm. The characteristic of the open bell stethoscope was obtained by picking up sound from the surface of a piece of fresh beef and measuring the relative intensity of sound on a condenser transmitter[2] with and

[1] The presence of any one of several types of lesions in or near the valves of the heart "gives rise to eddies in the blood current and thereby to the abnormal sounds to which we give the name murmurs." "No one of the various blowing, whistling, rolling, rumbling or piping noises to which the term refers, sounds anything like a 'murmur' in the ordinary sense of the word." (R. C. Cabot—Physical Diagnosis, pp. 182-3, 1923.)
"The term 'râles' is applied to sounds produced by the passage of air through bronchi (windpipes) which contain mucus or pus, or which are narrowed by swelling of their walls." (R. C. Cabot—Physical Diagnosis, p. 163) Râles may appear either as bubbling sounds, occurring singly or in showers, or as musical squeaks and groans.

[2] E. C. Wente, "The Sensitivity and Precision of the Electrostatic Transmitter for Measuring Sound Intensities," *Phys. Rev.* 19, No. 5, p. 498, 1922.

without the test stethoscope inserted in the sound path. In this experiment, it was impracticable to set up pure vibrations in the human body. A piece of fresh beef was a convenient substitute and one which for the purposes of such physical analysis appeared satisfactory.

The frequency characteristic of the open bell stethoscope is shown

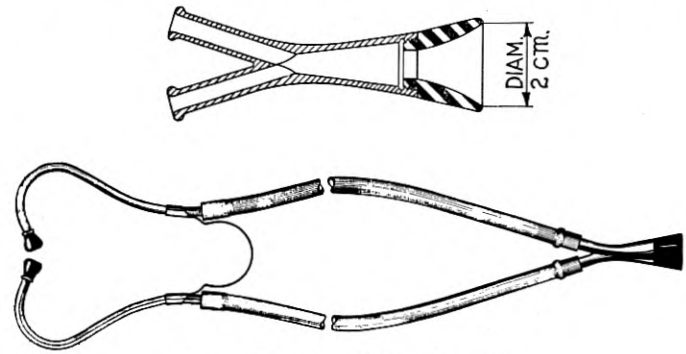


Fig. 1—The open bell stethoscope

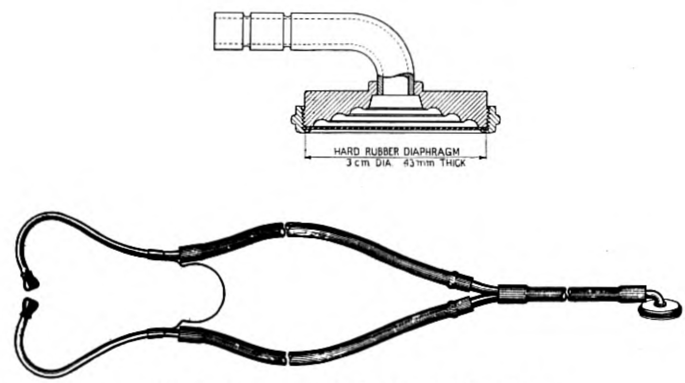


Fig. 2—One type of Bowles stethoscope

in Fig. 3, in which the "sensation value" as interpreted by the ear is plotted in transmission units,[3] TU,—convenient units used to

[3] The transmission unit used in this paper is a logarithmic function of power ratio. The number of transmission units N corresponding to the ratio of two amounts of power P_1 and P_2 is given by the relation $N=10 \log_{10}\frac{P_1}{P_2}$. The power ratio corresponding to N units is therefore $10^{\frac{N}{10}}$. For example, an increase of 10 TU signifies 10 times as much power; of 20 TU, 100 times as much power, etc. See W. H. Martin, "The Transmission Unit," *Journal A. I. E. E.*, Vol. 43, p. 504, 1924; *B. S. T. J.* Vol. 3, p. 400, 1924.

express relative loudness. A power ratio scale is also shown at the left and the power at 100 cycles is assumed equal to unity as a reference point.

This curve shows the relative efficiency of transmission for frequencies up to 2,000 cycles. The successive peaks are due primarily to resonance of the air columns and are partly determined by the length of the stethoscope tubing. Resonance thus increases the

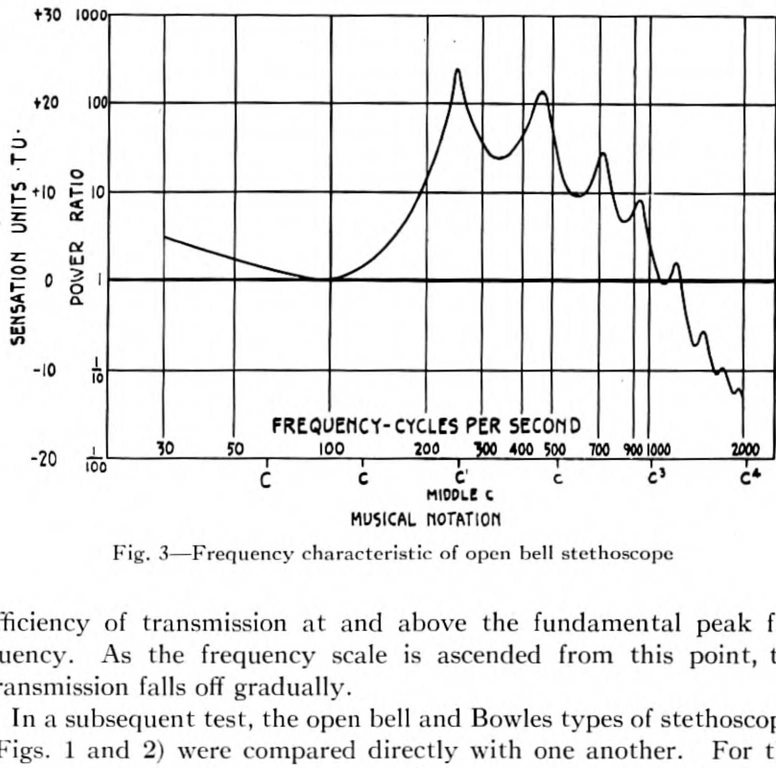


Fig. 3—Frequency characteristic of open bell stethoscope

efficiency of transmission at and above the fundamental peak frequency. As the frequency scale is ascended from this point, the transmission falls off gradually.

In a subsequent test, the open bell and Bowles types of stethoscopes (Figs. 1 and 2) were compared directly with one another. For this test, a vigorous sound was imparted to the sternum of a patient and the sound was picked up over the apex of the heart. Below 150 cycles, the Bowles stethoscope averaged approximately 15 *TU* less efficient, whereas, disregarding the somewhat different arrangement of the resonance peaks, between 300 and 1,000 cycles, it varied from 5 to 10 *TU* more efficient than the open bell type. These features of the Bowles stethoscope are due to the chest piece diaphragm. As will be shown in another paper, much of the energy of systolic and diastolic murmurs is made up of frequencies between 120 and 660 cycles per second. Thus, concurring with observations made by Dr. R. C. Cabot,[4] it is to be expected that many of the moderately high and high pitched murmurs can be heard more distinctly with the Bowles than with the open bell stethoscope. On the other hand, for many faint pathological sounds such as presystolic murmurs which are composed primarily of relatively low frequencies, the open bell stethoscope is more satisfactory for observation. The latter introduces less distortion so that with it all sounds are retained more nearly at their original relative intensities. These remarks are, of course, confined to the particular designs of stethoscopes shown in Figs. 1 and 2. It should be noted that as the length of the rubber tubing is increased, the fundamental peak of Fig. 3 moves downward in frequency, and the transmission at higher frequencies becomes poorer. In order to retain the very high pitched components of certain heart and chest sounds, the use of long rubber tubing should, therefore, be avoided.

The common stethoscope serves as a convenient means of observing body sounds. If the available energy from a single chest piece is subdivided in order to supply several individuals, however, the sounds observed by each are much fainter. In cases where the sounds of pathological interest are sufficiently near the threshold of audibility the use of a multiplicity of observing tubes renders these sounds inaudible. This is often the case.

For teaching purposes or for consultation, it is extremely desirable to have multiple listening units. In the past, it has been necessary to handle the students of large classes either singly or in small groups. This method naturally limits the number of cases that can be demonstrated and makes it impossible to give each student as much practice as has been found necessary for him to become familiar with the more obscure sounds. Aside from these factors, it has not been feasible for a large group to observe simultaneously with the instructor the peculiarities and changes in murmurs of a transient or evanescent character.

With the development of vacuum tube amplifiers, the possibilities of reproducing and magnifying body sounds electrically were considered. It appeared that a device might be provided which would be useful not only in teaching but also in diagnosis, as an aid to physicians of subnormal hearing, in the reproduction of the very faint fetal heart sounds or even in fields beyond the scope of the ordinary stethoscope.

[4] R. C. Cabot, "Physicial Diagnosis," Chap. VI, 1923.

2. Early Development of the Electrical Stethoscope

The earliest development work on electrical stethoscopes was naturally centered about the carbon transmitter and other microphonic contact devices. In 1907, Einthoven [5] made records of normal heart sounds and murmurs. In 1910,[6] heart sounds were reproduced by a tuned mechanical relay consisting of a single microphonic contact and an electromagnetic element. With this device, heart sounds were transmitted audibly but evidently with a considerable amount of distortion, over a commercial telephone line in London. The normal heart sounds were amplified by Squier [7] for a group of physicians by means of a carbon transmitter in 1921. It is readily possible to amplify the fluctuations in current in a carbon microphone by means of vacuum tube amplifiers. However, the carbon microphone also introduces a certain amount of noise inherent in the use of loose contacts. This noise is below the threshold of audibility for the normal use of the microphone, as in the telephone plant, but when it is amplified along with the faint sounds of interest, in auscultation it becomes very annoying and tends to obscure these other sounds. This "microphone roar" contains components throughout the range of audible frequencies and hence cannot be eliminated. Various experimenters have, however, attempted to perfect such a device.[8,9] As far as we have been able to determine, such devices have not satisfactorily reproduced faint heart murmurs or chest sounds.

Of the other possible types, the electromagnetic has thus far appeared to offer the greatest promise. In design, this resembles closely the ordinary telephone receiver. This type requires a more powerful amplifier than the carbon microphone but this is not a serious limitation. Such a combination has been used with promising results to obtain graphical records of heart murmurs.[10] The progress made with this type of equipment for teaching purposes has been outlined.[11] The successful application of the electromagnetic

[5] W. Einthoven, "Die Registrierung der menschlichen Herztöne mittels des Saitengalvanometers," *Arch. f.d. ges. Physiol.*, 117:461 April 1907; "Ein dritler Herzton," ibid. 120:31 Oct. 1907

[6] S. G. Brown, "A Telephone Relay," *Journal I. E. E.* May 5, 1910.

[7] S. W. Winters, "Diagnosis by Wireless," *Scient. Amer.* 124:465 June, 1921.

[8] R. B. Abbott, "Eliminating Interfering Sounds in a Telephone Transmitter Stethoscope," *Phys. Rev.* 21:200 Feb., 1923.

[9] Jacobsohn, "Amplified Audibility of Heart Sounds," *Berlin Letter J. A. M. A.*, 80:493 Feb. 17, 1923.

[10] H. B. Williams, "New Method for Graphic Study of Heart Murmurs," *Proc. Soc. Exper. Biol. and Med.*, 18:179 March 16, 1921.

[11] R. C. Cabot, "A Multiple Electrical Stethoscope for Teaching Purposes," *J. A. M. A.*, 81:298 July 28, 1923.

transmitter to teaching was due largely to the work of Dr. R. C. Cabot and Dr. C. J. Gamble at the Massachusetts General Hospital where a successful multiple electrical stethoscope was first employed for classroom lectures in June, 1923. The equipment consisted of an electromagnetic transmitter provided with a special form of mouthpiece for picking up the body sounds, a three-stage vacuum tube amplifier and a distribution system to accommodate as many as 125 students with single head receivers on which individual ordinary stethoscopes were held.[12]

The experience gained with this equipment indicated certain improvements to increase the sensitivity to body sounds, and at the same time decrease the disturbances caused by extraneous noises. Greater sensitivity required a better transference of sound energy from the body to the transmitter. Reduced room noise required that we couple the transmitter as closely as possible with the human body and at the same time make it insensititive to sound vibrations in the air. A preliminary analysis with electrical filters of the frequency characteristics of sounds of pathological interest to the physician showed that these sounds were composed largely of frequencies below 1,000 cycles. Inasmuch as the frequency characteristics of these various sounds are different, it has been found very useful to permit concentration on the sounds of interest by the use of electrical filters.

These factors led to the development of the electrical stethoscope called the "stethophone" which is described in the following paragraphs. This development was undertaken at the request and with the active cooperation of Dr. H. B. Williams of the College of Physicians and Surgeons, New York, Dr. Richard C. Cabot [11] of the Massachusetts General Hospital, Boston, and Dr. C. J. Gamble [12] of the School of Medicine of the University of Pennsylvania, Philadelphia. The cooperation of these physicians permitted the instrument to be given practical tests at every stage of its development.

3. General Description of the Stethophone

The stethophone consists essentially of the following elements:

1. An electromagnetic transmitter.
2. A three-stage amplifier with a potentiometer control.
3. A selected group of electric filters.
4. A multiplicity of output receivers for observers.

The whole is assembled in a substantial cabinet on wheels re-

[12] A detailed description of the apparatus used in this installation was presented in a recent paper. See Gamble and Replogle, "A Multiple Electric Stethoscope for Teaching," *J. A. M. A.*, Vol. 82, p. 387, 1924.

sembling a "tea-wagon." It requires for its operation a six-volt storage battery and a 130-volt "B" battery. These are housed in compartments in the lower part of the cabinet. Ten jack positions are provided to permit this number of persons to listen simultaneously

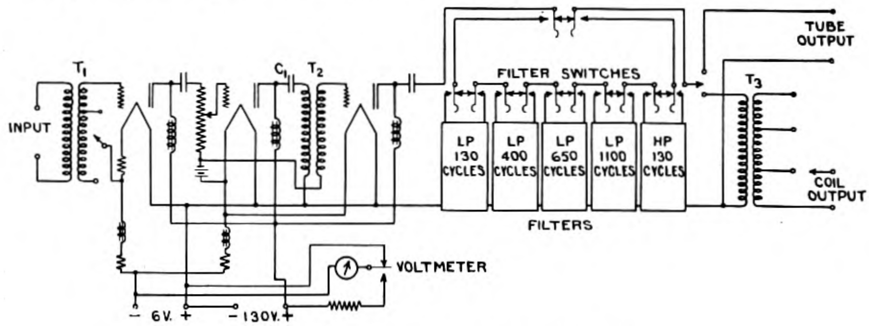


Fig. 4—Circuit diagram of stethophone

around the stethophone. All controls are conveniently placed on a single panel to facilitate operation.

A schematic circuit diagram is shown in Fig. 4.

4. Transmitter

The transmitter employed with greatest success with the stethophone thus far is of the electro-magnetic type equipped with a special vibratory element which is placed in direct contact with the flesh of the patient.

One of the features of the transmitter is its insensitiveness to sound waves in the air. Thus, the ratio of extraneous noise picked up by the transmitter to the body sounds is greatly reduced so that observations can be made with a minimum amount of interference from room noise.

The transmitter construction provides efficient transfer of vibrational energy from the flesh or bony framework of the body to the vibratory steel element. It provides a means for coupling which serves as a mechanical transformer for body sound energy and avoids an abrupt change in the path of the waves and large attendant losses by reflection. The system is highly damped and minimizes the distortion of the sounds of interest.

Since the transmitter is a contact device, the physician may vary the pressure of application at will. Firm but light contact is desirable. The human flesh contributes damping to the vibratory system of the transmitter. Undoubtedly this damping is not only variable for different individuals but depends upon the pressure and the nature of the flesh and bone structure in the vicinity of the point of application for any one individual. Thus the frequency characteristic of the transmitter is somewhat dependent on the conditions of use. The frequency of maximum response is slightly above 200 cycles, and the nature of the response-frequency curve indicates that the vibratory system is highly damped. A discussion of the overall frequency characteristic of the stethophone, including the transmitter, is given in a later section of the paper.

It is obvious that variations in the pressure of application will introduce disturbing noises in the audible frequency range. Suitable means have, therefore, been provided to eliminate the communication to the vibratory system of hand tremors, slight movements of the patient, and friction noises of the fingers on the case of the transmitter.

Another source of extraneous noise is the rubbing of the transmitter cord on the clothing or on other surfaces. A stiff cord is very objectionable from the standpoint of transmission of friction noises. Insulation from these noises has been provided by a very flexible section of cord at the transmitter end.

5. Amplifier

The three-stage amplifier employs one Western Electric 102-D and two Western Electric 101-D vacuum tubes. As shown in Fig. 4, the input transformer *T*1 connects the transmitter to the grid of the first tube which is coupled to the second tube through a resistance potentiometer. The second and third tubes are coupled through a transformer *T*2. The output circuit of the last tube may be connected to the load directly from its plate circuit for high impedance loads, or through an output transformer *T*3 for low impedance loads. The plate circuit of the second tube is tuned by means of a condenser *C*1 in order to retain high amplification at the low end of the frequency scale.

A very flat characteristic is obtained over the range of interest, the maximum variation being only about 3 *TU* (See Fig. 5). A total gain of about 80 *TU* is provided, that is, a power amplification of about one hundred million times. With an amplification of 50 *TU*, about the same loudness is observed in a single receiver in the output circuit of the stethophone as is heard by the direct use of the open bell stethoscope. This leaves a reserve amplification of about 30 *TU* available for obtaining greater intensity of sounds or for supplying a large number of individual listening units.

The potentiometer between the first and second tubes makes it possible to adjust the amplification in small steps, each step giving

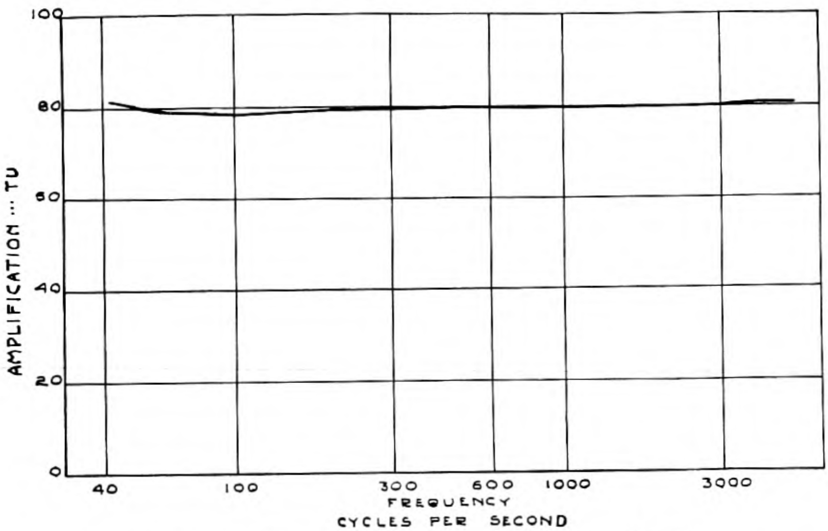


Fig. 5.—Amplifier characteristics—maximum amplification

approximately twice the energy of the preceding one. This is an essential element of a flexible system.

6. Electric Filters

An electric filter is a combination of coils and condensers capable of separating electrical waves characterized by a difference in frequency.[13]

The three fundamental forms of filters are commonly termed "low-pass," "high-pass," and "band-pass." A low-pass filter is one which passes currents of frequencies below a particular "cut-off frequency" and attenuates or weakens very greatly currents of higher frequencies. A high-pass filter does the opposite—attenuates below the cut-off frequency and passes above this frequency. A band-pass filter is one which passes currents of frequencies within a definite band fixed by two cut-off frequencies. A low-pass and a high-pass filter connected in series constitute one form of band-pass filter. For any type of filter, the sharpness of cut-off and the amount of attenuation can be controlled at will by suitable design constants.

[13] G. A. Campbell, "Physical Theory of Electrical Wave Filters," *Bell System Tech. Journal*, Nov., 1922.

The stethophone is equipped with five filters whose cut-off frequencies are based on careful analyses of about 100 hospital cases of heart murmurs, râles and breathing sounds. These analyses showed that the sounds of pathological interest to the physician can be grouped into fairly definite frequency regions. When sounds in a particular range of frequencies are of immediate importance, they may be emphasized by suppressing sounds outside of this band.

The frequency characteristics of the filters are shown in Fig. 6.

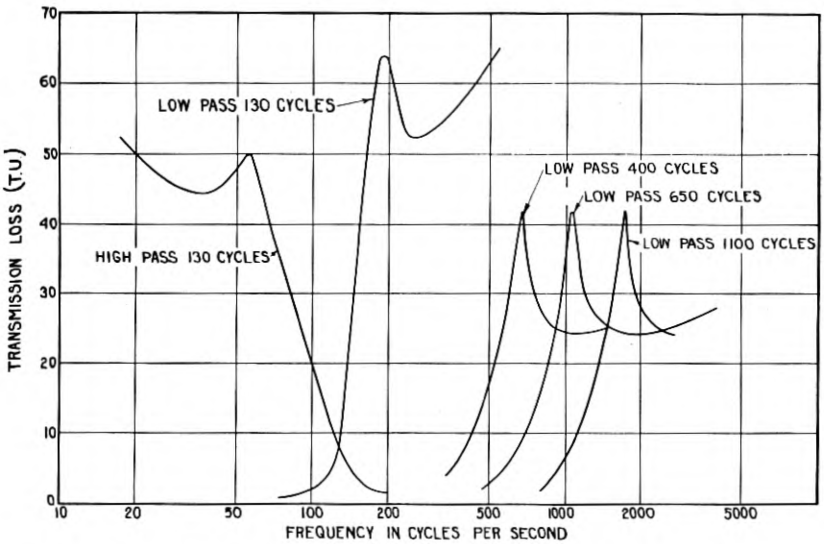


Fig. 6.—Loss characteristics of the five filters

For convenience, the cut-off frequency has been defined as that frequency at which the energy is reduced to approximately 1/10 of its original value.

The low-pass filter with a cut-off frequency of 130 cycles is of primary use for reproducing the normal heart sounds and fetal heart sounds in cases where the rate alone is desired. Most of the energy of these sounds is below 100 cycles. With this filter most of the common interfering noises, including the sounds of the human voice, are excluded.

The low-pass 400 cycle filter is particularly useful for observing presystolic and certain low-pitched systolic and diastolic murmurs.

The low-pass 650 cycle filter has been found the most valuable of all five filters. With it, most high-pitched murmurs, low-pitched râles and certain types of breathing sounds can be observed to the greatest advantage.

The low-pass 1,100 cycle filter passes the higher frequency components of very high-pitched murmurs and high-pitched râles in a majority of cases.

The high-pass 130 cycle filter serves a unique and important purpose. It may be regarded as in value second only to the low-pass 650 cycle filter. In many cases, the loud normal sounds tend to mask or obscure the faint higher-pitched murmurs. The high-pass 130 cycle filter serves to weaken greatly the normal heart sounds so that the murmur sounds occurring in the intervals between the beats appear with its use to be relatively much louder. In this filter, the amount of attenuation in the low frequency region has been made such that the residual low frequency energy and the higher frequency components of the normal heart sounds are just sufficiently audible so that the murmurs may be timed with relation to their positions in the cardiac cycle. This filter is also very useful for weakening the heart sounds when râles or pericardial friction sounds are to be observed in areas where the heart sounds are loud.

The high-pass 130 cycle filter may be connected into the circuit jointly with any one of the low-pass filters, thus making available a group of band-pass filters with a lower cut-off frequency of 130 cycles.

7. Output Receivers

When the stethophone is used for teaching or consultation purposes, a number of high impedance receivers are connected in parallel in the output circuit. Each observer is provided with a single receiver to which the ordinary stethoscope earpieces may be readily

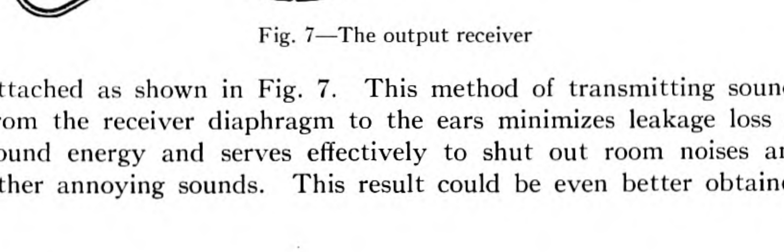


Fig. 7—The output receiver

attached as shown in Fig. 7. This method of transmitting sounds from the receiver diaphragm to the ears minimizes leakage loss of sound energy and serves effectively to shut out room noises and other annoying sounds. This result could be even better obtained and with less distortion by providing the receivers with small tips to insert in the ears but at a greater cost for additional receivers. It is perhaps better to use the tubing and earpieces of the ordinary stethoscope as this is the equipment to which physicians are most accustomed and to which the student must accustom himself for future practice. The receiver case is provided with a spring clip for attachment to the clothing. This allows full freedom of both hands for manipulating the transmitter and the control switches of the amplifier, taking notes, etc.

The impedance of the output circuit depends upon the number of receivers in use and, for parallel connection, decreases as the number of receivers is increased. To care for the variable number that may be used at different times, the output transformer has been tapped and a three-way switch provided. By operating this switch, the apparatus can be adjusted to a load varying from 1 to 600 receivers with a maximum transmission loss of 2.5 *TU*.

8. Frequency Characteristics of the Stethophone

The overall frequency characteristics of the stethophone, including the transmitter, the amplifier, and the output receivers, are given in Fig. 8. Two curves are shown. The solid line curve represents

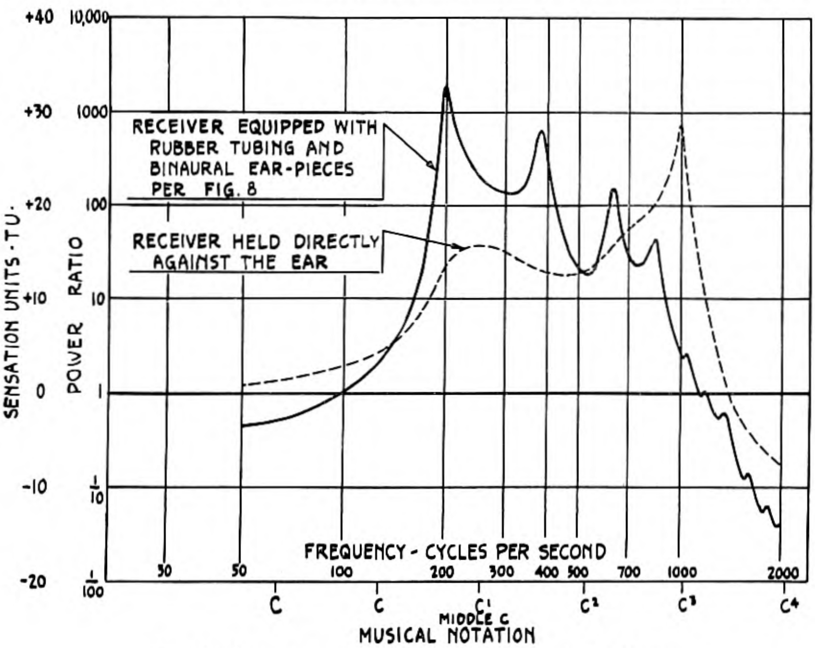


Fig. 8—Frequency characteristics of the stethophone

operating conditions when the output receivers are equipped with rubber tubing as in Fig. 7, and with the binaural ear-pieces held in the ears. The peaks in this curve are due principally to the resonance in the air columns of the rubber tubing, and correspond to the similar peaks of Fig. 3 for the open bell stethoscope. In order to point out the effect of the stethoscope attachment of the output receiver, a second characteristic is shown by a dotted curve which represents conditions when an output receiver of the same type is held directly against the ear. It is noted that the stethoscope attachment increases the transmission between 150 and 500 cycles per second, and damps the sharp resonant peak of the receiver.

The overall characteristic of the stethophone as employed for auscultation is quite similar to that of the open bell stethoscope. It is desirable that the body sounds as observed by the stethophone should appear the same as in the ordinary stethoscope, particularly in teaching work since the latter is used almost universally in regular practice. If it were deemed desirable for special purposes to avoid the distortion introduced by the stethoscope attachment, receivers with small tips to insert in the ears could be used. For such an arrangement, the overall characteristics could be further improved by using damped receivers which would practically eliminate the sharp peak of the dotted curve in Fig. 8.

9. Installation for Teaching Purposes

When the stethophone is to be used for teaching purposes a permanent wiring or distribution system should be installed with outlets distributed among the seats of the amphitheatre or lecture room.[11] A schematic diagram of such a system is shown in Fig. 9. A distributing pair of feeder wires, preferably shielded, is run between alternate rows of seats below the floor casing, or suitably sheathed to prevent damage. An outlet block "A" of six double contact jacks is mounted on the back of each third seat of alternate rows. Thus, one outlet block will supply six seats, three in front and three in back of the block. Substantial jacks should be used throughout and all receivers should be equipped with rugged plugs. In addition to furnishing jack outlets among the seats, two or three multiple outlet blocks may be installed at the center of the amphitheatre as shown at "B" for the use of guests or others on the floor of the amphitheatre. The output of the stethophone can be connected to the distributing system of the amphitheatre at any one of these outlets. Switch boxes should be installed at various points as at "C" to facilitate the localization of an accidental short circuit. If a short circuit should occur in any part of the system this section can thus be disconnected and the balance used independently.

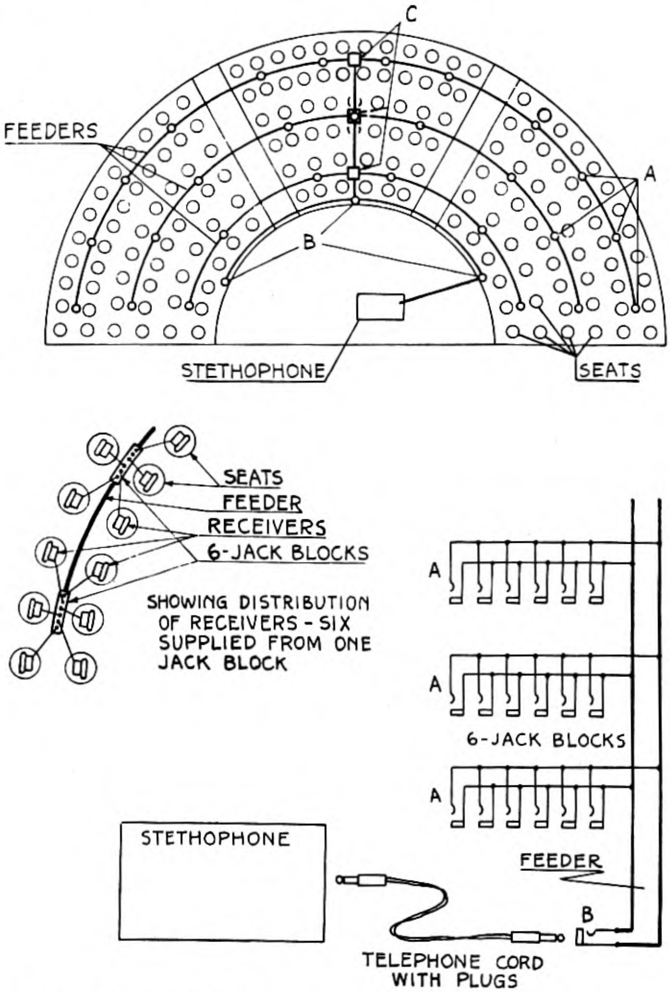


Fig. 9—Wiring installation in an amphitheatre for teaching purposes

In class room lectures, the instructor can make announcements to the students and point out features of particular interest in a convenient and somewhat novel manner without requiring the removal of the stethoscope tubes from their ears. The human body acts as a sounding board for sounds in air—that is, when words are spoken in the vicinity of a patient, the flesh and bone structure vibrates to these sounds. This is particularly true of the areas commonly used in auscultation. The transmitter, resting on the flesh, will pick up these vibrations together with those originating in the body of the patient. The instructor may, therefore, talk to his students by directing his words at that portion of the body to which the transmitter is applied. Best results are obtained with a talking distance of about ten inches. During such announcements, it is essential,

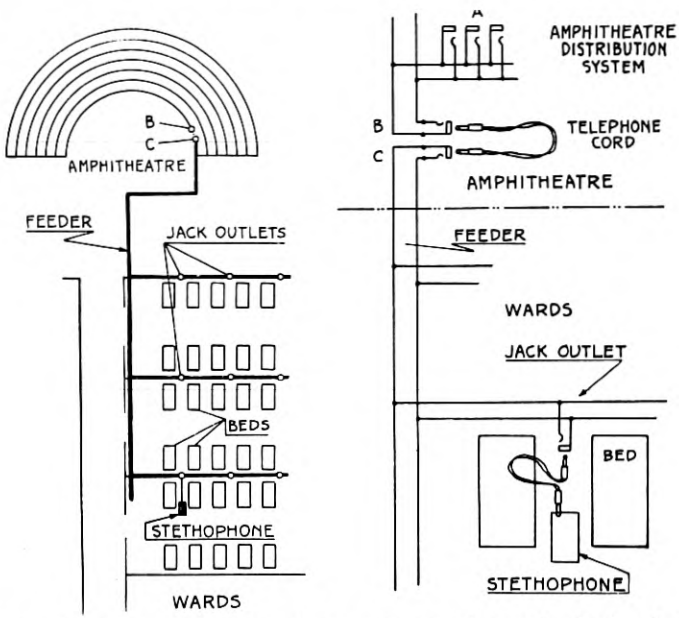


Fig. 10—Wiring installation for the rooms or the wards of a hospital

of course, that the electrical filters be removed from the circuit in order that the important higher frequency components of speech may be transmitted to the receivers. Because of this operating feature, it is obviously necessary to have the patient in a reasonably quiet place.

It is often desirable to reproduce in the lecture hall, the heart and chest sounds of confined patients too ill to be moved. For this purpose, the rooms or the wards of a hospital may be connected by a pair of wires to the lecture room. Such an installation is shown in Fig. 10. Terminal outlets are distributed throughout the rooms or the wards as desired and all are connected to the main feeder wires which communicate with the lecture hall. It is necessary to take the stethophone to the bedside. Long wires from the transmitter to the amplifier cannot be tolerated on account of inductive disturbances from neighboring telephone or other electrical circuits. If desirable, announcements may be made as before by talking close to the body of the patient under observation. In cases where exposure of a patient is inadvisable or where accurate statements pertaining to the seriousness of a disease are preferably withheld from the patient, announcements may be made by talking in a low tone of voice at about one inch distance from the transmitter itself. Reasonably satisfactory reproduction is obtained by this means.

10. Other Applications

Aside from its application to teaching purposes, the stethophone appears to have possibilities in fields which have not yet been thoroughly studied. Further experimental investigation by the medical profession can alone bring out these possibilities.

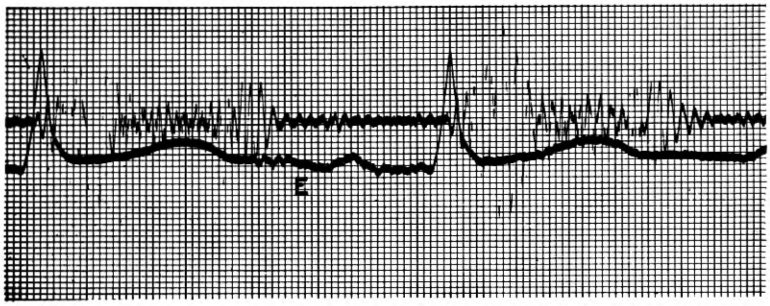
The possibility of substituting a loud speaker for the individual receivers in the output circuit has been investigated in a preliminary manner. This problem involves certain very fundamental factors relating to the sense of hearing which must be considered carefully. To a remarkable extent, the ear is capable of selective observation. Ordinarily we listen to sounds through a sea of noise to which we become so accustomed that we fail to notice it. However, when listening to sounds near the threshold of audibility, such as the body sounds under consideration, this noise may render the sounds of interest inaudible. In order to hear them, it therefore, becomes necessary to increase the loudness to a point well above that commonly observed by the physician with his stethoscope. This increase in loudness brings within the audible range, sound components ordinarily not heard and changes the quality of the whole as judged by the ear. Such alteration of quality is obviously very unsatisfactory for diagnosis or teaching purposes. Assuming that we had available a perfect loud speaker, one that would transmit the very low and the higher frequency components of faint body sounds without distortion, difficulties would still be presented by the acoustic characteristics of the room in which the loud speaker was placed. All rooms are more or less reverberant. When these sounds are reproduced by a good loud speaker in a small heavily damped sound-proof room, they appear quite natural, but such a room is seldom available practically. None of the ordinary loud speakers with horns will transmit the low frequencies here of interest and would sound very unnatural even with ideal room conditions. While a loud speaker has been used under proper acoustic conditions to reproduce faint pathological sounds, as murmurs and râles, this does not appear in general to be a practical arrangement. Most arguments, except perhaps that of economy, tend to favor the use of individual output receivers for practically all purposes where critical analysis of sounds is the objective.

Fetal heart sounds as heard through the mother's abdomen are much fainter and require considerably higher amplification than adult heart sounds. Preliminary data indicate that the energy of fetal heart sounds is approximately only 1/50 to 1/500 of the energy of average normal heart sounds. The low pass 130 cycle filter is not only useful for suppressing the extraneous sounds and electrical disturbances which usually attend the use of high amplification, but serves most effectively to eliminate the voice sounds of the patient. At Sloane Hospital in New York City it has been found possible to reproduce clearly on a loud speaker and with a negligible amount of interference, fetal heart sounds which were barely audible in the physician's stethoscope. In these cases no interference was experienced from the maternal heart sounds. However, even in surgical work where the rate of the adult or fetal heart is alone of importance, it is felt that the best plan is to equip an attendant with earpieces attached to a receiver and to make it his chief duty to observe the heart action.

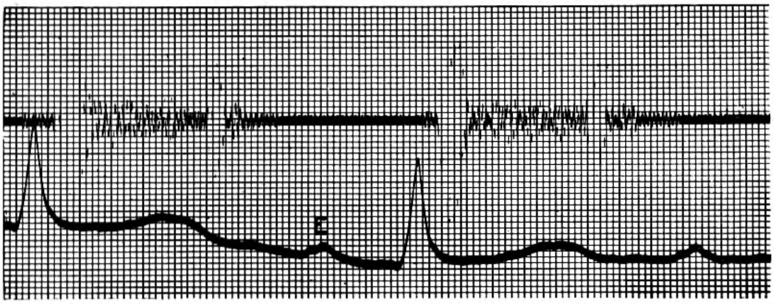
A very important application of the electrical stethoscope is its association with a recording galvanometer for making photographic records of heart and chest sounds. Permanent records of this sort might constitute a valuable addition to the history records of important cases in large hospitals. Some excellent graphical records have already been made.[10] It has been found that such records can be obtained much more easily with the stethophone, principally because of the part played by the electrical filters. The low-pass filter suppresses largely the current fluctuations caused by mechanical vibrations and noise at the apparatus. The high-pass 130-cycle filter is also valuable for bringing out very faint murmurs. By its use, the amplitude of the normal heart sounds can be greatly reduced. When this is done, the amplitude of the faint murmur sounds may be magnified relatively and hence shown very nicely on the record.

This is illustrated in the two charts of Fig. 11, which are presented through the courtesy of Dr. H. B. Williams. The stethophone records are accompanied by simultaneous electro-cardiograms (E) for timing the events. The first record was made with a low pass

650 cycle filter and the second with a band pass 130–1,100 cycle filter and increased amplification. The latter shows the systolic murmur very clearly. With the *LP* 650 filter, the murmur is more or less obscured by very low pitched sounds which may really be a part



Low pass 650 cycle filter



Band pass 130–1100 cycle filter

Fig. 11—Records of a systolic murmur taken with the stethophone and a recording galvanometer

of it, but certainly play a subordinate role in producing the audible sound. The effect of suppressing the low pitched sounds by using the high pass filter is more pronounced in charts of this sort for murmurs which have negligible sound components below 130 cycles per second.

Phonograph records of heart sounds have been made previously.[14] With the stethophone and a special electrical recorder, records of some 15 cases of murmurs and chest sounds have recently been made. The results are very encouraging. All of the characteristics of these sounds, such as the relative intensities of the different components

[14] G. F. Keiper, *Letter J. A. M. A.*, 81:679 Aug. 25, 1923.

and their quality, have been retained remarkably well. The problem of subsequent reproduction of these records has been met satisfactorily in two ways, in both of which the factor of particular concern is the elimination of "needle scratch" noises. First, an electromagnetic reproducer has been used in conjunction with the stethophone. In this case, low-pass filters serve to reduce the scratching noises. Second, the records have been reproduced by attaching the earpieces of the ordinary physician's stethoscope to a special adapter used with a commercial phonograph reproducer. To reduce needle scratch in this case it is only necessary to introduce some form of air passage between the reproducer and the binaural earpieces which acts as a low-pass acoustic filter. The ordinary commercial phonograph is quite unsatisfactory for reproducing these records partly for the same reasons mentioned above relative to the use of loud speakers.

Phonograph records of heart and chest sounds can be employed to some extent for preliminary teaching purposes and do not require much equipment if reproduced acoustically. No patients are required in this case, and the records can be accompanied by the analysis or diagnosis of an expert. It is suggested that phonograph records might be used to advantage as permanent records to follow the progress of disease in important cases.

11. Summary

A summary of the applications and limitations of a new form of electrical stethoscope has been given. However, the extent of its usefulness can be brought out only after it has been placed at the disposal of experienced men in the medical profession. With it, heart murmurs and râles can be magnified and observed with greater clearness than with the ordinary stethoscope. Extremely faint sounds may be heard clearly without great acuity of hearing by inexperienced observers, a thing which has not hitherto been possible. In several instances, murmurs have been discovered with the stethophone which were not discerned initially with the ordinary stethoscope although discernible after having heard them with the more powerful apparatus. In a few of these cases, very faint murmurs, although undoubtedly present could not be heard at all with the ordinary stethoscope. It is felt that the electrical filters have played an important part in such cases. These facts lead us to believe that the stethophone may have real value for diagnosis.

The field of physical research of body sounds has been touched upon but lightly. For special purposes, an endless variety of electrical filters can be used with the stethophone.

Mathematics in Industrial Research[1]

By GEORGE A. CAMPBELL

"Selling" Mathematics to the Industries

THE necessity for mathematics in industry was recognized at least three centuries ago when Bacon said: "For many parts of nature can neither be invented [discovered] with sufficient subtility nor demonstrated with sufficient perspicuity nor accommodated unto use with sufficient dexterity without the aid and intervening of mathematics." Since Bacon's time only a very small part of nature has been "accommodated unto use," yet even this has given us such widely useful devices as the heat engine, the telegraph, the telephone, the radio, the airplane and electric power transmission. It is impossible to conceive that any of these devices could have been developed without "the aid and intervening of mathematics." Present day industry is indeed compelled, in its persistent endeavors to meet recognized commercial needs, to make use of mathematics in all of the three ways pointed out by Bacon. The record of industrial research abundantly confirms his assertion that sufficient subtility in discovery, sufficient perspicuity in demonstration, and sufficient dexterity in use can be achieved only with the aid of mathematics.

There is throughout industry one vitally important common characteristic,—uncertainty. In one industry the uncertainty may be due to the supply of raw material, the supply of labor, the supply of brains or the supply of capital. In another industry the uncertainty may be due to the activity of competitors, to fluctuating public demand or to the passage and subsequent interpretations of statutory laws. Still other industries are the playthings of the weather. Whatever the sources of uncertainty it is of vital importance to the industry to reduce to a minimum the hazards due to each of the uncertainties to which it is subjected. To a limited extent hazards may be transferred by means of insurance; but most uncertainties cannot be disposed of in this manner—they must be met by the industry individually.

The practice of probabilities, therefore, has a place in every industry. In fact, it occupies the first place in industrial mathematics, barring only the elementary arithmetical operations. It is remarkable how subtle are the mathematical difficulties presented by ap-

[1] Paper read at the International Mathematical Congress, at Toronto, August 11, 1924.

parently innocent problems in the theory of probability. For this reason, mathematicians who are entrusted with the application of probability to industry must have great insight and acumen. Even so, in applying probability to any industry, a beginning should be made with the simpler problems, going on by gradual steps to more and more complicated ones.

Each industry has its own special mathematical problems, which must be considered individually in order to determine where mathematics should be applied. No industrial problem can seem much more hopeless, as a field for exact mathematics, than the subject of electricity as understood in the time of Bacon. It was then a mere collection of curious observations, such as the evanescent attraction of rubbed amber. Persistent observation and careful correlation have, however, brought a large domain of present day electricity under quantitative relations. Electricity is now preeminently a field for mathematics, and all advances in it are primarily through mathematics.

Industrial mathematics will achieve but little unless it is undertaken by persons with suitable aptitudes working under favorable conditions, on problems which have reached the mathematical stage. Industrial mathematical research involves much more than the mechanical application of established mathematical formulas. It involves cooperation in determining the problems to be attacked, in deciding what experimental data are necessary, in obtaining these data, in formulating the mathematical problem, in carrying through the analytical and numerical work, in applying the results to the physical actuality and in practically testing the commercial results achieved. In this cooperation many individuals may be involved and many tentative trials may be necessary in order to determine the solution which best meets all of the commercial conditions.

The cooperation must be effective; it must produce results, and these promptly. Mathematical deductions must be made intelligible and convincing, so that they will eventuate in action even when the indications of theory are apparently contrary to practical experience. This is important because the most valuable theoretical results are often revolutionary.

On the part of the industrial mathematician, powers of observation, clear physical concepts, quick resourcefulness, creative imagination and constant persistency are required. These are rare human qualities.. Unless industrial mathematical work is made attractive to men possessing these high talents, the full measure of success cannot be expected. Industrial mathematics must offer a career in itself, since specialization is required—specialization of a type which eventually disqualifies most men from undertaking other lines of work most effectively.

Mathematics in Electrical Communication

In order to make the foregoing observations somewhat more specific, I will refer to a few applications of mathematics in the industrial research of the Bell Telephone System. This field is selected because I am more familiar with it than with other industrial activities.

Certainty of prediction is the basic requirement in the development and operation of the telephone system; no vital need of the system can be left to chance or to fortuitous development. For this reason, the Bell System is highly organized under research control. The telephone situation is studied as a whole; all departments cooperate; each problem is considered from every point of view. Every attempt is made to master a situation in advance of the necessity of action, so that the most effective and economical means for electrical communication may be adopted with each expansion of the system. Much more than the immediate requirements of the hour must be known; preparation for all eventualities must be made. Fortunately, the executives have carried out this program with a prophetic appreciation of the value and necessity of mathematics.

The importance of the theory and practice of probabilities was recognized as soon as the telephone reached a thoroughly commercial basis. It has proved invaluable during the great expansion which has already carried the number of telephones in the city of New York to over a million. Meeting the peak load demand of the million-odd telephones in New York City, on a practically no-delay basis, with the minimum amount of equipment, is a highly complex and important problem. Without probability studies of the situation, the equipment installed at one point would be inadequate, while at other points it would be superabundant. The superfluous equipment would involve a waste of capital, while the inadequate equipment would mean inconvenience to the public and a loss of possible revenue. Equipment engineering involves a large number of probability problems which are novel, difficult, and financially most important. The aggregate cost of all such studies is large, but the resulting saving to the telephone-using public is much greater. Satisfactory telephone service in metropolitan areas is as dependent upon applied probability as is the success of life insurance.

The telephonic ideal, which is the perfect reproduction of speech, with articulation which is indistinguishable from face-to-face conversation, involves extensive and exhaustive investigations in many fields, in particular in mechanics, acoustics and electromagnetism, since each telephonic conversation involves oscillations in the air, in solids and in the ether. Fortunately, the foundations of the mathematical theory in these three fields had been securely laid by the time Alexander Graham Bell effected their harmonious cooperation in his first telephone. It is impossible for us to be too well informed concerning the consequences of the mathematical laws in these three fields.

It is characteristic of many problems encountered in industry that a great number of independent variables are involved, far too great a number for the best solution to be reached simply by trained judgment. Consider the transposition problem of the telephone system, which is this: on pole lines, long lines between cities, for example, several wires—sometimes a great number of wires—are strung along in close proximity. Each pair of wires receives inductive effects from the electric waves carried by every other pair, producing so-called crosstalk. To reduce such effects, the pairs of wires are transposed according to a set plan; that is, the positions of the two wires are interchanged, an expedient analogous to the twisting of a pair of wires. It is necessary to consider not only the ideal location of the transpositions in each pair of wires, but also the practical irregularities which occur in the actual placing of the transpositions. One of the practical problems, in fact, is to determine the allowable tolerances limiting the irregularities in the location of loading coils and transpositions, since these irregularities modify the crosstalk and also the transmission efficiency by an amount which must be determined by the laws of probability.

Transpositions were originally introduced with complete success about thirty years ago, and yet at the present time this subject is being more actively studied than ever; this is due to the extended use of phantom circuits and the new uses of carrier frequencies, that is, high-frequency speech-carrying currents which are superposed on ordinary telephony.

To illustrate the way in which problems in industrial mathematics become, step by step, more complex by the progressive inclusion of one factor after another, brief reference may be made to the loaded cable circuit. The first successful telephone cable circuits could be treated mathematically on the basis of Kelvin's simple cable diffusion theory. To allow for the ignored inductance and to determine the effect of added inductance, Heaviside's much more complete transmission formulas were employed somewhat later. The next stage was to allow for the effect of inductance which was not uniformly distributed, but lumped at regular intervals. Here the steady state solution for sinusoidal vibrations of a loaded string was employed, and the cutoff frequency due to internal reflections at the loading coils determined. But with loaded cables of great length, extending from New York to Chicago and beyond, the transient state may be of such duration as to require consideration. The loaded line does not transmit the impulse as a whole, but breaks it up by reflection and transmission at each loading coil. Therefore some of the impulses arrive after a few short backward reflections, while other impulses may travel many times the length of the line, due to reflections back and forth at many of the thousand loading coils in the circuit. The calculation of the transient state at the receiving end, due to the arrival of these impulses in groups, one after another, involved the calculation of Bessel functions up to order 2000 and subsequent integration by an application of the principle of stationary phase to Fourier's integral.

Industrial Mathematics as a Career

It is true that the mathematician who takes up industrial work is not entirely free to set his own problems; the industry which he has chosen provides these and it demands concentration upon them. Such problems are often less inviting than the clear-cut, tractable problem which the pure mathematician is at liberty to set himself. Industrial problems may be most complicated to frame and they may admit only of approximate solution by laborious numerical methods. In addition to delimiting the nature of his problems, the imperative needs of industry set time limits for their solution, and the nature of industry demands a financial profit from industrial mathematics. But these restrictions of industry should not make the work less attractive. On the contrary, restrictions disclose the master. There is an inspiration in overcoming even the humblest difficulty standing in the path of progress. Restrictions, even in the case of the most gifted, may be beneficial in concentrating activities, thereby making up in depth what may seem lacking in breadth.

The industrial mathematician may have a chance to attack many large-scale investigations which would be impossible, except under the patronage of industry, because of the exceptional material equipment and widely sustained cooperation required. Some of the opportunities offered by cheap electrical power from Niagara, by high-voltage electric power lines, and by large steam turbines may be mentioned. It is often left to the industrial mathematician to reap the harvest from seed sown under adverse circumstances by pure mathematicians.

The industrial mathematician may hope to make some return for the debt which he owes the pure mathematician. He may introduce new mathematical problems, of which industry is an inexhaustible source. He may point out the application of pure mathematical results, stimulating further investigations along the same lines. He may assist mathematicians generally by promoting the preparation of needed tables and by creating a commercial demand for calculating machines and other brain-saving devices.

The opportunities presented by industrial mathematics are boundless, because mathematics is the key to extrapolation in time, and industry is absolutely dependent upon prediction. The position of mathematicians in industry must eventually correspond with the importance of the function which they may perform.

Training for Industrial Mathematics

In industry we are concerned with mathematics not as an objective, but only as a tool. It follows that the required training in mathematics should develop a wide acquaintance with the available mathematical tools and practical skill in their use. It is important to note the distinction between the using of tools and the making of tools. Under primitive conditions the workman makes his own tools, but in a highly organized society the tools are made by specialists, who provide the workman with an endless variety of implements superior to anything which he himself could make. By long experience the tool designer has discovered how best to adapt the tool to its intended use in order to economize the workman's time and energy as much as possible. Furthermore, the substitution of one tool for another with the minimum number of motions is made possible by the use of interchangeable parts and systematically arranged cabinets.

But no complete line of mathematical tools is for sale across the counter; only a limited number of numerical and algebraic tables and a few types of calculating machines are supplied as ready-made tools. By far the larger part of known mathematical tools must be sought for in the literature of the subject, but there they may be difficult to find and isolate in the form best adapted for the purpose in hand. What is very greatly needed at the present time is a compendium or unabridged dictionary of mathematical results concisely and uniformly stated, and systematically classified for convenient reference. What I have in mind is not a mere handbook of applied mathematics, but a statement of theorems and formulas and tabulated results expressed in the language of pure mathematics, and comparable in scope and size with the "Encyklopädie der Mathematischen Wissenschaften." Preparation of such a compendium would be a tremendous undertaking, but it would also be of the greatest value. To such a collection of tools the industrial mathematician would turn for the appropriate tool as each new problem arises.

I would have the university training of the industrial mathematician based upon such a compendium by means of judicious sampling, at many points, under competent leadership. He would thus become familiar with his source book as a whole and thereafter turn to it instinctively and use it with confidence. At the present time, when the average text-book is held in low esteem and nothing has been substituted which adequately fills the gap, the student of mathematics leaves the university with a five-foot shelf of notebooks, together with what he carries in his head. Neither the memory nor the notebook is likely to be a reliable source of information when a particular result is needed for the first time, ten years later. It then becomes necessary for him to take the time to deduce the result from first principles, or to hunt up lecture notes, a text-book or original paper and waste much valuable time picking up the thread of the argument. The sampling to which I have referred should not be that of a dilletante; it should be an intensive grounding in the fundamental concepts and methods of mathematics, and the development *ab initio* of several well distributed branches of mathematics.

The combination of mathematical ability with an observant mind is as desirable as it is rare. The university training should include non-mathematical courses adapted for developing the powers of observation, or at least an appreciation of the necessity of cooperating with others who are observant. A study of the natural sciences, accompanied by experimental work, should be of great value. It is of course, difficult to be reasonable and not ask the impossible of the university in the training of any specialist. We recognize that, at best, only a beginning can be made at the university, but this beginning should include the fundamentals and should not attempt to impart details of current industrial practice. These details are best acquired in the industrial environment itself. Self-training in fundamentals, on the other hand, is much more difficult, and is not likely to go far, unless a start has been made under the favorable conditions afforded by the university.

What I have tried to emphasize is that industry can realize its greatest possibilities only with the aid of mathematicians, and that mathematicians can find opportunities in industry worthy of their powers, however great those powers may be. To ensure the success of industrial mathematics the industry must inaugurate mathematical research as early as possible, so that ample time may be afforded for the gradual accumulation of information upon which mathematics may be securely based, and for deriving quantitative results before the necessity for commercial action arrives. The industrialist must also be ready to give the mathematician's conclusions a sympathetic trial even though they run contrary to established precedent. Above all, industry needs mathematicians of an especially broad type—men whose interests naturally extend beyond their special field, and who are flexible enough to cooperate with non-mathematicians. These industrial mathematicians must inspire confidence by their firm grasp of physical realities, by the relevance of their mathematics, and by the ability to present their results clearly and convincingly.

The Building-up of Sinusoidal Currents in Long Periodically Loaded Lines

By JOHN R. CARSON

IMPORTANT information regarding the excellence of a signal transmission system is deducible from a knowledge of the mode in which sinusoidal currents "build-up" in response to suddenly applied sinusoidal electromotive forces, since on the character and duration of the "building-up" process depend the speed and fidelity with which the circuit transmits rapid signal fluctuations.[1] The object of this note is to disclose and discuss general formulas and curves which describe the building-up phenomena, as a function of the line characteristics and the frequency of the applied e.m.f., in the extremely important case of long periodically loaded lines. The formulas in question are approximate but give accurate engineering information and are applicable to all types of periodic loading under two restrictions: (1) the line must be fairly long, that is, comprise at least 100 loading sections, and (2) it must be approximately equalized, as regards *absolute* steady-state values of the received current, in the neighborhood of the applied frequency. Fortunately these conditions are usually satisfied in practice in those cases where the building-up phenomena are of practical engineering importance. Furthermore, the formulas to be discussed supply a means for the accurate and rapid comparison of different types of loading in correctly engineered lines.

The building-up process may be precisely defined and formulated as follows: Suppose that an e.m.f., $E\cos\omega t$, is suddenly applied, at reference time $t=o$, to a network of transfer impedance

$$Z(i\omega)=|Z(i\omega)|\cdot\exp\,[iB(\omega)]. \tag{1}$$

The resultant current, $I(t)$, may be written as

$$I(t)=\frac{1}{2}\frac{E}{|Z(i\omega)|}\left\{(1+\rho)\cos\,[\omega t-B(\omega)]+\sigma\sin\,[\omega t-B(\omega)]\right\}, \tag{2}$$

$$=\frac{1}{2}\sqrt{(1+\rho)^2+\sigma^2}\,\frac{E}{|Z(i\omega)|}\cos\,[\omega t-B(\omega)+\theta], \tag{3}$$

where

$$\theta=\tan^{-1}(\sigma/\rho).$$

Evidently the functions ρ and σ must be -1 and o respectively for negative values of t, and approach the limits $+1$ and o as $t\doteq\infty$.

[1] For published discussions of the "building-up" of sinusoidal currents in loaded lines, see Clark, *Journ. A.I.E.E.*, Jan., 1923, Kupfmuller, Telegraphen u. Fernsprech-Technik, Nov., 1923; Carson, Trans. A.I.E.E., 1919.

In an engineering study of the building-up process we are principally concerned with the *envelope* of the oscillations, which, by (3), is proportional to

$$\frac{1}{2}\sqrt{(1+\rho)^2+\sigma^2}.$$

The problem is therefore to determine the functions ρ and σ and to examine the effect of the applied frequency $\omega/2\pi$ and the characteristics of the circuit on their rate of building-up and mode of approach to their ultimate steady values.

Two propositions will now be stated which cover the building-up process in the practically important cases. Since the line is assumed to be approximately equalized, as regards the absolute value of the received current in the neighborhood of the applied frequency $\omega/2\pi$, the building-up process depends only on the total phase angle $B(\omega)$. The successive derivatives of the phase angle with respect to ω will be denoted by $B'(\omega)$, $B''(\omega)$, $B'''(\omega)$, $B^{iv}(\omega)$, etc.

Case I. $B''(\omega)\neq O$ and $\sqrt{B''(\omega)/2!}$ large compared with $\sqrt[3]{B'''(\omega)/3!}$.

The envelope of the oscillations in response to an e.m.f. E cos *ωt applied at time t=o, is proportional to*

$$\frac{1}{2}\sqrt{(1+\rho)^2+\sigma^2} \tag{4}$$

where

$$\rho=C(x^2)+S(x^2), \tag{5}$$

$$\sigma=C(x^2)-S(x^2), \tag{6}$$

$$x=\frac{t-B'(\omega)}{\sqrt{2B''(\omega)}}=\frac{t'}{\sqrt{2B''(\omega)}}, \tag{7}$$

and $C(x)$, $S(x)$ are Fressel's Integrals to argument x.

The envelope therefore reaches 50 per cent. *of its ultimate steady value at time $t=\tau=B'(\omega)$ and its rate of building-up is inversely proportional to $\sqrt{B''(\omega)}$.*

The curve of Fig. 1 is a plot of the envelope function $\frac{1}{2}\sqrt{(1+\rho)^2+\sigma^2}$ to the argument x and is therefore applicable to all types of loading and lengths of line, subject to the restrictions noted above.

Case II. $B''(\omega)=O$: $B'''(\omega)\neq O$ and $\sqrt[3]{B'''(\omega)/3!}$ large compared with $\sqrt[4]{B^{iv}(\omega)/4!}$.

The envelope of the oscillations is proportional to

$$\frac{1}{3}+\frac{1}{2}\int_0^y A(\mu)d\mu \tag{8}$$

where $A(\mu)$ is Airey's Integral[2] *and*

$$y=\left(\frac{2}{\pi}\right)^{2/3}\frac{t-B'(\omega)}{\sqrt[3]{B'''(\omega)/3!}} \tag{9}$$

$$=t'\sqrt[3]{\frac{24}{\pi^2 B'''(\omega)}}. \tag{10}$$

At time $t=B'(\omega)$ the envelope N has reached 1/3 of its ultimate steady value and its rate of building-up is inversely proportional to $\sqrt[3]{B'''(\omega)}$.

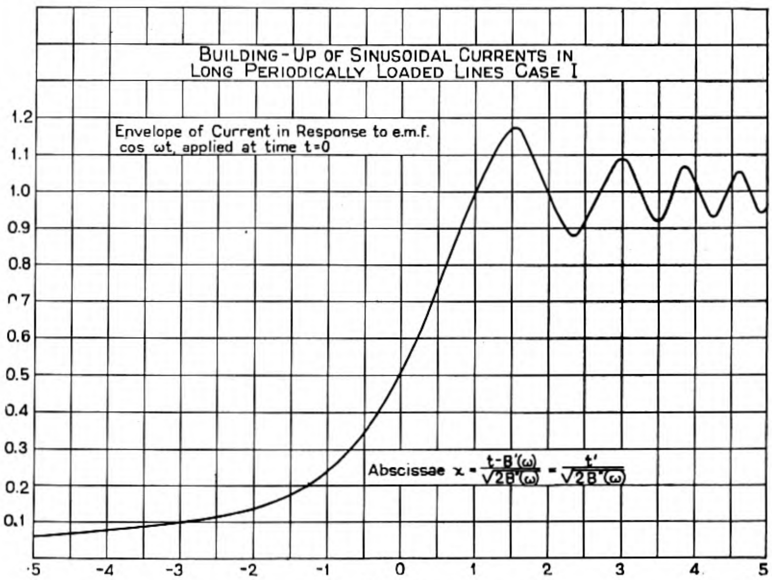


Fig. 1

The curve of Fig. 2 is a plot of the envelope function $\frac{1}{3}+\frac{1}{2}\int_0^y A(\mu)d\mu$ to the argument y and is therefore of general applicability under the circumstances where case II obtains.

The practical value of the foregoing propositions resides in the fact that they enable us to calculate two important criteria of the transmission properties of the line: (1) the variation with respect to frequency of the time interval τ required for the current to build-up to

[2] See Watson, Theory of Bessel Functions, p. 190.

its proximate steady-state value: and (2) its rate of building-up at time $t=\tau$.

As will be seen in connection with the proof given below, the formulas of the foregoing propositions are approximate. Provided, however, that the lines to which they are applied are long and provided that the

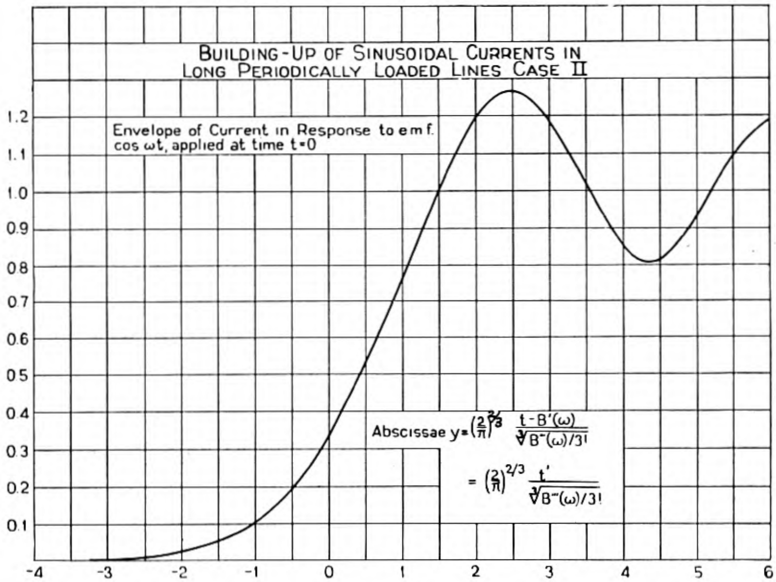


Fig. 2

applied frequency is such that the restrictions underlying either I or II are satisfied, their accuracy is quite sufficient for engineering purposes, such as the design of loading systems, or a study of the comparative merit of different types of loading.

Before proceeding with the mathematical proof, the formulas will be applied to the interesting and important case of an ideal non-dissipative periodically coil-loaded line of N sections in length and cut-off frequency $\omega_c/2\pi$. For this line it is easy to show that [3]

$$B'(\omega)=\frac{2N}{\omega_c}\frac{1}{\sqrt{1-w^2}}=N\beta'(\omega),$$

$$B''(\omega)=\frac{2N}{\omega_c^2}\frac{w}{(1-w^2)^{3/2}}=N\beta''(\omega),$$

$$B'''(\omega)=\frac{2N}{\omega_c^3}\frac{1+2w^2}{(1-w^2)^{5/2}}=N\beta'''(\omega),$$

[3] The following formulas assume that the line is closer to its characteristic impedance. $\beta(\omega)$ is then the phase angle per loading section of the line.

where w denotes ω/ω_c. It follows that

$$t' = t - \frac{2N}{\omega_c} \frac{1}{\sqrt{1-w^2}}$$

and that the oscillations build-up to the proximate steady-state in a time interval[4] $\tau = 2N/\omega_c \sqrt{1-w^2}$ after the voltage is applied.

Case I, it will be observed, does not hold for $\omega = o$ since $B''(o) = o$. The condition that Case I shall apply is that

$$\sqrt[6]{18N} \cdot (1-w^2)^{1/12} \frac{\sqrt{w}}{(1+2w^2)^{1/3}}$$

shall be substantially greater than unity. Hence Case I applies only when $1/\sqrt[3]{18N} < w < 1$. This however, includes the important part of the signalling frequency range in properly designed lines, provided that they are long ($N \geq 100$).

In the range of applied frequencies, therefore, corresponding to $1/\sqrt[3]{18N} < w < 1$, the current reaches 50 per cent. of its ultimate steady value in a time interval $\frac{2N}{\omega_c} \frac{1}{\sqrt{1-w^2}}$ after the voltage is applied and its rate of building-up at this time is proportional to

$$\frac{\omega_c}{\sqrt{4N}} \frac{(1-w^2)^{3/4}}{\sqrt{w}}.$$

For the non-dissipative coil-loaded line $B''(\omega) = o$ when $\omega = o$, and Case II applies. Consequently when $\omega = o$, the oscillations reach 1/3 of the ultimate steady value at time $t = 2N/\omega_c$, at which time their rate of building-up is proportional to

$$\omega_c \sqrt[3]{\frac{12}{\pi^2 N}}.$$

The foregoing formulas have been shown to be in good agreement with experimental results, and have been applied to the design of loaded lines in the Bell System.

Mathematical Discussion

The functions ρ and σ of equations (2) and (3) can be formulated as the Fourier integrals

[4] It will be noted that this formula breaks down at $\omega = \omega_c$ or $w = 1$.

$$\rho=\frac{1}{\pi}\int_0^\infty \frac{d\lambda}{\lambda}\sin t\lambda.\,[P_\omega(\lambda)+P_\omega(-\lambda)]$$
$$-\frac{1}{\pi}\int_0^\infty \frac{d\lambda}{\lambda}\cos t\lambda\,[Q_\omega(\lambda)-Q_\omega(-\lambda),] \tag{11}$$

$$\sigma=\frac{1}{\pi}\int_0^\infty \frac{d\lambda}{\lambda}\sin t\lambda\,[Q_\omega(\lambda)+Q_\omega(-\lambda)]$$
$$+\frac{1}{\pi}\int_0^\infty \frac{d\lambda}{\lambda}\cos t\lambda\,[P_\omega(\lambda)-P_\omega(-\lambda)], \tag{12}$$

where

$$P_\omega(\lambda)=\frac{A(\omega+\lambda)}{A(\omega)}\cos[B(\omega+\lambda)-B(\omega)], \tag{13}$$

$$Q_\omega(\lambda)=\frac{A(\omega+\lambda)}{A(\omega)}\sin[B(\omega+\lambda)-B(\omega)], \tag{14}$$

and $A(\omega)=1/|Z(i\omega)|$.

These formulas are directly deducible from the fact that the applied e.m.f., defined as zero for negative values of t and $E\cos\omega t$ for $t\geq o$, can itself be expressed as

$$\frac{E}{2}\cos\omega t\left[1+\frac{2}{\pi}\int_0^\infty \frac{d\lambda}{\lambda}\sin t\lambda\right].$$

In the practically important case where $B'(\omega)$ is finite, it is of advantage to introduce the transformation $t'=t-B'(\omega)$, and to write:

$$\rho=\frac{1}{\pi}\int_0^\infty \frac{d\lambda}{\lambda}\sin t'\lambda\,[U_\omega(\lambda)+U_\omega(-\lambda)]$$
$$-\frac{1}{\pi}\int_0^\infty \frac{d\lambda}{\lambda}\cos t'\lambda\,[V_\omega(\lambda)-V_\omega(-\lambda)], \tag{15}$$

$$\sigma=\frac{1}{\pi}\int_0^\infty \frac{d\lambda}{\lambda}\sin t'\lambda\,[V_\omega(\lambda)+V_\omega(-\lambda)]$$
$$+\frac{1}{\pi}\int_0^\infty \frac{d\lambda}{\lambda}\cos t'\lambda\,[U_\omega(\lambda)-U_\omega(-\lambda)], \tag{16}$$

where

$$U_\omega(\lambda)=\frac{A(\omega+\lambda)}{A(\omega)}\cos[B(\omega+\lambda)-B(\omega)-\lambda B'(\omega)], \tag{17}$$

$$V_\omega(\lambda)=\frac{A(\omega+\lambda)}{A(\omega)}\sin[B(\omega+\lambda)-B(\omega)-\lambda B'(\omega)]. \tag{18}$$

The foregoing formulas for ρ and σ are exact subject to certain restrictions on the impedance function $Z(i\omega)$ which are satisfied in the case of periodically loaded lines. Their useful application to the problem under consideration depends, however, on the following approximations.

First it will be assumed that the line is approximately equalized, as regards *absolute* value of steady state received currents in the neighborhood of the impressed frequency $\omega/2\pi$. By virtue of this assumption, which is more or less closely realized in practice, the ratio $A(\omega+\lambda)/A(\omega)$ may be replaced by unity in the integrals (15) and (16), and in equations (17) and (18). It is further assumed that the function

$$B(\omega+\lambda)-B(\omega)-\lambda B'(\omega)$$

admits of power series expansion, so that

$$U_\omega(\lambda)=\cos\,[(h_2\lambda)^2+(h_3\lambda)^3+\,.\,.],\tag{19}$$

$$V_\omega(\lambda)=\sin\,[(h_2\lambda)^2+(h_3\lambda)^3+\,.\,.],\tag{20}$$

where

$$h_n{}^n=\frac{1}{n!}\frac{d^n}{d\omega^n}\,B(\omega)=\frac{1}{n!}\,B^{(n)}(\omega).$$

By virtue of the foregoing ρ and σ are given by

$$\rho\doteq\frac{2}{\pi}\int_0^\infty\frac{d\lambda}{\lambda}\sin\,[t'\lambda-(h_3\lambda)^3-(h_5\lambda)^5\,.\,.]\cdot\cos\,[(h_2\lambda)^2+(h_4\lambda)^4+\,.\,.],\tag{21}$$

$$\sigma\doteq\frac{2}{\pi}\int_0^\infty\frac{d\lambda}{\lambda}\sin\,[t'\lambda-(h_3\lambda)^3-(h_5\lambda)^5\,.\,.]\cdot\sin\,[(h_2\lambda)^2+(h_4\lambda)^4+\,.\,.].\tag{22}$$

Now if the line is very long the integrals (11) and (12) may be replaced by the approximations

$$\rho\doteq\frac{2}{\pi}\int_0^\infty\frac{d\lambda}{\lambda}\sin\,[t'\lambda-(h_3\lambda)^3]\cdot\cos\,(h_2\lambda)^2,\tag{23}$$

$$\sigma\doteq\frac{2}{\pi}\int_0^\infty\frac{d\lambda}{\lambda}\sin\,[t'\lambda-(h_3\lambda)^3]\cdot\sin\,(h_2\lambda)^2.\tag{24}$$

In other words we retain only the leading terms in the expansion of the function

$$B(\omega+\lambda)-B(\omega)-\lambda B'(\omega).$$

The justification for this procedure depends on arguments similar to those underlying the Principle of Stationary Phase (see Watson, Theory of Bessel Functions, p. 229). Furthermore the upper limit

∞ may be retained without serious error, even when the line cuts off at a frequency $\omega_c/2\pi$, provided the line is sufficiently long, and the frequency $\omega/2\pi$ not too close to the cut-off frequency $\omega_c/2\pi$.

The formal solutions of the infinite integrals (23) and (24) can be written down by virtue of the following known relations:

$$\frac{2}{\pi}\int_0^\infty \frac{d\lambda}{\lambda} \sin t'\lambda \cdot \cos (h_2\lambda)^2 = C(x^2) + S(x^2), \tag{25}$$

$$\frac{2}{\pi}\int_0^\infty \frac{d\lambda}{\lambda} \sin t'\lambda \cdot \sin (h_2\lambda)^2 = C(x^2) - S(x^2), \tag{26}$$

where $C(x^2)$ and $S(x^2)$ are Fresnel's Integrals to argument x^2, and $x = t'/2h_2$.

$$\frac{2}{\pi}\int_0^\infty \frac{d\lambda}{\lambda} \sin [t'\lambda - (h_3\lambda)^3] = -\frac{1}{3} + \int_0^y A(y)dy \tag{27}$$

where $A(y)$ denotes Airey's Integral (see Watson, Theory of Bessel Functions) and $y = (2/\pi)^{2/3}(t'/h_3)$.

By aid of the preceding.

$$\rho = \left\{ 1 + \frac{\mu^3}{1!}\frac{d^3}{dx^3} + \frac{\mu^6}{2!}\frac{d^6}{dx^6} + \,.\,.\, + \,.\,. \right\} \cdot \left\{ C(x)^2 + S(x^2) \right\}, \tag{28}$$

$$\sigma = \left\{ 1 + \frac{\mu^3}{1!}\frac{d^3}{dx^3} + \frac{\mu^6}{2!}\frac{d^6}{dx^6} + \,.\,.\, + \,.\,. \right\} \cdot \left\{ C(x^2) - S(x^2) \right\}, \tag{29}$$

where $\mu = (h_3/2h_2)$.

This is the appropriate form of solution when (h_3/h_2) is less than unity.

On the other hand when (h_3/h_2) is greater than unity, the appropriate form of solution is

$$\rho = \left\{ 1 - \frac{\nu^4}{2!}\frac{d^4}{dy^4} + \frac{\nu^8}{4!}\frac{d^8}{dy^8} + \,.\,. \right\} \cdot \left\{ -\frac{1}{3} + \int_0^y A(y)dy \right\}, \tag{30}$$

$$\sigma = \left\{ \frac{\nu^2}{1!}\frac{d^2}{dy^2} - \frac{\nu^6}{3!}\frac{d^6}{dy^6} + \,.\,. \right\} \cdot \left\{ -\frac{1}{3} + \int_0^y A(y)dy \right\}, \tag{31}$$

where $\nu = \left(\frac{2}{\pi}\right)^{\frac{2}{3}} \left(\frac{h_3}{h_2}\right)$.

While no thorough investigation has been made, it appears probable that for all values of the ratio h_3/h_2, either (28), (29) or (30), (31) will be convergent. However, in practice it is sufficient for present purposes to deal only with the cases where h_3/h_2 is either small or large compared with unity, and to use the following approximations:

(1) (h_3/h_2) *small compared with unity.*

$$\rho = C(x^2) + S(x^2),$$
$$\sigma = C(x^2) - S(x^2),$$
$$x = (t'/2h_2)^2,$$
$$t' = t - B'(\omega).$$

(2) (h_3/h_2) *large compared with unity.*

$$\rho = -\frac{1}{3} + \int_0^y A(y)dy,$$
$$\sigma = O,$$
$$y = (2/\pi)^{2/3}(t'/h_3).$$

Transmission Characteristics of Electric Wave-Filters

By OTTO J. ZOBEL

SYNOPSIS: The transmission loss characteristic of a transmitting network as a function of frequency is an index of the network's steady-state selective properties. Methods of calculation heretofore employed to determine these characteristics for composite wave-filters are long and tedious. This paper gives a method for such determinations which greatly simplifies and shortens the calculations by the introduction of a system of charts. Account is taken of the effects of both wave-filter dissipation and terminal conditions. The method is based upon formulae containing new parameters, called "image parameters," which are the natural ones to use with composite wave-filters.

A detailed illustration of the use of this chart calculation method is given and the transmission losses so obtained are found to agree, except for differences which in practice are negligible, with those obtained by long direct computation.

In the Appendix are derived two sets of corresponding formulae which are applicable to a linear transducer of the most general type, namely, an active, dissymmetrical one; the one set contains image parameters and the other set recurrent parameters. An impedance relation is found to exist between the four open-circuit and short-circuit impedances of a linear transducer even in the most general case. Reduction of these formulae to the more usual case of a passive linear transducer is also made, those containing the image parameters being especially applicable to the case of composite wave-filters.

I. INTRODUCTION

ELECTRIC wave-filter characteristics and systematic methods of deriving them have been considered in previous numbers of this Journal.[1] This paper deals with a simple and rapid method of calculating the steady-state transmission losses of wave-filter networks over both the transmitting and attenuating frequency bands, including the effects of dissipation and wave-filter terminal conditions. Such transmission loss determinations are essential in showing the selective characteristics of these networks and serve as important guides in meeting given design requirements.

General formulae for any dissymmetrical linear transducer are derived in terms of new parameters, called image parameters. One of the formulae is fundamental to the solution of the present problem and is particularly well adapted to calculations in composite wave-filter structures. These parameters of such a composite structure, being readily obtainable from those of its parts, are the natural parameters to use in this case. The formula possesses, among others,

[1] Physical Theory of the Electric Wave-Filter, G. A. Campbell, B. S. T. J., Nov., 1922; Theory and Design of Uniform and Composite Electric Wave-Filters, O. J. Zobel, B. S. T. J., Jan., 1923; Transient Oscillations in Electric Wave-Filters, J. R. Carson and O. J. Zobel, B. S. T. J., July, 1923.

the advantage over other formulae and calculation methods of requiring for every alteration in a composite wave-filter only a partial recalculation rather than a more or less complete one. In addition, by its use much of the otherwise necessary calculation can be eliminated through the means of graphical representation.

The main object of this paper is to present this chart calculation method of determining composite wave-filter transmission losses, giving its theory, the necessary charts, and an application of its use.

Structure of Wave-Filter Networks

The ladder type of recurrent network having physical series and shunt impedances z_1 and z_2, respectively, as shown in Fig. 1, is the

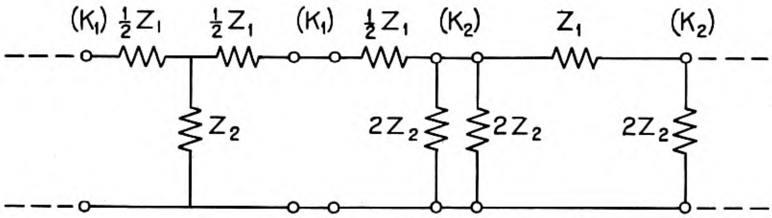


Fig. 1—Ladder Type Recurrent Network

one most frequently employed for wave-filters. Also any passive transducer having two pairs of terminals can theoretically be reduced to the form of the ladder type. Hence, in what follows the ladder type terminology will be used understanding, however, that other structural types may also be included; for example, such as are derivable from the ladder type by the substitution of an equivalent transformer with mutual impedance for T or Π connected inductances, or the lattice type. The figure illustrates, from left to right, one mid-series section, one mid-half section (a dissymmetrical half section terminated at mid-series and mid-shunt points), and one mid-shunt section all connected so as to give a uniform structure.[2] The characteristic impedances of the ladder type at mid-series and at mid-shunt points are K_1 and K_2, respectively.

The majority of wave-filter networks are not uniform throughout their length but have a composite structure designed as given in the paper (B. S. T. J., Jan., 1923) already mentioned. That is, the interior or *mid-part* of a composite wave-filter consists of mid-series,

[2] The same network may also be considered as made up in other ways; for example, two mid-series and one mid-half sections, one mid-half and two mid-shunt sections, or five mid-half sections.

mid-shunt, and mid-half sections, usually dissimilar, so connected serially and of such types that at any junction the terminations of the two adjacent types correspond to an equivalent image impedance. The use of dissimilar sections gives a resultant selective characteristic different from that possible with a uniform type. At the terminals of the network there need not be complete full or half sections; this

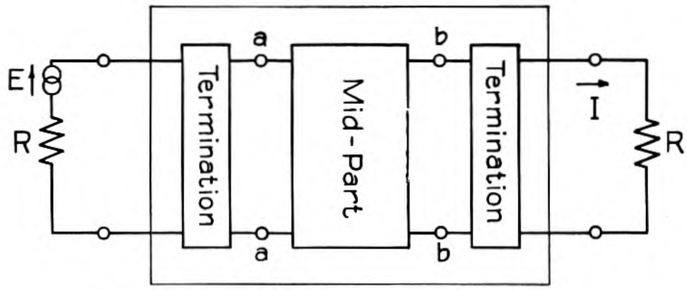


Fig. 2—General Composite Wave-Filter Network

is represented in Fig. 2 by the wave-filter parts external to the ***mid-part*** which latter is included between terminals a and b.

The terminations of wave-filter networks specifically considered in detail here include all terminations which have been found to be practical. In any particular class of wave-filter they are all closely related to the "constant k" wave-filter ($z_{1k}z_{2k}=k^2=$constant) of that class and are of the following four types,[3] being designated by their characteristic impedances in corresponding ladder type structures.

[3] It is assumed that the reader is familiar with the terms and notation used in the paper, B.S.T.J., Jan., 1923.

If z_{1k} and z_{2k} are the series and shunt impedances of the "constant k" wave-filter, the corresponding series and shunt impedances of the mid-series "const nt k" equivalent M-type are expressible as

$$z_{11}=mz_{1k},$$

and

$$z_{21}=\frac{1-m^2}{4m}z_{1k}+\frac{1}{m}z_{2k};$$

and of the mid-shunt "constant k" equivalent M-type

$$z_{12}=\frac{1}{\dfrac{1}{mz_{1k}}+\dfrac{1}{\dfrac{4m}{1-m^2}z_{2k}}},$$

and

$$z_{22}=\frac{1}{m}z_{2k}.$$

Here the condition $0<m\leqq 1$ is sufficient for a physical structure in all cases.

1, mid-shunt of a mid-series "constant k" equivalent M-type, ($K_{21}(m)$);
2, mid-series of a mid-shunt "constant k" equivalent M-type, ($K_{12}(m)$);
3, x-shunt of the "constant k" wave-filter, (K_{x2}); and
4, x-series of the "constant k" wave-filter, (K_{x1}).

The terminations in $K_{21}(m)$ and $K_{12}(m)$ are employed, as stated in a previous paper,[4] when it is desirable to obtain certain selective characteristics and to minimize reflection losses at the important frequencies to be transmitted, a minimum for the latter occurring where $m=.6$ approximately. The x-shunt "constant k" termination, designated by the characteristic impedance K_{x2}, is a "constant k" type termination in a shunt element, whose admittance is x times (x from 0 to 1) that of a full shunt "constant k" admittance, $\frac{1}{z_{2k}}$; that is, a shunt element whose impedance is $\frac{z_{2k}}{x}$. Similarly, the x-series "constant k" type termination corresponding to the characteristic impedance K_{x1} ends in a series element of impedance xz_{1k}. In the usual case where two or more wave-filter networks having different transmitting bands are associated together, either termination 1 or 2 is suitable for the unconnected terminals, while terminations 3 and 4 are adapted to the terminals connected in series or in parallel, respectively. For two complementary wave-filters, thus connected, minimum reflection losses occur at their junction with a transmission line if $x=.8$ approximately. A relation between this case and termination $K_{12}(m)$ and $K_{21}(m)$ has previously been pointed out, namely, that the series or parallel connected wave-filters have a combined impedance in the transmitting band of either wave-filter approximately like that of $K_{12}(m)$ or $K_{21}(m)$, respectively.

Where the termination is x-shunt or x-series we shall consider that the *mid-part* of the wave-filter begins at the mid-shunt or the mid-series point, respectively, irrespective of whether x is greater or less than .5. Also the *mid-part* need not here necessarily begin in the "constant k" type, but in any wave-filter having an equivalent characteristic impedance.

Transmission Loss

In the design of a wave-filter network the magnitude of k for the corresponding "constant k" wave-filter has been taken equal to the

[4] B. S. T. J., Jan., 1923, page 18, gives a diagram for the non-dissipative case of $R/K_{21}(m)$ and $K_{12}(m)/R$ in the transmitting band.

mean resistance, R, of the line with which the network is to be associated. If the network is closed at each end by a resistance of magnitude R, as in Fig. 2, we have not only a circuit arrangement which approximates more or less closely actual operating conditions,[5] but also a simple test circuit in which to determine the transmission loss of the network over the desired frequency range.

The transmission loss of a wave-filter network, defined with reference to Fig. 2, is the natural logarithm, with negative sign, of the ratio of the absolute value of the current transmitted from a source of resistance R to a receiving resistance R when the latter are connected through the network, to that transmitted when they are connected directly. Let E represent the electromotive force of the source, I the current transmitted to R through the network, and $E/2R$ that transmitted by direct connection. Then the *transmission loss L*, thus defined, is

$$L = -\log_e \left| \frac{I}{E/2R} \right|, \tag{1}$$

and

$$e^{-L} = |2RI/E|. \tag{2}$$

The unit in which L is expressed, the *attenuation unit*,[6] is the natural unit to use here and from the above relations it is seen that one attenuation unit of transmission loss corresponds to an absolute value of current ratio of $1/e$. The method of determining the transmission loss under various possible conditions will be presented in the next part of this paper.

II. Theory of Chart Calculation Method

The principles given here are basic and apply to composite wave-filters having any terminations. However, in all practical cases, as previously stated, the terminations belong to the four types: 1, mid-shunt M-type; 2, mid-series M-type; 3, x-shunt "constant k;" and 4, x-series "constant k," all related to the "constant k" wave-filter.

[5] It should be clearly borne in mind that the unique selective properties of a wave-filter of freely transmitting currents in continuous frequency bands and of attenuating others are those for the wave-filter terminated in its characteristic impedance. It is practical to have approximately such a termination in the transmitting band only, as when connecting the wave-filter to a transmission line, in which case the general properties still persist. *Correct termination* rather than *number of sections* is what brings out these properties although the degree of selectivity is naturally increased by the addition of sections.

[6] A synonym sometimes used is the *Napier*. One attenuation unit is equivalent to 9.174 "800-cycle miles of standard cable," and to 8.686 TU. The TU (transmission unit) is that unit which designates a power ratio of $10^{\cdot 1}$, and the number of TU is ten times the common logarithm of the power ratio.

These four cases will be developed in detail and equivalence relations for certain sets of terminal combinations shown.

Fundamental Formula

The formula which is general and fundamental to what follows is the one giving the current received through a passive transducer in terms of the sending electromotive force, the terminal impedances,

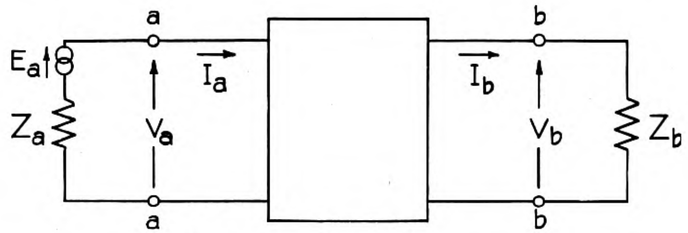


Fig. 3—General Linear Transducer

and the transfer constant[7] and image impedances of the transducer. Referred to Fig. 3 the received current is

$$I_b = \frac{2E_a \sqrt{W_a W_b}\, e^{-T}}{(W_a+Z_a)(W_b+Z_b)(1-r_a r_b e^{-2T})}, \tag{3}$$

where

E_a = sending electromotive force,

Z_a, Z_b = sending and receiving impedances,

$T = D + iS$ = transfer constant of the transducer,

D, S = diminution constant and angular constant, defined as the real and imaginary parts of the transfer constant,

W_a, W_b = image impedances of the transducer at terminals a and b,

r_a, r_b = current reflection coefficients at terminals a and b,

$$r_a = \frac{W_a - Z_a}{W_a + Z_a},$$

and

$$r_b = \frac{W_b - Z_b}{W_b + Z_b}.$$

[7] The terms *transfer constant*, T and *image impedances*, W_a and W_b, as applied to a dissymmetrical passive transducer, are defined in the Appendix. These three parameters are to be distinguished from another set, the propagation constant, Γ, and characteristic impedances, K_a and K_b. In a symmetrical structure $T = \Gamma$ and $W_a = W_b = K_a = K_b$.

Another form obtained by suitable transformation is

$$I_b = \frac{E_a\sqrt{W_a W_b}}{(W_a W_b + Z_a Z_b)\sinh T + (W_a Z_b + W_b Z_a)\cosh T}. \tag{4}$$

Formula (3), derived in the Appendix with several general transducer formulae and relations, is especially useful when applied to composite wave-filter networks, since, as we shall see, it contains the natural parameters for such structures. Upon comparing Fig. 2, which represents such a general network, with Fig. 3 we find that the two can be made to correspond exactly if the mid-part of the wave-filter, between terminals a and b in Fig. 2, is considered to be the transducer of Fig. 3, and if the wave-filter terminations combined with the resistances R are considered to be the terminal impedances Z_a and Z_b of Fig. 3. The relation between the electromotive force, E, applied in R and that, E_a, acting through Z_a depends upon the particular wave-filter termination at terminals a. Similarly, the relation between the currents, I and I_b, transmitted to R and Z_b, respectively, depends upon the termination at terminals b.

As already stated, the mid-part of the composite wave-filter consists in general of mid-series, mid-shunt, and mid-half sections, properly combined as to their impedance relations at the junction points. *The method of combination employed in a composite wave-filter consists in connecting two sections whose image impedances at their junction are equal.* (An analogy which might be given is the matching of dominoes in a line by the corresponding ends, numbers referring to image impedances.)

Let us assume for the moment that the mid-part, as thus made up, is terminated by impedances respectively equal to its image impedances. There is then an "image condition" for the impedances measured in the two directions not only at each of these terminal points but also at each junction point throughout the network; and in this case each section transmits under the "image condition" of its terminating impedances. As a result we obviously obtain the following properties for the mid-part.

1. *The transfer constant of the mid-part of a composite wave-filter, consisting of mid-series, mid-shunt, and mid-half sections, is the sum of the transfer constants of all the individual sections.*

2. *The image impedances of the mid-part of a composite wave-filter are the external image impedances of the two end sections.*

In addition we have the following important relations between the transfer constant and image impedances of a single section, and the propagation constant and mid-point characteristic impedances of the corresponding ladder network.

3. *The transfer constant of a symmetrical mid-series or mid-shunt section is equal to the propagation constant of the corresponding ladder type; that of a dissymmetrical mid-half section (having mid-series and mid-shunt terminations) is equal to one-half the above propagation constant.*

4. *The image impedance of a mid-point section at a mid-series or mid-shunt point termination is equal to the mid-series, K_1, or mid-shunt, K_2, characteristic impedance, respectively, of the corresponding ladder network.*

Formula (3) is for the present purpose superior to the well known formula for transmitted current (derived for comparison in the Appendix) which contains the transducer recurrent parameters in the form of its propagation constant, Γ, and characteristic impedances, K_a and K_b. The reason for this is that in a dissymmetrical composite wave-filter where K_a differs from K_b, the usual case, no simple relations exist between these latter parameters of the transducer and the corresponding parameters of the individual sections comprising the network. In the special case of symmetrical networks, however, the latter formula becomes identical with (3) which follows from what has already been said.

Another method of obtaining the transmitted current, which may be termed the "section-by-section elimination method," consists in calculating by the aid of the Kirchhoff laws the current ratios and total impedances from section to section back through the entire network beginning at the receiving impedance. From the standpoint of time economy certain objections may be raised to the possible use here of this general long hand method of calculation. The method carries with it the determination of the phase as well as the amplitude of the transmitted current; but since the amplitude only is required in the transmission loss formula, this method does more than is necessary. Again, an alteration in the composite network structure requires a more or less complete recalculation when this method is employed, whereas by the application of (3) it will be found that this is not necessary. However, this method is useful where irregularities exist in the network, or where the particular method of design which had been followed in obtaining the composite structure cannot readily be found, but its impedance elements and R are known.

General Form of Transmission Loss Formula

Formulae (2) and (3) corresponding to Figs. 2 and 3 may be combined. If (3) is written in the general form

$$2RI/E = F_t\, F_a\, F_b\, F_r, \tag{5}$$

we obtain with (2)

$$e^{-L} = |\,2RI/E\,| = e^{-(L_t+L_a+L_b+L_r)}, \tag{6}$$

where the four factors comprising the current ratio $2RI/E$ are

$F_t = e^{-T} =$ the transfer factor between terminals a and b;

$F_a = \dfrac{2\sqrt{W_aR}\,E_a}{(W_a+Z_a)\,E} =$ the terminal factor at terminals a;

$F_b = \dfrac{2\sqrt{W_bR}\,I}{(W_b+Z_b)\,I_b} =$ the terminal factor at terminals b;

$F_r = \dfrac{1}{1 - r_a r_b e^{-2T}} =$ the interaction factor due to repeated reflections at terminals a and b where the current reflection coefficients are

$$r_a = \frac{W_a - Z_a}{W_a + Z_a} \text{ and } r_b = \frac{W_b - Z_b}{W_b + Z_b};$$

and the transmission losses corresponding to the absolute values of these factors are called, respectively,

$L_t =$ the transfer loss;

L_a, $L_b =$ the terminal losses at terminals a and b;

and $L_r =$ the interaction loss.

The total transmission loss is the sum of these four losses, thus,

$$L = L_t + L_a + L_b + L_r. \tag{7}$$

The relative importance of the three types of losses, transfer, terminal, and interaction, is usually in the order given. Hence, as a first approximation the transmission loss of a composite wave-filter is given by the transfer loss, L_t, but the error due to the omission of the other losses is often considerable. A second approximation is obtained by including the terminal losses, L_a and L_b, and for many purposes this is sufficiently accurate. The final step for accuracy is the further addition of the interaction loss, L_r, whose effect on the total transmission loss is usually appreciable in the transmitting band of a wave-filter near the critical frequencies.

The three types of losses will now be considered separately and in detail.

1. *Transfer Losses*

The transfer loss, L_t, is by (6) equal to D, the diminution constant, which is the real part of the transfer constant, T, of the wave-filter mid-part taken between mid-points.

We have previously established the following:

> (1) *T is the sum of the transfer constants of all the individual sections, i.e., $T = \Sigma T_j$; and* (2) *the transfer constant of a mid-series or mid-shunt section is equal to the propagation constant, $\Gamma = A + iB$ per full section, of the corresponding ladder type; that of a mid-half section is $\Gamma/2$.*

Hence, to get the transfer loss we need to know only the attenuation constant, A, of each full mid-section, the half or whole of which forms a part of the composite wave-filter structure. However, since the interaction factor which is to be discussed later requires a knowledge of the phase constant, B, as well, we shall consider both parts of the propagation constant at this point.

Propagation Constant of Ladder Type Network. The relation between the propagation constant $\Gamma = A + iB$, and the series and shunt impedances, z_1 and z_2, respectively, of the ladder type in Fig. 1 is known to be

$$\cosh \Gamma = 1 + \tfrac{1}{2}\frac{z_1}{z_2}. \tag{8}$$

This applies as well to any recurrent structure if z_1 and z_2 correspond to the analytically equivalent ladder type.

Let us introduce two variables U and V by making the substitution

$$\frac{z_1}{4z_2} = U + iV. \tag{9}$$

The reason for this choice is that this ratio appears frequently in impedance formulae. Then in non-dissipative wave-filters, where $V = O$, the transmitting bands include all frequencies at which U satisfies the relation

$$-1 \leqq U \leqq O. \tag{10}$$

By (8) and (9)

$$\cosh(A+iB)=\cosh A\cos B+i\sinh A\sin B=1+2U+i2V, \tag{11}$$

whence

$$\cosh A\cos B=1+2U,$$

and

$$\sinh A\sin B=2V. \tag{12}$$

The solution of this pair of simultaneous equations leads to separate relations for A and B,

$$\left(\frac{1+2U}{\cosh A}\right)^2+\left(\frac{2V}{\sinh A}\right)^2=1, \tag{13}$$

and

$$\left(\frac{1+2U}{\cos B}\right)^2-\left(\frac{2V}{\sin B}\right)^2=1. \tag{14}$$

As is well known from (13) equal attenuation constant loci are represented in the U, V plane by confocal ellipses with foci at $U=-1$, $V=O$ and $U=O$, $V=O$, thus having symmetry about the U-axis. *The locus for $A=O$, the limiting case, is a straight line between the foci and it corresponds to the transmitting band in a non-dissipative wave-filter.* Similarly from (14) equal phase constant loci are represented by confocal hyperbolas which have the same foci as above and are orthogonal to the equal attenuation constant ellipses. It will be assumed that the phase constant, B, lies between $-\pi$ and $+\pi$, which amounts to neglecting multiples of 2π. Then from (12) B has the same sign as V, so that loci in the upper half of the plane correspond to a positive phase constant while those in the lower half correspond to a negative one.

It is possible, however, to represent all this in just the upper half of the plane using coordinates U and $|V|$. Put

$$V=c\,|V|, \tag{15}$$

where $c=\pm 1$, the sign being that of V. The attenuation constant is independent of the sign of V, i.e., of c. But for the phase constant we get from (12)

$$\sin cB=\frac{2\,|V|}{\sinh A}, \tag{16}$$

and

$$O\leqq cB\leqq +\pi.$$

Thus, as here considered, *the product cB, where $c=\pm 1$ has the sign of V, is always positive with a value less than or equal to π.*

Explicit formulae for A and B from (13) and (14) are

$$A = \sinh^{-1}\sqrt{2\left[\left|\sqrt{(U+U^2+V^2)^2+V^2}\right|+(U+U^2+V^2)\right]}, \qquad (17)$$

and

$$cB = \sin^{-1}\sqrt{2\left[\left|\sqrt{(U+U^2+V^2)^2+V^2}\right|-(U+U^2+V^2)\right]}. \qquad (18)$$

The above formulae are general and applicable to any ladder type structure or its equivalent.

In the case of wave-filters certain approximate formulae are often useful. At frequencies in the attenuating bands away from the critical frequencies and the frequencies of maximum attenuation, and *wherever* V^2 *is negligible compared with* $(U+U^2)>O$,

$$A = \sinh^{-1} 2\sqrt{U+U^2},$$

and (19)

$$cB = O \text{ or } \pi.$$

At the critical frequencies and the frequencies of maximum attenuation, *where* $(U+U^2)$ *is negligible compared with* V^2,

$$A = \cosh^{-1}(\sqrt{1+V^2}+|V|),$$

and (20)

$$cB = \cos^{-1}\pm(\sqrt{1+V^2}-|V|).$$

In the latter the positive sign applies to a critical frequency at which $U=O$, and the negative sign to one at which $U=-1$.

U and V for "Constant k" and M-type Wave-Filters. Since the wave-filter structures under consideration have "constant k" or derived M-type terminations, the U and V variables corresponding to these wave-filters will always be required. Hence, formulae for the variables are given here, limiting them to the four lowest wave-filter classes generally used.

Resistance in an inductance coil of inductance, L_1, is taken into account by expressing the total coil impedance as

$$(d+i)\ L_1 2\pi f,$$

where d, the "coil dissipation constant," is the ratio of coil resistance to coil reactance. The value of d is ordinarily between $d=.004$ and $d=.04$, and it does not vary rapidly with frequency. Similarly, dissipation in a condenser of capacity C_1 can be included by expressing the total condenser admittance as $(d'+i)\ C_1 2\pi f$, but since d' is usually negligible in practice it will here be omitted.

The formulae derived from (9) are based upon those given in this

Journal, Jan., 1923, pages 39 to 41, and contain the critical frequencies and frequencies of maximum attenuation. Subscripts $_k$ and $_m$ will be used to denote the "constant k" and M-type U and V variables. The "constant k" formulae for the four classes follow.

Low Pass.

$$U_k = -\left(\frac{f}{f_2}\right)^2,$$

and (21)

$$V_k = d\left(\frac{f}{f_2}\right)^2.$$

High Pass.

$$U_k = -\left(\frac{f_1}{f}\right)^2 / (1+d^2),$$

and (22)

$$V_k = -d\left(\frac{f_1}{f}\right)^2 / (1+d^2).$$

Low-and-High Pass.

$$U_k = -\frac{(f_1-f_0)^2}{f_0 f_1} \frac{\left[\frac{f_0 f_1}{f^2} - (1+d^2)(2 - \frac{f^2}{f_0 f_1})\right]}{\left[\frac{f_0 f_1}{f^2} + (1+d^2)\frac{f^2}{f_0 f_1} - 2\right]^2},$$

and (23)

$$V_k = d\frac{(f_1-f_0)^2}{f_0 f_1} \frac{\left[\frac{f_0 f_1}{f^2} - (1+d^2)\frac{f^2}{f_0 f_1}\right]}{\left[\frac{f_0 f_1}{f^2} + (1+d^2)\frac{f^2}{f_0 f_1} - 2\right]^2}.$$

Band Pass.

$$U_k = -\frac{f_1 f_2}{(f_2-f_1)^2}\left[\frac{f_1 f_2}{(1+d^2)f^2} + \frac{f^2}{f_1 f_2} - 2\right],$$

and (24)

$$V_k = -d\frac{f_1 f_2}{(f_2-f_1)^2}\left[\frac{f_1 f_2}{(1+d^2)f^2} - \frac{f^2}{f_1 f_2}\right].$$

At the mid-frequency, $\sqrt{f_1 f_2}$, the point of confluency of two bands in the transmitting band of this wave-filter, we obtain approximately from (19), when d is small,

$$A = 2d\frac{\sqrt{f_1 f_2}}{f_2 - f_1},$$

and (25)

$$B = O.$$

The derived M-type variables of any class are given directly in terms of the "constant k" variables of that class and the parameter m by the general relations

$$U_m=\frac{m^2[U_k+(1-m^2)(U_k^2+V_k^2)]}{[1+(1-m^2)U_k]^2+(1-m^2)^2V_k^2},$$

and (26)

$$V_m=\frac{m^2V_k}{[1+(1-m^2)U_k]^2+(1-m^2)^2V_k^2}.$$

This assumes that the M-type has the same grade of coils and condensers as its "constant k" prototype. The parameter m has a different formula determining its value for each class, the general relation being (neglecting dissipation)

$$m=\sqrt{1+\left(\frac{1}{U_k}\right)_{f_\infty}}, \tag{27}$$

where f_∞ is a frequency of maximum attenuation of the M-type. The particular relations for the above four classes follow.

Low Pass

$$m=\sqrt{1-\frac{f_2^2}{f_{2\infty}^2}}. \tag{28}$$

High Pass

$$m=\sqrt{1-\frac{f_{1\infty}^2}{f_1^2}}. \tag{29}$$

Low-and-High Pass

$$m=\frac{\sqrt{\left(1-\frac{f_0^2}{f_{1\infty}^2}\right)\left(1-\frac{f_{1\infty}^2}{f_1^2}\right)}}{1-\frac{f_0}{f_1}}. \tag{30}$$

Band Pass

$$m=\frac{\sqrt{\left(1-\frac{f_1^2}{f_{2\infty}^2}\right)\left(1-\frac{f_2^2}{f_{2\infty}^2}\right)}}{1-\frac{f_1f_2}{f_{2\infty}^2}}. \tag{31}$$

2. *Terminal Losses*

The general terminal losses L_a and L_b are determined by (6) from the absolute values of the terminal factors F_a and F_b, which factors we have assumed apply to the sending and receiving ends, respectively.

That either factor is dependent only upon its own type of termination and not upon its position at the sending or receiving end, can readily be shown. By the reciprocal theorem the product $F_t F_a F_b F_r$ is independent of the direction of current propagation, and from the forms of F_t and F_r the latter are also, whence the product $F_a F_b$ is independent of direction. Since in addition F_a and F_b are independent of each other they cannot depend upon position. This is equivalent to the statement that the ratios E_a/E and I/I_b which any particular termination would give at the sending and receiving ends, respectively, are equal. It will then be sufficient to consider the factor for a given termination at either end, say the receiving end.

The four terminations found practical give terminal losses which are reducible to two, namely, L_m and L_x now to be derived.

Terminal Losses, L_m, with Mid-M-type Terminations. These terminations, already mentioned, are

1, mid-shunt of a mid-series "constant k" equivalent M-type, $(K_{21}(m))$; and

2, mid-series of a mid-shunt "constant k" equivalent M-type, $(K_{12}(m))$.

The relations between the M-type characteristic impedances $K_{21}(m)$ and $K_{12}(m)$, the parameter m, and the variables U_k and V_k of the "constant k" prototype are, from formulae [8] in a previous paper

$$\frac{R}{K_{21}(m)}=\frac{K_{12}(m)}{R}=\frac{\pm\sqrt{1+U_k+iV_k}}{1+(1-m^2)(U_k+iV_k)}. \tag{32}$$

Since $K_{12}(m)\cdot K_{21}(m)=R^2$, $K_{12}(m)$ and $K_{21}(m)$ are inverse networks of impedance product R^2. As either of these terminations is at a mid-point, it forms an end for the wave-filter mid-part and in the terminal factor F_b, arbitrarily chosen, $Z_b=R$ and $I/I_b=1$, leaving

$$F_b=\frac{2\sqrt{W_b R}}{W_b+R}. \tag{33}$$

In this factor the image impedance W_b is either $K_{21}(m)$ or $K_{12}(m)$, depending upon the type of termination. By (32) the factor is the same for both types provided they have the same parameter m, so

[8] The radicals which occur in this and succeeding formulae are proportional to physical impedances with positive resistance components. Hence, in each case the double sign is to be interpreted such as to make the real part of the radical positive.

that we may put for either of them *the single terminal loss* L_m defined by (6) as

$$e^{-L_m}=\left|\frac{2\sqrt{K_{21}(m)R}}{K_{21}(m)+R}=\frac{2\sqrt{K_{12}(m)R}}{K_{12}(m)+R}\right|$$

which upon the substitution of (32) gives

$$L_m=\log_e\left(\frac{1}{2}\left|1\pm\frac{\sqrt{1+U_k+iV_k}}{1+(1-m^2)\,(U_k+iV_k)}\right|\cdot\left|\frac{1+(1-m^2)\,(U_k+iV_k)}{\sqrt{1+U_k+iV_k}}\right|^{\frac{1}{2}}\right).\tag{34}$$

Terminal Losses, L_x, *with x-"constant k" Terminations.* The terminations are

3, x-shunt of the "constant k" wave-filter, (K_{x2}); and

4, x-series of the "constant k" wave-filter, (K_{x1}).

The x-shunt and x-series characteristic impedances, K_{x2} and K_{x1}, are related by the formulae

$$\frac{R}{K_{x2}}=\frac{K_{x1}}{R}=\frac{K_{1k}+(x-.5)z_{1k}}{R}=\pm\sqrt{1+U_k+iV_k}\pm(2x-1)\sqrt{U_k+iV_k},\tag{35}$$

and $$K_{x1}K_{x2}=K_{1k}K_{2k}=z_{1k}z_{2k}=R^2,$$

where K_{2k} and K_{1k} are the mid-shunt and mid-series values corresponding to $x=.5$. With either termination K_{x2} or K_{x1} it is assumed that the mid-part of the wave-filter begins at the mid-point, i.e., at the position corresponding to K_{2k} or K_{1k}, respectively, even when x is less than .5. In the latter case an impedance is theoretically added which is sufficient to "build-out" the wave-filter to the mid-point, and an equal impedance is similarly subtracted from the terminal impedance.

For termination 3, that is K_{x2}, the elements of factor F_b in (6) have the values

$$W_b=K_{2k},$$

$$Z_b=z_{2k}\,R/(z_{2k}+(x-.5)R),\tag{36}$$

and $$I/I_b=z_{2k}/(z_{2k}+(x-.5)R).$$

For termination 4, K_{x1}, they are

$$W_b=K_{1k},$$

$$Z_b=R+(x-.5)z_{1k},\tag{37}$$

and $$I/I_b=1.$$

The substitution of (36) or (37) in F_b gives an identical result, as shown by relations (35), provided x is the same in both. *A single terminal loss L_x may then apply to either*, which is defined from (6) as

$$e^{-L_x}=\left|\frac{2\sqrt{(R^2/K_{2k})R}}{R^2/K_{x2}+R}=\frac{2\sqrt{K_{1k}R}}{K_{x1}+R}\right|,$$

giving by (35)

$$L_x=\log_e\left(\frac{1}{2}\left|1\pm\sqrt{1+U_k+iV_k}\pm(2x-1)\sqrt{U_k+iV_k}\right|\cdot\left|\frac{1}{1+U_k+iV_k}\right|^{\frac{1}{4}}\right). \quad (38)$$

A comparison of (34) and (38) shows that when $m=1$ and $x=.5$, $L_m=L_x$ as should be the case.

3. *Interaction Losses*

The interaction loss defined in (6) is expressible in its general form as

$$L_r=\log_e|1-r_a r_b e^{-2T}|. \quad (39)$$

It depends not only upon the transfer constant T, including both diminution and angular constants, but also upon the complex reflection coefficients, r_a and r_b, at the two ends. That is, it is a function both of the internal structure and of the terminations of the wave-filter. For this reason its determination offers the most complexity of all the three types of losses and, in fact, requires a knowledge of the transfer loss. On the other hand, it is usually the least important part of the total transmission loss and may usually be omitted except at frequencies within a transmitting band and near a critical frequency.

The transfer constant $T=D+iS$ is given by the relations and formulae developed when considering the transfer loss.

The multiplication of the reflection coefficients and the square of the transfer factor is simplified to a problem in addition by expressing each of these coefficients in the exponential form,

$$r_a=e^{-G_a-iH_a},$$

and

$$r_b=e^{-G_b-iH_b}.$$

Then, putting $r_a r_b e^{-2T}=e^{-P-iQ}$,

$$L_r=\tfrac{1}{2}\log_e(1+e^{-2P}-2e^{-P}\cos Q), \quad (40)$$

where $P=G_a+G_b+2D$,

and $Q=H_a+H_b+2S$.

The subscripts, as before, merely refer to the terminations. The G and H expressions which correspond to the reflection coefficient with each of the four particular types of terminations, 1, 2, 3, and 4, follow.

Reflection Coefficients, r_{m2} and r_{m1}, with Mid-M-type Terminations. For termination 1 arbitrarily assumed at b we have $W_b=K_{21}(m)$ and $Z_r=R$. Introducing for this case the subscript $_{m2}$, signifying M-type and mid-shunt, it follows by (6) and (32) that

$$r_{m2}=\frac{K_{21}(m)-R}{K_{21}(m)+R},$$

and its equivalent

$$e^{-G_{m2}-iH_{m2}}=r_{m2}=\frac{1+(1-m^2)(U_k+iV_k)-(\pm\sqrt{1+U_k+iV_k})}{1+(1-m^2)(U_k+iV_k)\ \pm\ \sqrt{1+U_k+iV_k}}. \quad (41)$$

With termination 2, $W_b=K_{12}(m)$ and $Z_b=R$, so that by (32) the corresponding coefficient r_{m1} becomes

$$r_{m1}=-r_{m2},$$

or (42)

$$e^{-G_{m1}-iH_{m1}}=-e^{-G_{m2}-iH_{m2}}.$$

Since $-1=e^{-i\pi}$,

$$G_{m1}=G_{m2},$$

and (43)

$$H_{m1}=H_{m2}+\pi.$$

Reflection Coefficients, r_{x2} and r_{x1}, with x-"constant k" Terminations. In the case of the x-shunt termination 3, K_{x2}, relations (36) give

$$r_{x2}=\frac{K_{2k}-z_{2k}R/(z_{2k}+(x-.5)R)}{K_{2k}+z_{2k}R/(z_{2k}+(x-.5)R)}.$$

Introducing (35) this is

$$e^{-G_{x2}-iH_{x2}}=r_{x2}=\frac{1\pm(2x-1)\sqrt{U_k+iV_k}-(\pm\sqrt{1+U_k+iV_k})}{1\pm(2x-1)\sqrt{U_k+iV_k}\pm\sqrt{1+U_k+iV_k}}. \quad (44)$$

The x-series termination 4, K_{x1}, has a coefficient r_{x1} determined by (37) which is related to r_{x2} through (35) as

$$r_{x1}=-r_{x2}. \quad (45)$$

It follows from the corresponding exponential expressions that

$$G_{x1}=G_{x2},$$

and (46)

$$H_{x1}=H_{x2}+\pi.$$

Hence, the two members of each pair of reflection coefficients, r_{m2}, r_{m1}, *and* r_{x2}, r_{x1}, *differ only in sign so that their G's are the same but their H's differ by* π.

4. *Wave-Filter Structures Having Equivalent Transmission Losses*

There are six groups of possible wave-filter networks involving the four terminations above, each group of which is made up of pairs having equivalent current ratios $2RI/E$ and hence equivalent transmission losses. By (5) this means that the members of such a pair have products for their four factors, $F_tF_aF_bF_r$, which are equal. It may readily be shown from preceding relations that these groups, represented symbolically by brackets enclosing the transfer constants of their mid-parts and the terminations, are the following:

$$\begin{aligned} &(a)\ [T,\ K_{21}(m),\ K_{21}(m')] = [T,\ K_{12}(m),\ K_{12}(m')],\\ &(b)\ [T,\ K_{21}(m),\ K_{12}(m')] = [T,\ K_{12}(m),\ K_{21}(m')],\\ &(c)\ [T,\ K_{21}(m),\ K_{x2}] = [T,\ K_{12}(m),\ K_{x1}],\\ &(d)\ [T,\ K_{21}(m),\ K_{x1}] = [T,\ K_{12}(m),\ K_{x2}],\\ &(e)\ [T,\ K_{x2},\ K_{x'2}] = [T,\ K_{x1},\ K_{x'1}],\\ &(f)\ [T,\ K_{x2},\ K_{x'1}] = [T,\ K_{x1},\ K_{x'2}]. \end{aligned} \qquad (47)$$

This symbolic representation in (*c*), for example, means that a composite wave-filter whose mid-part has a transfer constant, T, and whose terminations are those designated by $K_{21}(m)$ and K_{x2}, will give the same current ratio $2RI/E$ as another wave-filter whose mid-part has the same transfer constant, T, but whose terminations are those designated by $K_{12}(m)$ and K_{x1} where m and x are respectively the same in both networks.

III. Charts for Determining Transmission Losses

The accompanying charts apply to the three groups of transmission losses, transfer, terminal, and interaction, and are derived from the general formulae already given. The curves represent constant parameter loci for A, cB, L_m, L_x, G_{m2}, cH_{m2}, G_{x2}, cH_{x2}, and L_r as functions of several variables and include the most practical range; where further extension is required the original formulae may be consulted. The U and V variables for the ladder type of recurrent network (or its equivalent) which form the basis of this chart calculation method are to be found as a function of frequency, in the general case from formula (9),

$$z_1/4z_2 = U + iV,$$

and in the lower class "constant k" and M-type wave-filters from formulae (21) to (31). Owing to the large number of intermediate equations which it was necessary to obtain before direct computations could suitably be made for the charts, these equations will not be given here, but only a brief designation of the resulting charts together with the approximations involved, if any.

The units employed throughout are the *attenuation unit* and the *radian*. The former unit applies to A, D, L_m, L_x, G_{m2}, G_{x2}, P, L_r and L, and the latter unit to B, S, H_{m2}, H_{x2} and Q.

Transfer Loss

This is determined through the propagation constant, $\Gamma = A + iB$.

Charts 1, 2, *and* 3.—*A and cB in and about transmitting band*;
$c = \pm 1$ has the sign of V.

Chart 4.—*A in attenuating band*;
V^2 negligible compared with $(U + U^2) > O$.

Chart 5.—*A at maximum attenuation*;
$(U + U^2)$ negligible compared with V^2.

Terminal Losses, L_m and L_x

Chart 6.—*L_m in transmitting band*;
V_k neglected.

Chart 7.—*L_m at critical frequency*;
$U_k = -1$.

Chart 8.—*L_m in attenuating band*;
V_k neglected.

Chart 9.—*L_m at maximum attenuation of M-type*;
$$U_k = -\frac{1}{1-m^2}.$$

Chart 10.—*L_x in transmitting band*;
V_k neglected.

Chart 11.—*L_x at critical frequency*;
$U_k = -1$.

Chart 12.—*L_x in attenuating band*;
V_k neglected.

Reflection Coefficients

Note that

$$G_{m1}=G_{m2},$$

and $$H_{m1}=H_{m2}+\pi;$$

also that

$$G_{x1}=G_{x2},$$

and $$H_{x1}=H_{x2}+\pi.$$

Chart 13.—*G_{m2} and H_{m2} in transmitting band;*
V_k neglected.

Chart 14.—*G_{m2} and cH_{m2} at critical frequency;*
$U_k=-1$ and $c=\pm1$ has the sign of V_k.

Chart 15.—*G_{m2} and cH_{m2} in attenuating band;*
V_k neglected.

Chart 16.—*G_{x2} and cH_{x2} in transmitting band;*
V_k neglected and $c=\pm1$ has the sign of V_k.

Chart 17.—*G_{x2} and cH_{x2} at critical frequency;*
$U_k=-1$.

Chart 18.—*G_{x2} and cH_{x2} in attenuating band;*
V_k neglected.

Interaction Loss, L_r

Note that $T=D+iS$ = transfer constant of mid-part of wave-filter:

$$P=G_a+G_b+2D,$$

and $$Q=H_a+H_b+2S,$$

where a and b refer to the terminations.

Chart 19.—*L_r as a function of P and Q.*

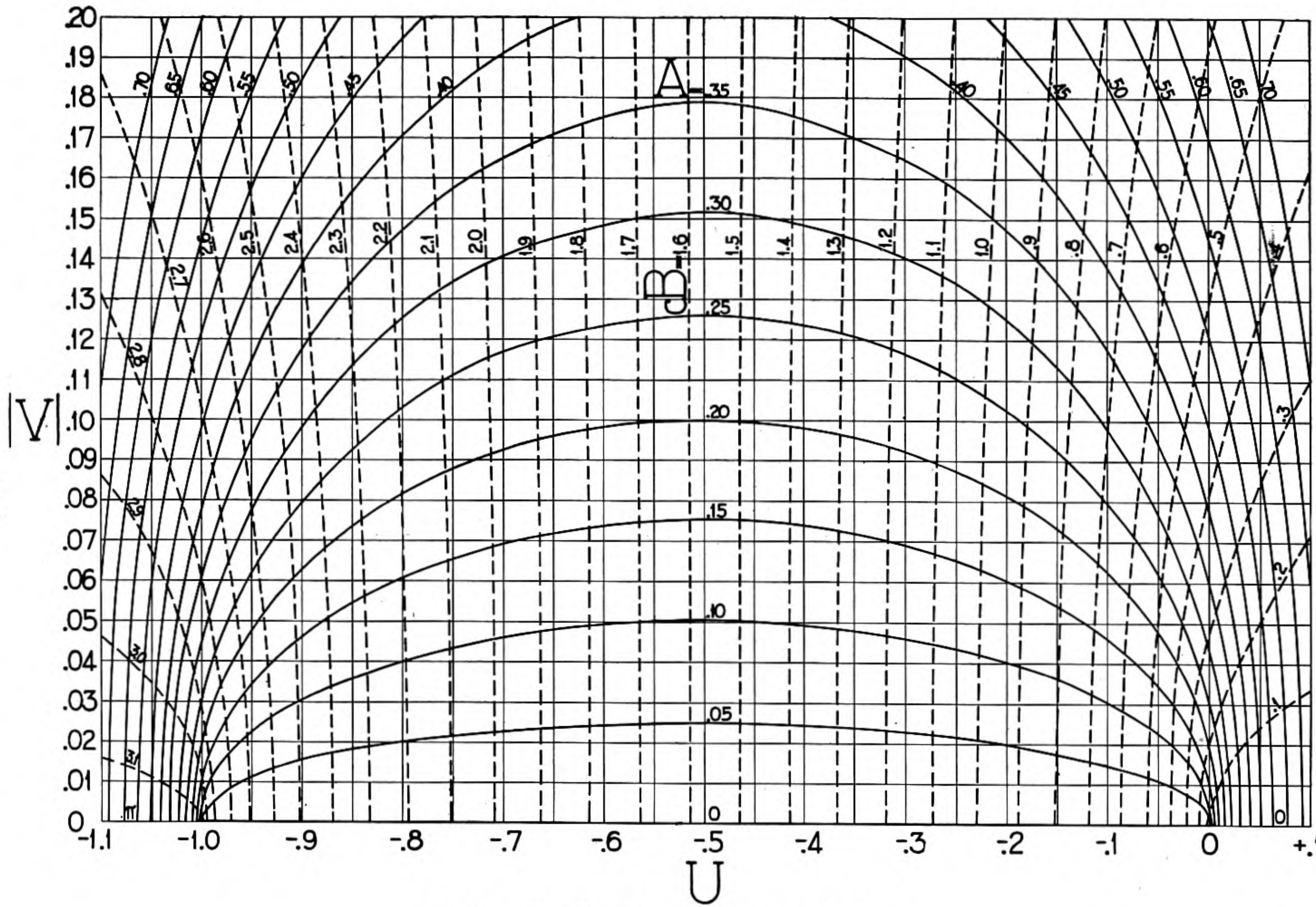


Chart 1.—A and cB in and about transmitting band;
$c = \pm 1$ has the sign of V.

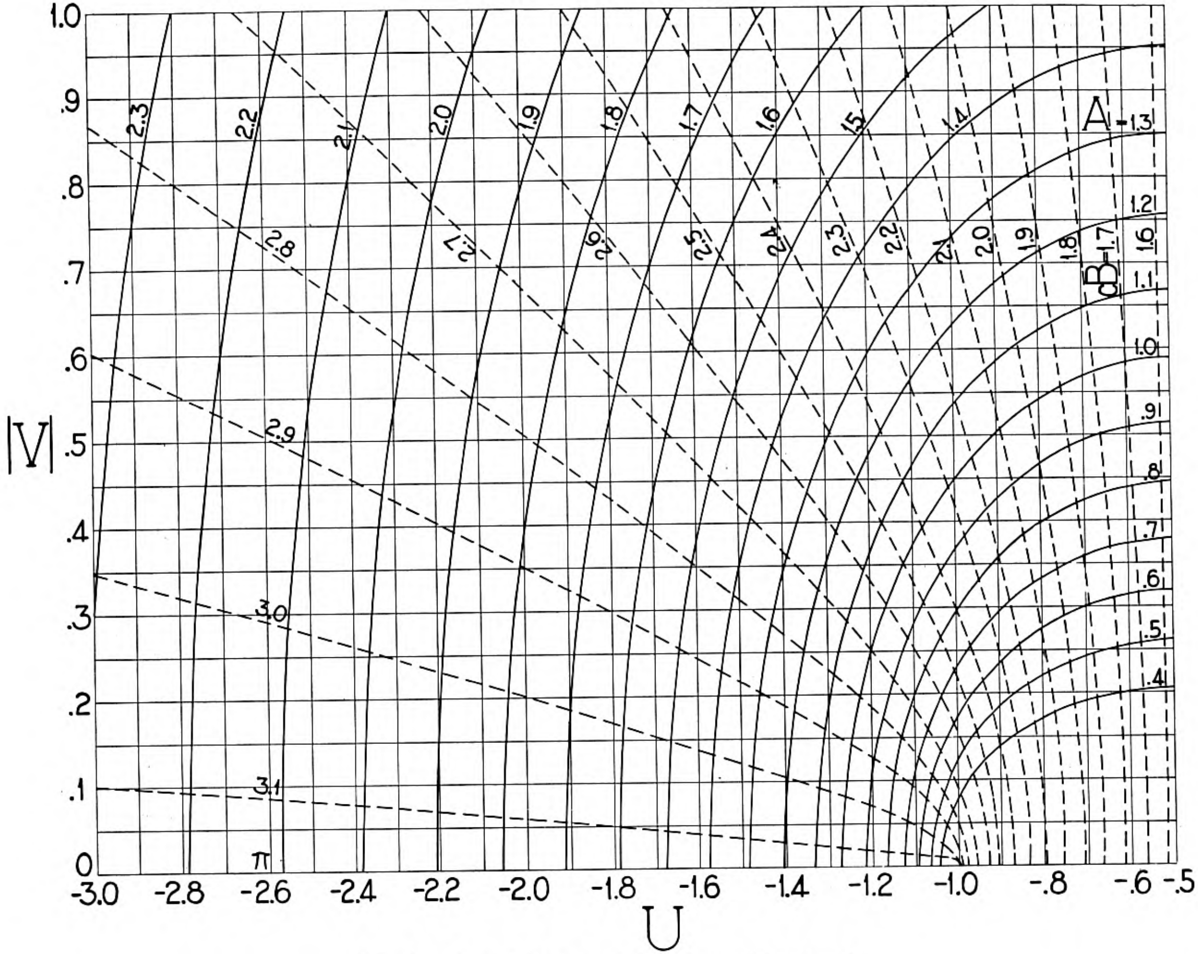


Chart 2.—A and cB in and about transmitting band;
$c = \pm 1$ has the sign of V.

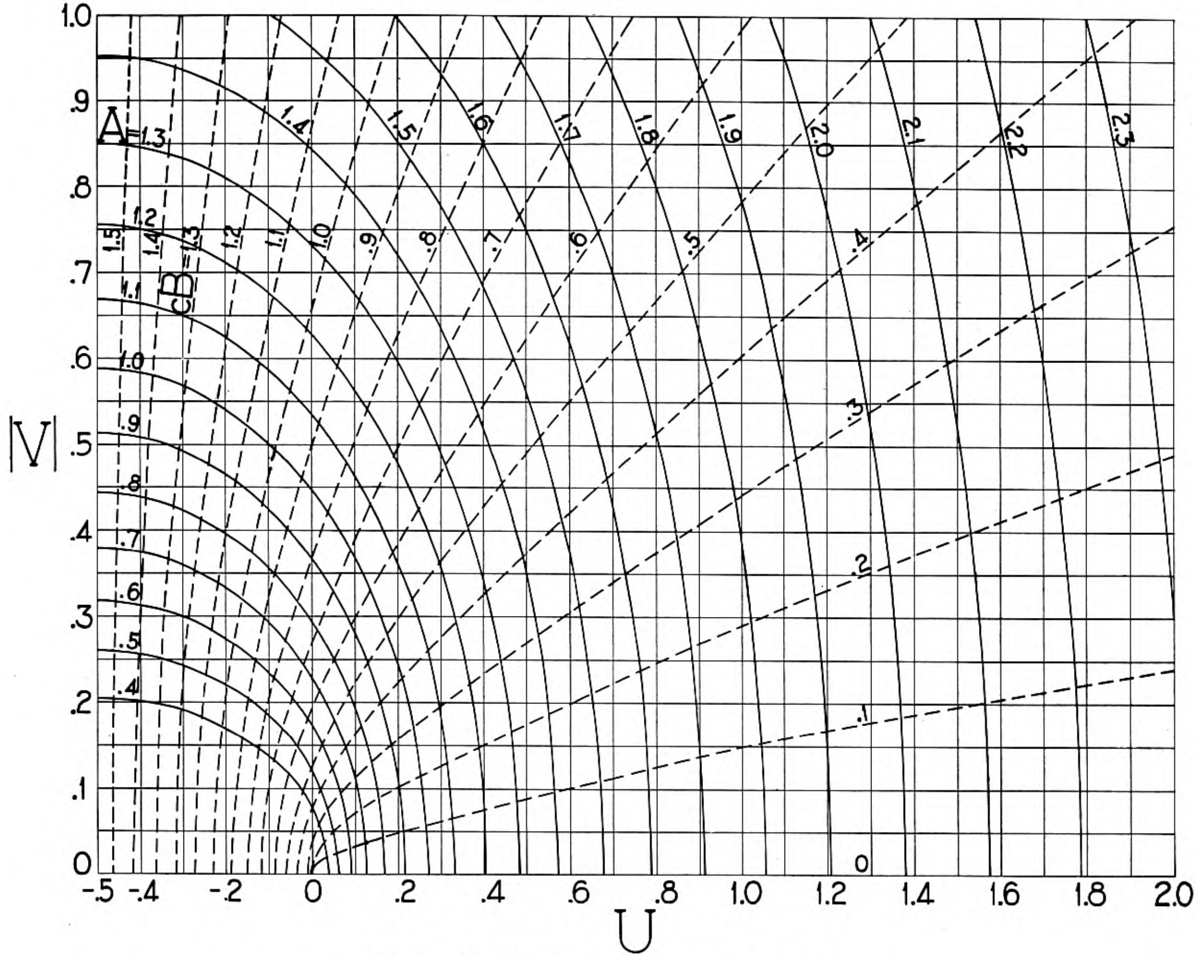


Chart 3.—A and cB in and about transmitting band;
$c = \pm 1$ has the sign of V.

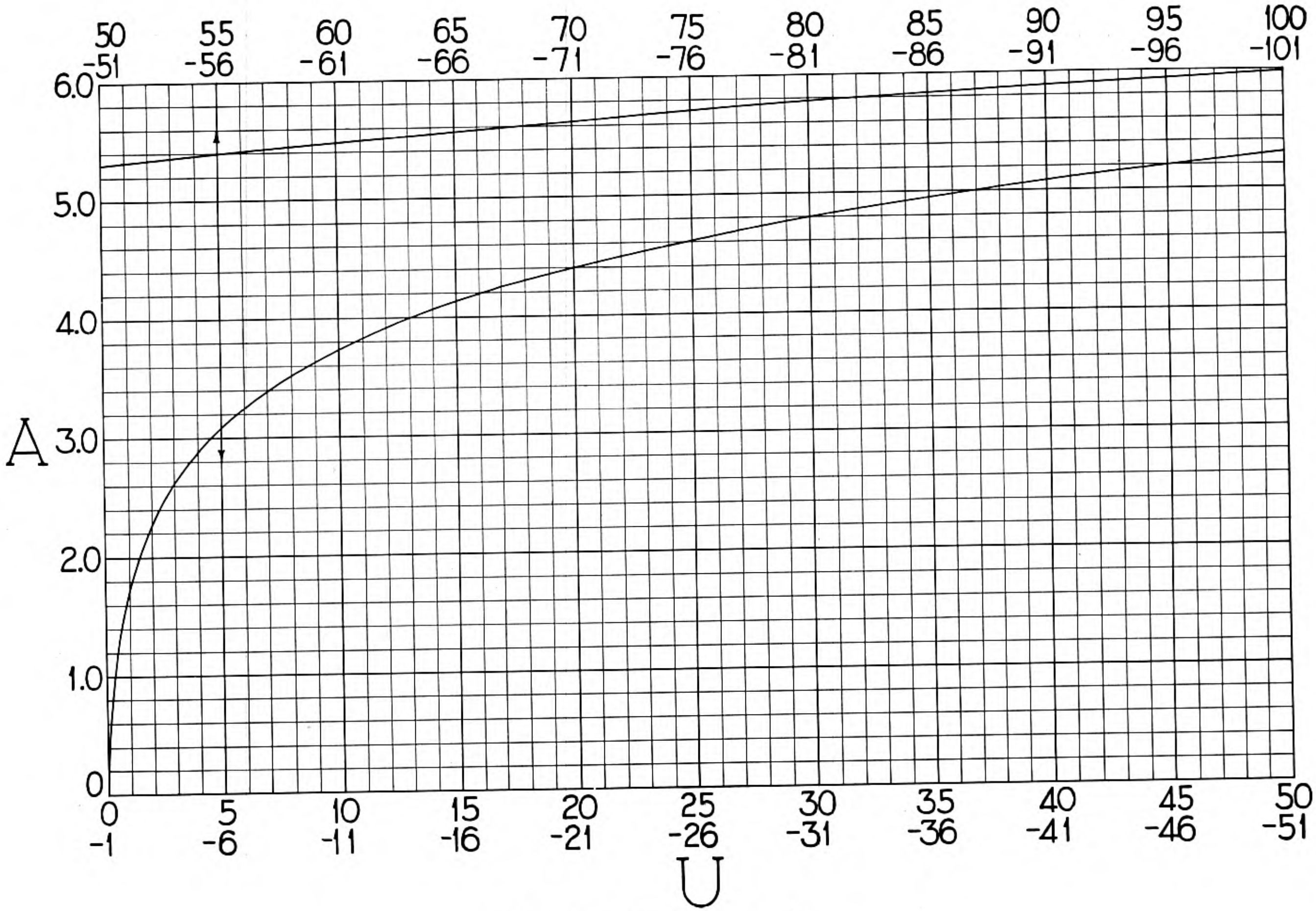


Chart 4.—*A in attenuating band;*
V^2 negligible compared with $(U+U^2)>O$.

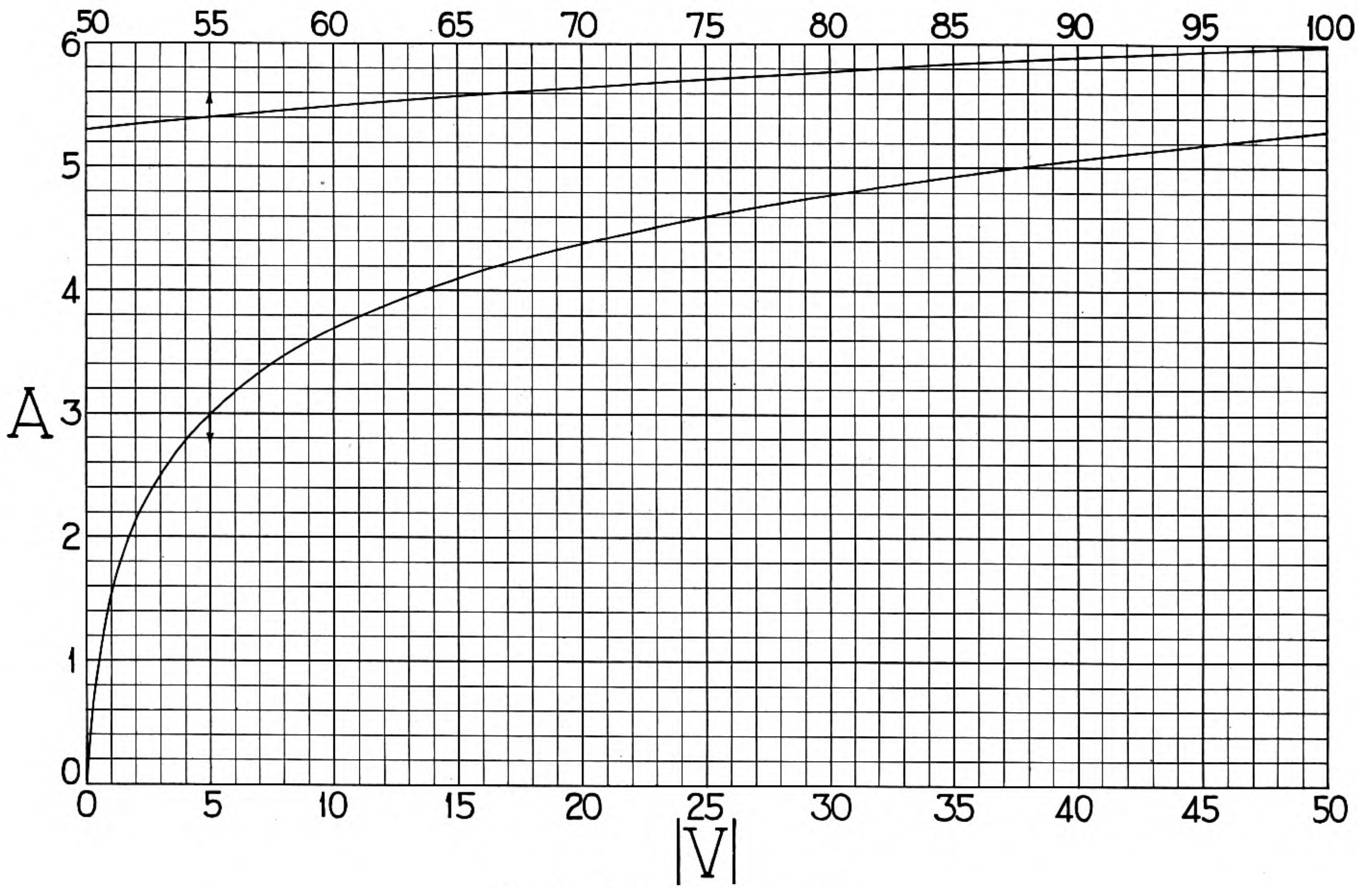


Chart 5.—*A at maximum attenuation;*
$(U+U^2)$ negligible compared with V^2.

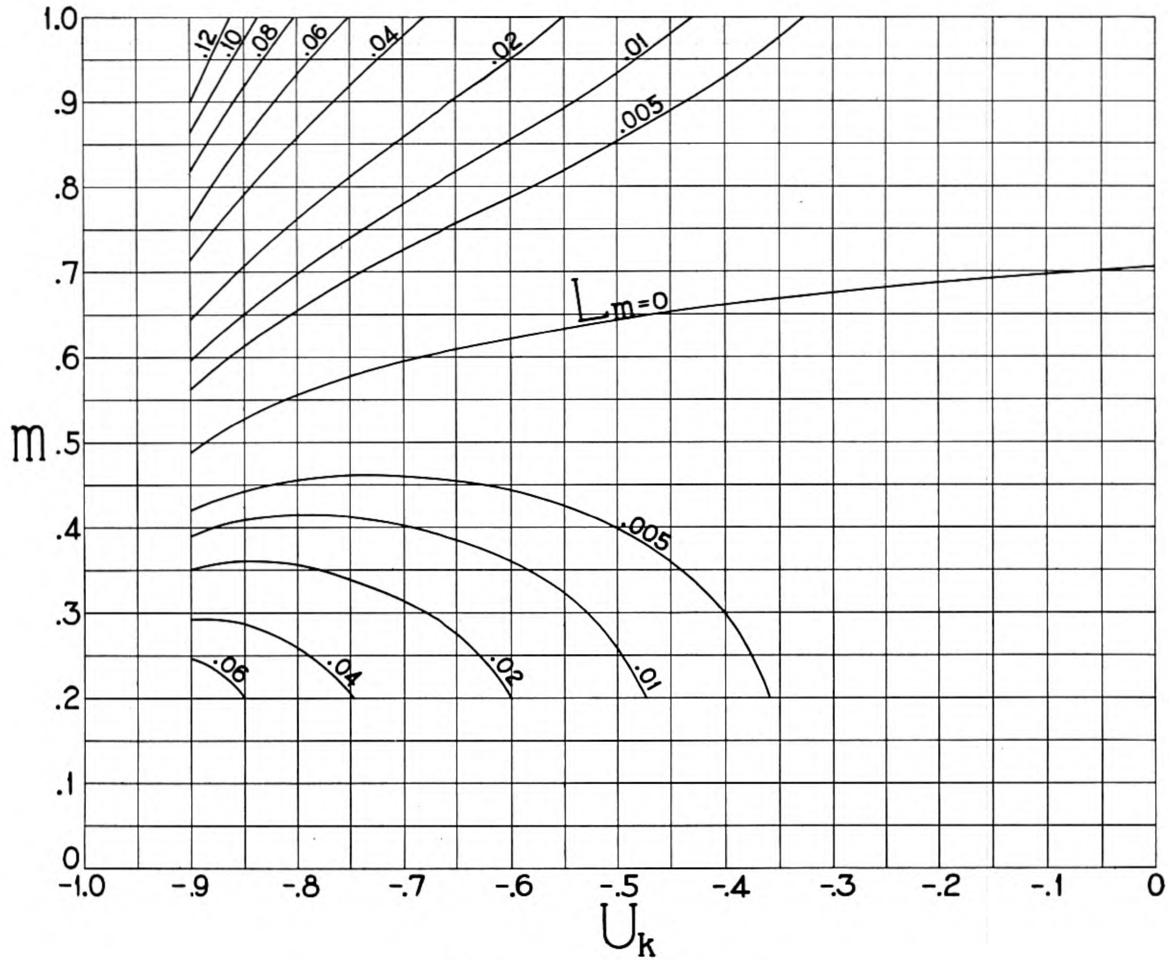


Chart 6.—L_m in transmitting band;
$V_k^{\prime}$ neglected.

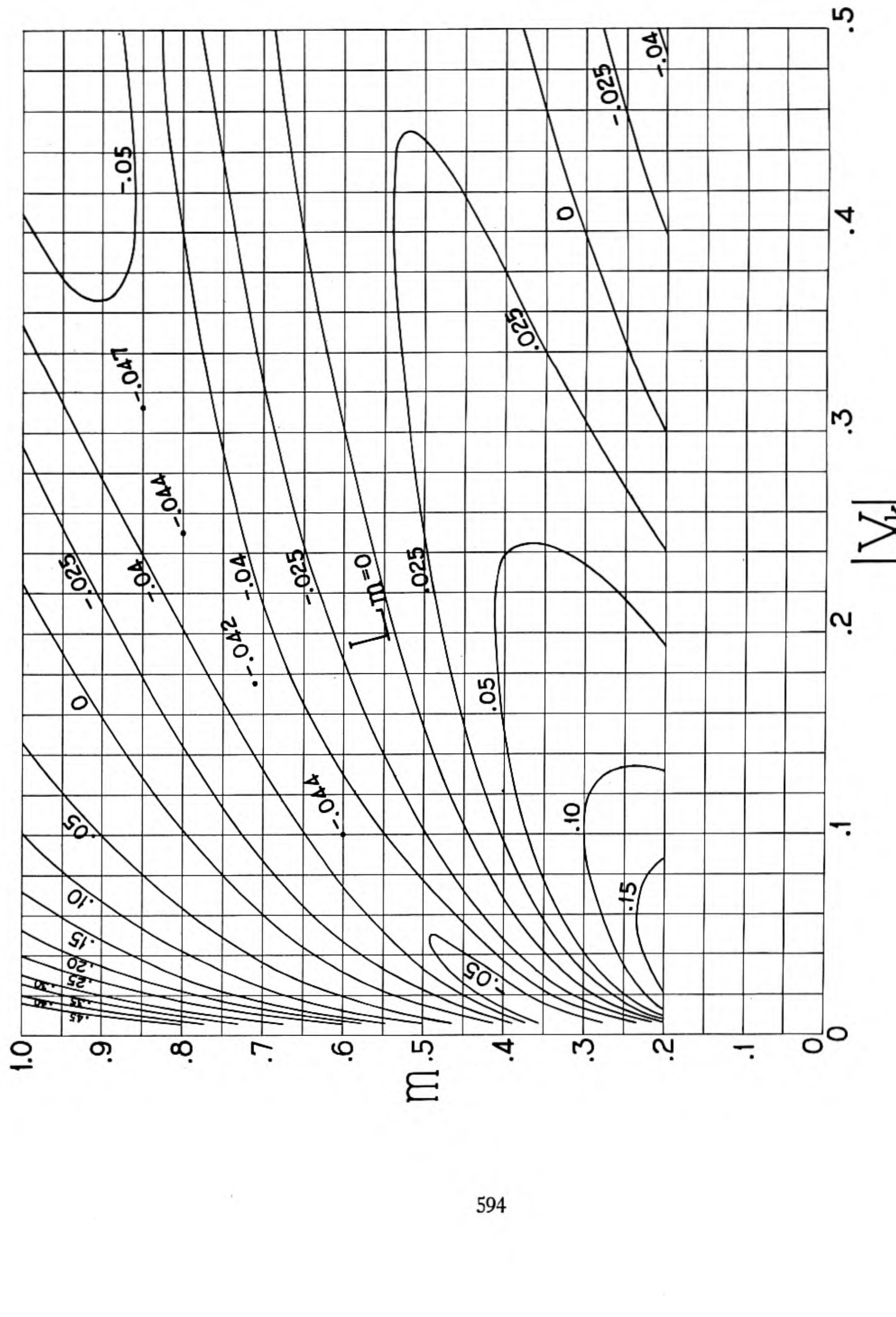


Chart 7.—L_m at critical frequency; $U_k=-1$.

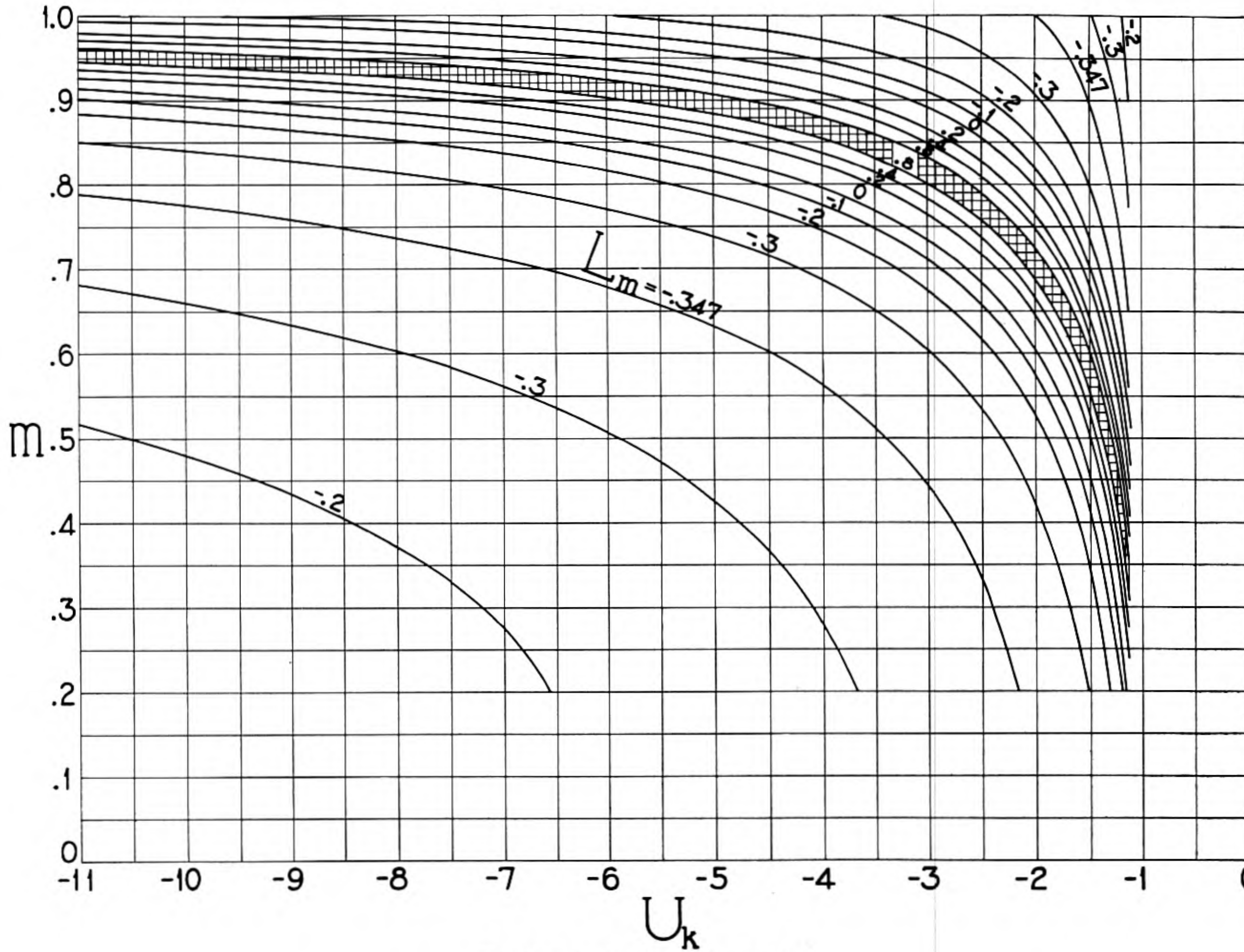


Chart 8.—L_m in attenuating band;
V_k neglected.

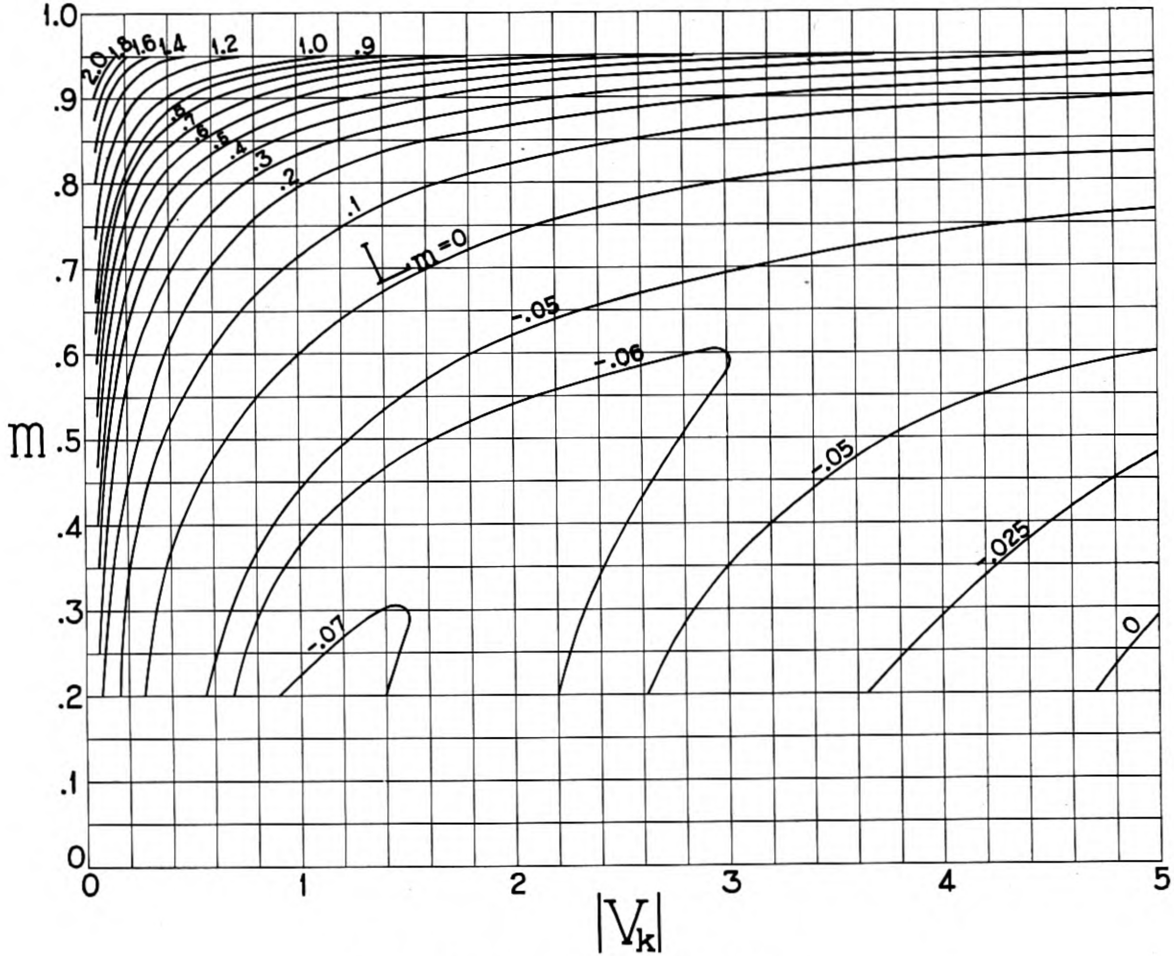


Chart 9.—L_m at maximum attenuation of M-type;

$$U_k = -\frac{1}{1-m^2}.$$

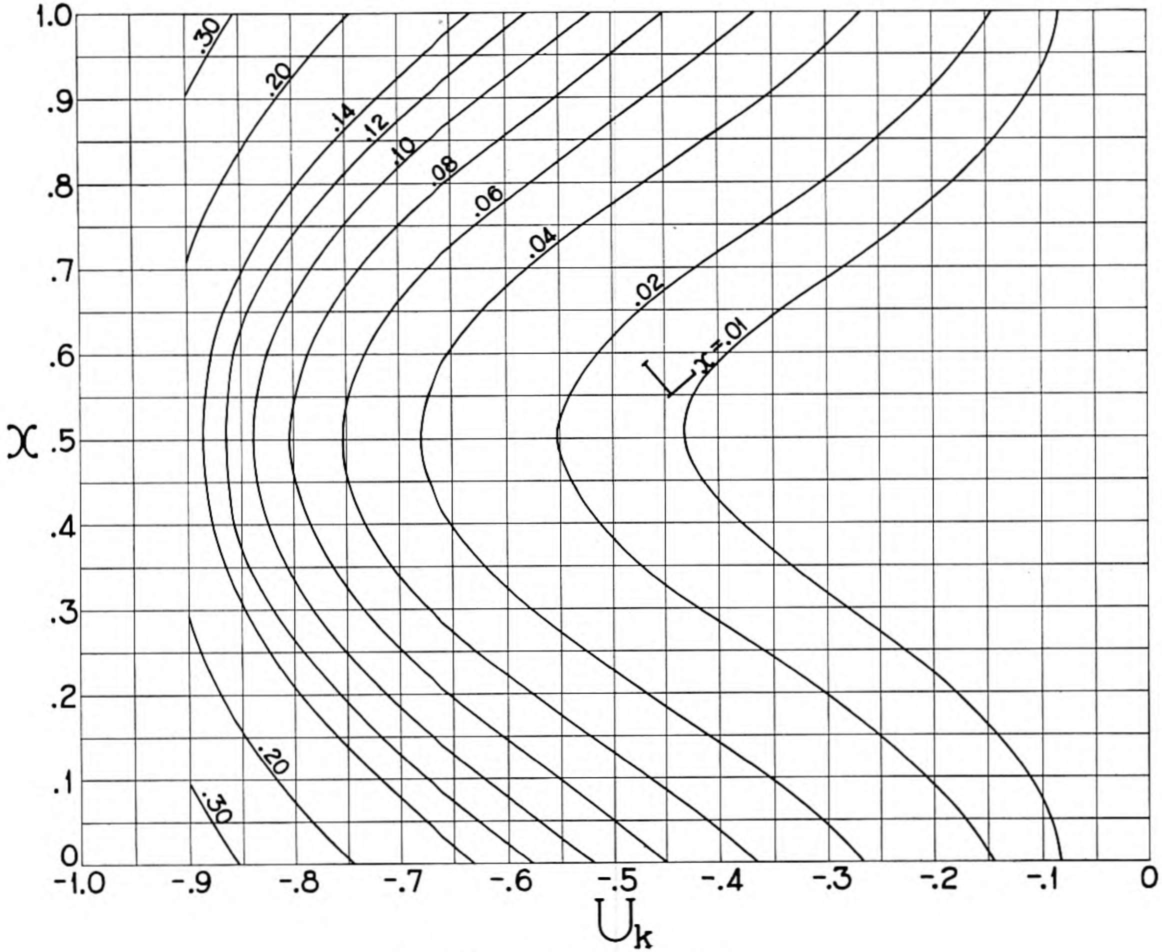


Chart 10.—L_x *in transmitting band;*
V_k neglected.

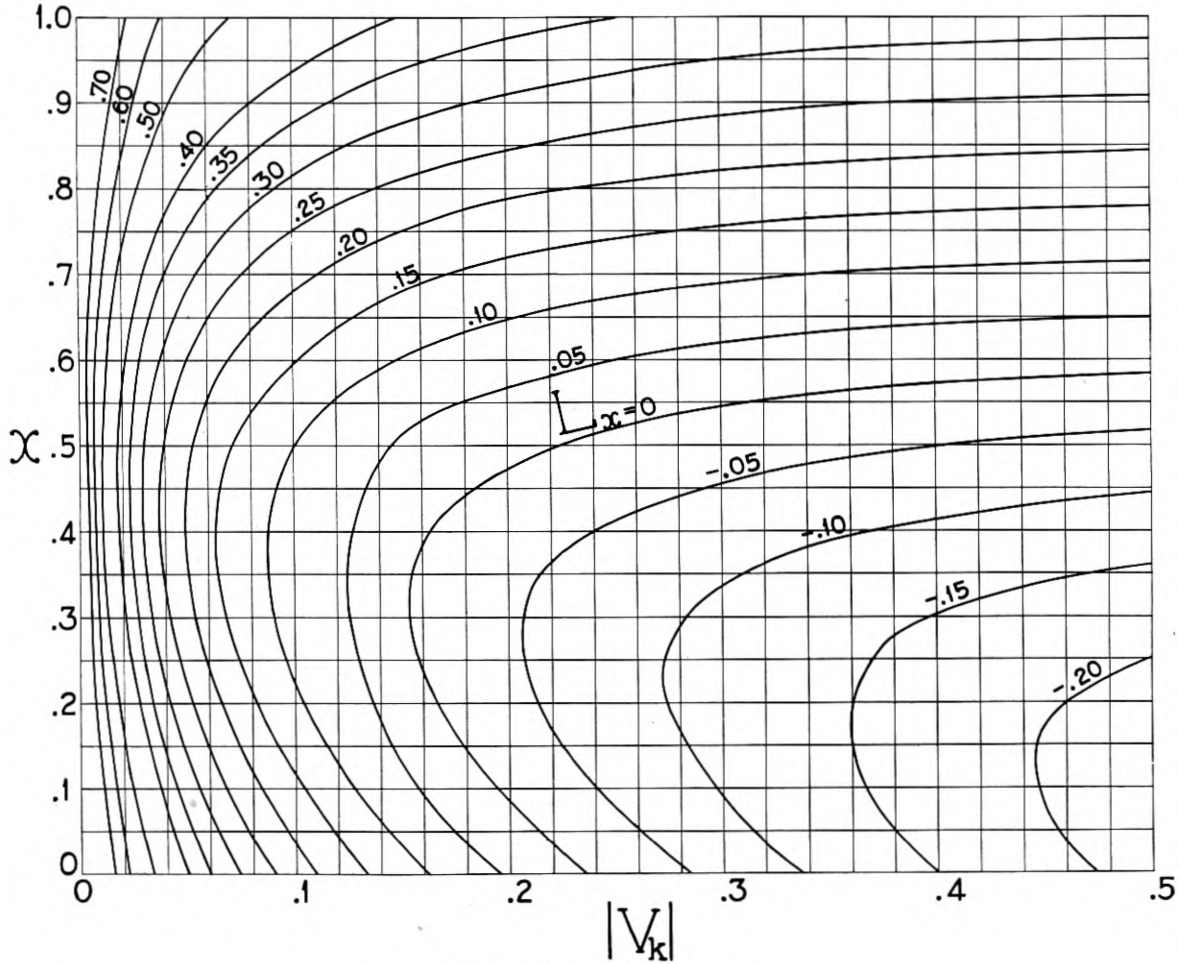


Chart 11.—L_x *at critical frequency;*
$U_k = -1$.

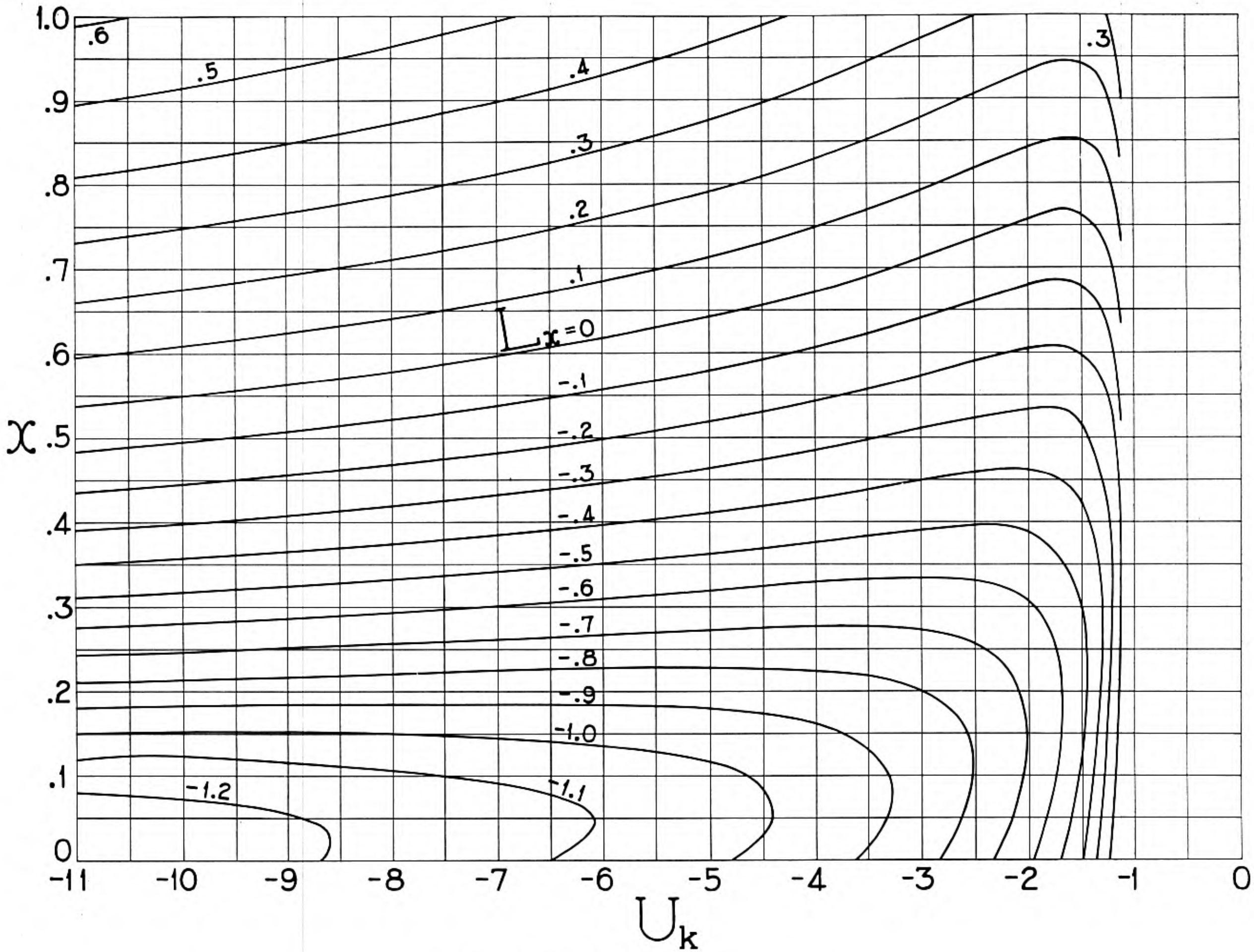


Chart 12.—L_x in attenuating band;
V_k neglected.

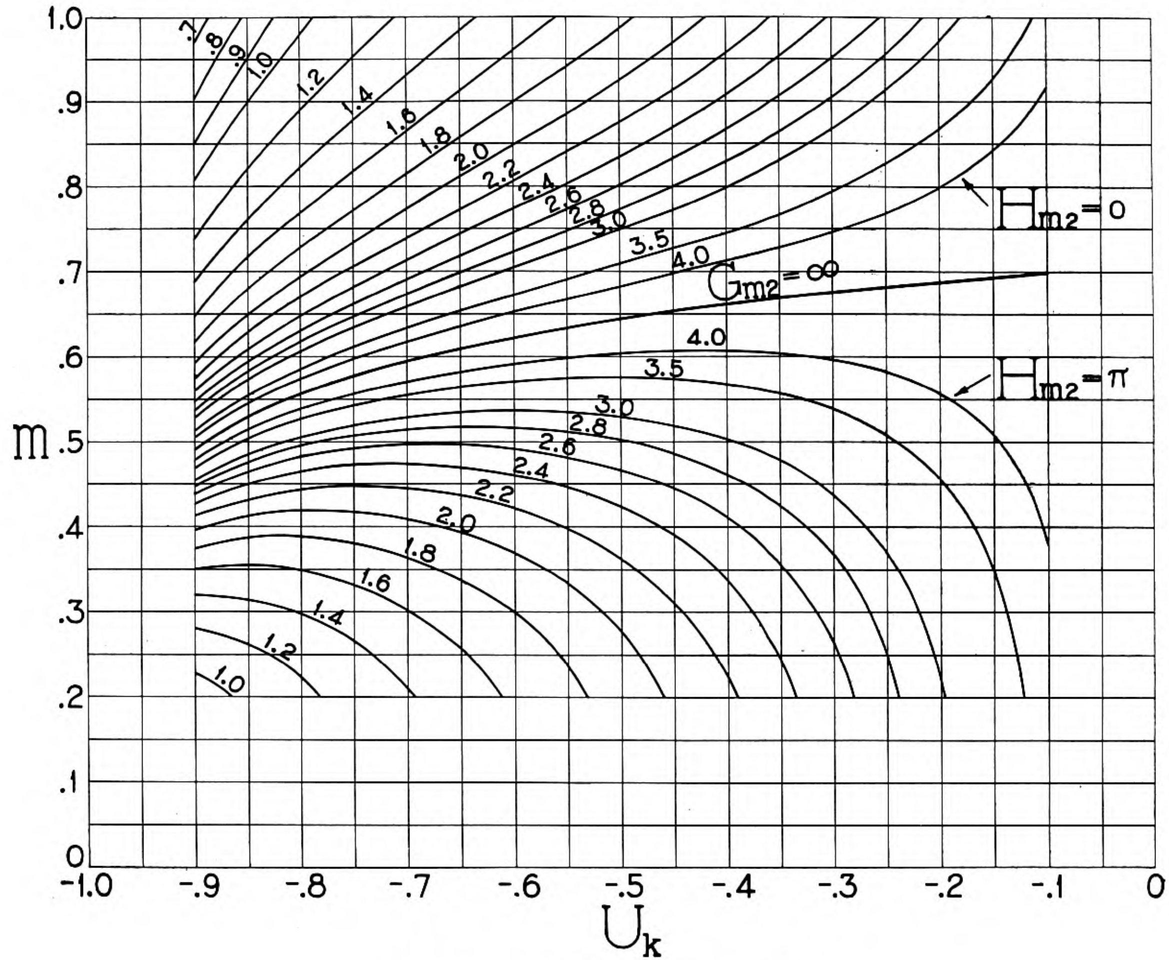


Chart 13.—G_{m2} *and* H_{m2} *in transmitting band;* V_k neglected.

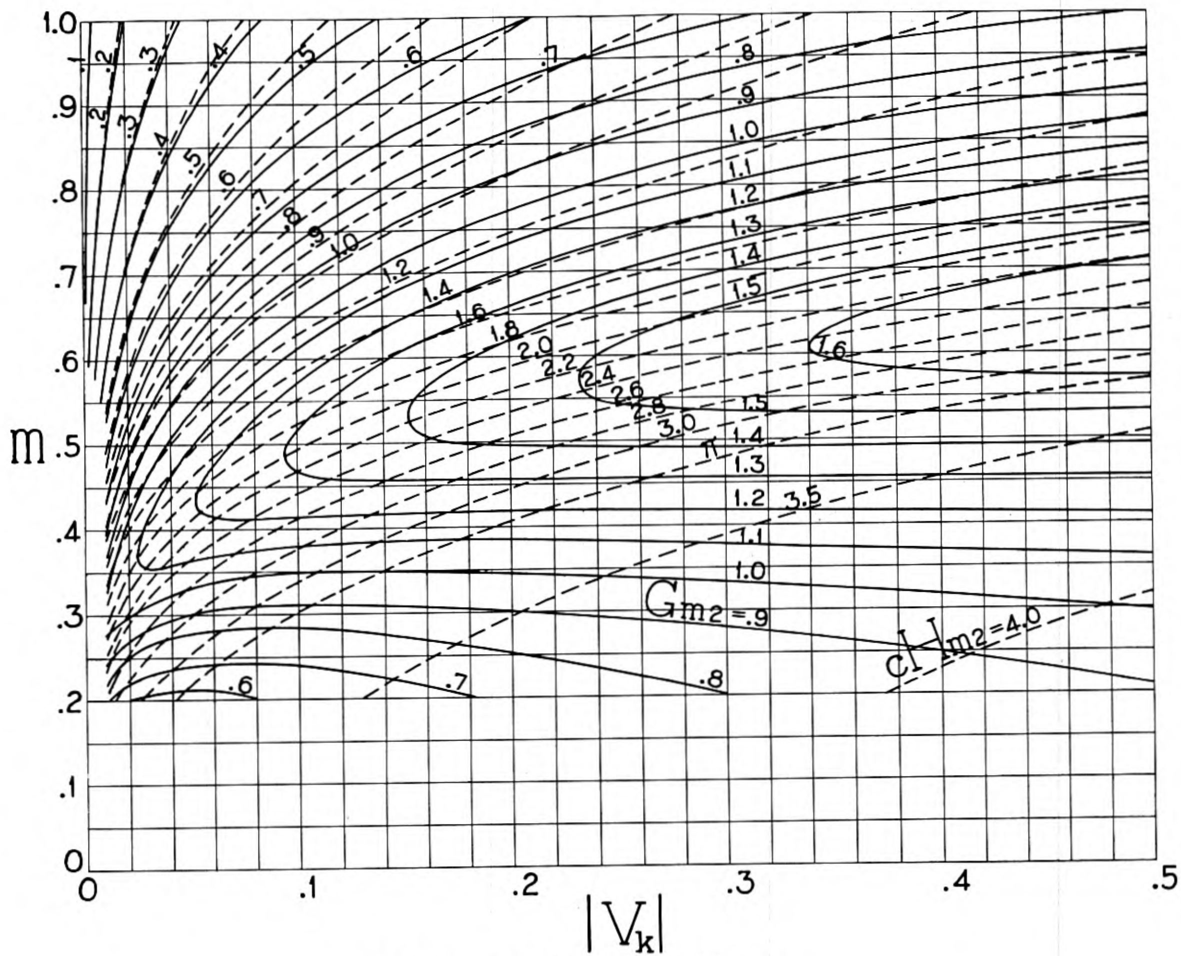


Chart 14.—G_{m2} *and* cH_{m2} *at critical frequency;*
$U_k = -1$ and $c = \pm 1$ has the signo f V_k.

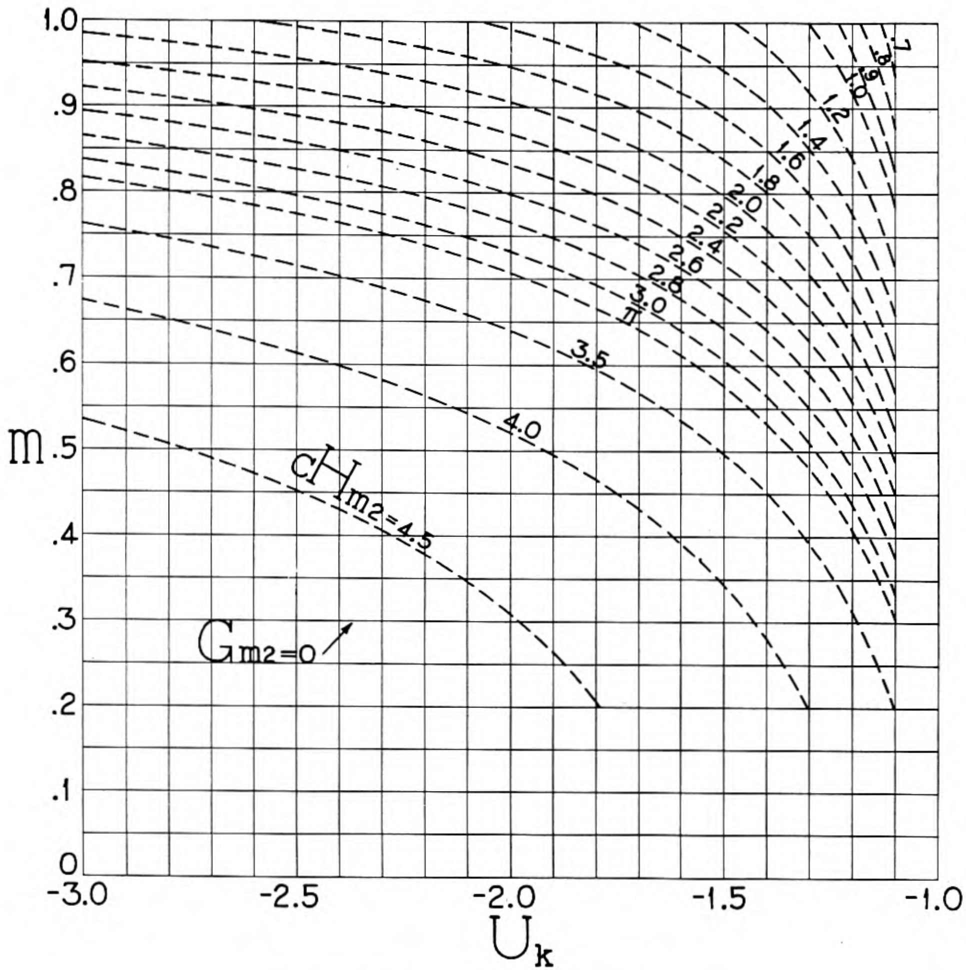


Chart 15.—G_{m2} *and* cH_{m2} *in attenuating band;* V_k neglected.

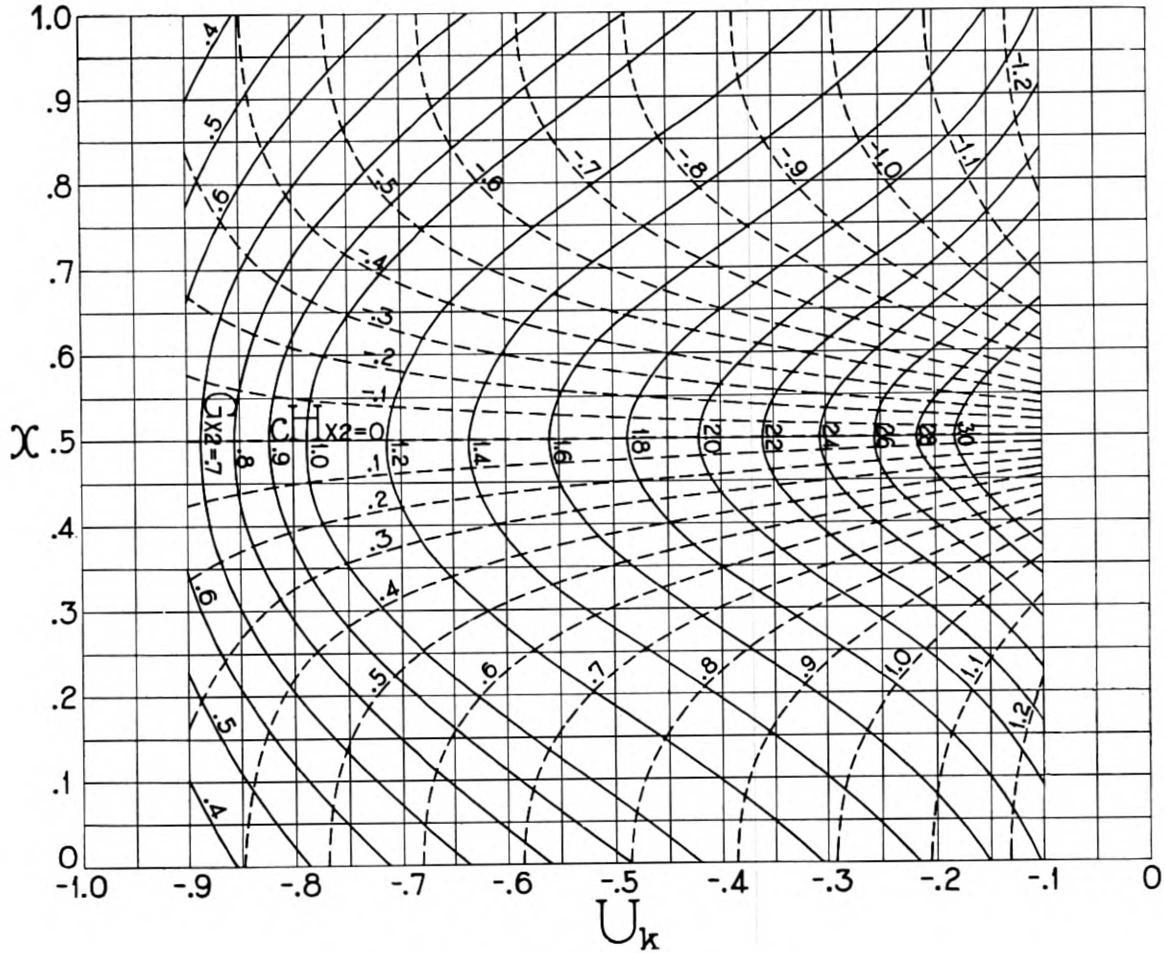


Chart 16.—G_{x2} and cH_{x2} in transmitting band;
V_k neglected and $c = \pm 1$ has the sign of V_k.

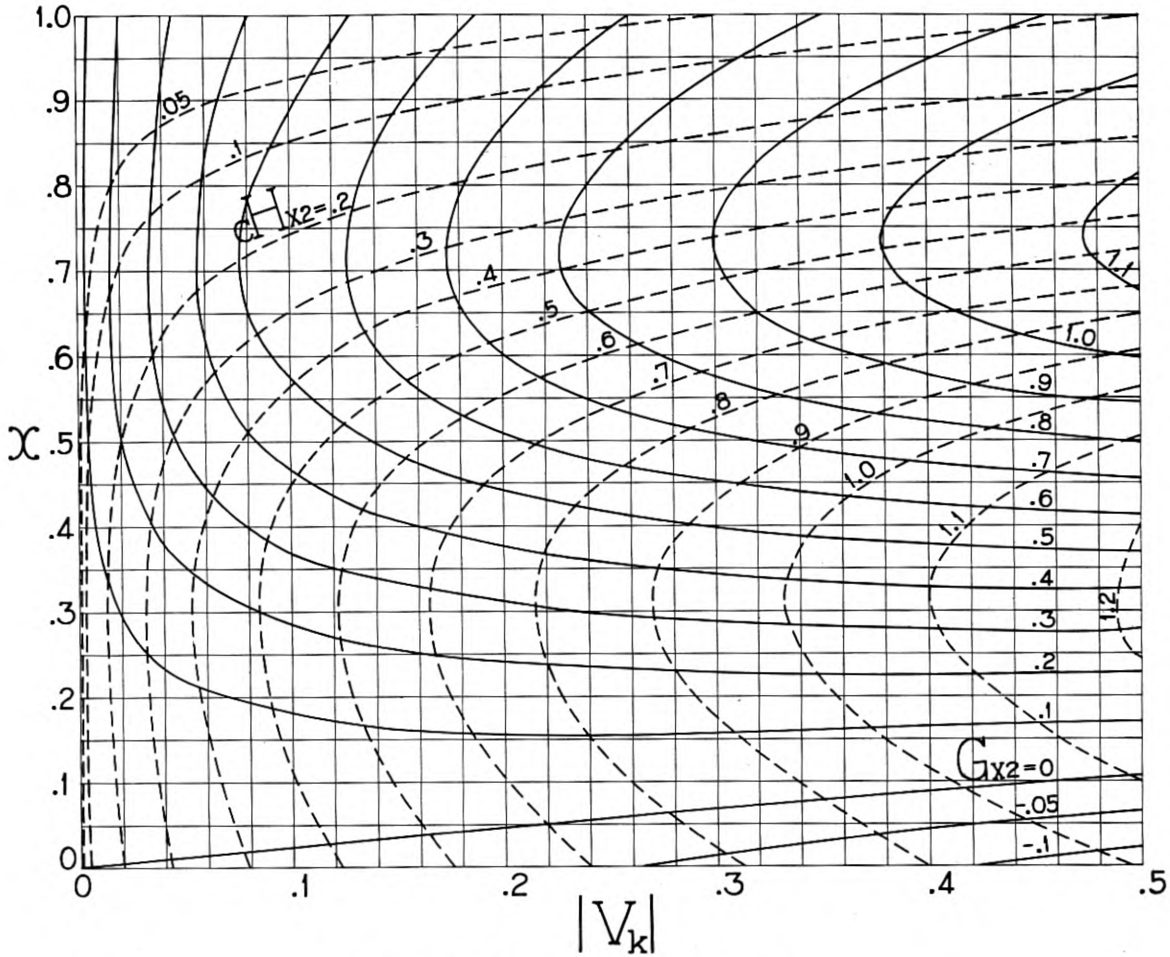


Chart 17.—G_{x2} *and* cH_{x2} *at critical frequency;*
$U_k = -1.$

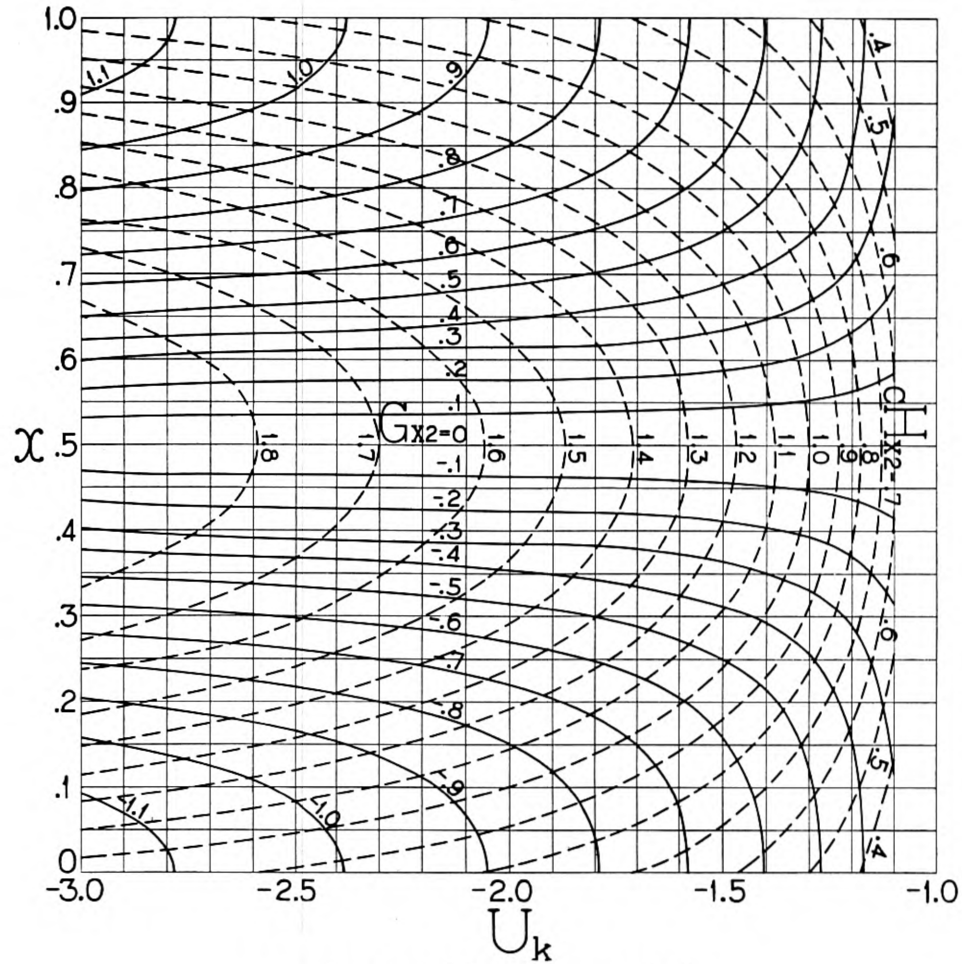


Chart 18.—G_{x2} and cH_{x2} in attenuating band; V_k neglected.

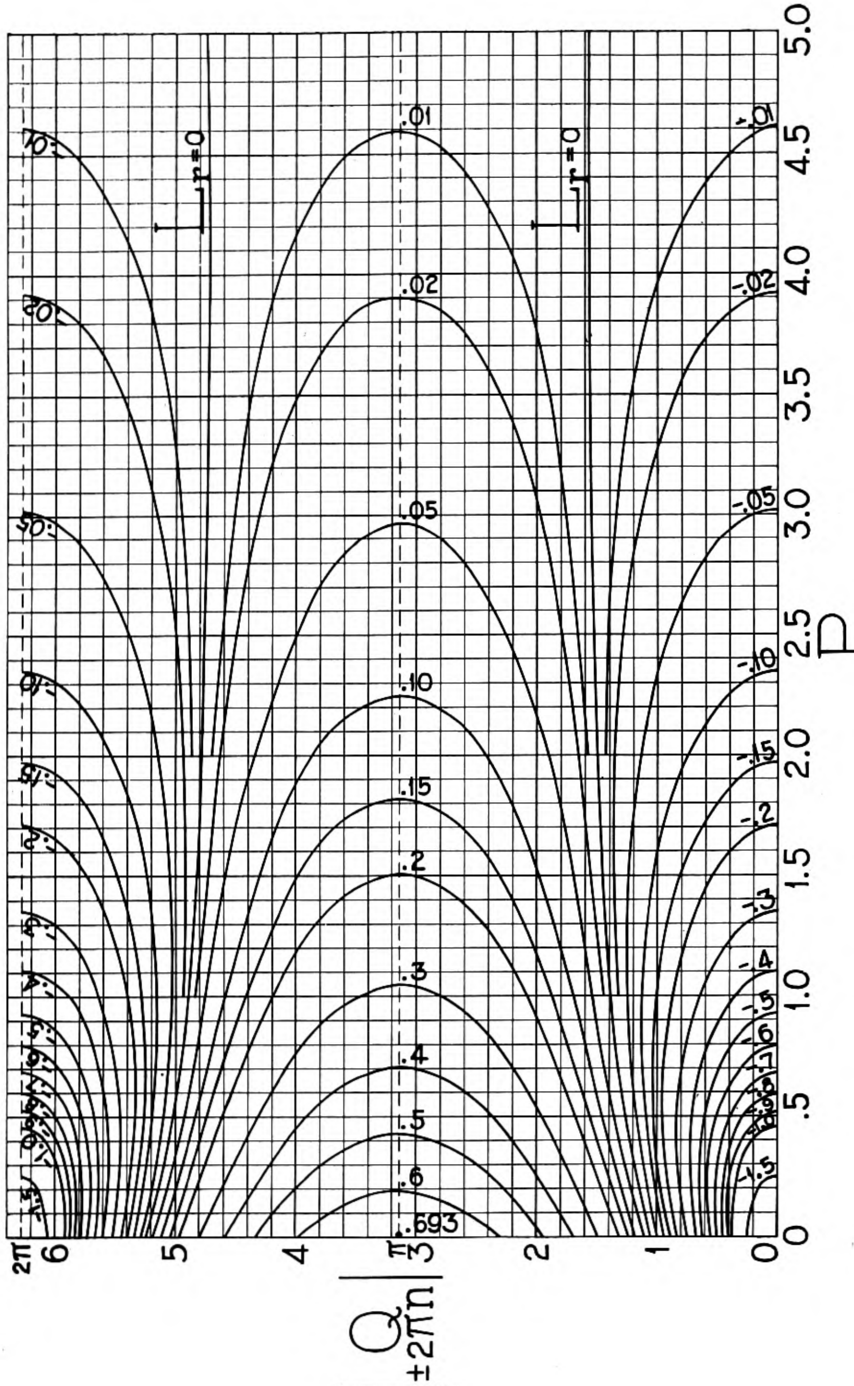


Chart 19.—L_r *as a function of P and Q.*

IV. Illustrative Application of the Method

In the following illustration a small number of band pass wave-filter sections having different characteristics is chosen purposely so as to allow an appreciable interaction factor and the use of all the charts.

The mid-part of the composite wave-filter is made up of one mid-series section of type VI_1 and one mid-half section of M-type IV_1, the designations being those of a previous paper. The termination at one end is made K_{x1} by adding $(x-.5)z_{1k}$ in series with type VI_1 and at the other is $K_{21}(m)$, as is diagrammatically represented at the top of Fig. 4. The values of all the parameters were chosen as follows:

$R=600$ ohms, $x=.80$,

$f_1=4{,}000\sim$, M-type, $f_{2\infty}=8{,}000\sim$,

$f_2=7{,}000\sim$,

and $d=.01$ (assumed constant for computation purposes).

With these values the magnitudes and locations of inductances and capacities are as shown in the center of Fig. 4, where the series impedance parts have been merged together.

The variables U_k and V_k for the "constant k" band pass wave-filter as well as U_m, V_m, and m of the M-type are given by formulae (24), (26), and (31). In the 3-element type VI_1

$$U_1+iV_1=\frac{1-(f/f_1)^2}{(f_2/f_1)^2-1}+id\,\frac{(f/f_1)^2}{(f_2/f_1)^2-1}. \tag{48}$$

These variables have been computed in the present case for frequencies on both sides of the transmitting band and are tabulated below. The other tables including that of transmission losses are based upon this table and the charts.

The next to the last, and the last columns give the total transmission losses as obtained by this chart method and by direct network computation, respectively. Comparison shows that there is a very satisfactory agreement between them, the differences at all frequencies being negligible in practice. The greatest differences of approximately .05 attenuation units at frequencies 3750 and 7500 cycles per second, just outside the transmitting band, can readily be explained as due to the omission of dissipation in the two terminal loss factors and the reflection coefficients. The transmission loss is shown graphically at the bottom of Fig. 4.

It is believed that the use of this chart method will result in considerable time economy with calculations of this kind.

TABLE I

U and V Variables

f Cycles/sec.	"Constant k"		Type VI_1		M-type, $m=.7454$	
	U_k	V_k	U_1	V_1	U_m	V_m
1000	−81.0	−.870	.455	.0003	1.29	− .0004
1500	−32.7	−.385	.417	.0007	1.34	− .0012
2000	−16.0	−.213	.364	.0012	1.45	− .0032
2500	− 8.41	−.132	.296	.0019	1.71	− .0098
3000	− 4.46	−.0868	.212	.0027	2.53	− .050
3250	− 3.20	−.0707	.165	.0032	4.21	− .220
3500	− 2.25	−.0575	.114	.0037	.36	−48.9
3750	− 1.53	−.0463	.0586	.0042	−2.67	− .253
4000	− 1.00	−.0367	0	.0049	− .997	− .0659
4250	− .607	−.0282	− .0625	.0055	− .461	− .0293
4500	− .329	−.0205	− .129	.0061	− .214	− .0156
5292	.00031	0	− .364	.0085	.00017	0
6500	− .534	.0263	− .796	.0128	− .389	.0252
6750	− .752	.0315	− .897	.0138	− .627	.0395
7000	− 1.00	.0367	−1.00	.0148	− .998	.0659
7500	− 1.58	.0470	−1.22	.0170	−2.90	.290
8000	− 2.25	.0575	−1.46	.0194	.85	48.9
8500	− 3.01	.0682	−1.70	.0219	4.92	.329
9000	− 3.85	.0792	−1.97	.0246	3.00	.0866
10000	− 5.76	.102	−2.55	.0303	2.05	.0234
11000	− 7.94	.127	−3.18	.0367	1.75	.0110
12000	−10.4	.154	−3.88	.0436	1.60	.0065

TABLE II

Transfer Constants

f Cycles/sec.	Mid-series Type VI_1		Mid-half M-type IV_1		Mid-part of Wave-filter	
	$T_1=A_1+iB_1$		$T_m=\frac{1}{2}(A_m+iB_m)$		$T=T_1+T_m=D+iS$	
1000	1.26	—	.97	—	2.23	—
1500	1.21	—	.99	—	2.20	—
2000	1.14	—	1.02	—	2.16	—
2500	1.04	—	1.08	—	2.12	—
3000	.89	—	1.23	—	2.12	—
3250	.79	—	1.46	—	2.25	—
3500	.66	—	2.64	—	3.30	—
3750	.480	$+i$ .02	1.077	$-i$1.51	1.557	$-i$1.49
4000	.100	$+i$ .10	.181	$-i$1.39	.281	$-i$1.29
4250	.025	$+i$ .51	.029	$-i$ .75	.054	$-i$ .24
4500	.019	$+i$ .73	.019	$-i$ .48	.038	$+i$ .25
5292	.018	$+i$1.30	.013	$+i$ 0	.031	$+i$1.30
6500	.032	$+i$2.20	.026	$+i$ .67	.058	$+i$2.87
6750	.043	$+i$2.48	.040	$+i$ .92	.083	$+i$3.40
7000	.173	$+i$2.97	.181	$+i$1.39	.354	$+i$4.36
7500	.910	$+i$3.10	1.125	$+i$1.51	2.035	$+i$4.61
8000	1.27	—	2.64	—	3.91	—
8500	1.52	—	1.53	—	3.05	—
9000	1.74	—	1.31	—	3.05	—
10000	2.09	—	1.15	—	3.24	—
11000	2.36	—	1.09	—	3.45	—
12000	2.59	—	1.06	—	3.65	—

TABLE III

Reflection Coefficients and Interaction Factor

f Cycles/sec.	$x=.80$		$m=.7454$		P	Q
	G_{x1}	H_{x1}	G_{m2}	H_{m2}		
3750	.58	2.18	0	−2.31	3.69	−3.11
4000	.30	3.06	.49	− .50	1.35	− .02
4250	1.04	3.75	2.58	0	3.73	3.27
4500	1.57	4.04	3.85	0	5.50	4.54
5292	∞	—	∞	—	∞	—
6500	1.17	2.45	2.90	0	4.19	8.19
6750	.78	2.67	1.92	0	2.87	9.47
7000	.30	3.22	.49	.50	1.50	12.44
7500	.60	4.14	0	2.39	4.67	15.75

TABLE IV

Transmission Losses

f Cycles/sec.	Transfer	Terminal		Interaction	Total $=L$	
	L_t	L_x	L_m	L_r	ΣL_j	Network Computation
1000	2.23	.88	.02	—	3.13	3.13
1500	2.20	.65	−.18	—	2.67	2.68
2000	2.16	.48	−.30	—	2.34	2.35
2500	2.12	.33	−.35	—	2.10	2.11
3000	2.12	.19	−.25	—	2.06	2.08
3250	2.25	.12	−.01	—	2.36	2.37
3500	3.30	.06	1.21	—	4.57	4.59
3750	1.557	.042	−.190	.025	1.434	1.487
4000	.281	.443	.082	−.300	.506	.508
4250	.054	.067	.004	.024	.149	.154
4500	.038	.023	.001	.001	.063	.068
5292	.031	.000	.000	.000	.031	.036
6500	.058	.052	.003	.005	.118	.127
6750	.083	.118	.011	.055	.267	.276
7000	.354	.443	.082	−.250	.629	.632
7500	2.035	.038	−.150	.009	1.932	1.987
8000	3.91	.06	1.21	—	5.18	5.19
8500	3.05	.11	.06	—	3.22	3.24
9000	3.05	.16	−.18	—	3.03	3.05
10000	3.24	.24	−.31	—	3.17	3.18
11000	3.45	.31	−.34	—	3.42	3.43
12000	3.65	.38	−.33	—	3.70	3.70

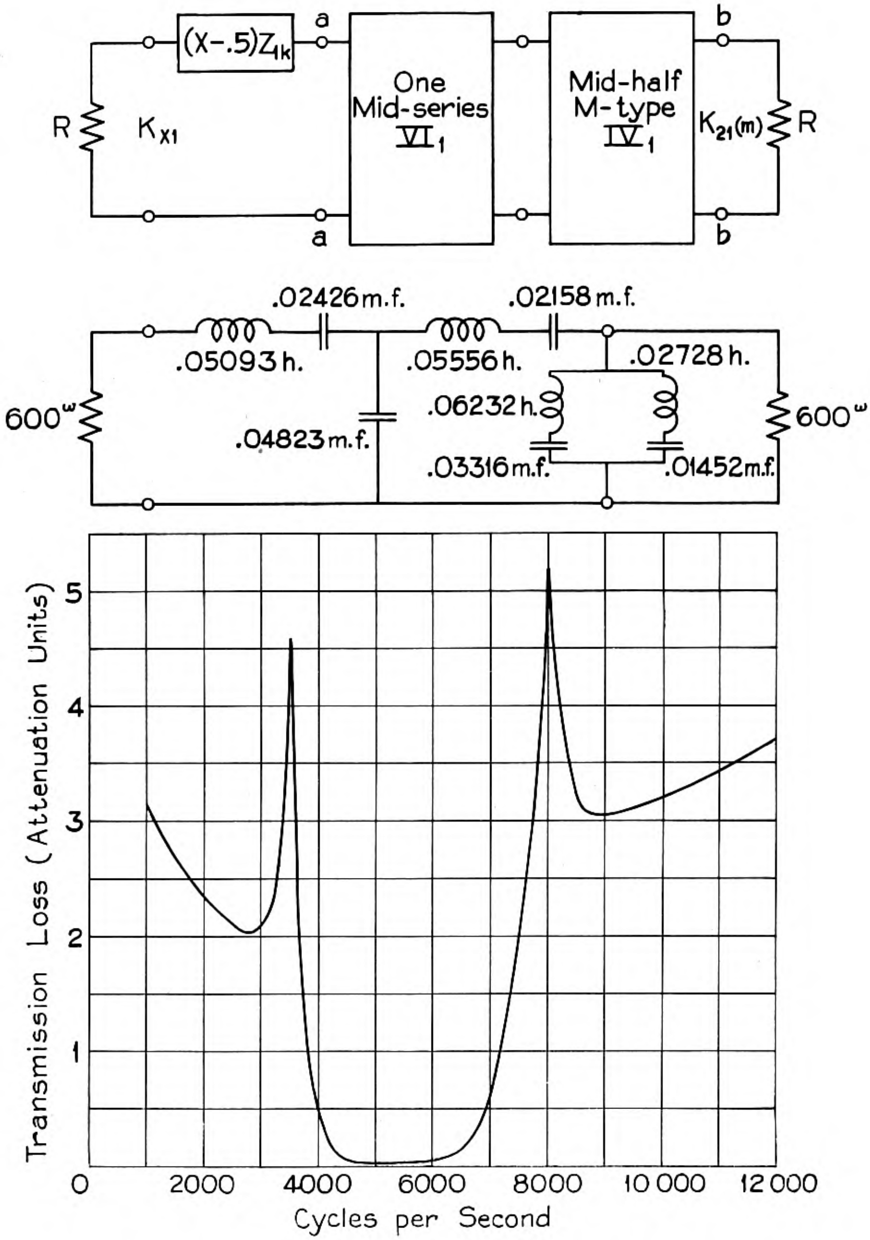


Fig. 4—Transmission Loss of Composite Band Pass Wave-Filter

APPENDIX

Derivation of Linear Transducer Formulae

The formula used in the text for a dissymmetrical composite wave-filter structure contains the image parameters [9] and is a special case of a general formula which is applicable to any linear transducer, active or passive. This general formula is derived here together with other useful ones.

A linear transducer will be defined as an electrical network which has two input and two output terminals and a structure such that so far as these terminals are concerned the currents are linear functions of the potential differences and therefore the principle of superposition holds. The structure may contain sources as well as sinks of energy; that is, the transducer may be active or passive. In the most general case, that of an active dissymmetrical linear transducer, four independent parameters are necessary to specify its electrical properties. Two sets of such parameters will be considered in deriving corresponding formulae, the image parameters and the recurrent parameters.

I. Image Parameters

1. *General Linear Transducer.* The parameters in this case are defined with reference to the single transducer in Fig. 3. Let the terminal impedances in this figure be so chosen that the impedances in the two directions from terminals a are equal, that is, the latter impedances are the "image" of each other, and at the same time a similar "image condition" holds with reference to terminals b. With the transducer so terminated, its *directional transfer constants* are here defined as $T_{ab}=\log_e(I_a/I_b)$ when transmitting from terminals a to terminals b, and $T_{ba}=\log_e(I_b/I_a)$ when transmitting from terminals b to terminals a. The *image impedance* W_a of the transducer is the impedance across terminals a in either direction, and the *image impedance* W_b is similarly defined at terminals b. In general, T_{ab} and T_{ba} are different, as are also W_a and W_b.

The transducer is now to be terminated by the general impedances Z_a and Z_b with an electromotive force E_a applied in series with Z_a.

[9] The relations among five other distinct sets of parameters for a transducer (such as a passive one) which can be specified by three complex parameters were given by G. A. Campbell in Cisoidal Oscillations, Trans. A. I. E. E., Vol. XXX, Part II, Table I, p. 885, 1911. The different sets correspond to the four normal networks designated as the T, the Π, the transformer, and the artificial line, and to the simple circuit one-point and two-point impedances. A sixth set, one-point and two-point admittances, was used in Appendix I of my paper in the B. S. T. J., Jan. 1923.

It is desired to obtain, among others, expressions for the sending end and receiving end currents I_a and I_b which contain the image parameters.

Each terminal impedance will be considered as equivalent to the image impedance at that end plus another impedance whose potential drop is to be replaced in the usual manner by an equal opposing electromotive force. In effect this equivalent electromotive force substitution reduces the system to one in which the transducer is terminated by its image impedances and in which determinate electromotive forces are acting at *both* ends. From this viewpoint, the total effective electromotive forces acting at the ends a and b of the transducer terminated by its image impedances W_a and W_b are, respectively,

$$E_a+(W_a-Z_a)I_a,$$

and

$$(W_b-Z_b)I_b. \tag{49}$$

Superposing the currents due to these electromotive forces at both ends we may write the current expressions immediately from the definitions of the parameters involved.

Thus

$$I_a=\frac{E_a+(W_a-Z_a)I_a}{2W_a}+\frac{(W_b-Z_b)I_b}{2W_b}e^{-T_{ba}},$$

and

$$I_b=\frac{E_a+(W_a-Z_a)I_a}{2W_a}e^{-T_{ab}}+\frac{(W_b-Z_b)I_b}{2W_b}. \tag{50}$$

Their solution gives the explicit formulae [10]

$$I_a=\frac{E_a}{W_a+Z_a}\;\frac{(1+r_b e^{-(T_{ab}+T_{ba})})}{(1-r_a r_b e^{-(T_{ab}+T_{ba})})},$$

and

$$I_b=\frac{E_a}{W_a+Z_a}\;\frac{(1+r_b)e^{-T_{ab}}}{(1-r_a r_b e^{-(T_{ab}+T_{ba})})}, \tag{51}$$

where r_a and r_b, the current reflection coefficients at terminals a and b, are

$$r_a=\frac{W_a-Z_a}{W_a+Z_a},$$

and

$$r_b=\frac{W_b-Z_b}{W_b+Z_b}.$$

[10] These formulæ may also be derived synthetically by the current reflection method.

Although the transducer has four independent parameters, it will be seen that the sending end current involves but three effective transducer parameters, the sum $(T_{ab}+T_{ba})$, W_a, and W_b. As a result, the four one-point impedance measurements which can be made upon the transducer itself, the open-circuit and short-circuit driving-point impedances at both ends, must have a relation between them. Let X_a and Y_a denote the driving-point impedances across terminals a when terminals b are open-circuited and short-circuited, respectively. Then if in (51) $Z_a=0$ and terminals b are open-circuited by putting $Z_b=\infty$, the impedance at terminals a, the *open-circuit impedance*, is

$$X_a=\frac{E_a}{I_a}=W_a \coth \tfrac{1}{2}(T_{ab}+T_{ba}). \tag{52}$$

Similarly for the *short-circuit impedance*, when $Z_a=0$ and $Z_b=0$,

$$Y_a=W_a \tanh \tfrac{1}{2}(T_{ab}+T_{ba}). \tag{53}$$

For the other end we get by interchanging subscripts

$$X_b=W_b \coth \tfrac{1}{2}(T_{ab}+T_{ba}), \tag{54}$$

and

$$Y_b=W_b \tanh \tfrac{1}{2}(T_{ab}+T_{ba}). \tag{55}$$

These give the necessary relation as

$$\frac{X_a}{Y_a}=\frac{X_b}{Y_b}. \tag{56}$$

Hence, *in the most general linear transducer the ratio of the open-circuit to short-circuit impedances at one end is equal to the corresponding ratio at the other end.*

Other important derived formulae are

$$T_{ab}+T_{ba}=2 \tanh^{-1}\sqrt{\frac{Y_a}{X_a}}=2 \tanh^{-1}\sqrt{\frac{Y_b}{X_b}}, \tag{57}$$

$$W_a=\sqrt{X_a Y_a}, \tag{58}$$

$$W_b=\sqrt{X_b Y_b}, \tag{59}$$

$$W_a-W_b=\sqrt{(X_a-X_b)\ (Y_a-Y_b)}, \tag{60}$$

and

$$W_a W_b=X_a Y_b=X_b Y_a. \tag{61}$$

Thus the open-circuit and short-circuit impedance measurements determine the *sum* of the directional transfer constants and both of the image impedances.

To obtain the separate values of T_{ab} and T_{ba}, it is necessary to make at least one two-point measurement, as seen from the formula for I_b which contains four distinct transducer parameters. For example, to find T_{ab} perhaps the simplest method is to terminate with the image impedance at terminals b, whence

$$T_{ab}=\log_e(I_a/I_b), \text{ where } Z_b=W_b. \tag{62}$$

The constant T_{ba} is the difference between the sum $(T_{ab}+T_{ba})$, obtained from (57), and T_{ab}; it may also be determined by a two-point measurement similar to the above for transmission in the opposite direction.

For some purposes it is convenient to have formulae involving the potential differences V_a and V_b across the two pairs of terminals, rather than the terminal impedances Z_a and Z_b and the series applied electromotive force E_a. Such formulae in combination with the above can be used to advantage in determining the currents and potential differences at points within a composite transducer. They are derived readily by making the substitutions in the above,

$$Z_a=\frac{E_a-V_a}{I_a},$$

and

$$Z_b=V_b/I_b. \tag{63}$$

For current transmission from terminals a to terminals b

$$I_a=I_b\left(\frac{e^{T_{ab}}+e^{-T_{ba}}}{2}\right)+\frac{V_b}{W_b}\left(\frac{e^{T_{ab}}-e^{-T_{ba}}}{2}\right),$$

and

$$V_a=V_b\frac{W_a}{W_b}\left(\frac{e^{T_{ab}}+e^{-T_{ba}}}{2}\right)+I_bW_a\left(\frac{e^{T_{ab}}-e^{-T_{ba}}}{2}\right). \tag{64}$$

Also,

$$I_b=I_a\left(\frac{e^{T_{ba}}+e^{-T_{ab}}}{2}\right)-\frac{V_a}{W_a}\left(\frac{e^{T_{ba}}-e^{-T_{ab}}}{2}\right),$$

and

$$V_b=V_a\frac{W_b}{W_a}\left(\frac{e^{T_{ba}}+e^{-T_{ab}}}{2}\right)-I_aW_b\left(\frac{e^{T_{ba}}-e^{-T_{ab}}}{2}\right). \tag{65}$$

Interchanging the subscripts and changing the signs of the currents in (64) will also lead to (65).

2. *Passive Linear Transducer.* Since the reciprocal theorem holds here one relation exists between the four parameters leaving three

independent ones. This relation is given directly by the theorem in the case where $Z_a = W_a$ and $Z_b = W_b$, the equivalent transfer currents being

$$\frac{e^{-T_{ab}}}{2W_a} = \frac{e^{-T_{ba}}}{2W_b}. \tag{66}$$

Although any three of these parameters might be assumed as independent, it is convenient to take as the independent parameters T, W_a, and W_b, where

$$T = D + iS = \tfrac{1}{2}(T_{ab} + T_{ba}) \tag{67}$$

is thus defined for the *passive transducer* as the *transfer constant. The transfer constant is the arithmetic mean of the two directional transfer constants.* The real and imaginary parts of T, namely D and S, will be called the *diminution constant* and the *angular constant* to distinguish them from the attenuation constant and the phase constant of the ordinary propagation constant to which they reduce in the case of a symmetrical transducer. Then these parameters are given by the formulae

$$T = \tanh^{-1}\sqrt{\frac{Y_a}{X_a}} = \tanh^{-1}\sqrt{\frac{Y_b}{X_b}},$$

$$W_a = \sqrt{X_a Y_a}, \tag{68}$$

and

$$W_b = \sqrt{X_b Y_b},$$

and are completely determined by the open-circuit and short-circuit driving-point impedances.

With these parameters the current formulae become

$$I_a = \frac{E_a}{W_a + Z_a} \frac{(1 + r_b e^{-2T})}{(1 - r_a r_b e^{-2T})},$$

and

$$I_b = \frac{E_a}{W_a + Z_a} \sqrt{\frac{W_a}{W_b}} \frac{(1 + r_b)e^{-T}}{(1 - r_a r_b e^{-2T})}, \tag{69}$$

$$= \frac{2E_a\sqrt{W_a W_b}\, e^{-T}}{(W_a + Z_a)(W_b + Z_b)(1 - r_a r_b e^{-2T})}.$$

the latter being the one used in the text. Other forms are

$$I_a = \frac{E_a(Z_b \sinh T + W_b \cosh T)}{(W_a W_b + Z_a Z_b)\sinh T + (W_a Z_b + W_b Z_a)\cosh T},$$

and

$$I_b = \frac{E_a\sqrt{W_a W_b}}{(W_a W_b + Z_a Z_b)\sinh T + (W_a Z_b + W_b Z_a)\cosh T}. \tag{70}$$

Introducing the potential differences, for current transmission from terminals a to terminals b

$$I_a = I_b\sqrt{\frac{W_b}{W_a}}\cosh T + \frac{V_b}{\sqrt{W_aW_b}}\sinh T,$$

and (71)

$$V_a = V_b\sqrt{\frac{W_a}{W_b}}\cosh T + I_b\sqrt{W_aW_b}\sinh T.$$

Also,

$$I_b = I_a\sqrt{\frac{W_a}{W_b}}\cosh T - \frac{V_a}{\sqrt{W_aW_b}}\sinh T,$$

and (72)

$$V_b = V_a\sqrt{\frac{W_b}{W_a}}\cosh T - I_a\sqrt{W_aW_b}\sinh T.$$

II. Recurrent Parameters

1. *General Linear Transducer.* Here four parameters [11] of the transducer in Fig. 3 are defined in terms of its properties when it is one section of an infinite recurrent structure which is made up of identical sections, similarly oriented. With such terminal conditions for the transducer, its *directional propagation constants* are defined as follows: $\Gamma_{ab} = \log_e (I_a/I_b)$ when transmitting from terminals a to terminals b, and $\Gamma_{ba} = \log_e (I_b/I_a)$ when transmitting from terminals b to terminals a. The *characteristic impedance* K_a is the impedance across terminals a in the direction from a to b, and the *characteristic impedance* K_b is similarly defined for the impedance across terminals b in the opposite direction.

Terminating the transducer by the general impedances Z_a and Z_b and applying an electromotive force E_a in series with Z_a, the current formulae containing the recurrent parameters may be derived in a manner analogous to that used with the image parameters. In this case the total effective electromotive forces acting at the ends a and b of the transducer terminated by its characteristic impedances K_b and K_a are, respectively,

$$E_a + (K_b - Z_a)I_a,$$

and (73)

$$(K_a - Z_b)I_b.$$

[11] These parameters may also be designated in the general case as those of a generalized artificial line.

Hence,

$$I_a=\frac{E_a}{K_a+Z_a}\;\frac{(1+\rho_b e^{-(\Gamma_{ab}+\Gamma_{ba})})}{(1-\rho_a\rho_b e^{-(\Gamma_{ab}+\Gamma_{ba})})},$$

and (74)

$$I_b=\frac{E_a}{K_a+Z_a}\;\frac{(1+\rho_b)e^{-\Gamma_{ab}}}{(1-\rho_a\rho_b e^{-(\Gamma_{ab}+\Gamma_{ba})})},$$

where the current reflection coefficients at terminals a and b are

$$\rho_a=\frac{K_b-Z_a}{K_a+Z_a},$$

and

$$\rho_b=\frac{K_a-Z_b}{K_b+Z_b}.$$

Introducing the open-circuit and short-circuit driving-point impedances X_a, X_b and Y_a, Y_b of the transducer it follows that

$$\Gamma_{ab}+\Gamma_{ba}=2\ \tanh^{-1}\frac{\sqrt{(X_a-X_b)^2+2(X_aY_b+X_bY_a)}}{X_a+X_b},\tag{75}$$

$$\left.\begin{matrix}K_a\\K_b\end{matrix}\right\}=\frac{1}{2}\Big[\sqrt{(X_a-X_b)^2+2(X_aY_b+X_bY_a)}\pm(X_a-X_b)\Big],\tag{76}$$

$$K_a-K_b=X_a-X_b,\tag{77}$$

and

$$K_aK_b=X_aY_b=X_bY_a.\tag{78}$$

Any three of these measured impedances are sufficient, because of relation (56), to obtain the *sum* $(\Gamma_{ab}+\Gamma_{ba})$, K_a, and K_b.

A directional propagation constant may be obtained separately from one two-point measurement; thus

$$\Gamma_{ab}=\log_e\,(I_a/I_b),\ \text{where } Z_b=K_a.\tag{79}$$

The current and potential difference at one pair of terminals in terms of those at the other are given by the following.

For current transmission from terminals a to terminals b

$$I_a=I_b\left(\frac{K_be^{\Gamma_{ab}}+K_ae^{-\Gamma_{ba}}}{K_a+K_b}\right)+\frac{V_b}{K_a+K_b}\,(e^{\Gamma_{ab}}-e^{-\Gamma_{ba}}),$$

and (80)

$$V_a=V_b\left(\frac{K_ae^{\Gamma_{ab}}+K_be^{-\Gamma_{ba}}}{K_a+K_b}\right)+I_b\,\frac{K_aK_b}{K_a+K_b}\,(e^{\Gamma_{ab}}-e^{-\Gamma_{ba}}).$$

Also,

$$I_b = I_a\left(\frac{K_a e^{\Gamma_{ba}} + K_b e^{-\Gamma_{ab}}}{K_a+K_b}\right) - \frac{V_a}{K_a+K_b}\left(e^{\Gamma_{ba}} - e^{-\Gamma_{ab}}\right),$$

and (81)

$$V_b = V_a\left(\frac{K_b e^{\Gamma_{ba}} + K_a e^{-\Gamma_{ab}}}{K_a+K_b}\right) - I_a\frac{K_a K_b}{K_a+K_b}\left(e^{\Gamma_{ba}} - e^{-\Gamma_{ab}}\right).$$

2. *Passive Linear Transducer.* Because of the reciprocal theorem the directional propagation constants become equal giving a single *propagation constant,*

$$\Gamma = A + iB = \Gamma_{ab} = \Gamma_{ba}, \tag{82}$$

which is obtainable from the general formula (75). Here A is the *attenuation constant* and B is the *phase constant.*

The current formulae become

$$I_a = \frac{E_a}{K_a+Z_a}\,\frac{(1+\rho_b e^{-2\Gamma})}{(1-\rho_a\rho_b e^{-2\Gamma})},$$

and (83)

$$I_b = \frac{E_a}{K_a+Z_a}\,\frac{(1+\rho_b)e^{-\Gamma}}{(1-\rho_a\rho_b e^{-2\Gamma})}.$$

In the other form they are

$$I_a = \frac{E_a[(-K_a+K_b+2Z_b)\sinh\Gamma + (K_a+K_b)\cosh\Gamma]}{[(2(K_aK_b+Z_aZ_b)-(K_a-K_b)(Z_a-Z_b))\sinh\Gamma + (K_a+K_b)(Z_a+Z_b)\cosh\Gamma]},$$

and (84)

$$I_b = \frac{E_a(K_a+K_b)}{[(2(K_aK_b+Z_aZ_b)-(K_a-K_b)(Z_a-Z_b))\sinh\Gamma + (K_a+K_b)(Z_a+Z_b)\cosh\Gamma]}.$$

Introducing the terminal potential differences, when transmitting from terminals a to terminals b

$$I_a = I_b\left(\cosh\Gamma - \frac{K_a-K_b}{K_a+K_b}\sinh\Gamma\right) + \frac{V_b}{K_a+K_b}\,2\sinh\Gamma,$$

and (85)

$$V_a = V_b\left(\cosh\Gamma + \frac{K_a-K_b}{K_a+K_b}\sinh\Gamma\right) + I_b\frac{K_aK_b}{K_a+K_b}\,2\sinh\Gamma;$$

and at the other terminals

$$I_b = I_a\left(\cosh \Gamma + \frac{K_a - K_b}{K_a + K_b}\sinh \Gamma\right) - \frac{V_a}{K_a + K_b} 2 \sinh \Gamma,$$

and (86)

$$V_b = V_a\left(\cosh \Gamma - \frac{K_a - K_b}{K_a + K_b}\sinh \Gamma\right) - I_a\frac{K_a K_b}{K_a + K_b} 2 \sinh \Gamma.$$

Comparison shows that the general formulae for the currents I_a and I_b given by (51) and (74) in terms of the two sets of parameters are of the same functional form involving their respective reflection coefficients; the latter are of slightly different functional forms. This similarity is what one expects when deriving the formulae synthetically by the current reflection method.

In all cases by (61) and (78)

$$W_a W_b = K_a K_b. \tag{87}$$

The sum $(T_{ab} + T_{ba})$, W_a, and W_b of any transducer are obviously also equal to the propagation constant and respective characteristic impedances of the two symmetrical transducers which can individually be formed with two such identical transducers.

If $T_{ab} = T_{ba}$, the reciprocal theorem holds only when $W_a = W_b$, for which case the transducer is symmetrical. On the other hand if $\Gamma_{ab} = \Gamma_{ba}$, this theorem holds irrespective of the values of K_a and K_b. In each of these cases which satisfies the theorem the transducer may be active or passive.

In an electrically symmetrical transducer, whether active or passive, two parameters specify its properties where

$$T_{ab} = T_{ba} = \Gamma_{ab} = \Gamma_{ba},$$

and (88)

$$W_a = W_b = K_a = K_b,$$

in which case the corresponding formulae are identical in the parameters. Structural symmetry is not necessary here as may be seen, for example, in the case of a composite wave-filter made up of different mid-series sections whose characteristic impedances are equivalent.

In a passive dissymmetrical transducer the formulae containing hyperbolic functions are of simpler form with the parameters T, W_a, and W_b than with the parameters Γ, K_a, and K_b. The image parameter formulae are readily applicable where the transducer is made up of parts whose image impedances at the junctions are equivalent, as in the present case of a composite wave-filter. Simple relations exist here between these parameters of the transducer and of its parts, as shown in the text, which is not true with the other parameters. The recurrent parameter formulae, on the other hand, apply more naturally when dealing with a succession of identical dissymmetrical sections, or of different dissymmetrical sections whose characteristic impedances in one direction are equivalent, in which cases the propagation constant of the transducer is equal to the sum of the progagation constants of the parts. In conclusion, it is seen that the set of parameters most suitable for use in any case depends upon the particular structure of the transducer.

Some Contemporary Advances in Physics—V

By KARL K. DARROW

Electricity in Solids

IN considering such topics as the flow of electricity through solids and the outflow of electricity across their boundaries, we have to forego the assistance of the great system of laws, models, and word-pictures which constitutes the contemporary theory of the structure of the atom. This imposing and truly powerful theory, which nowadays seems to bulk larger than all of the rest of physics, is after all limited to certain restricted fields; it deals successfully with particular properties of isolated atoms, and also with certain qualities of atoms which seem to be localized in their inner regions; but it avails little or nothing in the study of the behavior of liquids and solids. Much of the present-day theory of electrical conduction in solids is based only on the very simplest assumptions as to the nature of the atoms of which they are built, some would even remain valid under the old-fashioned ideas of continuous electrical fluids; and profoundly as we may believe that solids are built of atoms resembling Bohr's famous model, it is highly doubtful whether that model has ever helped to interpret a single one of the phenomena of conduction or done more than to provide a new language for old ideas.

We have first to make the distinction between the substances in which atoms migrate along the path of the flowing current and apparently carry the moving charge, and the substances in which the atoms stand still while the current flows past them. It is universally conceded that elements, and likewise the alloys of metals and a number of solid compounds, belong to the latter class; whatever it is that carries the current flows through and past the substance, leaving it at the end as it was at the beginning. Weber said in 1858, "In the metals there are electrically-charged particles as well as atoms; some of the former are freely mobile and others vibrate about the atoms; they are the cause of the conduction of electricity and of heat, and of magnetic phenomena as well." Considering that in Weber's day electricity had never been observed apart from ponderable matter and electrons were unknown, this is entitled to rank as a daring anticipation.

Next we have to distinguish between conduction by metals and conduction by non-metallic elements. Strictly we should begin by defining a "metal"; but this task had better be left to the chemists, as being really their affair; and they have found it no easy affair to

PERIODIC CLASSIFICATION OF THE ELEMENTS[1]

	I	II	III	IV	V	VI	VII	VIII	O
1	1 H								2 He
2	3 Li	4 Be	5 B	6 C	7 N	8 O	9 F		10 Ne
3	11 Na	12 Mg	13 Al	14 Si	15 P	16 S	17 Cl		18 A
4	19 K	20 Ca	21 Sc	22 Ti	23 V	24 Cr	25 Mn	26 Fe 27 Co 28 Ni	
	29 Cu	30 Zn	31 Ga	32 Ge	33 As	34 Se	35 Br		36 Kr
5	37 Rb	38 Sr	39 Yt	40 Zr	41 Nb	42 Mo	43—	44 Ru 45 Rh 46 Pd	
	47 Ag	48 Cd	49 In	50 Sn	51 Sb	52 Te	53 I		54 Xe
6	55 Cs	56 Ba	Rare Earths	72 Hf	73 Ta	74 W	75—	76 Os 77 Ir 78 Pt	
	79 Au	80 Hg	81 Tl	82 Pb	83 Bi	84 Po	85—		86 Nt
7	87—	88 Ra	89 Ac	90 Th	91 Pa	92 U			

[1] The atomic numbers are given in front of the abbreviations of the elements.

set up a definition by which every element can be confidently assigned to one class or to the other. In fact there is a tendency to begin by defining metallic conduction, and then define metals as the elements which display it! The difficulty, as usual, is to make the definition sharp enough to decide a few intermediate or transitional cases. Anyone even slightly acquainted with chemistry or physics would instantly recognize as metals the elements in the first column of the Periodic Table, and those at the bottom of the table in all the columns; and as non-metals, with the same ease, the elements in the topmost row of the table and down the right-hand side. The first element of every column after the first two is non-metallic, and the non-metallic character advances farther and farther down the columns as one proceeds across the Table from left to right. One might say that the elements which are not metals occupy the north-east sector of the Table, and the debatable ones cross in a diagonal band from northwest to southeast. The elements which are gases under the usual circumstances of temperature and pressure are extreme instances of non-metals; but some of the definitely non-metallic elements, and all of the debatable ones, are solid or liquid under the usual conditions.

Very little could be said about the elements which under ordinary conditions are gases, for very little is known about the manner in which they conduct electricity when liquefied or frozen. Probably the reason is that the experimental conditions would be unusually difficult, and the substances probably very bad conductors; it is not easy to imagine solid hydrogen moulded into a cylinder, drawn into a wire, clamped or sealed between electrodes, or filled into a sheath less conductive than the hydrogen itself. The difficulties may not be insuperable; but they have not been generally overcome.

As for the solid elements which are definitely not metals, or which belong to the debatable group, there is an abundance of data in print, and yet not nearly so much as we need. In general their resistances are tremendously greater than the resistances of metals; "tremendously" for once is not an extravagant word, for the conductivities of the elements are spread over a sweeping range of orders of magnitude which few if any other qualities of theirs can rival. The mass of the heaviest known atom differs from the mass of the lightest only by a factor of 240; the densities of the solidified elements, their compressibilities, their other mechanical and thermal properties range over not more than one or two, at the most three orders of magnitude; even the energy required to extract the innermost electron of an atom rises by a factor of only 10^5 in passing from the first to the last element of the series; but the conductivity of silver stands to the conductivity of sulphur in the ratio 10^{21}. The distance from the sun to the nearest star is some 10^{18} cm.; we see that a sheet of sulphur a thousandth of an inch thick would offer more of an obstacle to the passage of electricity than a cable of silver of the same diameter, extending from the earth to Alpha Centauri. The variations of conducting-power from element to element are thus as fantastically great as the variations in scale from the world of common life to the world of interstellar spaces. The conductivities of the metals, however, are confined within a narrow fraction of this range; it is between the metals and the non-metals, and between one non-metallic element and another, that the leaps are surprisingly great.

In general, too, the resistance of a non-metallic element decreases as its temperature is raised; the curve of resistance versus temperature (I shall often call it *characteristic*, henceforward) slants downward, the derivative and the temperature-coefficient of resistance are negative. Near room-temperature this is the usual behavior, but not always over the entire accessible range; of some elements it is observed that the resistance declines less rapidly as the temperature is raised, the curve is concave upward; eventually the decline ceases, the resistance passes through a minimum value at a certain characteristic temperature, and thereafter increases with the temperature as the resistances of metals do. At least one element of the debatable class (germanium) exhibits a characteristic curve that slants upward instead of downward at room-temperatures; but when the curve is followed towards lower temperatures, it too is found to be concave upward with a minimum of resistance below $-100°$ C. This suggests that for all of the non-metals the resistance-temperature curve may be a loop bulging downward, with a minimum at a certain temperature that varies from element to element; on this generalization one of the contemporary theories is founded.

These rules can be illustrated by mentioning briefly the behavior of the non-metallic elements one by one. Beginning at the foot of the procession of elements, we pass over hydrogen (no data), lithium and beryllium (metals), and commence with boron. *Boron* has a very high resistance at room temperature, which drops a hundredfold when it is heated to 180° C. and ten-million-fold when it is raised to a red heat. On *carbon* a tremendous amount of work has been done, which unfortunately largely goes to show that the word "carbon" usually signifies a framework of carbon atoms packed with occluded gases, organic compounds, and impurities of divers kinds, which no known mode of treatment avails to expel entirely, although almost anything which is done to the substance alters its constitution enough to affect its resistance. (We shall later see that the situation with many of the metals is almost as bad.) Most of the experiments reveal a steady decline of resistance as the temperature is raised, whether the sample used be amorphous or crystalline (graphitic) and whatever its history; but Noyes recently traced several very concordant curves for several samples of graphite (all however of the same provenience) showing a minimum of resistance near 800° C. Diamonds have exceedingly high resistances, which fall when they are heated.

Passing over four gases and three metals, we come next to *silicon*; the curve traced by Koenigsberger shows the resistance descending as the temperature is increased, until at a certain critical temperature it leaps sharply upward; from the new high value it descends again as the silicon is further warmed, only to make a second upward jump; from this second maximum it drops steadily away, at least as far as the highest temperature attained in the experiment. This illustrates another perplexing property of some elements; they have several distinct "allotropic" forms, each of them more or less stable over a distinct range of temperature which may or may not overlap with the ranges of the others; each must be regarded, so far as its conducting-power is concerned, as a distinct element. In some instances the several forms of an element are vividly contrasted in appearance and in general behavior; such is the case with *phosphorus*, all of the forms of which have high resistances, but little is known about their trends with temperature. In other cases the anomalous changes of temperature with resistance are not accompanied by other striking changes; and there is a tendency to explain any deviation from an expected trend—such as, for example, a maximum in a resistance-temperature curve—by saying that the substance is gradually changing from one form into another.

Sulphur is the extreme case of high resistance. I know of no data for *scandium*, which is to be regretted, as there is some reason from general atomic theory for supposing that this element stands at a turning-point of the Periodic Table. *Titanium*, like silicon, has several modifications, in some of which the characteristic rises while in others it descends. *Germanium* has been studied lately by Bidwell; it is the element mentioned above which displays a minimum of resistance at −116° C. *Arsenic* resembles the metals. *Selenium* in the dark has an extremely high resistance; its character when illuminated is too much of a subject to be discussed in this place. *Zirconium* was found, at least by one observer, to display a minimum of resistance at 70° C., though in conductivity it compares favorably with the accepted metals. *Antimony*, although ranked among the metals, is usually to be found among the exceptions to any rules laid down for them; the same can be said of *bismuth*. *Tellurium* is an outstanding instance of an element with two modifications, and a sample taken at random is likely to be a mixture of them in unpredictable proportions, which change when it is heated; the characteristics are correspondingly crooked, and rarely agree. *Iodine* has a very high resistance.

Comparing the metals as a group with the non-metals, the first striking rule is that their conductivities are much higher and rather close together; from silver (the most conductive of all substances at room-temperature), to bismuth, the most resistant of the elements commonly accepted as metals, the conducting-power descends in the relatively small ratio of 75 to 1. The next and familiar rule is, that increasing temperature and increasing resistance always go together; the characteristic always slants upward to the right, the derivative and the temperature-coefficient of resistance are positive. It is customary to say that the resistance is always approximately proportional to the temperature, and that the temperature-coefficient of resistance always has approximately the one universal value, which is the value of the temperature-coefficient of volume of an ideal gas at constant pressure (or its temperature-coefficient of pressure at constant volume). That is to say, when the temperature of a piece of metal is increased by a given amount, its resistance increases approximately in the same proportion as would the pressure of a fixed quantity of an ideal gas, enclosed in a non-expanding container and raised from the same initial to the same final temperature as the metal. Were these statements literally true, all the resistance-temperature curves for metals would be straight lines intersecting the axis of temperatures at absolute zero. But the second statement cannot even be considered a good approximation, unless one is willing to confer the title "good approximation" on a numerical value .00365 which is expected to agree with a set of observed values which ranges upwards to .0058 (potassium) and .0063 (iron). (I refrain from giving a lower limit for the range, for a reason which will presently be made clear.) Also the characteristic curves are not rigorously straight lines, although it is not unreasonable to call some of them *approximately* straight, when one considers how wide is the interval of temperature over which some of them have been traced. In some cases a quadratic term added to the linear expression, resulting in a formula $R=R_o+at+bt,^2$ is sufficient to express the data. Usually, but not always, the extra coefficient b is positive; the characteristic is concave upward. "Usually but not always" is a phrase much in demand when one is laying down rules for conducting bodies. In this case metals of the platinum triad furnish the exceptions. In other instances cubic terms must be added to the formulæ, and in still others even these are inadequate. One of the longest characteristics ever traced, the one determined by Worthing and Forsythe for tungsten from 1400° to 3250° C., conforms to the equation $R = const.\ T^{1.2}$.

All these details about values of resistances and shapes of resistance-temperature curves are sedate and commonplace enough; but there is one quite extraordinary phenomenon in this field, one of the strange discontinuities which appear here and there in the theatre of nature and contribute more of dramatic interest to the spectacle than any amount of smooth correlations between continuous variables. Extensions of the characteristics downwards toward the absolute zero have to follow upon improvements in the art of producing and maintaining very low temperatures; and for the last twenty years the advances in this art have been made in the Cryogenic Laboratory of the University of Leyden, and there the curves have been extended downwards step by step as additional ranges of cold were made accessible. The temperatures down to 14° K. attained with liquefied hydrogen did not affect the resistances of metals in any very startling way, although the characteristics are generally more sharply curved there than at ordinary temperatures; but when with the aid of liquefied helium Kamerlingh Onnes penetrated to within five degrees of the absolute zero, something astonishing took place.

Kamerlingh Onnes had been experimenting with platinum wire, and he had found that over the interval of temperatures newly made available, the interval from 4.3° to 1.5° K. (a small range when measured in degrees, but a great one when considered in terms of the distance between its lower limit and the absolute zero) the resistance of the wire did not change. This he thought might mean that the proper resistance of the metal had become exceedingly small, leaving as the chief component of the observed resistance a term unaffected by temperature and due possibly to some such thing as discontinuities in the wire, for example between the platinum and bits of impurities mixed into it. To have a purer metal he replaced the platinum by repeatedly-distilled mercury. It was contained in a slender glass capillary tube, forming so fine a filament that the resistance at room-temperature was rather considerable; in one specified instance, 173 ohms. When he lowered this filament of mercury to the temperature of frozen helium, at a certain point the resistance suddenly vanished. Literally it vanished; the word is justified, for the value to which it had dropped was, if not truly zero, at all events not so much as one five-billionth of its value at room-temperature, and not so much as one ten-millionth of its value just before, at about 4.1° K., it suddenly disappeared. The mercury had altogether lost what had always seemed to be as inseparable a quality of matter as its inertia or its weight.

A few other elements were later found to share this property; tin, of which the resistance vanishes at 3.78°; lead, having its threshold-temperature at 7.2°; thallium, at 2.3°. Three of these four are consecutive in the procession of elements. Other elements were definitely found not to become "supra-conductive" within the accessible range: gold, cadmium, platinum, copper and iron. In the vicinity of the absolute zero each of these metals has a constant resistance independent of temperature. This as I mentioned was interpreted to mean that these metals, or at least these samples, behaved thus because they were impure—that impurities prevented the vanishing of resistance—but since mercury contaminated intentionally with gold or with cadmium was found to become supraconductive, and tin amalgam likewise, it has become necessary to save this interpretation, if at all, by assuming that in the five specified metals the impurities coalesce with the metal in some particular way. It is interesting to note that the threshold-temperature of tin amalgam lies above that of either of its components—at 4.29° K., to be compared with the 4.1° of mercury and the 3.78° of tin. These thresholds are not entirely independent of circumstances; they diminish when a large current-density is used, and also when a magnetic field is applied, possibly from the same reason in both cases.

A number of fantastic things could happen in a world from which electrical resistance had vanished, and one of them was actually realized by Kamerlingh Onnes within the compass of his helium-cooled chamber, when a current of three hundred and twenty amperes flowed for half-an-hour around and around a leaden ring with no applied E.M.F. whatever to maintain it, and did not lose as much as one one-hundredth of its initial strength. In another experiment a current of forty-nine amperes flowed for an hour around a coil of lead wire of a thousand turns, wound upon a brass tube, and did not lose quite one per cent. of the intensity with which it had been started by removing a magnet of which the field had interlaced the coil. At this rate it would have taken over four days for the current to drop to the $1/e$th part of its initial value, if the coil could have been kept cold so long. This corresponds to a resistance lower than 3.10^{-7} ohms; the resistance of the coil at room-temperature was 734 ohms. Few discoveries in physics can have been so exciting as this one, and further news from Leyden is awaited with keen anticipation. Until the present liquefied helium has been made nowhere else, but from now on the process will be carried on at Toronto also.

Pressure affects the resistance of a metal much less than temperature; that is to say, doubling the hydrostatic pressure upon a metal makes no perceptible difference with its resistance if the initial pressure is one atmosphere or less, and usually alters it only by a few per cent. if the intial pressure amounts to thousands of atmosphere. The art of applying enormous pressures under controllable conditions has been developed furthest by Bridgman in the Physical Laboratory of Harvard University, which through his work holds the same unique rank in high-pressure investigations as Kamerlingh Onnes' laboratory at Leyden in low-temperature research. The highest pressure which Bridgman has applied to metals during resistance-measurements exceeds 12,000 kg/cm^2, which amounts practically to twelve thousand atmospheres. No one has ever applied temperatures twelve thousand times as great as room-temperature, nor even four thousand times as great as the lowest accessible temperature; but when the pressure is altered in this enormous ratio the resistance changes only by a few per cent. The volume likewise changes by only a small fraction, which rather suggests that it is the change in closeness of packing of atoms rather than the creation of intense internal stresses which is responsible for the change in conductivity; however, there is no close correlation between relative change in volume and relative change in resistance; sometimes the two are of opposite signs. Usually, but not always, the conductivity increases with the pressure; as if squeezing the atoms together facilitated the flow of electricity across the metal. The rule applies to thirty-five elements, distributed as follows in the Periodic Table: in the first column, 11 Na, 19 K, 29 Cu, 47 Ag, 79 Au; second column, 12 Mg, 30 Zn, 48 Cd, 80 Hg; third, 13Al, 31 Ga, 49 In, 81 Tl; fourth, 6C, 22 Ti, 40 Zr, 50 Sn, 82 Pb; fifth, 15 P, 33 As, 73 Ta; sixth, 42 Mo, 52 Te, 74 W, 92 U; seventh, 53 I; eighth, 26 Fe, 27 Co, 28 Ni, 45 Rh, 46 Pd, 77 Ir, 78 Pt; rare earths, 57 La, 60 Nd. Several of the non-metallic elements are found in the list. The exceptions are the five curiously assorted metals 3 lithium, 20 calcium, 38 strontium, 51 antimony, 83 bismuth—five elements distributed over three columns of the Periodic Table, each of which contains several other elements which conform to the rule. One modification of 55 caesium belongs under the rule, another among the exceptions. This illustrates how the behavior of metals in conducting electricity is liable to cut across the classification of the Periodic System, which controls nearly all of the properties of elements except those that vary uniformly from one element to the next all along the series.

As for the magnitude of the effect, the resistances of most metals are decreased through less than 10% by applying a pressure of ten thousand atmospheres, some only through one or two per cent.; but the decrease is 40% for sodium, 70% for potassium, 70% also for the "debatable" element tellurium, and 97% for black phosphorus; bismuth gains about 25% in resistance and antimony about 10%. The curves representing resistance as function of pressure are somewhat curved, but not greatly so; however the curvature frequently varies along the curve to such an extent that a two-constant formula is not sufficient to express the data. It is an interesting fact that the percentage by which a given pressure changes the resistance of a metal is approximately independent of its temperature, and consequently the percentage by which a given rise in temperature changes the resistance is approximately independent of the pressure; so that the combined effects of a pressure-change Δp and a temperature-change ΔT on a metal change its resistance from R_o to $R_o\ (1+a\Delta p)(1+b\Delta T)$.

Tension, which is equivalent to negative pressure acting along a particular direction (there is no way of applying a negative hydrostatic pressure) results in lengthening the metal along one direction, shortening it along all directions perpendicular to that one, and dilating it as a whole. Most of the information about what it does to electrical resistance is owed to Bridgman. Usually, but not always, tension increases the resistance to current-flow along the direction of the stress. The exceptions are bismuth and strontium. Comparing the data about the effects of pressure and of tension, we see that Bi and Sr are exceptions to the common rules for both, while Li, Ca and Sb are exceptions to the usual rule for pressure but not to the usual rule for tension. This helps to show why it is so difficult to set up a thoroughly satisfactory theory of conduction in metals.

By melting a substance its density can be altered without altering either its temperature or its pressure; of course, the balance of interatomic forces is also altered in some mysterious but very potent way. Melting a solid usually brings about a decrease in density; the solid sinks in the liquid; but there are exceptions (bismuth, antimony, gallium). The conductivity always changes in the same sense as the density; hence for most metals the solid is more conductive than the liquid, but bismuth, antimony, and gallium have greater resistances frozen than molten. This is one of the few rules in this field to which no exceptions have yet been discovered. The observed values of the ratio (resistance of liquid)/(resistance of solid), when tabulated and examined, show a tendency to cluster about values which are ratios of simple integers, such as 2:1, 1:3, 1:4. It would probably require a careful and expert analysis to show whether this tendency is more pronounced than a quite random distribution might reasonably be expected to display. Mercury has the highest ratio of all, 4:1.

Other agencies which are harder to measure or control may have distressingly great effects on the conductivity of a metal. The various metallurgical processes, annealing, cold-working and the rest, affect the resistance; sometimes the sign of the change can be explained by saying that the process has caused the many small crystals forming the metal to fuse into a few large ones, diminishing the resistance offered by the intercrystalline partitions; sometimes this explanation fails to work. Impurities may have a serious effect; for example Bridgman remarks of bismuth that "a fraction of a per cent. of lead or tin may change the temperature-coefficient from positive to negative and increase the specific resistance severalfold." Often impurities betray themselves by an abnormally low temperature-coefficient of the metal; this means that the absolute rate of increase is unusually small compared to the value of the resistance itself. This is so generally the case that a value of temperature-coefficient which (at 0° C.) is much below, say, .004 is usually taken to mean that the sample of metal under investigation is impure; and the "standard" values for individual metals set down in tables have often taken sudden jumps upward, when better-purified samples became available for measurements. For this reason I laid more stress, in a preceding paragraph, on the values which far exceed .00365 rather than the values which fall far below it. A metal contaminated by a small admixture of another metal may be regarded as the limiting case of an alloy. There is an enormous literature of the electrical behaviour of alloys, and some of the results can be extended to this limiting case. It is found, for example, that if two metals A and B form mixed crystals with one another, an alloy formed by mixing a small percentage or a fraction of one per cent of A into B, has a surprisingly greater resistance than B; and vice versa. The temperature-coefficient of the alloy is on the other hand much smaller than that of the metal, and may even be negative. Thus, although an alloy of this type may seem to be as thoroughgoing a metal as either of the pure elements of which it is made, it has a thoroughly anomalous electrical behaviour; and the alloys as a whole, instead of assisting us to understand conduction in metals, contribute generously to the already abundant supply of difficulties. It remains to be seen whether the measurements upon single crystals of metals, which are being published at a steadily-increasing rate, are going to clarify the situation or increase the perplexity.

While I have left unmentioned a large number of the phenomena which a theory of conduction must be required to explain, the few which I have described will give quite an adequate basis for beginning a discussion of some of the extant theories. It must be conceded at once that the situation is bad. Perhaps there is some set of assumptions or of postulates by which the whole chaotic crowd of phenomena can be unified into a harmonious system; but if so, no one has yet formulated it. The theories, such as they are, may be divided into two groups: theories in which the electrons are supposed to move freely within the atoms and be stopped when they reach an interspace, and theories in which the electrons are assumed to move freely within the interspaces and be stopped when they collide with atoms. Those of the first kind start out with the advantage of being better adapted to the usual effect of pressure on resistance; most metals become more conductive when compressed, as if conduction were assisted by squeezing the atoms closer together. Still the oldest, the best-known, and the most highly elaborated of all the theories belongs to the second kind. This is the one formally known as the electron theory of metallic conduction, or more briefly as the electron theory of metals, and quite commonly as the "classical" theory of conduction (it does not take an idea so long to become "classical" in physics as it does in the arts). Founded by Riecke and by Drude in the closing years of the last century, it was developed by Lorentz and has since been worked over by Planck, Wien, Bohr, and other savants of the first eminence. Its popularity is largely due, I suspect, to the fact that it can be formulated with great if specious exactness: that is to say, as soon as a few definite assumptions are made (such as the simple, if unplausible, assumptions that the atoms are big elastic spheres and the electrons little ones), numercial consequences can be calculated with any degree of precision. In this respect most of the competing theories are sadly defective. Two or three of the numerical deductions made from simple auxiliary assumptions have agreed rather well with experimental data; and they have contributed to the feeling that there must be some kernel of truth in the mathematics, even if not in the physics of the thing, although it breaks down in so many other comparisons with experiment.

Fundamentally the theory is very simple, and has not been helped to any great extent by the more sophisticated mathematics which its emendators have introduced into it. What is observed in electrical conduction is this: when a potential-difference is established across a piece of metal, the electrons do not fall freely clear across it and emerge at the positive end with all the kinetic energy which the P.D. should have communicated to them; they ooze gradually through the metal, heating it as they go along and emerging with no unusual amount of energy, as if they had rubbed along through the metal like heavy particles dropping at constant speed through a gas. "Rubbing along" being a concept foreign to the atomic scale, we have to interpret that each electron falls freely through a small distance, collides with something to which it gives up the energy acquired from the field during its fall, falls again across another short distance, gives up its new quota in another collision, and so forth from side to side of the metal. Furthermore the energy which it gives up at each stoppage must find its way directly or indirectly into the heat of the metal, i.e., into thermal agitation of its atoms. Representing by T the time-interval between two consecutive collisions, by E the fieldstrength in the metal, by e and m the charge and mass of the electron, by U the average kinetic energy acquired by the electron from the field in its free fall between two collisions, we have

$$U=\tfrac{1}{2}(eET/m)^2m. \tag{1}$$

If there are n electrons in unit cube of the metal, and each is stopped $1/T$ times in unit time, the rate at which heat appears in the unit cube is nU/T; but this rate is by definition the product of the conductivity σ by the square of the fieldstrength E, hence

$$\sigma=\tfrac{1}{2}\,ne^2T/m \tag{2}$$

The same equation (2) can be reached, if one prefers to think of conductivity as the ratio of current-density to fieldstrength, by considering that during each free fall, the field augments the speed of each electron in the direction of the field-vector by the amount eET/m, which on the average is lost at the collision terminating the fall; so that the result is as if the field imprinted a constant drift-speed equal to $\tfrac{1}{2}eE/m$ upon all the electrons. Multiplying by ne to get the current-density and dividing by E to get the conductivity, we arrive again at (2).

Equation (2) is the fundamental equation of the electron theory of conduction, and indeed of most of the other theories. Let us begin by trying the supposition that the electrons are at rest until the field is applied, and are brought to a full stop at each collision. Represent by l the average distance traversed between collisions. The proposed assumption leads to $T=\sqrt{2ml/eE}$. The conductivity therefore would depend on the fieldstrength, which would violate Ohm's law. Ohm's law being rigorously valid except under extreme conditions (Bridgman found the first slight deviations from it, in gold and silver, at current-densities of the order of 10^6 *amps/cm*^{2}) we have to discard the idea. The lesson is, that the electrons must be supposed to be normally in motion at speeds enormously greater than the speed imparted by the field during a free fall. Let u stand for the natural average speed of the electrons; we have $T=l/u$, and

$$\sigma=\tfrac{1}{2}ne^2l/mu, \tag{3}$$

provided always that $u \gg eET/m$.

This condition is abundantly fulfilled if we make the obvious and appealing assumption that the electrons are moving with the same average kinetic energy as atoms of a gas at the same temperature; in fact, if the free path l is no longer than the average distance between atom-centres, the deviations from Ohm's law should not appear even under such extreme circumstances as those of Bridgman's experiments. Making therefore this assumption, which in symbols is $\tfrac{1}{2}mu^2=\frac{3}{2}kT$, we find

$$\sigma=\tfrac{1}{2}\frac{e^2}{\sqrt{3km}}\;\frac{nl}{\sqrt{T}}. \tag{4}$$

Not much attention should be paid to the numerical factor, which would be slightly different if we should assume Maxwell's law of distribution for the velocities of the electrons; the essential factor is the last one, $nl/\sqrt{T}$. Examining (4) in the light of the fact that the conductivity of most metals decreases distinctly more rapidly than $1/\sqrt{T}$=in fact, as rapidly as $1/T$ or still more so—as the temperature increases, we see that the product nl will have to be supposed to vary with temperature. It seems natural to suppose that l depends altogether on the distance between adjacent atoms, which varies comparatively little with temperature, and anyway varies in the wrong direction for the purpose of the theory; so that the burden of accounting for the proportionality of σ to the first or a higher power of $1/T$ must be laid upon n.

Now it has occurred to a number of people that the free electrons are dissociated from the atoms, and the number of free electrons is given by the degree of dissociation, which in turn should vary with the temperature in a manner prescribed altogether by the amount of work necessary to remove an electron from an atom into the (presumed) interspace where it plays about freely. But we should certainly expect that this work would be positive, as it is for the extraction of electrons from free atoms; in which case the degree of dissociation and the number of free electrons should increase with temperature. The theory is therefore adapted to explain a resistance which decreases steadily with increasing temperature, as do the resistances of some non-metallic elements; it is adapted to explain a resistance which at first diminishes and then, as the temperature increases further, goes through a minimum and rises, for the decrease in the factor $l/\sqrt{T}$ finally predominates over the increase in the factor n; it is not adapted to explain a resistance increasing with temperature over the whole range, as do those of the metals. One might assume that the work of extracting an electron from an atom inside the metal is negative. This is essentially the alternative embraced by Waterman, who postulates that the work in question is a function of temperature, of the form $W=W_o-cT$, $c>0$. For metals W_o is to be chosen negative or zero, so that W shall be negative throughout; for non-metallic elements W_o is to be given some positive value, so that W shall change in sign at some point in the temperature-range. This unusual theory must be judged by its effectiveness; that it should reduce conduction in all elements, metallic and non-metallic alike, to a phenomenon of a single type is a feature appealing strongly in its favor; but Noyes' curves of resistance versus temperature for graphite did not agree with its demands in a satisfactory manner.

The assumption underlying (4) has however involved us in a collateral difficulty. If we believe that the n free electrons per cc. of the metal have an average energy $\frac{3}{2}kT$ and a total kinetic energy $\frac{3}{2}nkT$, we are certainly forced to admit that when the unit cube of metal is heated through 1° the electrons must take their share $\frac{3}{2}nk$ of the heat imparted to it; but the specific heat of most metals is such that it seems that the atoms must take it all and leave none over for the electrons. If we evade this difficulty by assuming n to be quite small compared with the number of atoms per cc., a few per cent. of it or less, we lose certain numerical agreements which will be mentioned later, and we have also to make l quite large, amounting to several times the distance between adjacent atoms;

yet all the tendency of modern atomic theory is to make it seem likely that the atoms fill almost the whole space within the metal.

Another way to avoid the difficulty with the specific heats consists in assuming that the high natural speed with which the electrons fly about is altogether independent of temperature; the burden of making σ as expressed in (3) vary in the proper manner with temperature is then laid upon l, which, Wien suggested, should be supposed to vary inversely as the amplitude of vibration of the atoms—that is, a free electron collides with an atom only if and when it is in vibration, and the chance of a collision increases with the amplitude of the vibration. The variation of resistance with pressure may then be explained, so far as the usual sign goes, by saying that when an ordinary metal is compressed the amplitude of oscillation of its atoms diminishes, though the temperature remain the same; the frequency of oscillation must then vary inversely as the amplitude, to keep the average energy of oscillation constant; there is some reason for expecting this to happen. Bridgman's theory somewhat resembles this one, except that the electrons are supposed to glide through the atoms and collide with the gaps; gaps between atoms are comparatively unusual, and occur chiefly when two atoms are vibrating with great amplitudes in opposite senses, so that the variation of conductivity with pressure again has the proper sign. But to explain the behavior of the three metals of which the resistance increases with pressure and with tension, Bridgman went back to the idea that in these the electrons glide through the interspaces.

As I have given only the phenomena of conduction which the electron-theory explains with difficulty, I must in justice mention the ones on which its reputation chiefly depends. In the first place it is a theory of thermal conduction as well as electrical conduction; the electrons in the hotter part of a metal maintained at an uneven temperature are assumed to have a greater average energy than the electrons in the cooler part, so that they diffuse down the temperature-gradient and realize a convection-current of heat. The theory leads to as definite a numerical value of the one conductivity as of the other, and the ratio of electrical to thermal conductivity is predicted as

$$\frac{\lambda}{\sigma}=2\left(\frac{k}{e}\right)^2.T, \tag{5}$$

a universal constant for all metals, multiplied into the absolute temperature, and devoid of the quantities n and l which have caused us so much trouble. This is one of the predictions which is nearly enough true to be impressive; the ratio $\lambda/\sigma T$ does indeed vary surprisingly little over a wide range of metals at room-temperature and over a fairly wide range of temperatures for each of many metals. It is usually somewhat larger than the predicted value (5); but this can be conveniently explained by saying that there must be an additional mechanism for transmitting heat, something in the nature of the elasticity of the substance, which superposes its conducting-power upon the conducting-power of the electrons, and so inflates the numerator of the ratio in (5). The reason for supposing such an extra mechanism is primarily that there must be some such mechanism to perform the thermal conduction in substances which are electrical insulators. No element conducts heat as badly as sulphur and boron conduct electricity; and if we imagine a special elastic mechanism for conducting heat in boron and sulphur, we can hardly deny it to copper and silver. Bridgman found that for six metals out of eleven tested, the thermal conductivity decreased when high pressure was applied, although the electrical conductivity increased. We must hope to find an explanation for this anomaly in the behaviour of the elastic mechanism; likewise an explanation for the deviations from (5) which occur at high and at low temperatures. In theories such as the one mentioned over Wien's name in the last paragraph, in which the average *vis viva* of the electrons is supposed not to vary from a hotter place in a metal to a cooler place, we have to lay the entire burden of thermal conduction upon the elastic mechanism. This makes it difficult to explain the universal relation (5).

Another striking feature of the theory is that Lorentz succeeded in deducing the Rayleigh-Jeans radiation-law from it. He obtained from it an expression for E, the radiant emissivity of a thin stratum of metal, as a function of temperature T of the metal and wavelength λ of the radiation; another for A, the absorbing-power of the metal, likewise a function of T and λ; divided the first by the second, and obtained a definite quotient. By Kirchhoff's thermodynamic laws, E/A is equal to E_o, the radiant emissivity of a perfectly black body. The expression deduced by Rayleigh and by Jeans for E_o and the expression deduced by Lorentz for E/A are identical. Lorentz assumed that the collisions of the electrons with the atoms (or whatever it is they collide with) are very short in duration compared with the intervals of free unaccelerated flight from one collision to the next, and that the speeds of the electrons are distributed according to Maxwell's law about the mean value corresponding to the mean energy $3kT/2$. He also made certain assumptions which restrict the validity of his expression for E/A to radiations of great wave-length; the Rayleigh-Jeans expression for E_o is restricted in exactly the same way. At least as much, it seems, should be demanded from any theory of conduction offered in competition with the "classical" one.

The conception of free electrons in metals also gives a beautiful qualitative explanation of the thermoelectric effects, although unfortunately it does not do very well as a quantitative theory. If in two metals at a certain temperature the densities of free electrons are different—n_1 free electrons per cc. in one and n_2 in the other—and these two metals are brought into contact with one another, electrons will flow from the one where the density is greater into the one where it is less; and this flow will continue until arrested by a counter-electromotive-force V, of which the equilibrium-value can be shown, in any one of a variety of ways, to be

$$V=\frac{kT}{e}ln(n_1/n_2)$$

Such an electromotive force would account for the Peltier effect; and conversely, if the theory were correct, measurements of the Peltier effect between two metals at a given temperature and pressure would give the ratio between the densities of free electrons in the two metals under the specified conditions. Such data, combined with data on conductivity interpreted by such an equation as (4), should give information about the free paths l_1 and l_2 in the metals. The Thomson effect is more difficult to deal with, as thermal equilibrium does not prevail; however it can be seen that there will be a counter E.M.F. in an unevenly-heated metal. Measurements on the Peltier and Thomson coefficients for many metals, over wide ranges of temperature and pressure, would be very valuable; but they are so extremely hard to make even under the best of conditions, that the outlook for obtaining a really extensive set is unpromising. Possibly there is a better chance with the indirect method (determining the first and second derivatives of the curve of thermal electromotive force versus temperature). Such data of the Thomson effect as exist are not helpful to the simple theory.

Another phenomenon which lends itself very readily to explanation by the theory, and so contributes a certain amount of support to it, is the thermionic effect—the spontaneous outflow of electrons through the surfaces of hot metals. (But carbon likewise exhibits it very efficiently, and we must beware of formulating any theory of it which reposes on specific properties of metals not shared by carbon!) To interpret the thermionic effect only one new feature need be added to the theory, and this a feature which in fact was all the time latent in it—the idea that there is a certain fixed potential-difference between the interior of a metal and the region outside of it, resulting in a potential-drop localized in a thin stratum at the surface, which an electron within the metal must surmount in order to escape from the metal into a contiguous vacuum. Such a potential-drop would for instance result from a "double layer" along the surface of the metal, a sheet of positive charges within and a sheet of negative charges opposite, parallel, and close to the positive sheet on the outside. It has been pointed out that, since probably half of the orbital electrons belonging to the atoms at the frontier of a metal lie outside the plane containing the nuclei of these atoms, they with the nuclei constitute a sort of double-layer; it has also been suggested that after a certain number of electrons issue from the metal, they are held as an electron-atmosphere above it by the forces due to the distribution of residual positive charge within the metal (Kelvin's electrical-image conception), and the electron-atmosphere with the positive surface-charge together form a double-layer. However we may conceive this double-layer, it is obvious that if we postulate free electrons within the metal, we must also postulate a barrier in the shape of an opposing potential-drop between the metal and the exterior world to keep the electrons from wandering away.

Designate this potential-drop by b, so that eb is the energy which an electron must give up in traversing it from inside to outside. Assume further (disregarding the old specific-heat difficulty) that the velocities of the electrons inside the metal are distributed isotropically in direction, and according to Maxwell's distribution-law in speed, with the mean kinetic energy $\frac{3}{2}kT$ appropriate to the temperature T of the metal. Imagine the metal surface to occupy the plane $x=0$, metal to the left and vacuum to the right. Consider the electrons which come from within the metal and strike unit area of the boundary in unit time; those of them which have velocities of which the x-component lies between u and $u+du$ are in number equal to

$$dI=\frac{nu}{\sqrt{2\pi kT/m}}e^{-\frac{mu^2}{2kT}}du, \tag{6}$$

n meaning as heretofore the number of electrons per unit volume of metal. The total number which strike unit area of the boundary from within is equal to the integral of this expression from $u=0$ to $u=\infty$, which is

$$I_o=n\sqrt{kT/2\pi m}. \tag{7}$$

Those which escape are those for which $\frac{1}{2}mu^2$ exceeds eb; we obtain the number of them by integrating (6) from $u=\sqrt{2eb/m}$ to $u=\infty$, and find

$$I_e=n\sqrt{kT/2\pi m}\, e^{-eb/kT}. \tag{8}$$

This, supposing n and b to be independent of temperature, is Richardson's well-known formula for the saturation-current from a hot body as function of temperature. All of the multitudinous observations agree with it; but this does not mean so much as might be thought, for the experts inform us that all the data, no matter how accurately taken, would agree quite as well with a formula in which T, or T^2, or even T^0, stood in the place of the factor $T^{1/2}$ by which the exponential is multiplied. Incidentally this would permit us to make n vary as some small power of temperature, such as the inverse square root, if we chose to make the resistance-temperature relation in (4) agree with experiment at such a price. Or if we assume n independent of temperature, we can calculate it from measurements on thermionic saturation-currents. The measurements usually give for n values of the order of magnitude of the number of atoms per unit volume.

What is more definitely significant is, that the velocities of the emerging electrons are actually distributed in a manner compatible with the assumptions made. Let us enquire how many of the electrons issuing from unit area of the metal have velocities of which the x-component lies between u and $u+du$. These are the very same electrons which struck the surface from within, having velocities of which the x-component lay between u' and $u'+du'$; u' and du' being related to u and $u+du$ by the equations:

$$\tfrac{1}{2}mu^2+eb=\tfrac{1}{2}m(u')^2, \quad u'du'=udu. \tag{9}$$

The number of these electrons is by (6)

$$dI'=\frac{nu'}{\sqrt{2\pi kT/m}}\, e^{-\frac{(mu')^2}{2kT}}\, du', \tag{10}$$

which by virtue of the relations (9) reduces to

$$e^{-\frac{eb}{kT}}\cdot \frac{nu}{\sqrt{2\pi kT/m}}\, e^{-\frac{mu^2}{2kT}}\, du, \tag{11}$$

which is identical with (6) except for a constant factor; which means in turn that the distribution-function of the emerging electrons is identical with the distribution-function of the internal electrons, being in fact the Maxwell distribution-function with the same mean kinetic energy $\frac{3}{2}kT$. The argument as given proves the point only for the distribution in the velocity-component u; but the distribution-functions in v and w, the velocity-components parallel to the boundary of the metal, are unaffected by the double-layer, since v and w for any particular electron are unaffected by the passage through it; and since it is the essential feature of the Maxwell distribution-law that the distributions in v and w are identical for each and every value of u, the conclusion follows as stated. Nevertheless it does sound paradoxical.

This conclusion has been verified repeatedly by experiment. Richardson began by simulating the simple mathematical conditions of infinite plane electrodes as closely as practicable; he inserted a small flat incandescent surface in an aperture in the middle of a large flat cold plate, charged the two to the same potential, and placed opposite and parallel to them a large flat collecting-electrode. Charging this latter to various potentials V inferior to the potential of the emitting surface, he plotted the electron-current which it received as function of V; this is the distribution-function of the speed u of equation (6) and the following equations translated into terms of the corresponding kinetic energy $\frac{1}{2} mu^2$ as independent variable. To ascertain the distribution-functions in v and w he isolated a small area of the collecing-electrode, moved it to and fro in a plane parallel to the plane of the emitting surface, and measured the current into it in its various positions. Many measurements have since been made upon the currents into cylindrical collectors from hot wires stretched along the axes of the cylinders; it is somewhat more difficult to write out the formula for the expected relation between current and retarding-potential, but the experimental conditions are much more under the experimenter's control. All these investigations have confirmed the theorem, except a single discordant one which was later explained away; the strongest verification is furnished by the experiments of Germer, whose precautions of preparation and accuracy of measurements far surpassed everything that had gone before.

The evidence thus is quite favorable to the idea of an electron-gas within the metal with its electrons moving with velocities as prescribed by Maxwell's distribution-law, and kept from diffusing away by a double-layer covering the surface. Other evidence for the existence of a double-layer is furnished by the photoelectric effect and by the existence of contact-potential-differences. When light of frequency ν falls upon a metal, electrons emerge from it with velocities which are distributed in a manner quite distinct from Maxwell's distribution and have nothing to do with the temperature of the metal. The kinetic energies of some of the electrons attain a certain upper limit W_m, but none surpasses it; W_m is a linear function of ν given by the equation

$$W_m = h\nu - P, \tag{12}$$

h being Planck's constant, P a positive constant characteristic of the metal. This is an exceedingly strong intimation that each of the emerging electrons, while still inside the metal, suddenly absorbed a quantum of energy $h\nu$ from the light and departed with it, giving up a fixed quantity P in passing through the surface. (Those which issue with energies clearly less than W_m can be supposed to have started distinctly beneath the surface and to have lost additional energy in struggling through the metal to it). Translating P into potential-drop, we see that it represents the potential-difference or the "strength" of the surface double layer. It may be determined by measuring W_m for light of various frequencies, plotting it against frequency, and extrapolating the resulting straight line to its intersection with the axis of frequencies. Or it may, in principle, be determined by plotting the photoelectric current as a function of frequency, and extrapolating the curve to its intersection with the axis of frequencies, where no electrons escape and the photosensitiveness ceases; but curves are not so easy to extrapolate as straight lines, and there are some anomalous results which are still unexplained.

It would seem an easy matter to measure the strength of the double-layer by both photoelectric and thermionic methods upon a single substance. But it is rather difficult; for one reason, the substances for which the photoelectric currents are easy to produce and measure are precisely the metals upon which good thermionic measurements are next to impossible, and vice versa. The best photoelectric measurements have been made upon the alkali metals, which are very sensitive to visible light; but they cannot be formed into wires, and volatilize furiously when heated enough to produce an important thermionic effect, filling the evacuated tube with dense vapors which ruin the accuracy of the measurements. The best thermionic measurements have been made upon platinum and tungsten, which are not sensitive at all to visible light, and begin to be sensitive far out in the ultraviolet where experiments with radiation are difficult. Furthermore there is the capital difficulty that the photoelectric measurements must be confined to temperatures where the thermionic current is imperceptible; if one were to irradiate an incandescent tungsten filament the extra current of photoelectrons would be too small to notice. If we assume outright that P does not vary greatly from room-temperature up to the temperatures of incandescence, and therefore compare photoelectric data upon cool metals with thermionic data upon the same metals when hot, we find that there is a fairly good agreement. Values of the thermionic constant b between 4 and 5 volts correspond to photoelectric sensitiveness commencing between 3,100 and 2,500 Angstrom units, and this correctly describes the behavior of several of the heavy high-melting-point metals; photoelectric sensitiveness extending well up into the visible spectrum, such as the alkali metals display, corresponds to values of P/e of the order of 2 volts and lower, and such values are indicated by the thermionic experiments made upon sodium and potassium by Richardson under the inevitably bad conditions.

Contact-potential-difference, one of the longest known of all electrical phenomena—Volta discovered it—agrees admirably with this interpretation of the photoelectric constant P. Imagine that we have pieces of two metals, potassium and silver for example, which are drawn out and welded together at one end, and at their other ends are spread out into plates and face one another across a vacuous space. We know that the opposing faces behave as if they were at essentially different potentials, the potential-difference V between them being characteristic of the two metals and independent of the size or separation of the opposing faces. Yet this potential-difference V is not equal to, is indeed usually much greater than the potential-difference between the interiors of the metal across the welded joint, which is deduced from the Peltier effect. The only way to resolve the contradiction is to assume that it is the region just outside the potassium which differs by V from the region just outside the silver; the metals themselves are at nearly the same potential, but there is a double-layer at the surface of each which establishes a fixed potential-drop between it and the vacuum. Representing by P_1/e and by P_2/e the voltage-drops at these two double-layers, by M the potential-difference between the interiors of the metals as inferred from the Peltier effect, by V the potential-difference between the regions just outside the metals which we identify with the contact potential difference, we find

$$P_2/e - P_1/e = V + M \qquad (13)$$

in which M is so small compared with the other terms that henceforth we will leave it out.

Now imagine that light of a high frequency ν_o falls upon the potassium; it elicits electrons of which the maximum energy at emergence is $h\nu_o - P_1$; these highest-speed electrons arrive at the silver plate with energy $(h\nu_o - P_1/e - V)$, having had to overcome the additional potential-drop V in passing from the region just outside the potassium to the region just outside the silver. (The reader can make the changes in language required if V happens to be of the sign corresponding to a potential-rise). From (13) we see that this energy of arrival is equal to $(hv_o—P_2/e)$—an expression from which P_1, the only quantity characterizing the irradiated metal, has fallen out! Therefore the electrons arrive at the silver plate with the same maximum speed, whether the irradiated metal be potassium, sodium, silver, or any other metal! (unless we hit upon a metal for which $h\nu_o < P_1/e$, in which case we shall never get any at all).

This experiment is usually performed by putting a battery between the silver and the irradiated metal, and adjusting its E.M.F. until the fastest electrons are just turned back before reaching the silver; this is known as "determining the stopping potential." If our interpretation of contact-potential-difference is correct, the stopping-potential must be independent of the irradiated metal, and depend only on the material of which the collecting-electrode is made; further, the difference between the stopping-potentials observed with two different metals as collecting-electrodes should be equal to their contact potential difference. These predictions have been verified in several sets of experiments, notably by Richardson and Compton. Millikan developed the interesting theoretical consequences which they suggest. There should be similar relations involving thermionic currents; observations confirming them have been made, but not so extensively published; they are more difficult to make with accuracy because the thermionic electrons have no definite recognizable maximum velocity.

We seem to have marshalled a formidable amount of evidence in favor of the electron-theory of conduction with the associated idea of the surface double-layer. Yet it would be misleading not to point out that an equation quite as satisfactory as (8) in representing the thermionic current as function of temperature can be deduced by reasoning in an entirely different fashion from entirely different postulates. This, the thermodynamical method of speculating about the thermionic effect, was originated by H. A. Wilson; it consists essentially in assuming a thoroughgoing analogy between the outflow of electrons from a hot metal and the evaporation of molecules from a solid or a liquid. We know that if an evacuated chamber is partly filled with liquid water or solid $CaCO_3$, the remaining space inside the chamber is quickly pervaded with H_2O or CO_2 molecules composing a gas, its pressure and density being determined absolutely by the temperature T. We infer that if an evacuated chamber, with its walls made of some insulating substance, contains a piece of metal and is heated to a high temperature, the whole evacuated space will be pervaded with electrons composing a gas, its pressure p and density n being determined absolutely by the temperature of the system, T. We must assume that the electron-gas outside the metal conforms to the ideal-gas law

$$p = nkT \tag{14}$$

and we shall also presently assume that its specific heats have the values characteristic of monatomic ideal gases,

$$C_v = \frac{3}{2}Nk, \qquad C_p = \frac{5}{2}Nk. \tag{15}$$

I use n to represent the number of electrons per unit volume of the gas, as the number within the metal no longer enters in any way into the reasoning; N to represent the number in a gramme-molecule (Avogadro's constant). These are the only assumptions which involve a kinetic theory in any way.

Imagine now a wire of which one end projects into an evacuated chamber of the sort described, maintained at T, and the other into another such chamber maintained at $T+dT$. We consider a process which consists of increasing the volume of the first chamber by just enough to require N additional electrons to come out of the wire to fill the additional space, and simultaneously decreasing the volume of the second chamber by just enough to crowd N electrons into the wire; so that in effect N electrons are transferred from the one chamber to the other through a wire of which the two ends are at temperatures $T+dT$ and T. This process will be carried on reversibly. Designate by L the heat which must be imparted to the metal at T, to remove one electron from it under the circumstances of the experiment; by sdT the heat which is absorbed when one electron is transferred through the metal from a point where the temperature is T to a point where the temperature is $T+dT$. s is the coefficient of the Thomson effect, referred to a single electron instead of a coulomb. L contains a term kT, which corresponds to the mechanical work done in forcing back the walls of the chamber to make place for the evaporated electron-gas. Subtracting it we obtain (L—kT), to be called $e\phi$, as the actual energy expended in putting the electron across the boundary of the metal.*

In the process which I have just described, the *input of heat* consists of the following terms: NL which goes to extract the N electrons from the metal in the first chamber, $-N\left(L+\frac{dL}{dT}dT\right)$ which is liberated when N electrons condense into the metal in the second chamber, and $-NsdT$ which is absorbed by the electrons in travelling through the wire. The *output of work* is NkT during the expansion of the first chamber, $-NkT-NkdT$ during the contraction of the second chamber. The *input of entropy* is NL/T during the evaporation in the first chamber, $-N\left(L/T+\frac{d(L/T)}{dT}\right)dT$ during the condensation in the second chamber, and $(-Ns/T)dT$ during the flow of electrons through the wire.

We now complete the cycle by changing the pressure and temperature of the gramme-molecule of electron-gas in the first chamber from p, T to $p+dp$, $T+dT$, after which it becomes equivalent with the gramme-molecule in the second chamber at the beginning of the process. Calculated in the usual way—isothermal contraction at T from p to $p+dp$, isobaric expansion at $p+dp$ from T to $T+dT$—we find: *input of heat*, $\frac{5}{2}NkdT-NkT[d(\ln p)/dT]dT$; *output of work*, $NkdT-NkT[d(\ln p)/dT]dT$; *input of entropy*, $\frac{5}{2}(Nk/T)dT-Nk[d(\ln p)/dT]dT$.

The two processes together constitute a complete reversible cycle. We therefore equate the sum of the inputs of entropy to zero, and obtain:

$$\frac{e\phi}{T}-e\frac{d\phi}{dT}-s+\frac{5}{2}k-kT\frac{d(\ln p)}{dT}=O. \tag{16}$$

and equate the difference of the inputs of heat and the outputs of work to zero, which gives:

$$-e\frac{d\phi}{dT}-s+\frac{3}{2}k=O \tag{17}$$

* This definition suggests a thermal method of measuring L, which has several times been put into practice. The experiments are difficult and the data must be corrected for many influences, but the best results indicate that $(L-kT)$ is approximately equal to eb of (8). The data of Darisson and Germer indicate a slight difference, which may be an important test of suggested theories of conduction.

and combine the equations into

$$\frac{e\phi}{T}+k=kT\frac{d(lnp)}{dT} \tag{18}$$

which integrated, yields

$$p=ATe^{\int\frac{e\phi}{kT^2}dT}. \tag{19}$$

We still have to make the bridge between this formula, which relates to the pressure of the electron-gas in equilibrium with the metal, and the quantity actually observed, which is the saturation-current out of the metal surface in an accelerating field. In the equilibrium-state, the number of electrons which issue from the metal is equal to the number which, coming from the external electron-gas, strike its boundary and do not rebound. This is indisputable; to make it useful we have to make two new assumptions: one, that the number of electrons which issue from the metal is the same in an accelerating field as in the equilibrium-state; the other, that no electrons rebound from the surface. The first assumption had to be made in the preceding deduction—that is, we had to assume tacitly that the uncompensated outflow of electrons through the surface of the metal did not appreciably distort the Maxwell distribution within; the second is a drawback peculiar to the thermodynamic method. Accepting these two assumptions along with all their predecessors, we finally reach the expression for the number of electrons emitted per unit area per unit time from the surface of the hot metal:

$$I=CT^{1/2}e^{\int\frac{e\phi}{kT^2}dT}. \tag{20}$$

This is the equation for the thermionic saturation-current attained by the thermodynamical reasoning.

Let us finally try some hypotheses about the variation of ϕ with temperature: for a first one, the hypothesis ϕ=constant. The general equation becomes

$$I=CT^{1/2}e^{-\frac{e\phi_0}{kT}}, \tag{21}$$

which is perfectly identical with (8) which was deduced from the electron-theory with the additional assumption of a double-layer independent of temperature. We cannot however freely make an assumption like this, for our equation (17) shows that an assumption about $d\phi/dT$ implies, and conversely is implied by, an assumption about the value of the Thomson coefficient s. In making ϕ independent of temperature we in effect assumed that the Thomson coefficient has the value $s=\frac{3}{2}k$ (per electron), which happens to be precisely the value demanded (and vainly demanded) by the electron-theory of conduction. If on the other hand we choose to accept from the experiments the fact that s is extremely small compared to $\frac{3}{2}k$, the equation (16) compels us to set

$$\phi=\frac{3}{2}kT/e+\phi_o. \tag{22}$$

Inserting this into (19) we obtain

$$I=CT^2e^{-\frac{e\phi_o}{kT}}, \tag{23}$$

which is commonly known as the T^2-law, and is at the moment the favorite way of expressing the variation of thermionic current with temperature. As I said earlier, experiment is thus far powerless to distinguish between (8), (20) and (22).

This brief and superficial sketch of the thermodynamic argument is meant partly to familiarize the reader with the T^2 formula, and partly to show that the observations upon the dependence of thermionic current on temperature do not necessarily sustain the particular type of theory which has figured most in these pages, as against its rivals actual or conceivable. Of course it would be unjustifiable to say that any argument of the thermodynamical type is *ipso facto* stronger than any argument based on a physical model. It may be true that the laws of thermodynamics are valid everywhere without exception; but it is certainly true that in any particular case it is extremely difficult to feel sure just how they should be applied to arrive at absolutely binding conclusions. In this case, for instance, we have assumed as both possible and reversible a process which no one has ever carried through, and no one, in all likelihood, ever will; and in the course of analyzing the transfers of energy between the system and the external world in this imagined process, we have classified some as transfers of heat and some as transfers of mechanical work, and possibly ignored yet others, so that the analysis requires careful thought and has in fact been made in different ways by different authorities. There is for example the problem of the allowance to be made for work done in transferring the electrons from place to place against electromotive forces, which might or might not be *nil* when summed around the complete cycle; H. A.

Wilson has recently made a specific assumption regarding these. For still further subtleties Bridgman's theoretical articles may be consulted. I must however add that an extension of the thermodynamical argument, with the assistance of Nernst's "third law of thermodynamics," leads to the conclusion that the constant C of equation (23) should have for all elements, if not indeed for all substances, the same universal value, calculable in terms of certain universal constants. There is some evidence that this may be true for emission from pure elements. Were it so, the result would be of fundamental importance; but another article almost as long as this one would be required to explain it properly.

The general tone and character of this article will probably leave the final impression that the electrical behaviour of solids is an utterly confused and chaotic department of physics, a hopeless entanglement of incongruous rules diversified by numberless exceptions. I fear that this impression—except perhaps for the hopelessness of the situation—is substantially the correct one. In fact this presentation has put the state of affairs in rather too favorable a light, for I have passed over a number of the perplexities. I have scarcely mentioned the thermoelectric effects, or spoken of the complexities of the photoelectric effect, or of the emission of electrons from metals which are bombarded by other electrons or by ionized atoms; and I have not mentioned at all the galvanomagnetic and thermomagnetic effects, the most baffling and bewildering of all. In fact it seems only too probable that if one should succeed in erecting a theory by which all the phenomena I have described could be brought into one coherent system, some galvanomagnetic effect would be lying in wait for it to bring it to the dust. Clairaut is said to have been saddened by feeling that Newton had discovered all the laws of celestial mechanics, leaving nothing for men born after him to do except to improve the methods of calculation. Ambitious students of physics who, through too exclusive a study of the radiations from atoms, may have come to feel in the same way about Bohr, should find consolation in contemplating the present status of the Theory of Conduction in Solids.

LITERATURE

The chief recent compilation of data upon conduction in solids is Koenigsberger's article in Graetz' *Handbuch der Elektrizitat*. K. Baedeker wrote an excellent short account of the data and the theories, entitled *Die elektrischen Erscheinungen in metallischen Leitern*, which although published in 1911 is not yet superseded. Bidwell's paper on germanium is in Phys. Rev. (2) *19*, pp. 447–455 (1922); Noyes' article on carbon in Phys. Rev. (2) *24*, pp. 190–199 (1924). The investigations on supraconductivity are reported chiefly in the *Leyden Communications*; Crommelin

has given a comprehensive account of them in Phys. ZS. *21* (1920), with a bibliography of all of the work; two or three subsequent communications are reviewed in Science Abstracts. Bridgman's work on the effect of pressure and of tension on the electrical and thermal conductivities of the elements is printed chiefly in the *Proceedings of the American Academy of Arts and Sciences* from 1917 onward, with occasional announcements in the Physical Review, where also his theoretical papers are published (Phys. Rev. *14*, pp. 306–347 (1919); *17*, pp. 161–195 (1921) and *19*, pp. 114–134 (1922). For the effect of melting, consult Bridgman's papers, and one by von Hauer, in Ann. d. Phys. *51*, pp. 189–219 (1916).

The "classical" theory of conduction is presented in Lorentz' book *The Theory of Electrons*, which bears his signature as of 1915. Bohr wrote a dissertation upon it which is highly praised by those who have succeeded in reading it in the Danish. Wien's and Planck's modifications of it are published in the *Sitzungsberichte* of the Berlin Academy for 1912 and 1913. In the Philosophical Magazine of 1915 there are a number of articles on the theory by G. H. Livens, like Baedeker a victim of the war. The conception of quantity of free electrons determined by dissociation of atoms is presented by Koenigsberger in Ann. d. Phys. *32*, pp. 170–230 (1910) and Waterman's extension of it is in Phys. Rev. *22*, pp. 259–270 (1923). Some chapters in J. J. Thomson's *Corpuscular Theory of Matter* deal with the theories; in an article in Phil. Mag. *29* (1915) he offers a theory involving an attempt on supra-conductivity, which the others do not touch.

The field of thermionics is thoroughly covered in Richardson's *Emission of Electricity from Hot Bodies* (2d edition, 1921). Subsequent theoretical papers by Richardson are in Proc. Roy. Soc. *A105*, pp. 387–405 (1924) and Proc. Phys. Soc., London, *36*, pp. 383–399 (1924), and one by H. A. Wilson on what I have called the "thermodynamical argument" in Phys. Rev. (2) *24*, pp. 38–48 (1924). For various interpretations of the surface double-layers see Debye, Ann. d. Phys. *33*, pp. 440–489 (1910); Schottky, ZS. f. Phys. *14*, pp. 63–106 (1923); and Frenkel, Phil. Mag. 33, pp.297–322 (1917). Germer's investigation of the distribution-in-energy of thermionic electrons is briefly reported in Science, *42*, 392 (1923), and a fuller account is to be published; Davisson and Germer's determination of L in Phys. Rev. *20*, pp. 300–330 (1922). For the photoelectric measurements establishing equation (12), consult Millikan, Phys. Rev. *7*, pp. 355–388 (1916); for the relation between values of P and contact-potential-difference consult Page, Am. Journ. Sci. *36*, pp. 501–508 (1913) and Millikan, Phys. Rev. *7*, pp. 18–32 (1916). Values of the thermionic constant b are tabulated in Richardson's book and in Dushman's article, Phys. Rev. (2) *21*, pp. 623–636 (1923). Values of the photoelectric constant P are tabulated by Kirchner, Phys. ZS. *25*, pp. 303–306 (1924) and by Hamer (Journ. Opt. Soc. *9*, pp. 251–257 (1924). For the arguments that the constant C of equation (23) is a universal constant, consult the references given by Dushman (*l. c. supra*) and Richardson, Phys. Rev. (2) *23*, pp. 153–155 (1924); for the data, Dushman in Phys. Rev. (2) *23*, p. 156 (1924).

Theorems Regarding the Driving-Point Impedance of Two-Mesh Circuits*

By RONALD M. FOSTER

Synopsis: The necessary and sufficient conditions that a driving-point impedance be realizable by means of a two-mesh circuit consisting of resistances, capacities, and inductances are stated in terms of the four roots and four poles (including the poles at zero and infinity) of the impedance. The roots and the poles are the time coefficients for the free oscillations of the circuit with the driving branch closed and opened, respectively. For assigned values of the roots, the poles are restricted to a certain domain, which is illustrated by figures for several typical cases; the case of real poles which are not continuously transformable into complex poles is of special interest. All driving-point impedances satisfying the general conditions can be realized by any one of eleven networks, each consisting of two resistances, two capacities, and two self-inductances with mutual inductance between them; these are the only networks without superfluous elements by which the entire range of possible impedances can be realized; the three remaining networks of this type give special cases only. For each of these eleven networks, formulas are given for the calculation of the values of the elements from the assigned values of the roots and poles.

1. Statement of Results

The object of this paper is, first, to determine the necessary and sufficient conditions that a driving-point impedance [1] be realizable by means of a two-mesh circuit consisting of resistances, capacities, and inductances, and second, to determine the networks [2] realizing any specified driving-point impedance staisfying these conditions.

These necessary and sufficient conditions are stated in the form of the following theorem:

Theorem I. Any driving-point impedance S of a two-mesh circuit consisting of resistances, capacities, and inductances is a function of the time coefficient $\lambda = ip$ of the form

$$S = H\frac{(\lambda-\alpha_1)(\lambda-\alpha_2)(\lambda-\alpha_3)(\lambda-\alpha_4)}{\lambda(\lambda-\beta_2)(\lambda-\beta_3)} \tag{1a}$$

$$= \frac{a_0\lambda^4+a_1\lambda^3+a_2\lambda^2+a_3\lambda+a_4}{b_1\lambda^3+b_2\lambda^2+b_3\lambda}, \tag{1b}$$

* Presented by title at the International Mathematical Congress at Toronto, August 11–16, 1924, as "Two-mesh Electric Circuits realizing any specified Driving-point Impedance."

[1] The driving-point impedance of a circuit is the ratio of an impressed electromotive force at a point in a branch of the circuit to the resulting current at the same point.

[2] The networks considered in this paper consist of any arrangement of resistances, capacities, and inductances with two accessible terminals such that, if the two terminals are short-circuited, the resulting circuit has two independent meshes. Thus the impedance measured between the terminals of the network is the same as the driving-point impedance of the corresponding two-mesh circuit. Throughout the paper this distinction will be made in the use of the terms "network" and "circuit."

where $$H \geq 0,\ \alpha_1+\alpha_2 \leq 0,\ \alpha_1\alpha_2 \geq 0,\ \alpha_3+\alpha_4 \leq 0,\ \alpha_3\alpha_4 \geq 0,\ \beta_2+\beta_3 \leq 0,\ \beta_2\beta_3 \geq 0, \tag{2}$$

and $$\begin{aligned} &b_1^2(a_3^2-4a_4d)+b_2^2[(a_2-d)^2-4a_0a_4]+b_3^2(a_1^2-4a_0d) \\ &-2b_1b_2[a_3(a_2-d)-2a_1a_4]-2b_1b_3[a_1a_3-2d(a_2-d)] \\ &-2b_2b_3[a_1(a_2-d)-2a_0a_3]=0, \end{aligned} \tag{3}$$

for all values of $d \geq 0$, *provided*

$$-a_4b_2^2+a_3b_2b_3-db_3^2 \geq 0, \tag{4}$$

$$-a_0b_3^2+(a_2-d)b_3b_1-a_4b_1^2 \geq 0, \tag{5}$$

$$-db_1^2+a_1b_1b_2-a_0b_2^2 \geq 0, \tag{6}$$

and, conversely, any impedance S of the form (1) *satisfying these conditions* (2)–(6) *can be realized as the driving-point impedance of a two-mesh circuit consisting of resistances, capacities, and inductances.*

Theorem I thus gives the most general form of this type of impedance, showing that it is a rational function of the time coefficient,[3] completely determined, except for a constant factor, by assigning four roots and two poles, in addition to the poles at zero and infinity, subject to certain conditions. The assigned roots and poles are the time coefficients for the free oscillations of the circuit with the driving branch closed and opened, respectively. That is, the roots and poles correspond to the resonant and anti-resonant points of the impedance.

The conditions are as follows: The real part of each root and pole is negative or zero; the roots and poles occur in pairs of real or conjugate complex quantities; certain additional restrictions must be satisfied, as stated in terms of the symmetric functions of the roots and poles by formulas (3)–(6).

By virtue of these restrictions, the pair of poles, for assigned values of the two pairs of roots, is limited to a certain domain of values. This domain is conveniently illustrated by plotting, in the upper half of the complex plane, the locus of one pole, the other pole being its conjugate. For real poles, a device is used to indicate pairs of points on the real axis. Figs. 3–5 show the domain of the poles, plotted in this manner, for several typical cases.

Provided the roots are not all real, this domain consists of a connected region of values, so that it is possible to pass from one pair of poles to any other pair satisfying the same conditions by a continuous transformation. In the case of four real roots, however, the domain consists, in general, of two non-connected regions, as illustrated in Fig. 5. Under these circumstances there is a region of real poles which are not continuously transformable into complex poles.

The networks realizing any specified driving-point impedance are

[3] All electrical oscillations considered in this paper are of the form $e^{\lambda t}$, where the time coefficient $\lambda = ip$ may have any value, real or complex.

Table I

Networks	1	2	3	4	5	6	7	8	9	10	11
M	$\frac{a_0T_1^2-dT_2^2-a_4T_3^2-a_3T_2T_3}{b_1T_1^2}$.	$\frac{T_1T_3}{b_2^2b_3}$	$\frac{(a_3b_2-db_3)T_3}{b_2^2T_1}$	$\frac{(cb_3-a_4b_1)T_1+(a_3b_3-a_4b_2)T_2}{b_3^2T_1}$	$\frac{T_1T_2}{b_2b_3^2}$	$\frac{(a_3b_3-a_4b_2)T_2}{b_3^2T_1}$	$\frac{(a_1b_2-db_1)T_1+(a_3b_2-db_3)T_3}{b_2^2T_1}$	$\frac{-b_1U_1^2\pm(a_1b_3-cb_2)U_1}{2T_1^2}$	$\frac{(db_3-a_3b_2)(a_3b_1+cb_2-a_1b_3)}{2b_2T_1^2}$	$\frac{(a_3b_3-a_4b_2)(-a_3b_1+cb_2-a_1b_3)}{2b_3T_1^2}$	$\frac{a_0b_2b_3-T_2T_3}{b_1b_2b_3}$
L_1	0	$\frac{T_3^2+a_0b_2^2}{b_1b_2^2}$	$\frac{T_3^2+a_0b_2^2}{b_1b_2^2}$	0	$\frac{T_2^2+a_0b_3^2}{b_1b_3^2}$	$\frac{cb_3-a_4b_1}{b_3^2}$	0	$\frac{\mp U_1(a_1b_3-cb_2)+c(a_3b_2-2db_3)+a_1(a_3b_3-2a_4b_2)}{2T_1^2}$	$\frac{c(a_3b_2-db_3)-a_4(a_1b_2-db_1)}{T_1^2}$	$\frac{d(a_4b_1-cb_3)-a_1(a_4b_2-a_3b_3)}{T_1^2}$	0
L_2	$\frac{U_1(a_1b_3-cb_2)+c(a_3b_2-2db_3)+a_1(a_3b_3-2a_4b_2)}{2T_1^2}$	$\frac{b_1T_1^2}{b_2^2b_3^2}$	$\frac{b_1(a_3b_2-db_3)^2}{b_2^2T_1^2}$	$\frac{d(a_4b_1-cb_3)+a_1(a_3b_3-a_4b_2)}{T_1^2}$	$\frac{b_1T_1^2}{b_2^2b_3^2}$	$\frac{b_1(a_3b_3-a_4b_2)^2}{b_3^2T_1^2}$	$\frac{c(a_3b_2-db_3)-a_4(a_1b_2-db_1)}{T_1^2}$	$\frac{b_1U_1^2}{T_1^2}$	$\frac{b_1(a_3b_2-db_3)^2}{b_2^2T_1^2}$	$\frac{b_1(a_4b_2-a_3b_3)^2}{b_3^2T_1^2}$	$\frac{a_0b_3^2+T_2^2}{b_1b_3^2}$
L_3	$\frac{-U_1(a_1b_3-cb_2)+c(a_3b_2-2db_3)+a_1(a_3b_3-2a_4b_2)}{2T_1^2}$	0	0	$\frac{cb_3-a_4b_1}{b_3^2}$	0	0	$\frac{a_1b_2-db_1}{b_2^2}$	0	0	0	$\frac{a_0b_2^2+T_3^2}{b_1b_2^2}$
R_1	0	$\frac{d}{b_2}$	$\frac{d}{b_2}$	0	$\frac{d}{b_2}$	0	$\frac{d}{b_2}$	0	$\frac{d}{b_2}$	0	$\frac{d}{b_2}$
R_2	$\frac{b_2U_1^2-(a_3b_2-2db_3)U_1}{2T_1^2}$	$\frac{T_1^2}{b_2b_3^2}$	$\frac{(a_3b_2-db_3)^2}{b_2T_1^2}$	$\frac{d(a_3b_3-a_4b_2)}{T_1^2}$	0	$\frac{d(a_3b_3-a_4b_2)}{T_1^2}$	$\frac{(a_3b_2-db_3)^2}{b_2T_1^2}$	$\frac{b_2U_1^2\mp(a_3b_2-2db_3)U_1}{2T_1^2}$	0	$\frac{a_3b_3-a_4b_2}{b_3^2}$	$\frac{T_1^2}{b_2b_3^2}$
R_3	$\frac{b_2U_1^2+(a_3b_2-2db_3)U_1}{2T_1^2}$	0	0	$\frac{a_3b_3-a_4b_2}{b_3^2}$	$\frac{T_1^2}{b_2b_3^2}$	$\frac{a_3b_3-a_4b_2}{b_3^2}$	0	$\frac{b_2U_1^2\pm(a_3b_2-2db_3)U_1}{2T_1^2}$	$\frac{(a_3b_2-db_3)^2}{b_2T_1^2}$	$\frac{d(a_3b_3-a_4b_2)}{T_1^2}$	0
C_1	∞	$\frac{b_3}{a_4}$	∞	$\frac{b_3}{a_4}$	$\frac{b_3}{a_4}$	$\frac{b_3}{a_4}$	∞	∞	∞	$\frac{b_3}{a_4}$	$\frac{b_3}{a_4}$
C_2	$\frac{2T_1^2}{b_3U_1^2+(a_3b_3-2a_4b_2)U_1}$	∞	$\frac{T_1^2}{a_4(a_3b_2-db_3)}$	$\frac{b_3T_1^2}{(a_3b_3-a_4b_2)^2}$	$\frac{b_2^2b_3}{T_1^2}$	$\frac{b_3T_1^2}{(a_3b_3-a_4b_2)^2}$	$\frac{T_1^2}{a_4(a_3b_2-db_3)}$	$\frac{2T_1^2}{b_3U_1^2\pm(a_3b_3-2a_4b_2)U_1}$	$\frac{b_2^2}{a_3b_2-db_3}$	∞	∞
C_3	$\frac{2T_1^2}{b_3U_1^2-(a_3b_3-2a_4b_2)U_1}$	$\frac{b_2^2b_3}{T_1^2}$	$\frac{b_2^2}{a_3b_2-db_3}$	∞	∞	∞	$\frac{b_2^2}{a_3b_2-db_3}$	$\frac{2T_1^2}{b_3U_1^2\mp(a_3b_3-2a_4b_2)U_1}$	$\frac{T_1^2}{a_4(a_3b_2-db_3)}$	$\frac{b_3T_1^2}{(a_3b_3-a_4b_2)^2}$	$\frac{b_2^2b_3}{T_1^2}$

$$S = \frac{a_0\lambda^4+a_1\lambda^3+a_2\lambda^2+a_3\lambda+a_4}{b_1\lambda^3+b_2\lambda^2+b_3\lambda}, \quad \lambda = ip,$$

$$d^2(b_2^2-4b_1b_3)-2d(2a_4b_1^2+a_2b_2^2+2a_0b_3^2-a_3b_1b_2-2a_2b_1b_3-a_1b_2b_3)+\left[a_3^2b_1^2+(a_2^2-4a_0a_4)b_2^2+a_1^2b_3^2-2(a_2a_3-2a_1a_4)b_1b_2-2a_1a_3b_1b_3-2(a_1a_2-2a_0a_3)b_2b_3\right]=0,$$

$$c = a_2-d,$$

$$T_1 = \pm\sqrt{-a_4b_2^2+a_3b_2b_3-db_3^2}, \qquad U_1 = \sqrt{a_3^2-4a_4d},$$

$$T_2 = \pm\sqrt{-a_0b_3^2+cb_3b_1-a_4b_1^2}, \qquad U_2 = \sqrt{c^2-4a_0a_4},$$

$$T_3 = \pm\sqrt{-db_1^2+a_1b_1b_2-a_0b_2^2}, \qquad U_3 = \sqrt{a_1^2-4a_0d}.$$

with signs chosen so that
$b_1T_1+b_2T_2+b_3T_3 = 0$,

determined by the arrangement and magnitudes of the elements, as given by the following theorem:

Theorem II. All driving-point impedances satisfying the necessary and sufficient conditions, as stated in Theorem I, can be realized by any one of the eleven networks shown by Fig. 1, *upon assigning to the elements of each network the values given by Table I. These eleven networks are the only networks without superfluous elements by which the entire range of possible impedances can be realized.*

By Theorem II, any network obtained from a two-mesh circuit

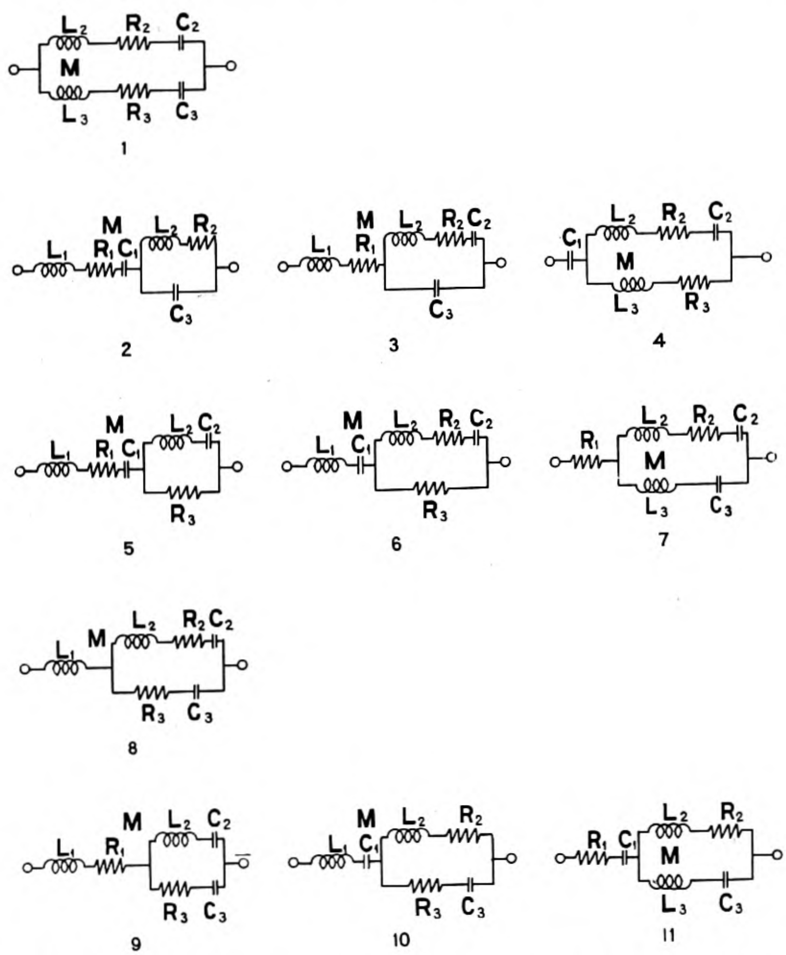


Fig. 1—Networks realizing any driving-point impedance of a two-mesh circuit consisting of resistances, capacities, self-inductances, and mutual inductances

consisting of resistances, capacities, and inductances can be replaced, in so far as the impedance between terminals is concerned, by any one of the eleven networks shown by Fig. 1, upon assigning the proper values to the elements. Each of these networks consists of two resistances, two capacities, and two self-inductances with mutual inductance between them.

Each of these eleven networks realizes impedances with arbitrarily assigned roots and with poles anywhere in the entire domain of possibilities, subject to the general conditions stated in Theorem I. Special cases of these networks realize, for arbitrarily assigned roots, only critical lines and points in the domain. All these special cases are listed in Table III, with a specification of the lines or points in the domain realizable by each, as illustrated by Figs. 4 and 5.

Certain limited regions of the domain can be realized by networks which contain no mutual inductance and which are not special cases of the networks given by Theorem II. These networks are given by the following theorem:

Theorem III. Any driving-point impedance of a two-mesh circuit consisting of resistances, capacities, and self-inductances can be realized by at least three and not more than five of the twelve networks shown by Fig. 2, upon assigning to the elements of each network the values given by Table II. These twelve networks are the only networks without mutual inductance and without superfluous elements by which any impedance can, in general, be realized.

These twelve networks, taken together, cover that portion of the domain realizable without mutual inductance. Networks with mutual inductance are needed in order to cover the entire domain. These twelve are the only networks, without superfluous elements, realizing limited regions in the domain. Each of these networks consists of two resistances, two capacities, two self-inductances, and one additional resistance, capacity, or self-inductance. The twelve networks, with their special cases, are all listed in Table III, with a specification of the regions, lines, or points realizable by each.

In addition to the specific formulas for the networks of Figs. 1 and 2, it is convenient to have general formulas for the computation of all networks meeting the given conditions, including those networks with superfluous elements as well as all special cases. The most general two-mesh circuit is shown by Fig. 6; accordingly, the most general network under consideration is that shown by Fig. 7. Formulas

Table II

Networks	12	13	14	15	16	17
L_1	$\frac{a_0T_1^2-dT_2^2-a_4T_3^2-a_3T_2T_3}{b_1T_1^2}$	$\frac{a_0b_2b_3-T_2T_3}{b_1b_2b_3}$	$\frac{(a_1b_2-db_1)T_1+(a_3b_2-db_3)T_3}{b_2^2T_1}$	$\frac{(cb_3-a_4b_1)T_1+(a_3b_3-a_4b_2)T_2}{b_3^2T_1}$	0	$\frac{a_0}{b_1}$
L_2	$\frac{b_1U_1^2+(a_1b_3-cb_2)U_1}{2T_1^2}$	$-\frac{T_1T_2}{b_2b_3^2}$	$\frac{(a_3b_2-db_3)(a_3b_1+cb_2-a_1b_3)}{2b_2T_1^2}$	$\frac{(a_3b_3-a_4b_2)(a_3b_1-cb_2+a_1b_3)}{2b_3T_1^2}$	$\frac{b_1U_2^2-(cb_1-2a_0b_3)U_2}{2T_2^2}$	$\frac{T_2^2}{b_1b_3^2}$
L_3	$\frac{b_1U_1^2-(a_1b_3-cb_2)U_1}{2T_1^2}$	$-\frac{T_1T_3}{b_2^2b_3}$	$\frac{(db_3-a_3b_2)T_3}{b_2^2T_1}$	$\frac{(a_4b_2-a_3b_3)T_2}{b_3^2T_1}$	$\frac{b_1U_2^2+(cb_1-2a_0b_3)U_2}{2T_2^2}$	0
R_1	0	$\frac{d}{b_2}$	$\frac{d}{b_2}$	0	$\frac{dT_2^2-a_4T_3^2-a_0T_1^2-cT_1T_3}{b_2T_2^2}$	$\frac{db_1b_3-T_1T_3}{b_1b_2b_3}$
R_2	$\frac{b_2U_1^2-(a_3b_2-2db_3)U_1}{2T_1^2}$	$\frac{T_1^2}{b_2b_3^2}$	$\frac{(a_3b_2-db_3)^2}{b_2T_1^2}$	$\frac{d(a_3b_3-a_4b_2)}{T_1^2}$	$\frac{b_2U_2^2+(a_1b_3-a_3b_1)U_2}{2T_2^2}$	$-\frac{T_1T_2}{b_1b_3^2}$
R_3	$\frac{b_2U_1^2+(a_3b_2-2db_3)U_1}{2T_1^2}$	0	0	$\frac{a_3b_3-a_4b_2}{b_3^2}$	$\frac{b_2U_2^2-(a_1b_3-a_3b_1)U_2}{2T_2^2}$	$-\frac{T_2T_3}{b_1^2b_3}$
C_1	∞	$\frac{b_3}{a_4}$	∞	$\frac{b_3}{a_4}$	∞	$\frac{b_3}{a_4}$
C_2	$\frac{2T_1^2}{b_3U_1^2+(a_3b_3-2a_4b_2)U_1}$	∞	$\frac{T_1^2}{a_4(a_3b_2-db_3)}$	$\frac{b_3T_1^2}{(a_3b_3-a_4b_2)^2}$	$\frac{2T_2^2}{b_3U_2^2+(cb_3-2a_4b_1)U_2}$	∞
C_3	$\frac{2T_1^2}{b_3U_1^2-(a_3b_3-2a_4b_2)U_1}$	$\frac{b_2^2b_3}{T_1^2}$	$\frac{b_2^2}{a_3b_2-db_3}$	∞	$\frac{2T_2^2}{b_3U_2^2-(cb_3-2a_4b_1)U_2}$	$\frac{b_1^2b_3}{T_2^2}$

Networks	18	19	20	21	22	23
L_1	$\frac{a_0}{b_1}$	0	0	$\frac{a_0}{b_1}$	$\frac{a_0}{b_1}$	0
L_2	$\frac{(cb_1-a_0b_3)^2}{b_1T_2^2}$	$\frac{a_0(cb_3-a_4b_1)}{T_2^2}$	$\frac{b_1U_3^2+(a_1b_1-2a_0b_2)U_3}{2T_3^2}$	$\frac{T_3^2}{b_1b_2^2}$	$\frac{(a_1b_1-a_0b_2)^2}{b_1T_3^2}$	$\frac{a_0(a_1b_2-db_1)}{T_3^2}$
L_3	0	$\frac{cb_3-a_4b_1}{b_3^2}$	$\frac{b_1U_3^2-(a_1b_1-2a_0b_2)U_3}{2T_3^2}$	0	0	$\frac{a_1b_2-db_1}{b_2^2}$
R_1	$\frac{(a_1b_1-a_0b_2)T_2+(cb_1-a_0b_3)T_3}{b_1^2T_2}$	$\frac{(cb_3-a_4b_1)T_1+(a_3b_3-a_4b_2)T_2}{b_3^2T_2}$	0	$\frac{d}{b_2}$	0	$\frac{d}{b_2}$
R_2	$\frac{(cb_1-a_0b_3)(cb_2+a_3b_1-a_1b_3)}{2b_1T_2^2}$	$\frac{(cb_3-a_4b_1)(-a_3b_1+cb_2+a_1b_3)}{2b_3T_2^2}$	$\frac{b_2U_3^2-(a_1b_2-2db_1)U_3}{2T_3^2}$	0	$\frac{d(a_1b_1-a_0b_2)}{T_3^2}$	$\frac{(a_1b_2-db_1)^2}{b_2T_3^2}$
R_3	$\frac{(a_0b_3-cb_1)T_3}{b_1^2T_2}$	$\frac{(a_4b_1-cb_3)T_1}{b_3^2T_2}$	$\frac{b_2U_3^2+(a_1b_2-2db_1)U_3}{2T_3^2}$	$\frac{T_3^2}{b_1^2b_2}$	$\frac{a_1b_1-a_0b_2}{b_1^2}$	0
C_1	∞	$\frac{b_3}{a_4}$	$\frac{b_3T_3^2}{a_4T_3^2-dT_2^2-a_0T_1^2-a_1T_1T_2}$	$\frac{b_1b_2b_3}{a_4b_1b_2-T_1T_2}$	$\frac{b_1^2T_3}{(cb_1-a_0b_3)T_3+(a_1b_1-a_0b_2)T_2}$	$\frac{b_2^2T_3}{(a_3b_2-db_3)T_3+(a_1b_2-db_1)T_1}$
C_2	$\frac{T_2^2}{a_4(cb_1-a_0b_3)}$	$\frac{b_3T_2^2}{(cb_3-a_4b_1)^2}$	$\frac{2T_3^2}{b_3U_3^2+(a_3b_1-cb_2)U_3}$	$-\frac{b_1b_2^2}{T_1T_3}$	$\frac{2b_1T_3^2}{(a_1b_1-a_0b_2)(a_1b_3-cb_2+a_3b_1)}$	$\frac{2b_2T_3^2}{(a_1b_2-db_1)(a_1b_3+cb_2-a_3b_1)}$
C_3	$\frac{b_1^2}{cb_1-a_0b_3}$	∞	$\frac{2T_3^2}{b_3U_3^2-(a_3b_1-cb_2)U_3}$	$-\frac{b_1^2b_2}{T_2T_3}$	$\frac{b_1^2T_3}{(a_0b_2-a_1b_1)T_2}$	$\frac{b_2^2T_3}{(db_1-a_1b_2)T_1}$

for the computation of the elements of this general network can be stated in the form of the following theorem:

Theorem IV. Any driving-point impedance satisfying the necessary and sufficient conditions, as stated in Theorem I, can be realized

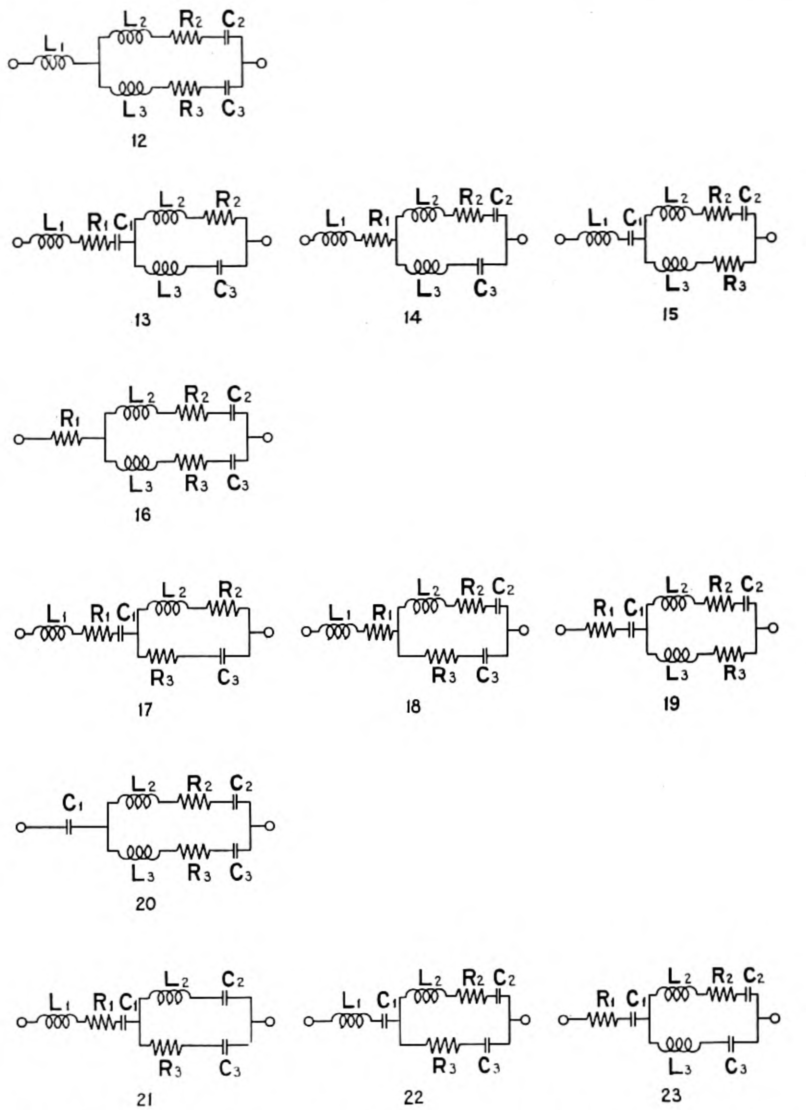


Fig. 2—Networks without mutual inductance realizing any driving-point impedance of a two-mesh circuit consisting of resistances, capacities, and self-inductances.

by any network of the form of Fig. 7, provided the elements of the network satisfy the following relations:

$$L_1'L_2'+L_1'L_3'+L_2'L_3'=a_0k^2, \tag{7}$$

$$R_1R_2+R_1R_3+R_2R_3=dk^2, \tag{8}$$

$$D_1D_2+D_1D_3+D_2D_3=a_4k^2, \tag{9}$$

$$L_2'+L_3'=b_1k^2, \tag{10}$$

$$R_2+R_3=b_2k^2, \tag{11}$$

$$D_2+D_3=b_3k^2, \tag{12}$$

$$R_2D_3-R_3D_2=\pm k^3(-a_4b_2^2+a_3b_2b_3-db_3^2)^{1/2}, \tag{13}$$

$$D_2L_3'-D_3L_2'=\pm k^3[-a_0b_3^2+(a_2-d)b_3b_1-a_4b_1^2]^{1/2}, \tag{14}$$

$$L_2'R_3-L_3'R_2=\pm k^3(-db_1^2+a_1b_1b_2-a_0b_2^2)^{1/2}, \tag{15}$$

where

$$D_1=C_1^{-1},\ D_2=C_2^{-1},\ D_3=C_3^{-1}, \tag{16}$$

and

$$L_1'=L_1+M_{12}+M_{13}+M_{23}, \tag{17}$$

$$L_2'=L_2+M_{12}-M_{13}-M_{23}, \tag{18}$$

$$L_3'=L_3-M_{12}+M_{13}-M_{23}, \tag{19}$$

the positive directions in Fig. 7 all being assigned arbitrarily to the right. The signs of (13)–(15) *are chosen so as to satisfy the identity*

$$(R_2D_3-R_3D_2)(L_2'+L_3')+(D_2L_3'-D_3L_2')(R_2+R_3) \\ +(L_2'R_3-L_3'R_2)(D_2+D_3)=0. \tag{20}$$

The value of d is given by equation (3), *which may be written in the form*

$$d^2(b_2^2-4b_1b_3)-2d(2a_4b_1^2+a_2b_2^2+2a_0b_3^2-a_3b_1b_2-2a_2b_1b_3-a_1b_2b_3) \\ +[a_3^2b_1^2+(a_2^2-4a_0a_4)b_2^2+a_1^2b_3^2-2(a_2a_3-2a_1a_4)b_1b_2-2a_1a_3b_1b_3 \\ -2(a_1a_2-2a_0a_3)b_2b_3]=0. \tag{21}$$

The parameter k may have any real value other than zero.

In these formulas the value of k is independent of the impedance, but can be chosen so as to give particular forms of the network. If the necessary and sufficient conditions as stated by Theorem I are satisfied, the values of the elements given by these formulas are positive or zero, and the values of the inductances satisfy the usual restrictions. The formulas of Tables I and II, for example, can all be computed by means of Theorem IV.

2. The Driving-Point Impedance of a Two-Mesh Circuit

Previous investigations of the two-mesh circuit have been directed, for the most part, toward the determination of the free periods (resonant frequencies and associated damping constants) of the circuit from the known values of the elements. This problem is intimately related to the determination of the driving-point impedance of the circuit, since the free periods of the circuit can be found by setting the driving-point impedance in any one mesh equal to zero.[4] By this method the free periods are found as the roots of an equation of the fourth degree,[5] the exact solution of which involves, in general, cumbersome formulas. In order to obtain formulas which are better adapted to numerical computation, various approximations are usually made.[6]

This electrical problem of the free oscillations of a circuit is formally the same as the dynamical problem of the small oscillations of a system about a position of equilibrium. The determination of the free periods of a circuit can be made directly from the solution of this dynamical problem.[7]

The first part of this paper treats a much more general problem than the determination of the driving-point impedance of a particular circuit from the given values of the elements, namely, the determination of the entire range of possibilities, together with the inherent limitations, of such an impedance. The method employed is to find the general form of the impedance as a function of the time coefficient, and then to investigate the restrictions which must be satisfied by a function of this character in order that it may represent an impedance realizable by means of a circuit consisting of resistances, capacities, and inductances. In the present paper, this investigation is limited to the driving-point impedance of a two-mesh circuit; the driving-point impedance of an n-mesh circuit will be treated in a future paper.

The driving-point impedance of any circuit containing no resistances has been investigated in a previous paper,[8] where it has been shown that any such impedance is a pure reactance with a number of resonant and anti-resonant frequencies which alternate with each other, and

[4] G. A. Campbell, *Transactions of the A. I. E. E.*, 30, 1911, pages 873–909.

[5] An exhaustive discussion of this fourth degree equation has been given by J. Sommer, *Annalen der Physik*, fourth series, 58, 1919, pages 375–392.

[6] For typical methods of solution see the papers of L. Cohen, *Bulletin of the Bureau of Standards*, 5, 1908–9, pages 511–541; B. Macků, *Jahrbuch der drahtlosen Telegraphie und Telephonie*, 2, 1909, pages 251–293; V. Bush, *Proceedings of the I. R. E.*, 5, 1917, pages 363–382.

[7] Representative investigations of this dynamical problem are those of Lord Rayleigh, *Proceedings of the London Mathematical Society*, 4, 1873, pages 357–368, *Philosophical Magazine*, fifth series, 21, 1886, pages 369–381, and sixth series, 3, 1902, pages 97–117 ("Scientific Papers," I, 170–181, II, 475–485, and V, 8–26); E. J. Routh, "Advanced Rigid Dynamics," sixth edition, 1905, pages 232–243; A. G. Webster, "Dynamics," second edition, 1912, pages 157–164.

[8] R. M. Foster, *Bell System Technical Journal*, 3, 1924, pages 259–267.

that any such impedance may be realized by a network consisting of a number of simple resonant elements (inductance and capacity in series) in parallel or a number of simple anti-resonant elements (inductance and capacity in parallel) in series.

With resistances added to the circuit, the impedance is, in general, complex; that is, it has both resistance and reactance components. For a two-mesh circuit the impedance is expressed as a function of the time coefficient by Theorem I.

Formula (1) gives the driving-point impedance of a two-mesh circuit for any electrical oscillation of the form $e^{\lambda t}$, where the time coefficient λ may have any value, real or complex. The time coefficients for the free oscillations of the circuit with the driving branch closed are the roots of the numerator (α_1, α_2, α_3, α_4), as given by (1a); the free periods of the circuit with the driving branch opened are the roots of the denominator (β_2, β_3), that is, the poles of the impedance function. For a complex value of the time coefficient, $\lambda=\lambda_1+i\lambda_2$, λ_1 is the damping factor and λ_2 is the frequency multiplied by 2π.

The two forms of formula (1) are equivalent, but each has its special advantages. Sometimes one, sometimes the other, form is more convenient; they will be used interchangeably throughout the paper.

Formula (1a) gives the impedance directly in terms of the roots and poles. Formula (1b) gives the impedance in terms of the symmetric functions of the roots and poles, with the addition of an arbitrary factor. Thus, without changing the impedance, all the coefficients of the numerator and denominator of (1b) may be multiplied by the same constant factor having any value other than zero. Formulas stated in terms of the coefficients of (1b) are in homogeneous and symmetrical form, and have the added advantage of involving real quantities only.

The special case of one root equal to zero is obtained by setting $\alpha_1=0$ in (1a) and $a_4=0$ in (1b). For one root infinite, however, in (1a) it is necessary to set $\alpha_1=\infty$ and $H=0$, with the provision that $H\alpha_1$ be finite; whereas in (1b) it is simply necessary to set $a_0=0$.

It is sometimes convenient to add the notation $\beta_1=0$ and $\beta_4=\infty$, corresponding to the poles at zero and infinity. In formula (1b) the corresponding addition to the notation consists of the coefficients $b_0=0$ and $b_4=0$.

By the general restrictions (2) the constant H is positive or zero, and the roots and poles are arranged in three pairs, (α_1, α_2), (α_3, α_4), and (β_2, β_3), each pair being the roots of a quadratic equation with positive real coefficients. Thus each pair of the roots and poles is either a pair of conjugate complex quantities or a pair of real quantities, with the added provision that the real part of each root and pole is negative or zero.

Stated in terms of (1b), these general restrictions (2) require all the coefficients to be real and to have the same sign. Throughout this paper these signs will always be taken positive; thus all the a's and b's are positive or zero. In order to provide that the real part of each root be negative or zero, the coefficients of the numerator must satisfy the additional requirement

$$-a_4a_1^2+a_1a_2a_3-a_0a_3^2\geq 0, \tag{22}$$

and also

$$a_2^2-4a_0a_4\geq 0. \tag{23}$$

The second condition (23) is satisfied automatically by virtue of the first condition (22), unless both a_1 and a_3 are zero; in that case (23) is required. These are precisely the necessary and sufficient conditions that the numerator of (1b) be factorable into two real quadratic factors with positive coefficients.

In addition to the general restrictions (2) upon the individual roots and poles, there are certain additional conditions which must be satisfied by all the roots and poles together. These conditions are more conveniently stated in terms of the coefficients by prescribing a certain domain of values of the eight coefficients (a_0, a_1, a_2, a_3, a_4, b_1, b_2, b_3) such that the coefficients of any driving-point impedance of a two-mesh circuit lie in this domain, and, conversely, any set of values in this domain can be realized as the coefficients of a driving-point impedance of a two-mesh circuit.

By a realizable circuit is understood a circuit consisting of resistances, capacities, and self-inductances, with positive or zero values, together with mutual inductances with values such that every principal minor of the determinant of the inductances is positive or zero. In the case of two self-inductances with mutual inductance between them, this reduces to the well known condition $L_1L_2-M^2\geq 0$.

The domain is defined analytically by formulas (3)–(6), in terms of a parameter d. This parameter is intimately related to the resistances in the circuit, as will be shown later. In order that this domain may contain real values, the following relation must be satisfied:

$$-d^3+2a_2d^2-(a_1a_3+a_2^2-4a_0a_4)d+(-a_4a_1^2+a_1a_2a_3-a_0a_3^2)\geq 0, \tag{24}$$

or in equivalent form,

$$-[d-a_0(\alpha_1+\alpha_2)(\alpha_3+\alpha_4)][d-a_0(\alpha_1+\alpha_3)(\alpha_2+\alpha_4)]$$
$$[d-a_0(\alpha_1+\alpha_4)(\alpha_2+\alpha_3)]\geq 0. \tag{25}$$

Provided there is one pair of conjugate complex roots of the numerator of the impedance, α_1 and α_2, the value of d is restricted to the range from zero to the smallest real root of (24), that is,

$$0 \leq d \leq a_0(\alpha_1+\alpha_2)(\alpha_3+\alpha_4). \tag{26}$$

In the case of four real roots, $\alpha_1 \geq \alpha_2 \geq \alpha_3 \geq \alpha_4$, the parameter d is restricted to the values

$$\left.\begin{aligned} 0 \leq d \leq a_0(\alpha_1+\alpha_2)(\alpha_3+\alpha_4), \\ a_0(\alpha_1+\alpha_3)(\alpha_2+\alpha_4) \leq d \leq a_0(\alpha_1+\alpha_4)(\alpha_2+\alpha_3). \end{aligned}\right\} \tag{27}$$

Thus there are, in general, two distinct ranges for the value of d in this case. The corresponding domain of values of the roots and poles consists of two non-connected regions, so that it is impossible to pass by a continuous transformation from a set of values in one region to a set in the other.

Formulas (3)–(6) are symmetrical in three different respects, since they remain unaltered upon interchanging certain pairs of elements, which may be any one of the three following sets:

$$\left.\begin{aligned} &\text{(a) } b_1 \text{ and } b_2,\ a_0 \text{ and } d,\ a_3 \text{ and } (a_2-d), \\ &\text{(b) } b_1 \text{ and } b_3,\ a_0 \text{ and } a_4,\ a_1 \text{ and } a_3, \\ &\text{(c) } b_2 \text{ and } b_3,\ a_4 \text{ and } d,\ a_1 \text{ and } (a_2-d). \end{aligned}\right\} \tag{28}$$

These three sets correspond to interchanging resistances and inductances, inductances and capacities, and resistances and capacities, respectively.

Since d is always positive or zero, formulas (4)–(6) lead to simple necessary conditions, namely,

$$a_3b_3-a_4b_2 \geq 0, \tag{29}$$

$$-a_4b_1^2+a_2b_1b_3-a_0b_3^2 \geq 0, \tag{30}$$

$$a_1b_1-a_0b_2 \geq 0. \tag{31}$$

The first and third of these conditions are conveniently interpreted in terms of the roots and poles: the sum of the reciprocals of the poles is algebraically greater than or equal to the sum of the reciprocals of the roots; and the sum of the poles is algebraically greater than or equal to the sum of the roots.

3. Domain of Poles for Assigned Roots

The conditions (2)–(6) define a domain of values for the roots and poles without distinguishing in any way those roots and poles which may be chosen independently. For many purposes it is convenient to specialize the problem to the extent of assigning definite values to the roots, subject, of course, to the restrictions (2), and then to investigate the domain of the poles which can be associated with these assigned roots.

For the mathematical analysis of the problem it is convenient to assign values of the coefficients $a_0 \ldots a_4$, subject to the restrictions stated in the preceding section, and then to plot the domain for the coefficients b_1, b_2, b_3,—treating the latter as homogeneous coordinates [9] in the plane, with $x = b_2/b_1$ and $y = b_3/b_1$.

With this method of representation, equation (3) is, for any fixed value of d, the equation of a conic. Considering d as a variable parameter, (3) represents a one-parameter family of conics. Each curve of this family is tangent to the four lines

$$\alpha_j^2 b_1 + \alpha_j b_2 + b_3 = 0, \qquad (j = 1, 2, 3, 4). \tag{32}$$

These lines are real lines in the plane if, and only if, the corresponding roots are real. They are all tangent to the parabola

$$b_2^2 - 4b_1 b_3 = 0, \tag{33}$$

which is the limiting case of the conic (3) as d becomes infinite. This parabola is a critical curve for the poles; every point in the plane above the parabola corresponds to a pair of conjugate complex poles, every point below the curve to a pair of real and distinct poles, and every point on the curve to a pair of real and equal poles.

The complete family of conics, that is, the set of curves for all real values of d, might be defined as the family of conics tangent to these four lines, which are the four lines tangent to the critical parabola (33) corresponding to the four roots of the impedance.

Not all the curves of this family lie in the domain of poles, however, since the conditions (4)–(6) must also be satisfied. For any fixed value of d, each of the three equations (4)–(6) is a degenerate conic, that is, a pair of straight lines. The six lines defined by these conditions are all tangent to the conic (3) corresponding to this same value of d. The inequalities (4)–(6) thus demand, in general, that the domain of poles lie within the area bounded by these six lines. Thus only those conics of the family (3) which are real ellipses, or their limiting cases, lie within the domain.

The condition that the conic (3) be an ellipse is precisely the necessary restriction on the value of d already stated, formula (24). Ellipses are obtained for all negative values of d, but these are not in the

[9] For some purposes the other choices of x and y might be used; this choice is more convenient here inasmuch as $-x$ is the sum and y the product of the poles.

domain, since by the conditions of the electrical problem d must be positive or zero. Ellipses for values of d from zero up to the smallest real root of the equation (24) are in the domain. If the roots of the impedance are all complex, equation (24) has three real roots, and thus there is a range of values of d from the second to the third root, arranged in the order of magnitude, for which the curves are ellipses, but these ellipses are imaginary, that is, there are no real points on them; thus there is only the one range of d which gives points in the domain. If two roots of the impedance are real and two complex, equation (24) has only the one real root, and thus there is only the one range of d. If all four roots of the impedance are real, however, equation (24) has again three real roots, and both ranges of d give real ellipses. In this case the two sets of ellipses are separate and distinct.

For the limiting values of d, that is, for the roots of equation (24), the corresponding conic (3) degenerates into a pair of coincident straight lines. Only those segments of these lines which satisfy the corresponding inequalities (4)–(6) are in the domain. Such segments are the limiting cases of the real ellipses for values of d above or below the critical values, as the case may be.

The domain of poles, plotted in terms of the coefficients in the manner described, consists of that domain covered by these real ellipses for $d \geq 0$, a domain bounded by the envelope of the curves. The env elopeconsists of the conic for $d=0$ and the four lines (32). For the case of four complex roots of the impedance, therefore, the domain consists simply of the region bounded by the ellipse (3) for $d=0$. For two complex and two real roots, the domain consists of the region bounded by the ellipse with the addition of the corner bounded by the ellipse and the two tangent lines to the ellipse corresponding to the two real roots. For four real roots, the domain consists of the region bounded by the ellipse together with the two corners bounded by the ellipse and the tangent lines, one by the two lines corresponding to the two smallest roots and the other the two largest roots; and a second region consisting of the quadrilateral bounded by the four tangent lines.

All points in the domain lying on or above the critical parabola lie on a single curve of the family of conics composing the domain, points below the parabola on two curves of the family. The corner regions and the quadrilateral are entirely below the critical parabola. Where there is a corner region, the ellipse goes below the parabola, otherwise not.

The foregoing discussion has all been for the general case of unrestricted roots. For special cases of zero, pure imaginary, or infinite roots, the corresponding domains are the limiting cases of the general domain, described above. Such limiting cases may reduce to a single segment or to a region bounded in part by the line at infinity. The homogeneous coordinates employed are very useful in dealing with these special cases.

4. Figures Illustrating the Domain of Poles

The preceding section presented a discussion of the domain of the poles associated with any four assigned roots, the domain being plotted in terms of the coefficients of the denominator of the impedance, that is, in terms of symmetric functions of the poles. In order to show the mutual relations between the actual values of the roots and the poles, it is convenient to plot, in the upper half of the complex plane, the domain of one pole, the other pole being its conjugate. This provides a complete representation for the case of complex poles. In order to include the domain of real poles, an auxiliary graph can be provided to indicate pairs of points on the real axis.

The mathematical analysis for this form of representation can be obtained from that of the preceding section by substituting $\beta_2+\beta_3=-b_2/b_1$ and $\beta_2\beta_3=b_3/b_1$. For complex poles, $\beta_2=u+iv$ and $\beta_3=u-iv$, this transformation from the x, y plane to the u, v plane is simply $2u=-x$ and $u^2+v^2=y$. Thus a conic in the x, y plane becomes, in general, a curve of the fourth degree in the u, v plane. The analysis of the curves obtained in the u, v plane is not so simple as in the other plane, but there is a decided advantage in the interpretation of the results in this plane, since the coordinate u, the real part of the pole, corresponds to the damping factor, and the coordinate v, the imaginary part of the pole, corresponds to the frequency factor.

In the complex plane, the necessary conditions (29)–(31) require the domain of complex poles to lie entirely within the region bounded by the vertical axis, a vertical line to the left of the axis, two circles about the origin as center, and a circle through the origin with its center on the real axis. Furthermore, the boundary curve of the domain must be tangent to each of these lines and circles, since the corresponding conic (3) for $d=0$ is tangent to the corresponding lines (4)–(6) for $d=0$.

For the special case of one root a positive pure imaginary, the second root being its conjugate, the domain in the upper half of the complex plane reduces merely to the points on an arc of a circle with its center on the real axis. If the third root is complex with a positive imaginary part, the fourth root being its conjugate, the domain is the circular arc extending from the first root to the third root. For a pure imaginary value of the third root the radius of the circle becomes infinite, and the domain is the segment of the vertical axis between the first and third roots. This is precisely the result already obtained for the resistanceless circuit.

For the limiting case of the third root real, with the fourth root equal to it, the domain is the circular arc extending from the root on the imaginary axis to the double root on the real axis. When the third and fourth roots are real and distinct, the domain is the circular arc from the first root to the point on the real axis midway between the two real roots. The complete domain also includes real poles in the segment between the two real roots, equally spaced about the midpoint of the segment.

This case of one pair of roots on the axis of imaginaries is illustrated by Fig. 3a, with the first root fixed at the point a, and the third root lying on any one of the family of circular arcs drawn through a, the fourth root being its conjugate; or the third and fourth roots lying on the real axis equally spaced about the end-point of one of the arcs.

Starting with one pair of roots on the axis of imaginaries, it is interesting to investigate the changes made in the domain by moving this pair of roots off the axis. The domain broadens out into a region lying about the circular arc, as shown by Fig. 3b for four typical cases. The first case is for the third root also near the axis ($\alpha_1=-0.5+i3$, $\alpha_3=-0.5+i9$); and the second case is for the third root some distance from the axis ($\alpha_1=-0.1+i3$, $\alpha_3=-5+i8$). The third section of Fig. 3b shows the domain when the third and fourth roots are real and equal ($\alpha_1=-0.1+i3$, $\alpha_3=\alpha_4=-9$); in this case the region has a cusp at this double root. The fourth section shows the domain of complex poles when the third and fourth roots are real and distinct ($\alpha_1=-0.1+i3$, $\alpha_3=-6$, $\alpha_4=-10$); in this case the region of complex poles terminates along a segment of the real axis lying in the interval between the two real roots, there is also a domain of real poles which is not shown.

It is interesting to note that, when both pairs of roots are near the axis of imaginaries, that is, for small damping, the frequency factor of the pole may always be taken outside the range of the frequency factors of the roots; whereas for zero damping the pole must lie between the roots, as noted above.

Fig. 3c shows the domain of the poles for two pairs of equal roots. If the first and third roots are equal, the second and fourth roots being their conjugates and thus also equal, the domain is bounded by a circle tangent to the vertical axis with its center vertically above the double root. If, for example, the double root describes a circle about the origin through the point a on the vertical axis, the corresponding circle is tangent to the vertical axis at a. Thus in Fig. 3c,

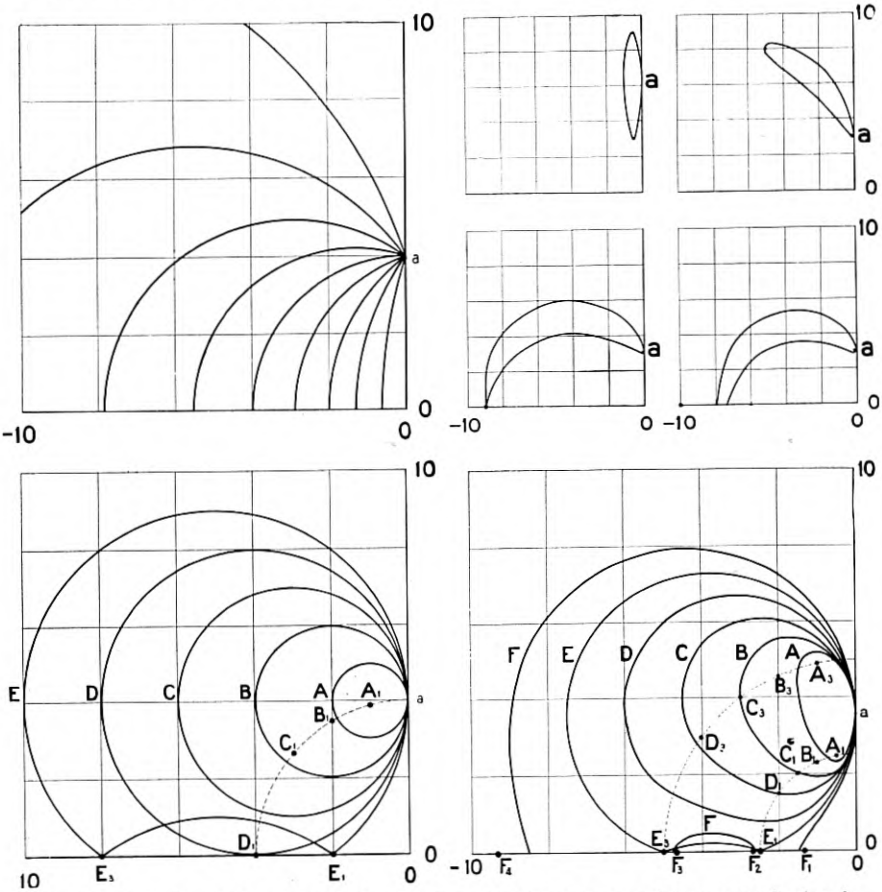


Fig. 3—Domain of the poles of the driving-point impedance of a two-mesh circuit with (a) one pair of roots on the axis of imaginaries, (b) one pair of roots near the axis of imaginaries, (c) two pairs of equal roots, and (d) two pairs of roots with equal angles.

for double roots at A_1, B_1, C_1, the corresponding domain is bounded by the circles A, B, C, respectively. The centers of these circles are all on the horizontal line through a, and the double roots are selected so as to space the centers uniformly. If all four roots are real and equal, the domain is bounded by a circle D tangent to the vertical axis at a and to the horizontal axis at this fourfold root D_1. If the roots are all real and equal in pairs the domain is bounded by a circle E, tangent to the vertical axis and passing through the two double roots, E_1 and E_3, and by the reflection of this circle in the real axis. Thus the domain has cusps at the double roots. For two pairs of

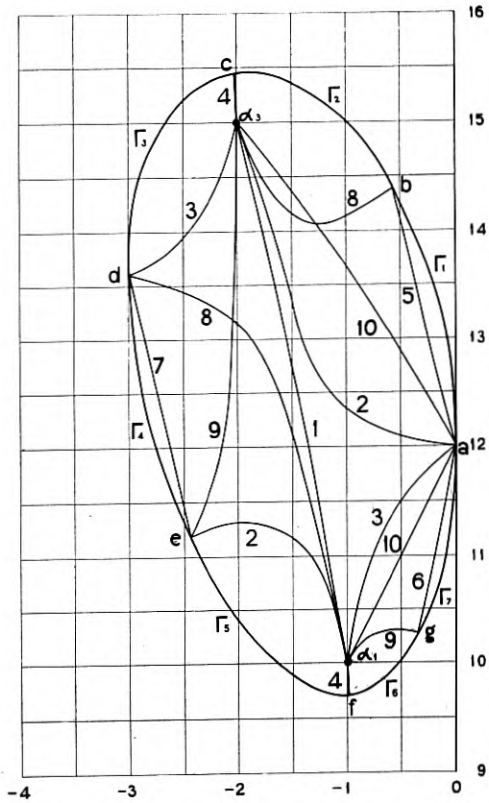


Fig. 4—Domain of the poles of the driving-point impedance of a two-mesh circuit with two pairs of complex roots, showing the portions of the domain realizable by each network listed in Table III.

equal roots, whether real or complex, the distance Oa is the geometrical mean value of all four roots.

Another kind of special case is shown by Fig. 3d, the case of two pairs of roots with equal angles. The first and third roots are on a line with the origin, so that the second and fourth roots, being their conjugates, are also on a line with the origin. Fig. 3d shows the boundary curves ($A \ldots E$) for five sets of roots ($A_1, A_3 \ldots E_1, E_3$) satisfying these conditions and with the same absolute values of the roots in each set, so that the roots lie on two circles about the origin.

The fifth set of roots (E_1, E_3) has a domain of the same type as the corresponding set of roots on Fig. 3c, since this set, being on the real axis, is a double set. The sixth curve F is the boundary of the domain for four real roots so chosen that $F_1F_3=E_1^2$ and $F_2F_4=E_3^2$. This is the same type of domain as will be described later under Fig. 5. The curves of Fig. 3d are all tangent to the vertical axis at the same point a; for each of these sets of roots the distance Oa is the geometrical mean value of all four roots.

The general case of four complex roots is illustrated by Fig. 4 for the numerical values $\alpha_1=-1+i10$, $\alpha_2=-1-i10$, $\alpha_3=-2+i15$, $\alpha_4=-2-i15$. For all complex roots the poles must also be complex; the pole with positive imaginary part must lie in the region bounded by the curve $\Gamma=\Gamma_1+\Gamma_2+\ldots+\Gamma_7$. This curve is tangent to the vertical axis at the point a, and tangent to a vertical line at the left at the point d. The largest absolute value of any point in the domain occurs at the point c, and the smallest at f; these two points are the points of tangency of the curve Γ with circles about the origin as center. The curve Γ is tangent at the point e to a circle through the origin having its center on the real axis. The coordinates of these points are all given in Table V.

The general case of four real roots is illustrated by Fig. 5 for the numerical values $\alpha_1=-1$, $\alpha_2=-2$, $\alpha_3=-5$, $\alpha_4=-7$. The domain of complex poles is bounded by the curve Γ, with the critical points defined and labeled as in Fig. 4. The domain of complex poles is bounded in part by two segments on the real axis, one lying in the interval between α_1 and α_2, the other between α_3 and α_4. Approximately, these segments are from -1.13 to -1.93 and from -5.13 to -6.70, for this numerical example. The points on these segments are in the domain of poles, corresponding to double real poles. The domain of real poles is shown by the graph below the axis, each point of this graph representing two real values, the two points on the real axis reached by following the $\pm 45°$ lines through the point. The domain of real poles is bounded by the continuation of the curve Γ and the tangent lines corresponding to the four roots. This gives two corners associated with the two segments on the real axis, and an isolated rectangle. Corresponding to the points in the rectangle, one pole may be chosen anywhere in the range from α_1 to α_2, and the second pole anywhere in the range from α_3 to α_4. Both poles may be chosen in the range from α_1 to α_2, or in the range from α_3 to α_4, with certain restrictions as shown by the figure, since the curve Γ cuts off the points of the triangles. The two corners and the rectangle are shown by Fig. 5a on a larger scale, with greater accuracy.

In some respects, the case illustrated by Fig. 5 is the most general case, from which all other cases can be obtained by a continuous transformation of the roots. Two of the adjacent real roots may be brought together to a single double root; the corresponding boundary curve then shrinks to a cusp at this point on the real axis, and the rectangle

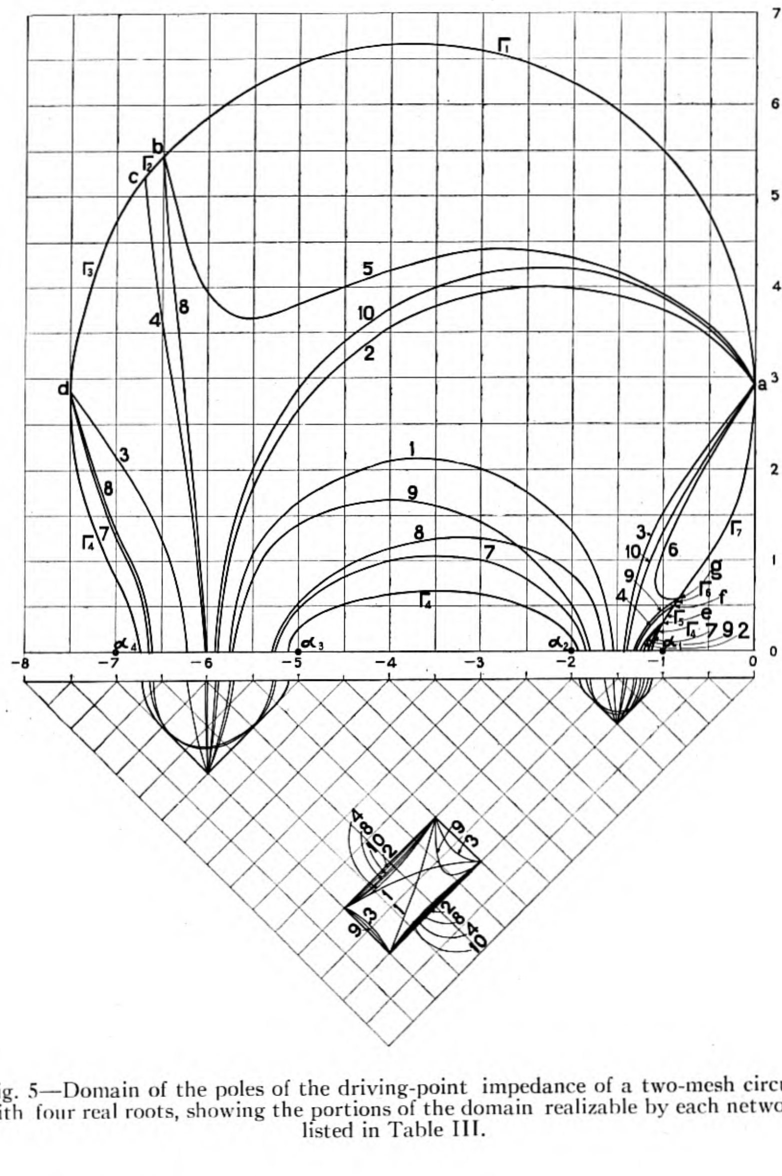


Fig. 5—Domain of the poles of the driving-point impedance of a two-mesh circuit with four real roots, showing the portions of the domain realizable by each network listed in Table III.

in the auxiliary diagram narrows down to a single line segment. Then if the other two real roots are brought together, the boundary curve has a second cusp and the domain in the auxiliary diagram shrinks to a single isolated point. If, now, one of the pairs of equal real roots is separated into a pair of conjugate imaginary roots, the

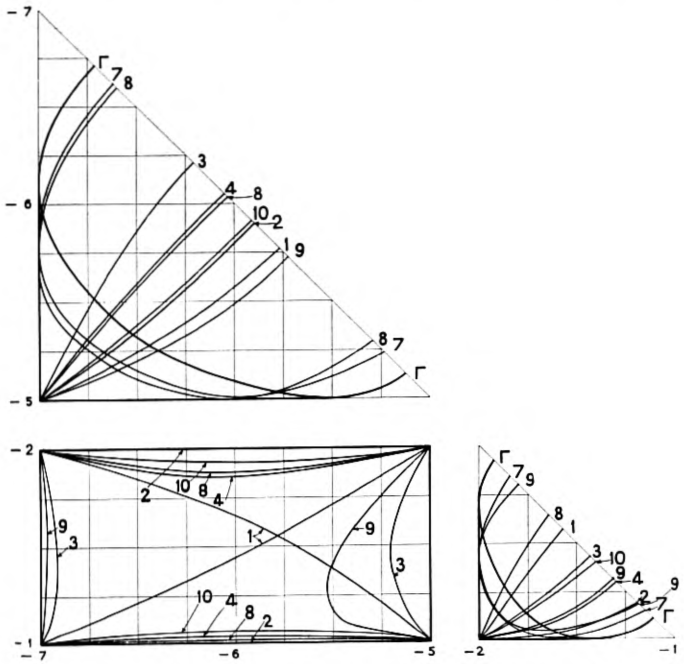


Fig. 5a—Domain of real poles of Fig. 5, on larger scale.

corresponding cusp is rounded off away from the axis, and the point in the auxiliary diagram vanishes. When the other pair of equal real roots separates into conjugate complex roots, the case illustrated by Fig. 4 is obtained. As one pair of complex roots approaches the imaginary axis, the domain narrows until, for one pair of roots on the vertical axis, the domain shrinks to a circular arc as illustrated by Fig. 3a. This sort of transformation may be followed through in different ways in order to obtain any desired distribution of the roots.

The complete domains are unique, that is, any one domain is given by only one set of roots.

Every domain includes the points corresponding to the roots for which the domain is defined. For these points, that is, for a pole coinciding with a root, the impedance expression has a common factor in numerator and denominator. When both poles coincide with roots the corresponding impedance expression can be obtained by means of a one-mesh circuit.

5. Two-Mesh Circuits and Associated Networks

The second object of this paper is the determination of the networks realizing any specified driving-point impedance which satisfies the conditions established in the first part of the paper. It is necessary to find the number, character, and arrangement of the elements in these networks, as well as to find the values of these elements.

Thus the problem met in this investigation differs from the usual network problem in that it calls for the determination of the elements of a network which has a certain specified impedance, instead of calling for the determination of the impedance of a network which has certain specified elements.

The most general two-mesh circuit has three branches connected in parallel, each branch containing resistance, capacity, and self-

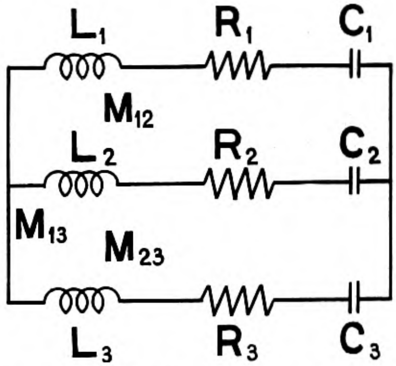


Fig. 6—Most general two-mesh circuit consisting of resistances, capacities, and inductances.

inductance, with mutual inductance between each pair of branches, as shown by Fig. 6.

The most general network under consideration is, therefore, the network obtained by opening one branch of this two-mesh circuit, as shown by Fig. 7. All the networks considered are special cases of this general network, obtained by making a sufficient number of the elements either zero or infinite. If, in particular, all the elements in one branch are replaced by a short circuit, the network splits up into two separate sections connected essentially only by mutual inductance, as shown by Fig. 7a.

It is convenient to limit this investigation to the determination of those networks which, without superfluous elements, realize any driving-point impedance having arbitrarily assigned roots. A network is considered to have superfluous elements if there exist other

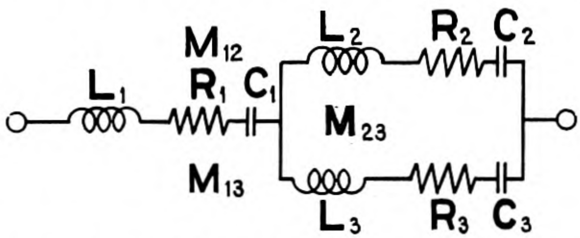


Fig. 7—Most general network obtained by opening one branch of a two-mesh circuit.

networks with fewer elements which, individually or collectively, realize the same range of possible impedances. Impedances with zero, pure imaginary, or infinite roots can be realized by the limiting cases of these networks.

A network realizing an impedance with arbitrarily assigned roots must consist of at least five elements,—one resistance, two capacities,

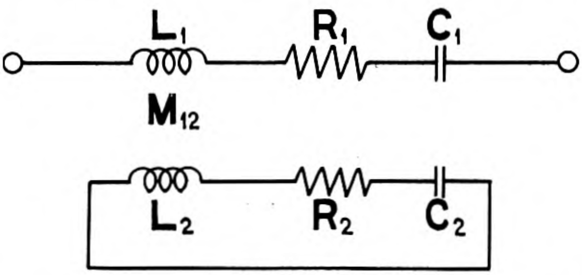


Fig. 7a—Special case of Fig. 7, obtained by replacing the elements of one branch by a short circuit

and two self-inductances, in order that the numerator of the impedance expression (1b) may contain odd powers of λ, a constant term, and a term in λ^4, respectively.

Since the general expression for the driving-point impedance contains essentially seven constants which may be assigned arbitrarily, subject to the restrictions already established, it is to be expected that the entire range of possible impedances can be realized by one or more networks consisting of seven elements only. This proves to be the case. Hence all networks with more than seven elements contain superfluous elements. It is also to be expected that one additional condition must be satisfied by the roots and poles in order that an impedance may be realized by a six-element network, and two additional conditions for a five-element network.

Accordingly, a census has been made of all networks consisting of not more than seven elements, each network containing at least one resistance, two capacities, and two self-inductances. This census is shown by Table III.

Each two-mesh circuit meeting these requirements as to the number of elements is represented in symbolical form in Table III. The letters L, R, and C, printed in the first, second, or third lines of the symbol, indicate the presence of self-inductance, resistance, and capacity in the first, second, or third branches of the circuit, respectively. The letter M is printed in the two lines of the symbol corresponding to the two branches which are connected by a mutual inductance. Thus the first circuit in the table is represented by the symbol

$$\begin{matrix} LRCM \\ L\ \ CM \\ L\qquad \end{matrix}$$

which indicates self-inductance, resistance, and capacity in the first branch, self-inductance and capacity in the second branch, and self-inductance in the third branch, with mutual inductance between the first two branches.

Three networks, in general, are obtained from each of these circuits by opening each of the three branches. If two of these branches are alike, only two distinct networks are obtained. If one branch of a circuit is a short-circuit, there being no elements assigned to that branch, the network obtained by opening one of the other branches is of the type shown by Fig. 7a; if the short-circuited branch is opened, the network consists simply of the parallel combination of the other two branches.

With circuits represented in this symbolical manner, there is, opposite each line of the symbol, a reference to the domain of poles indicating the portion of the domain realizable by the network obtained by opening the corresponding branch. Two like branches in a circuit are bracketed together with a single reference mark, since they each give the same network. The entire domain is indicated by Σ; the boundary curve of the domain by Γ, this being divided into seven segments, Γ_1, Γ_2, . . . , Γ_7; ten critical lines in the domain by the

TABLE III

Networks Realizing the Driving-Point Impedance of a Two-Mesh Circuit

1.	*LRCM* *L CM* *L*	((*a*)) ((*e*)) (Γ)
2.	*LRCM* *L C* *L M*	((*a*)) ((*e*)) (Γ)
3.	*LRC* *L CM* *L M*	((*a*)) ((*e*)) (Γ)
4.	*LR M* *L CM* *L C*	((*a*)) (Γ) (Γ)
5.	*LR* *L CM* *L CM*	((*a*)) } (Γ)
6.	*LRCM* *LRCM* —	} (2) Σ
7.	*LRCM* *LR M* *C*	Σ Σ Σ
8.	*LRCM* *L CM* *R*	Σ Σ Σ
9.	*LRCM* *L M* *RC*	(2) Σ (2)
10.	*LR M* *L CM* *RC*	Σ Σ Σ
11.	*LRCM* *L CM* *C*	((*a*)) (Γ) (Γ)
12.	*L CM* *L CM* *RC*	} (Γ) ((*a*))
13.	*LRC* *LRC* *L*	} (2) 1 − 8 − Γ₂ − Γ₃
14.	*LRC* *LR* *L C*	3 − 4 − 6 − Γ₃ − Γ₆ 3 − 5 − 10 − Γ₂ − Γ₃ 4 − 7 − 9 − Γ₃ − Γ₆
15.	*LRC* *L C* *L C*	((*a*)) } (Γ₂, Γ₃, Γ₅, Γ₆)
16.	*LRC* *LRC* *R*	} (4) 1 − 10
17.	*LRC* *LR* *RC*	2 − 3 − 7 3 − 5 − 8 2 − 6 − 9
18.	*LRC* *LRC* *C*	} (3) 1 − 9 − Γ₅ − Γ₆
19.	*LRC* *L C* *RC*	2 − 4 − 5 − Γ₂ − Γ₅ 4 − 7 − 8 − Γ₂ − Γ₅ 2 − 6 − 10 − Γ₅ − Γ₆
20.	*LRCM* *L CM* —	(*a*) (*e*) Γ
21.	*LRCM* *L M* *C*	(*a*) Γ (*e*)
22.	*LR M* *L CM* *C*	(*a*) Γ Γ
23.	*L CM* *L CM* *R*	} Γ (*a*)
24.	*L CM* *L M* *RC*	(*e*) Γ (*a*)
25.	*LRC* *L C* *L*	(*a*) (*e*) Γ₂, Γ₃
26.	*LR* *L C* *L C*	(*a*) } Γ₃, Γ₆
27.	*LRC* *LRC* —	} * 1
28.	*LRC* *LR* *C*	3 3 9
29.	*LRC* *L C* *R*	4 4 10
30.	*LRC* *L* *RC*	2 8 2
31.	*LR* *L C* *RC*	5 7 6
32.	*LRC* *L C* *C*	(*a*) (*d*) Γ₅, Γ₆
33.	*L C* *L C* *RC*	} Γ₂, Γ₅ (*a*)
34.	*LRC* *L C* —	* * **
35.	*LRC* *L* *C*	*a* *d* *e*
36.	*LR* *L C* *C*	*a* *d* *g*
37.	*L C* *L C* *R*	} *c*, *a*
38.	*L C* *L* *RC*	*e* *b* *a*

* Equivalent to a single-mesh network.

** Network by which only impedances having one pair of imaginary roots can be realized.

( ) The impedances realized by this network can be realized by other networks with one less element.

(( )) The impedances realized by this network can be realized by other networks with two less elements.

numbers 1, 2, . . . , 10; and seven critical points by the letters *a*, *b*, . . . , *g*, as illustrated by Figs. 4 and 5.

Networks with superfluous elements are indicated by placing parentheses around the corresponding reference mark, single parentheses for one superfluous element and double parentheses for two. In order that a seven-element network may contain no superfluous elements it must give the entire domain or a region in it, a six-element network a critical line, and a five-element network a critical point.

That is, an impedance with arbitrarily assigned roots, and with a pole chosen arbitrarily in the domain corresponding to these assigned roots, can be realized with the minimum number of elements only by a seven-element network. If the pole is chosen so as to satisfy one additional condition, namely, chosen at a point on one of the critical lines of the domain (including the boundary curve), the impedance can be realized by the six-element network giving that line. If the pole is chosen so as to satisfy two additional conditions, namely, chosen at one of the critical points, the impedance can be realized by the corresponding five-element network.

The conditions for the critical lines and for the critical points are given by Tables IV and V, respectively, in terms of the coefficients of the impedance.

TABLE IV

Critical Lines

Γ. $a_3^2b_1^2+(4a_1a_4-2a_2a_3)b_1b_2-2a_1a_3b_1b_3-(4a_0a_4-a_2^2)b_2^2+(4a_0a_3-2a_1a_2)b_2b_3+a_1^2b_3^2=0.$

1. $(8a_1a_4^2-4a_2a_3a_4+a_3^3)b_1^3-(16a_0a_4^2+2a_1a_3a_4-4a_2^2a_4+a_2a_3^2)b_1^2b_2$
$+(8a_0a_3a_4-4a_1a_2a_4+a_1a_3^2)b_1^2b_3+(8a_0a_3a_4-4a_1a_2a_4+a_1a_3^2)b_1b_2^2$
$-6(a_0a_3^2-a_1^2a_4)b_1b_2b_3-(8a_0a_1a_4-4a_0a_2a_3+a_1^2a_3)b_1b_3^2$
$-(a_0a_3^2-a_1^2a_4)b_2^3-(8a_0a_1a_4-4a_0a_2a_3+a_1^2a_3)b_2^2b_3+(16a_0^2a_4+2a_0a_1a_3$
$-4a_0a_2^2+a_1^2a_2)b_2b_3^2-(8a_0^2a_3-4a_0a_1a_2+a_1^3)b_3^3=0.$

2. $2a_4b_1b_2b_3-a_3b_1b_3^2-a_4b_2^3+a_3b_2^2b_3-a_2b_2b_3^2+a_1b_3^3=0.$

3. $a_3b_1^3-a_2b_1^2b_2-a_1b_1^2b_3+a_1b_1b_2^2+2a_0b_1b_2b_3-a_0b_2^3=0.$

4. $a_4b_1^2b_2-a_3b_1^2b_3+a_1b_1b_3^2-a_0b_2b_3^2=0.$

5. $a_3a_4b_1^3b_2+a_3^2b_1^3b_3+(a_2a_4-a_3^2)b_1^2b_2^2-(a_1a_4-a_2a_3)b_1^2b_2b_3$
$-2a_1a_3b_1^2b_3^2-a_1a_4b_1b_2^3+a_1a_3b_1b_2^2b_3+(a_0a_3-a_1a_2)b_1b_2b_3^2$
$+a_1^2b_1b_3^3+a_0a_4b_2^4-a_0a_3b_2^3b_3+a_0a_2b_2^2b_3^2-a_0a_1b_2b_3^3=0.$

6. $a_3a_4b_1^3b_2-a_3^2b_1^3b_3-a_2a_4b_1^2b_2^2-(a_1a_4-a_2a_3)b_1^2b_2b_3+2a_1a_3b_1^2b_3^2$
$+a_1a_4b_1b_2^3-a_1a_3b_1b_2^2b_3+(a_0a_3-a_1a_2)b_1b_2b_3^2-a_1^2b_1b_3^3-a_0a_4b_2^4$
$+a_0a_3b_2^3b_3-(a_0a_2-a_1^2)b_2^2b_3^2-a_0a_1b_2b_3^3=0.$

7. $a_3a_4b_1^3b_2-a_3^2b_1^3b_3-a_2a_4b_1^2b_2^2+(a_1a_4+a_2a_3)b_1^2b_2b_3-2a_1a_3b_1^2b_3^2$
$+a_1a_4b_1b_2^3-a_1a_3b_1b_2^2b_3+(a_0a_3+a_1a_2)b_1b_2b_3^2-a_1^2b_1b_3^3-a_0a_4b_2^4$
$+a_0a_3b_2^3b_3-a_0a_2b_2^2b_3^2+a_0a_1b_2b_3^3=0.$

8. $(8a_1a_4^2-4a_2a_3a_4+a_3^3)b_1^5-(8a_0a_4^2+2a_1a_3a_4-4a_2^2a_4+a_2a_3^2)b_1^4b_2$
$+(4a_0a_3a_4+a_1a_3^2-4a_1a_2a_4)b_1^4b_3+(2a_0a_3a_4+a_1a_3^2-4a_1a_2a_4)b_1^3b_2^2$
$-(4a_0a_2a_4-6a_1^2a_4)b_1^3b_2b_3+(4a_0a_1a_4-a_1^2a_3)b_1^3b_3^2-(a_0a_3^2-4a_0a_2a_4$
$-a_1^2a_4)b_1^2b_2^3-(8a_0a_1a_4+a_1^2a_3)b_1^2b_2^2b_3+(2a_0a_1a_3+a_1^2a_2)b_1^2b_2b_3^2$
$-a_1^3b_1^2b_3^3-2a_0a_1a_4b_1b_2^4+(2a_0^2a_4+2a_0a_1a_3)b_1b_2^3b_3$
$-(a_0^2a_3+2a_0a_1a_2)b_1b_2^2b_3^2+2a_0a_1^2b_1b_2b_3^3+a_0^2a_4b_2^5-a_0^2a_3b_2^4b_3$
$+a_0^2a_2b_2^3b_3^2-a_0^2a_1b_2^2b_3^3=0.$

9. $a_3a_4^2b_1^3b_2^2-2a_3^2a_4b_1^3b_2b_3+a_3^3b_1^3b_3^2-a_2a_4^2b_1^2b_2^3+(a_1a_4^2+2a_2a_3a_4)b_1^2b_2^2b_3$
$-(2a_1a_3a_4+a_2a_3^2)b_1^2b_2b_3^2-(4a_0a_3a_4-a_1a_3^2)b_1^2b_3^3+a_1a_4^2b_1b_2^4$
$-(2a_0a_4^2+2a_1a_3a_4)b_1b_2^3b_3+(8a_0a_3a_4+a_1a_3^2)b_1b_2^2b_3^2$
$+(4a_0a_2a_4-6a_0a_3^2)b_1b_2b_3^3-(4a_0a_1a_4-4a_0a_2a_3+a_1^2a_3)b_1b_3^4$
$-a_0a_4^2b_2^5+2a_0a_3a_4b_2^4b_3-(4a_0a_2a_4+a_0a_3^2-a_1^2a_4)b_2^3b_3^2$
$-(2a_0a_1a_4-4a_0a_2a_3+a_1^2a_3)b_2^2b_3^3+(8a_0^2a_4+2a_0a_1a_3-4a_0a_2^2+a_1^2a_2)b_2b_3^4$
$-(8a_0^2a_3-4a_0a_1a_2+a_1^3)b_3^5=0.$

10. $a_3^2a_4b_1^4b_2+a_3^3b_1^4b_3-a_3^3b_1^3b_2^2-2a_1a_3a_4b_1^3b_2b_3-3a_1a_3^2b_1^3b_3^2$
$+(4a_0a_4^2-a_2^2a_4+a_2a_3^2)b_1^2b_2^3+(4a_0a_3a_4+3a_1a_3^2-a_2^2a_3)b_1^2b_2^2b_3$
$-(a_0a_3^2-a_1^2a_4)b_1^2b_2b_3^2+3a_1^2a_3b_1^2b_3^3$
$-(4a_0a_3a_4-2a_1a_2a_4+a_1a_3^2)b_1b_2^4-(4a_0a_1a_4+3a_1^2a_3-a_1a_2^2)b_1b_2^2b_3^2$
$+2a_0a_1a_3b_1b_2b_3^3-a_1^3b_1b_3^4+(a_0a_3^2-a_1^2a_4)b_2^5+(4a_0a_1a_4-2a_0a_2a_3$
$+a_1^2a_3)b_2^4b_3-(4a_0^2a_4-a_0a_2^2+a_1^2a_2)b_2^3b_3^2+a_1^3b_2^2b_3^3-a_0a_1^2b_2b_3^4=0.$

TABLE V

Critical Points

Point	Coordinates	
	$\frac{b_2}{b_1}$	$\frac{b_3}{b_1}$
a	0	$\frac{a_3}{a_1}$
b	$\frac{a_0a_3^2+a_1^2a_4-a_1a_2a_3}{a_0(a_1a_4-a_2a_3)}$	$\frac{a_0a_3a_4+a_1a_2a_4-a_2^2a_3}{a_0(a_1a_4-a_2a_3)}$
c	$\frac{a_1}{2a_0}-\frac{2a_0a_3-a_1a_2}{2a_0\sqrt{a_2^2-4a_0a_4}}$	$\frac{1}{2a_0}(a_2+\sqrt{a_2^2-a_0a_4})$
d	$\frac{a_1}{a_0}$	$\frac{a_1a_2-a_0a_3}{a_0a_1}$
e	$\frac{a_3^2}{a_2a_3-a_1a_4}$	$\frac{a_3a_4}{a_2a_3-a_1a_4}$
f	$\frac{a_1}{2a_0}+\frac{2a_0a_3-a_1a_2}{2a_0\sqrt{a_2^2-4a_0a_4}}$	$\frac{1}{2a_0}(a_2-\sqrt{a_2^2-4a_0a_4})$
g	$\frac{a_0a_3^2+a_1^2a_4-a_1a_2a_3}{a_0a_1a_4+a_0a_2a_3-a_1a_2^2}$	$\frac{a_4(a_0a_3-a_1a_2)}{a_0a_1a_4+a_0a_2a_3-a_1a_2^2}$

These critical lines and points are illustrated, for numerical cases, by Figs. 4 and 5. The graph showing the domain of real poles in Fig. 5 is inaccurate to the extent that the critical lines have been spread somewhat apart from each other in order to show the sequence in which they occur. The actual curves are shown accurately drawn and on a larger scale in Fig. 5a. Even on this scale, Curve 2 cannot be distinguished from the side of the rectangle.

The diagrams for the domain of complex poles, as illustrated by Figs. 4 and 5, are approximately symmetrical with respect to the interchanging of inductances and capacities, with corresponding interchanges in all the curves and formulas. Thus b and g correspond, c and f, d and e, 2 and 3, 5 and 6, 8 and 9, α_1 and α_4, α_2 and α_3; while a, 1, 4, 7, and 10 remain unchanged. In the domain of real poles shown by Fig. 5, this symmetry does not appear. The explanation of this apparent discrepancy is as follows: Upon interchanging inductances and capacities, the values of the roots are changed to their reciprocals. Thus Fig. 5 is symmetrical with the corresponding figure drawn for the case of roots equal to -1, $-1/2$, $-1/5$, and $-1/7$, and thus symmetrical with the figure drawn for roots at -7, $-7/2$, $-7/5$, and -1, since the relative distribution of the roots is the same. This set of roots differs not very considerably from the original set of roots, in reverse order. In the main, therefore, the two figures may be expected to be approximately the same, that is, the original figure symmetrical with itself. In the rectangle, however, very small numerical changes in the constants make relatively large changes in the curves; so it is not surprising to find a lack of symmetry here. If the roots are assigned so that the product of two roots is equal to the product of the other two, there will be true symmetry in the corresponding diagram.

Table III lists 38 circuits, giving a total of 102 networks. Of these networks, three are essentially the equivalent of networks obtained from a one-mesh circuit, one realizes only those impedances which have one pair of pure imaginary roots, and, of the 98 remaining, 41 have superfluous elements. This leaves a total of 57 networks, of which 11 realize the entire domain as given by Theorem II, 12 realize regions in the domain as given by Theorem III, 23 realize critical lines in the domain, and 11 realize critical points.

The eleven networks of Theorem II are included in the first column of Table III and shown in detail by Fig. 1. Formulas for the computation of their elements are given by Table I. Thus the values of these elements can be computed directly in terms of the coefficients of the impedance expression as stated in the form (1b). The following method of computation is convenient:—First compute d as the root of the quadratic equation (21), which is repeated at the bottom of the table. Then find c by subtracting this value of d from a_2. Next compute T_1, T_2, and T_3, assigning signs so that the identity $b_1T_1+b_2T_2+b_3T_3=0$ is satisfied; this is possible since the equation for d was obtained by rationalizing this relation among the T's. There are, in general, two sets of signs for which this identity is satisfied; it is immaterial which set is chosen since the signs of all the T's may be changed without changing the values of any of the elements. Then compute U_1, U_2, and U_3, assigning positive values to each of these. With the values of all these quantities determined, the values of the elements of the networks can be calculated directly from the formulas given in the body of the table. If this solution turns out to be impossible, that is, if the value of an element is found to be negative or complex or if the value of a mutual inductance is found to be greater than the square root of the product of the associated self-inductances, it means that the conditions upon the roots and poles are not satisfied. If the conditions established in the first part of this paper are satisfied, the solution is possible.

These formulas give all the special cases of the eleven networks automatically, that is, the values of the appropriate elements will turn out to be zero or infinite, as the case may be. Since each of these eleven networks covers the entire domain, they are all mutually equivalent at all frequencies. These are the only networks without superfluous elements which cover the entire domain, that is, any network covering the entire domain must be one of these eleven or a network obtained from one of these by introducing additional elements. Each of the eleven contains just seven elements; thus the prediction that a seven-element network would cover the entire domain is verified. The three remaining networks of this same type, one from Circuit 6 and two from Circuit 9 of Table III give special cases only, in the sense that each of these can realize only those impedances which have a pole lying on Line 2; thus each of these three contains a superfluous element, since all the points on Line 2 can be realized by six-element networks, as shown in the fourth column of the table.

Network 1 of Fig. 1 is of particular interest since it consists simply of two branches in parallel, each containing resistance, capacity, and self-inductance, with mutual inductance between them.[10] By Theorem II, this network can be made equivalent to any network whatsoever obtained from a two-mesh circuit.

[10] It will be shown in a subsequent paper that any driving-point impedance of an n-mesh circuit can be realized by a network of n branches in parallel, each branch containing resistance, capacity, and self-inductance, with mutual inductance between each pair of branches.

The twelve networks of Theorem III are included in the second column of Table III and shown in detail by Fig. 2. Formulas for the computation of their elements are given by Table II. The values of the elements can be computed by the same rule as that given above for Table I.

Each of these twelve networks realizes those impedances which have poles lying in a certain restricted area or region of the entire domain of possibilities, as indicated for each network in the table by a specification of the boundary curves of the area. For each particular impedance in the domain various sets of these twelve networks are mutually equivalent. Some points in the domain cannot be realized by networks without mutual inductance. Of the remaining points, each is realizable, in general, by at least three, and by not more than five, of these twelve networks. This region of the domain which is realizable without mutual inductance is covered, with no overlapping, by each of the four following sets of networks: 13, 17, and 21; 13, 18, and 22; 14, 17, and 23; 15, 19, and 21; the numbers refer to the networks of Fig. 2.

That portion of the domain which cannot be realized by networks without mutual inductance comprises the three regions bounded by Γ_1 and 5, Γ_4 and 7, and Γ_7 and 6, respectively, as illustrated by Figs. 4 and 5.

The third and fourth columns of Table III show a total of 23 networks, each with six elements, realizing lines in the domain. Of these, eleven are derived as special cases of the networks of both Figs. 1 and 2, six as special cases of Fig. 1 but not of Fig. 2, and six as special cases of Fig. 2 alone. The fifth column of the table shows the eleven networks, each with five elements, realizing points in the domain.

6. Formulas for Calculation of General Network

Formulas for the calculation of the values of the elements of the general network of Fig. 7 are given in Theorem IV. These are given in the form of nine equations (7)–(15), inclusive, involving the twelve elements of the network and two parameters, d and k. The parameter d, however, is fixed by the impedance, since the left-hand members of equations (13)–(15) satisfy the identity (20). Upon substituting the right-hand members in the identity and rationalizing, equation (21) is obtained, this being a quadratic equation in d with coefficients which are functions of the known coefficients of the impedance. Since d is fixed in this way, there are essentially eight equations in thirteen variables,—the twelve elements and the arbitrary parameter k. In general, therefore, five of the elements may be specified, or five relations among the elements; whereupon the equations can be solved. Thus it is to be expected that a seven-element network will realize, in general, any specified driving-point impedance.

This method of solution is best illustrated by considering a particular case. Take, for example, the derivation of the formulas for Network 1 of Fig. 1, as given by Table I. This is the special case of the general network of Fig. 7 obtained by making $L_1=R_1=C_1^{-1}=M_{12}=M_{13}=0$. Substituting these values, together with the notation of Table I, equations (7)–(15) become

$$\begin{aligned}
L_2L_3-M_{23}^2&=a_0k^2,\\
R_2R_3&=dk^2,\\
D_2D_3&=a_4k^2,\\
L_2+L_3-2M_{23}&=b_1k^2,\\
R_2+R_3&=b_2k^2,\\
D_2+D_3&=b_3k^2,\\
R_2D_3-R_3D_2&=T_1k^3,\\
D_2L_3-D_3L_2-(D_2-D_3)M_{23}&=T_2k^3,\\
L_2R_3-L_3R_2-(R_3-R_2)M_{23}&=T_3k^3.
\end{aligned}$$

Eliminating R_2, R_3, D_2, and D_3 from the second, third, fifth, sixth, and seventh of these equations, the value of k is found to be equal to $\pm U_1/T_1$. Knowing the value of k, the equations may then be solved for the seven elements, obtaining the results given in Table I. The two sign choices for k in this example correspond to the possibility of interchanging branches 2 and 3 in the network. The values given in Table I are computed for k taken with the negative sign.

In the general solution, the parameter d is obtained from the quadratic equation (21). The explicit solution of this equation is

$$d=\frac{2a_4b_1^2+a_2b_2^2+2a_0b_3^2-a_3b_1b_2-2a_2b_1b_3-a_1b_2b_3\pm2\Delta}{b_2^2-4b_1b_3} \tag{34}$$

where

$$\begin{aligned}
\Delta^2=\;&a_4^2b_1^4+a_0a_4b_2^4+a_0^2b_3^4-a_3a_4b_1^3b_2-(2a_2a_4-a_3^2)b_1^3b_3-a_1a_4b_1b_2^3\\
&-a_0a_3b_2^3b_3-(2a_0a_2-a_1^2)b_1b_3^3-a_0a_1b_2b_3^3\\
&+a_2a_4b_1^2b_2^2+(a_2^2+2a_0a_4-2a_1a_3)b_1^2b_3^2+a_0a_2b_2^2b_3^2\\
&+(3a_1a_4-a_2a_3)b_1^2b_2b_3-(4a_0a_4-a_1a_3)b_1b_2^2b_3\\
&+(3a_0a_3-a_1a_2)b_1b_2b_3^2, && (35)\\
=\;&a_0^2(\alpha_1^2b_1+\alpha_1b_2+b_3)(\alpha_2^2b_1+\alpha_2b_2+b_3)\\
&(\alpha_3^2b_1+\alpha_3b_2+b_3)(\alpha_4^2b_1+\alpha_4b_2+b_3), && (36)\\
=\;&a_0^2b_1^4(\alpha_1-\beta_2)(\alpha_1-\beta_3)(\alpha_2-\beta_2)(\alpha_2-\beta_3)\\
&(\alpha_3-\beta_2)(\alpha_3-\beta_3)(\alpha_4-\beta_2)(\alpha_4-\beta_3). && (37)
\end{aligned}$$

In the case of real and distinct poles, formula (34) gives, in general, two positive values of d satisfying the necessary conditions (4)–(6), and thus two solutions for any particular network. For complex poles, only one such value of d is obtained, and there is thus a unique solution in each case. For real and equal poles, $b_2^2-4b_1b_3=0$, and so formula (34) does not apply directly; in this case, however, (21) reduces to a linear equation in d, so that the solution can be readily found.

An obvious necessary condition for a solution is that $\Delta^2\geq 0$, for otherwise the value of d would be complex. This condition is satisfied for any choice of poles provided there is not an odd number of real roots lying between two real poles. Thus for the case of all complex roots or for the case of complex poles with any choice of roots this condition is automatically satisfied. It is interesting to note that an impedance expression with poles failing to satisfy this condition cannot be realized by any network with positive or negative resistances, capacities, and inductances; it can be realized only by a network with elements having complex values.

7. Networks with Negative Resistances

If negative resistances are allowed in the two-mesh circuit, the only change necessary in the statement of the results of this investigation, as given in Theorems I–IV, is the removal of the restrictions $\alpha_1+\alpha_2\leq 0$, $\alpha_3+\alpha_4\leq 0$, $\beta_2+\beta_3\leq 0$, and $d\geq 0$. This removes the restriction of the real part of each root and pole to negative or zero values. The removal of the restriction on d adds to the domain of poles, considered in the x, y plane, all the ellipses of the family $-\infty<d<0$, thus filling out the region above the critical parabola (33), together with the corners in the case of real roots. In the u, v plane the domain comprises the entire upper half of the complex plane and, in the auxiliary diagram, the complete triangular corners and the rectangle, with the provision that the rectangle is not included in the case of two roots positive and two negative.

By means of a two-mesh circuit employing negative resistances, any impedance expression of the form (1) can be realized, with roots arbitrarily assigned in conjugate pairs or in real pairs, subject only to the condition that the number of positive roots is even, and with any pair of complex poles or with a pair of real poles lying anywhere in the ranges from the first to the second real roots and from the third to the fourth real roots, arranged in order of magnitude, subject only to the condition that both poles must be positive or both negative.

The network diagrams and all the formulas for the calculation of the elements remain unchanged.

8. Mathematical Proof

The circuits treated in this investigation are special cases of the general circuit which has any number of terminals m connected in pairs by $m(m-1)/2$ branches, each of which consists of a self-inductance, a resistance, and a capacity in series, with mutual inductance between each pair of branches. The only restrictions imposed are those inherent in all electrical circuits, namely, that the magnetic energy, the dissipation, and the electric energy are each positive for any possible distribution of currents in the branches. Circuits with any arrangement of elements in series or in parallel or in separated meshes can be derived as limiting cases of this general circuit by making a sufficient number of the inductances, resistances, and capacities either zero or infinite.

This general circuit connecting m terminals or branch-points has $n=(m-1)(m-2)/2$ degrees of freedom, that is, n independent meshes. The discriminant[11] of the circuit is the determinant A having the element Z_{jk} in the jth row and kth column, Z_{jk} being the mutual impedance between meshes j and k (self-impedance when $j=k$), the determinant including n independent meshes of the circuit.

The driving-point impedance in the qth mesh S_q is equal to the ratio A/A_{qq}, where A_{qq} is the cofactor of the element in the qth row and qth column of the determinant A. In general, the cofactor of the product of the elements located at the intersection of rows $j, q, s, \ldots$ with columns $k, r, t, \ldots$, respectively, will be denoted by $A_{jk.qr.st\ldots}$.

The determinant A for the general circuit described above is of order n with the element

$$Z_{jk}=iL_{jk}p+R_{jk}+(iC_{jk}p)^{-1} \tag{38}$$

where L_{jk}, R_{jk}, and C_{jk} are the inductance, the resistance, and the capacity, respectively, common to the two meshes j and k. The inductance L_{jk} includes, therefore, the self-inductances of the branches common to the two meshes together with the mutual inductances connecting each branch of one mesh with each branch of the other mesh. The determinant is symmetrical, that is $Z_{jk}=Z_{kj}$, since $L_{jk}=L_{kj}$, $R_{jk}=R_{kj}$, and $C_{jk}=C_{kj}$.

[11] A complete discussion of the solution of circuits by means of determinants has been given by G. A. Campbell, *loc. cit.*, pages 883–886.

These coefficients L_{jk}, R_{jk}, and C_{jk} are subject to the energy conditions stated above, namely, that the magnetic energy, the dissipation, and the electric energy,

$$\frac{1}{2}\sum_{j=1}^{n}\sum_{k=1}^{n} L_{jk}i_j i_k,\ \sum_{j=1}^{n}\sum_{k=1}^{n} R_{jk}i_j i_k,\ \text{and}\ \frac{1}{2}\sum_{j=1}^{n}\sum_{k=1}^{n}\frac{1}{C_{jk}}\int i_j dt\int i_k dt, \tag{39}$$

respectively, are each positive for any possible distribution of the currents (i_j, i_k, . . .) in the branches of the circuits.[12] In other words, the coefficients L_{jk}, R_{jk}, and $1/C_{jk}$ are subject to the condition that the three quadratic forms of which these are the coefficients must be positive for all real values of the variables. All the principal minors of the determinants

$$\begin{vmatrix} L_{11} & L_{12} & \dots & L_{1n} \\ L_{21} & L_{22} & \dots & L_{2n} \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ L_{n1} & L_{n2} & \dots & L_{nn} \end{vmatrix},\ \begin{vmatrix} R_{11} & R_{12} & \dots & R_{1n}' \\ R_{21} & R_{22} & \dots & R_{2n} \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ R_{n1} & R_{n2} & \dots & R_{nn} \end{vmatrix},\ \text{and}\ \begin{vmatrix} \frac{1}{C_{11}} & \frac{1}{C_{12}} & \dots & \frac{1}{C_{1n}} \\ \frac{1}{C_{21}} & \frac{1}{C_{22}} & \dots & \frac{1}{C_{2n}} \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \frac{1}{C_{n1}} & \frac{1}{C_{n2}} & \dots & \frac{1}{C_{nn}} \end{vmatrix} \tag{40}$$

are positive or zero by virtue of this condition.[13] This same condition holds for the inductances if the coefficients L_{jk} apply to branches instead of meshes.

By expanding the determinants in the numerator and denominator of the expression for the driving-point impedance given above, we find

$$S_q=\frac{A}{A_{qq}}=\frac{a_0(ip)^n+a_1(ip)^{n-1}+a_2(ip)^{n-2}+\dots+a_{2n-1}(ip)^{-n+1}+a_{2n}(ip)^{-n}}{b_1(ip)^{n-1}+b_2(ip)^{n-2}+\dots+b_{2n-1}(ip)^{-n+1}} \tag{41}$$

[12] For a recent statement of the energy conditions in this form see L. Bouthillon, *Revue Générale de l'Electricité*, 11, 1922, pages 656–661.

[13] A necessary and sufficient condition that the real quadratic form in n variables

$$\sum_{j=1}^{n}\sum_{k=1}^{n} a_{jk}x_j x_k \qquad (a_{jk}=a_{kj}),$$

be positive for all real values of the variables is that each of the n determinants,

$$a_{11},\ \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix},\ \dots,\ \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \dots & \cdot \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix},$$

be positive. For a proof of this see, for example, H. Hancock, "Theory of Maxima and Minima," 1917, pages 82–91.

Upon substituting $\lambda = ip$, multiplying numerator and denominator by λ^n, and dropping the subscript q, formula (41) becomes

$$S = \frac{a_0\lambda^{2n} + a_1\lambda^{2n-1} + a_2\lambda^{2n-2} + \ldots + a_{2n-1}\lambda + a_{2n}}{b_1\lambda^{2n-1} + b_2\lambda^{2n-2} + \ldots + b_{2n-1}\lambda} \qquad (42)$$

which may be taken as the most general form of a driving-point impedance. This formula, therefore, gives the impedance of the circuit for any electrical oscillations of the form $e^{\lambda t}$, where λ may have any value, real or complex. Formula (42) may be written in the alternative form

$$S = H\frac{(\lambda-\alpha_1)(\lambda-\alpha_2)(\lambda-\alpha_3)\ldots(\lambda-\alpha_{2n-1})(\lambda-\alpha_{2n})}{\lambda(\lambda-\beta_2)(\lambda-\beta_3)\ldots(\lambda-\beta_{2n-1})}. \qquad (43)$$

Thus there are $2n$ roots of S, regarded as a function of λ, which are the $2n$ resonant points of the circuit. There are also $2n$ poles of S, which are the $2n$ anti-resonant points of the circuit, namely, zero, infinity, and the $2n-2$ resonant points of the circuit obtained by opening the branch in which the driving-point impedance is measured.

Upon setting $n=2$ in equations (43) and (42), formulas (1a) and (1b) are obtained, respectively.

From the fact that the coefficients L_{jk},R_{jk}, and $1/C_{jk}$ satisfy the quadratic form conditions (39), it can be shown mathematically that the coefficients a_0, a_1, . . . , a_{2n} of (42) are all positive and that the roots α_1, α_2, . . . , α_{2n} of (43) have negative real parts.[14] This can also be shown from the fact that the free oscillations of the circuit are of the forms $e^{\lambda_1 t}$, $e^{\lambda_2 t}$, . . ., $e^{\lambda_{2n} t}$. Thus the roots occur in pairs each of which has negative real values or conjugate complex values with negative real parts.

The same restrictions hold for the coefficients b_1, b_2, . . . , b_{2n-1} and the poles β_2, β_3, . . . , β_{2n-1} since the denominator of S, with the exception of the factor λ^n, is also the discriminant of a circuit. Thus the general restrictions (2) are obtained.

In order to obtain the necessary and sufficient conditions that a function of the type (1b) represent a driving-point impedance realizable by a two-mesh circuit, set this function equal to the impedance of the most general two-mesh circuit and investigate the conditions which must hold upon the coefficients in order that the two forms may be equivalent.

[14] The mathematical work is identical with the mathematics of the corresponding dynamical problem. A detailed proof is given by A. G. Webster, *loc. cit.*

The discriminant of the most general two-mesh circuit is of the form

$$A = \begin{vmatrix} L_{11}\lambda+R_{11}+D_{11}\lambda^{-1}, & L_{12}\lambda+R_{12}+D_{12}\lambda^{-1} \\ L_{12}\lambda+R_{12}+D_{12}\lambda^{-1}, & L_{22}\lambda+R_{22}+D_{22}\lambda^{-1} \end{vmatrix}, \tag{44}$$

where the three sets of coefficients, using D_{jk} instead of $1/C_{jk}$, are subject to the restriction that the three determinants

$$\begin{vmatrix} L_{11} & L_{12} \\ L_{12} & L_{22} \end{vmatrix}, \quad \begin{vmatrix} R_{11} & R_{12} \\ R_{12} & R_{22} \end{vmatrix}, \text{ and } \begin{vmatrix} D_{11} & D_{12} \\ D_{12} & D_{22} \end{vmatrix} \tag{45}$$

are all positive or zero, as well as L_{11}, R_{11}, and D_{11}. This condition requires L_{22}, R_{22}, and D_{22} also to be positive or zero.

The most general driving-point impedance of a two-mesh circuit may be taken as the impedance in the first mesh of the circuit defined by the discriminant (44). Set A/A_{11} equal to the value of S given by (1b). Expanding into polynomials in λ, and equating coefficients of the numerators and denominators of the two expressions, the following relations are obtained:

$$L_{11}L_{22}-L_{12}^2=a_0k^2, \tag{46}$$
$$L_{11}R_{22}+L_{22}R_{11}-2L_{12}R_{12}=a_1k^2, \tag{47}$$
$$L_{11}D_{22}+L_{22}D_{11}+R_{11}R_{22}-2L_{12}D_{12}-R_{12}^2=a_2k^2, \tag{48}$$
$$R_{11}D_{22}+R_{22}D_{11}-2R_{12}D_{12}=a_3k^2, \tag{49}$$
$$D_{11}D_{22}-D_{12}^2=a_4k^2, \tag{50}$$
$$L_{22}=b_1k^2, \tag{51}$$
$$R_{22}=b_2k^2, \tag{52}$$
$$D_{22}=b_3k^2, \tag{53}$$

where k has any real value other than zero. Introduce the notation

$$R_{11}R_{22}-R_{12}^2=dk^2, \tag{54}$$

where d is positive or zero. Then, using (46), (54), and (50), eliminate L_{11}, R_{11}, and D_{11} from equations (47)–(49), obtaining

$$(L_{12}R_{22}-L_{22}R_{12})^2=k^2(-dL_{22}^2+a_1L_{22}R_{22}-a_0R_{22}^2), \tag{55}$$
$$(D_{12}L_{22}-D_{22}L_{12})^2=k^2[-a_0D_{22}^2+(a_2-d)D_{22}L_{22}-a_4L_{22}^2], \tag{56}$$
$$(R_{12}D_{22}-R_{22}D_{12})^2=k^2(-a_4R_{22}^2+a_3R_{22}D_{22}-dD_{22}^2). \tag{57}$$

Using (51)–(53), eliminate L_{22}, R_{22}, and D_{22} from the right-hand members of (55)–(57); extract the square root; rearrange the order of the equations, obtaining

$$R_{12}D_{22}-R_{22}D_{12}=\pm k^3(-a_4b_2^2+a_3b_2b_3-db_3^2)^{1/2}, \tag{58}$$
$$D_{12}L_{22}-D_{22}L_{12}=\pm k^3[-a_0b_3^2+(a_2-d)b_3b_1-a_4b_1^2]^{1/2}, \tag{59}$$
$$L_{12}R_{22}-L_{22}R_{12}=\pm k^3(-db_1^2+a_1b_1b_2-a_0b_2^2)^{1/2}. \tag{60}$$

Thus conditions (4)–(6) are obtained directly from (58)–(60). The left-hand members of (58)–(60) satisfy the identity

$$(R_{12}D_{22}-R_{22}D_{12})L_{22}+(D_{12}L_{22}-D_{22}L_{12})R_{22} \\ +(L_{12}R_{22}-L_{22}R_{12})D_{22}=0. \qquad (61)$$

Substituting (51)–(53) and (58)–(60) in this identity (61), and rationalizing, equation (3) and its equivalent (21) are obtained.

For the general network of Fig. 7,

$$\left.\begin{array}{lll} L_{11}=L_1'+L_2', & L_{12}=L_2', & L_{22}=L_2'+L_3', \\ R_{11}=R_1+R_2, & R_{12}=R_2, & R_{22}=R_2+R_3, \\ D_{11}=D_1+D_2, & D_{12}=D_2, & D_{22}=D_2+D_3, \end{array}\right\} \qquad (62)$$

where L_1', L_2', and L_3' are defined by (17)–(19). For this set of constants, branch 2 is made the branch common to the two meshes; the choice of branch 3 as the common branch would not affect the final formulas. Substituting these values (62) in (46), (54), (50)–(53), and (58)–(60), equations (7)–(15) are obtained directly.

Thus Theorems I and IV are completely proved. Theorems II and III are verified by the actual formulas for the elements given in Tables I and II, and by the census of networks presented in Table III.

I am indebted to Dr. George A. Campbell for inspiring the writing of this paper and for specific advice upon many points, and to Miss Frances Thorndike for the preparation of the tables and figures.

Contributors to this Issue

Halsey A. Frederick, B.S., Princeton University, 1910; E.E., Princeton University, 1912; Engineering Department Western Electric Company, 1912—. Mr. Frederick has been engaged in research and development problems connected with telephone receivers, transmitters, and allied subjects in the investigation of the transmission and reproduction of speech and music.

H. F. Dodge, S.B., Mass. Inst. Tech., 1916; Instructor in electrical engineering, Mass. Inst. Tech., 1916–1917; A.M. in Physics and Mathematics, Columbia University, 1922; Engineering Department of Western Electric Company, 1917—. Mr. Dodge has been engaged in development studies on telephone transmitters.

George A. Campbell, B.S., Massachusetts Institute of Technology, 1891; A.B., Harvard, 1892; Ph.D., 1901; Gottingen, Vienna and Paris, 1893–96; Mechanical Department, American Bell Telephone Company, 1897; Engineering Department, American Telephone and Telegraph Company, 1903–19; Department of Development and Research, 1919—; Research Engineer, 1908—. Dr. Campbell has published papers on loading and the theory of electric circuits, including electric wave-filters, and is also well known to telephone engineers for his contributions to repeater and substation circuits.

John R. Carson, B.S., Princeton, 1907; E.E., 1909; M.S., 1912; Research Department, Westinghouse Electric and Manufacturing Company, 1910–12; instructor of physics and electrical engineering, Princeton, 1912–14; American Telephone and Telegraph Company, Engineering Department, 1914–15; Patent Department, 1916–17; Engineering Department, 1918; Department of Development and Research, 1919—. Mr. Carson's work has been along theoretical lines and he has published several papers on theory of electric circuits and electric wave propagation.

Otto J. Zobel, A.B., Ripon College, 1909; A.M., Wisconsin, 1910; Ph.D., 1914; instructor in physics, 1910–15; instructor in physics, Minnesota, 1915–16; Engineering Department, American Telephone and Telegraph Company, 1916–19; Department of Development and Research, 1919—. Mr. Zobel has made important contributions to circuit theory in branches other than the subject of wave-filters.

KARL K. DARROW, S.B., University of Chicago, 1911; University of Paris, 1911–12; University of Berlin, 1912; Ph.D., in physics and mathematics, University of Chicago, 1917; Engineering Department, Western Electric Company, 1917—. At the Western Electric, Mr. Darrow has been engaged largely in preparing studies and analyses of published research in various fields of physics.

RONALD M. FOSTER, S.B., Harvard, 1917; American Telephone and Telegraph Company, Engineering Department, 1917–19; Department of Development and Research, 1919—.