The Bell System Technical Journal

April, 1924

High Frequency Amplifiers

By H. T. FRIIS and A. G. JENSEN

IN this paper, a simplified mathematical treatment of the theory of high frequency amplifiers is presented, and the theory is verified by experiment. This method of mathematical analysis provides a ready means of predicting the performance and action of an amplifier from a knowledge of the fundamental constants of its circuit and places the design of high frequency amplifiers on a precise and rational basis. The paper also includes a description of various methods for quantitatively determining the amount of amplification at high frequencies.

In order to make the discussion more easily followed, it is started with the simple close-coupled non-resonant transformer amplifier of the kind used at audio frequencies, and it is pointed out that good efficiency at higher frequencies requires that the transformer be used

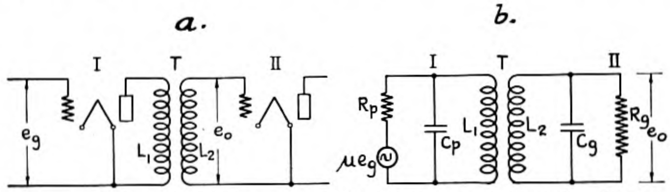


Fig. 1—Schematic of One Stage Transformer Coupled Amplifier

at its natural frequency, i.e., the transformer inductance and distributed capacity must be in resonance. We have therefore, next treated the simplest type of resonance circuit amplifier, namely, a single tuned circuit amplifier, and it is shown that exactly the same method can be used for a choke coil amplifier or a close coupled transformer amplifier. Finally, it is shown that a loosely coupled transformer amplifier can be treated like two coupled tuned circuits.

Considering a low frequency transformer-coupled amplifier, in Fig. 1 (a) there is shown an amplifier tube I with its output transformer *T* working into another tube II and in Fig. 1 (b) is given the corresponding equivalent circuit. The equivalent circuit is obtained by the theorem, that the plate circuit of a vacuum-tube may be treated as an ordinary a.c. circuit, consisting of the external impedance in series with a resistance R_p, and in which the impressed emf. is μe_g, R_p being the internal plate impedance of the tube, μ the amplification constant of the tube and e_g the voltage applied to the grid.

In Fig. 1 (b) C_p is the plate to filament capacity of tube I, and the input impedance of tube II is represented by a resistance R_g in parallel with a condenser C_g.

The maximum amplification which can be obtained by this amplifier is given by the well-known expression

$$K = \frac{e_o}{e_g} = \tfrac{1}{2}\mu\sqrt{\frac{R_g}{R_p}}, \tag{1}$$

but this maximum amplification can only be obtained when

$$\left.\begin{aligned} &\omega L_1 >> R_p, \\ &\omega L_2 >> R_g, \\ \text{and}\qquad &\frac{\omega L_1}{R_p} = \frac{\omega L_2}{R_g}. \end{aligned}\right\} \tag{2}$$

Large reactances $\omega\, L_1$ and $\omega\, L_2$ can only be obtained at low frequencies because at higher frequencies the effects of internal tube capacities and the distributed capacity of the coil become large. This may best be illustrated by means of the table given below:

TABLE I

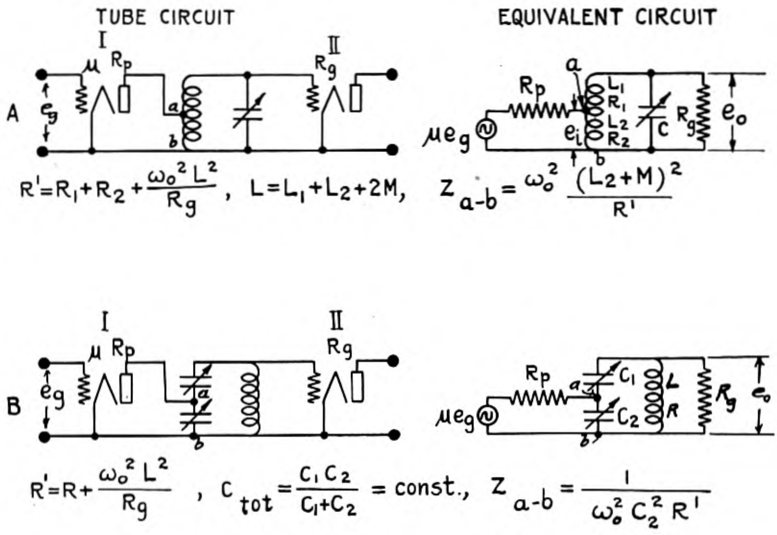
Coil No.	Inductance L	Natural Tuning Frequency f	Reactance at Half Natural Tuning Frequency $\pi f L$
1	.0025 henries	10^6 cycles	8,000 ohms
2	.25 "	10^5 "	80,000 "
3	25 "	10^4 "	800,000 "

The tube capacity plus the distributed capacity of each of the three coils for which these data are given is assumed to be 10 $\mu\mu f$. Since transformers in order to give a flat band must work below their natural frequency a much higher impedance than given by $\pi\, f\, L$ in the Table can therefore not be obtained. It is thus seen that only at audio frequencies is it possible to build a transformer with an impedance which is high compared with the tube resistances, the plate resistance being of the order of 6,000–50,000 ohms for ordinary receiving tubes and the grid resistance R_g being as high as 4×10^6 ohms but often limited to 500,000 ohms by an added resistance.

At higher frequencies sufficiently high impedances can only be obtained by working at the natural frequency of the transformer, and to illustrate this we shall in the following give some results of experiments made with ordinary tuned circuit amplifiers, choke coil amplifiers and loosely coupled transformer amplifiers at high frequencies.

Tuned Circuit and Choke Coil Amplifiers

In Fig. 2 there are shown to the left two different ways of connecting up a tuned circuit amplifier, and to the right are given the corresponding equivalent circuits. The input impedance to the next tube is assumed to be a pure resistance R_g, thus neglecting the grid-



$$R' = R_1 + R_2 + \frac{\omega_o^2 L^2}{R_g}, \quad L = L_1 + L_2 + 2M, \quad Z_{a-b} = \frac{\omega_o^2 (L_2 + M)^2}{R'}$$

$$R' = R + \frac{\omega_o^2 L^2}{R_g}, \quad C_{tot} = \frac{C_1 C_2}{C_1 + C_2} = \text{const.}, \quad Z_{a-b} = \frac{1}{\omega_o^2 C_2^2 R'}$$

$$\omega_o = 2\pi f_o, \quad f_o = \text{resonant frequency}$$

Fig. 2—Schematic of Tuned Amplifier Circuits

filament capacity and the grid-plate capacity of this tube. The effect of the grid-filament capacity, however, will only be to detune the circuit a little and can, therefore, be compensated for by retuning the condenser C (or C_1 and C_2) and the effects of the coupling through the grid-plate capacity of the second tube will be treated specially later.

Fig. 2 gives the well-known formulas for the equivalent series resistance R' of the circuit at resonance and for the impedance of the circuit Z_{a-b} measured between points a and b at resonance.

From this we then get in Case (A)

$$e_i = \mu e_g \frac{Z_{a-b}}{Z_{a-b}+R_p} = \mu e_g \frac{\omega_o^2(L_2+M)^2}{\omega_o^2(L_2+M)^2+R_pR'}, \tag{3}$$

and, assuming that $R_p >> \omega_o (L_2+M)$,

$$e_o = e_i \frac{L}{L_2+M}.$$

Hence, defining the voltage amplification K of the first stage as the voltage impressed upon the grid of tube II divided by the voltage impressed upon the grid of tube I, we have

$$K = \frac{e_o}{e_g} = \mu \frac{\omega_o^2 L(L_2+M)}{\omega_o^2(L_2+M)^2+R_pR'}. \tag{4}$$

In order to find the step-up ratio, which gives maximum amplification we have to solve for L_2+M in the equation $\delta K/\delta(L_2+M)=0$, which gives

$$R_p = \frac{\omega_o^2(L_2+M)^2}{R'} = Z_{a-b}, \tag{5}$$

and by inserting this in equation (4) we get

$$K_{max} = \frac{\mu}{2} \frac{1}{\sqrt{R_p}} \frac{\omega_o L}{\sqrt{R'}}. \tag{6}$$

From equation (5) it is seen that the condition for maximum voltage amplification is exactly the same as the well-known condition for maximum power amplification; namely, that the external impedance Z_{a-b} inserted in the plate circuit must be equal to the internal tube impedance R_p.

By repeating the calculations given above for Case (B) in Fig. 2, it will be found that this condition again holds good, and also it will be found that the expression for K_{max} is the same.

As already mentioned the resistance R' in the formulas above includes the equivalent series resistance introduced in the tuned circuit by the impedance of the input circuit of tube II, but in many cases this extra resistance will be negligible as compared to the resistance of the coil itself, and equation (6) thus gives us the very interesting information that the maximum amplification obtainable with a tuned circuit amplifier is proportional to the *ratio of the inductive reactance to the square root of the resistance.* In the case of an ordinary selective circuit such as a tuned loop antenna the output voltage developed is proportional to the ratio of the inductive reactance to the *first power of the resistance.* This does not mean that low resistance is less desirable in amplifier coils than in ordinary tuned circuits but it does mean that the penalty exacted by increasing the resistance is not as great.

In order to test the formulas given by equations (5) and (6) a series of experiments have been carried out.

For measurements of the maximum amplification of a tuned circuit amplifier a circuit as shown in Fig. 3 was used.[1] The grid of the am-

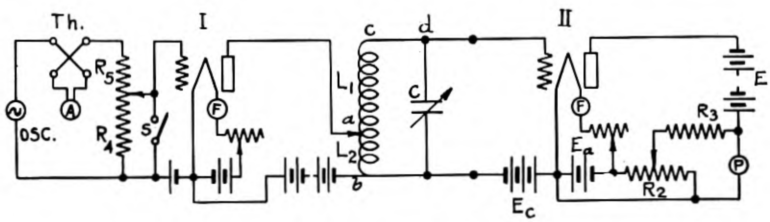


Fig. 3—Method of Measurement of Tuned Amplifier

plifier tube I is connected up to a known resistance, through which is passed a known current, and the voltage across the tuned circuit is measured by means of the tube-voltmeter II. The inductance L_1L_2 is made up of a single layer solenoid closely wound with 173 turns of solid wire and its value was 1.63×10^{-3} henries.

Keeping the frequency and the input from the oscillator constant the circuit is tuned to resonance by means of the variable condenser C and the lead from the plate to the coil is then moved along the coil until a point is reached which gives maximum deflection of the tube-voltmeter. During this process it is necessary to retune the circuit for each new point tried. Having thus obtained the right step-up for a certain frequency we then measure the amplification for different frequencies and get the amplification curves shown in the upper half of Fig. 4. On the lower half of Fig. 4 are given the number of turns (L_2) across the plate of the amplifier tube and also the capacity of the condenser C for each of the four cases shown.

[1] For a more detailed description of the method of measurement, see section entitled "Measurements" below.

In order to calculate the maximum amplification from formula (6) it is necessary to know the resistance R' of the circuit, the voltage amplification factor μ, and the internal plate impedance R_p of the amplifier tube. R' was obtained by running resonance curves for the circuit with the tubes connected up as usual, but with no filament current in the amplifier tube, in which case R_p may be regarded as being infinite.

These resonance curves are shown in the lower part of Fig. 4, and the resistance is then calculated from the well-known formula

$$R' = 2\pi(f_1 - f_2)L, \tag{7}$$

in which f_1 and f_2 are the frequencies, for which $E = E_{max}/\sqrt{2}$.

The resistance R' may also be obtained from the amplification curves as these can be regarded as resonance curves for the tuned circuit with the resistance R_p across part of the coil, and since this

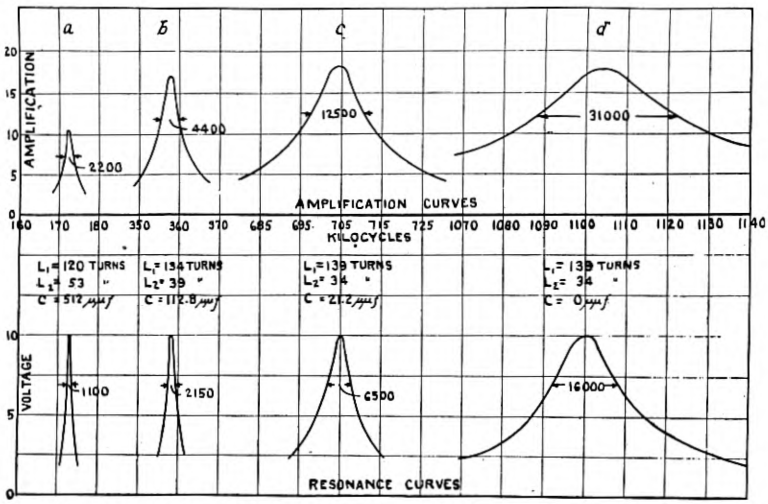


Fig. 4—Experimental Amplification and Resonance Curves of Tuned Circuit Amplifier

part of the coil is chosen so as to give $Z_{a-b} = R_p$, the equivalent series resistance should have increased to exactly twice the value found in formula (7). By comparing the widths of the amplification curves in Fig. 4 with the widths of the corresponding resonance curves it is seen that this actually was the case.

The internal plate impedance R_p and the amplification factor μ were obtained from the slope of the static characteristic of the amplifier tube used (a Western Electric 215-A or "peanut" tube).

The results of the calculations are given in Table II and the calculated values of K_{max} are seen to agree very well with the measured values given in the last column.

TABLE II

$L=1.63\times10^{-3}$ henries, $R_p=22{,}000$ ohms, $\mu=6{,}1$

Frequency	f_1-f_2	$R'=2\pi L(f_1-f_2)$	Calculated Maximum Amplification $K_{max}=\frac{\mu}{2}\frac{\omega L}{\sqrt{R_p\cdot R'}}$	Measured Maximum Amplification
172,000	1,110	11.4	10.7	10.3
357,200	2,150	22.1	15.9	16.6
704,000	6,500	66.6	18.1	18.1
1,100,000	16,000	164	18.1	17.8

The amplification of the amplifier was also measured with no step-up, i.e., with the plate of the amplifier tube connected across the whole

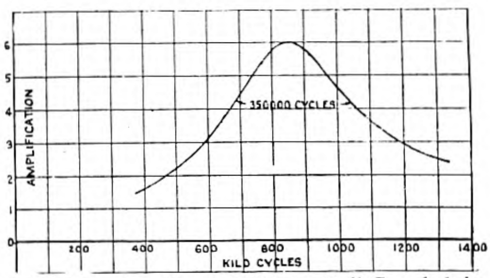


Fig. 5—Amplification Curve of Choke Coil Coupled Amplifier

coil and the tuning condenser C omitted (choke coil amplifier) the amplification curve shown in Fig. 5 being obtained.

From a resonance curve the following value is obtained:

$$R'=2\pi L(f_1-f_2)=2\pi\times1.63\times10^{-3}\times15{,}000=153 \text{ ohms},$$

and, therefore,

$$R_{tot}=R'+\frac{\omega^2L^2}{R_p}=153+\frac{4\pi^2 853{,}000^2\times163^2\times10^{-6}}{22{,}000}$$

$$=153+3320=3523 \text{ ohms},$$

which inserted in formula (7) gives

$$f_1-f_2=\frac{R_{tot}}{2\pi L}=\frac{3523}{2\pi\times1.63\times10^{-3}}=345{,}000 \text{ cycles}$$

while the amplification curve gives $f_1-f_2=350{,}000$ cycles.

As a final check of formula (6) by means of this tuned circuit, the maximum amplification was measured at 170,000 cycles with different values of extra resistance, R_{ext}, inserted in the circuit between c and d in Fig. 3. The results of these measurements agree very well with the formula as will be seen from Table III. For $R_{ext}=160$ it was found necessary to connect the plate across the entire coil in order to get maximum amplification and thus a further increase of R_{ext} beyond 160 ohms will make it impossible to obtain maximum amplification with this circuit.

TABLE III

$f=170{,}000$ cycles, $L=1.63\times10^{-3}$ henries, $R_p=22{,}000$ ohms, $\mu=6.1$.

R_{ext}	Total R'	Calculated Maximum Amplification $K_{max}=\frac{\mu}{2}\frac{\omega L}{\sqrt{R_p \cdot R'}}$	Measured Maximum Amplification
0	11.6	10.2	10.7
10	21.6	7.8	7.7
20	31.6	6.4	6.4
40	51.6	5.2	5
80	91.6	3.95	3.7
160	171.6	2.85	2.7

The variation with frequency of the resistance of the coil is shown on Fig. 6. These resistance values are obtained from the resonance curves in Fig. 4, and hence indicate also the losses in the variable condenser and the loss due to the input impedance R_g.

The curve in Fig. 6 gives what may be called the "true" resistance of the circuit, which is to be distinguished from the "apparent" resistance of the circuit as measured for instance by the well-known resistance variation method. By this latter method, the resistance of the coil is assumed to be equal to such an amount of extra resistance, as inserted in the circuit will decrease the resonance current to half its former value, but this assumption is only true when the distributed capacity of the coil is negligible as compared with the capacity of the variable condenser C, or when the resistance is introduced in the center of the coil.

It follows from formula (6) that for a given coil the maximum amplification is proportional to $\frac{\omega L}{\sqrt{R'}}$, and the measurements mentioned above seem to indicate that the maximum of this ratio has already been passed in the last case (d, Fig. 4) when the coil is used simply as a choke coil or auto-transformer (without any extra condenser). This, however, will depend upon the kind of wire used in making the coil. The coil used in the measurements above was made of No. 28

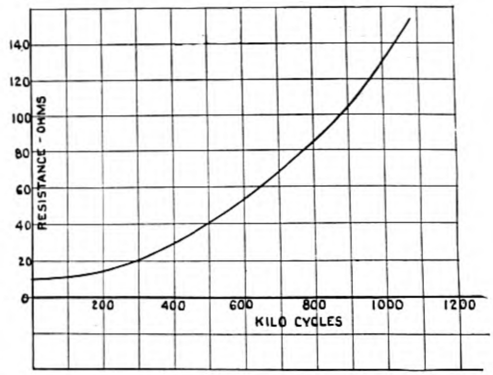


Fig. 6—Effective Resistance of Choke Coil

solid wire, but earlier results obtained by other investigators have shown that solid wire is superior to stranded wire at high frequencies, and thus it may be expected that the maximum of the ratio $\frac{\omega L}{\sqrt{R'}}$ for a given inductance will occur at a lower frequency when the coil is made of stranded wire.

For constant frequency the maximum amplification is proportioned to the ratio $\frac{L}{\sqrt{R'}}$ as already mentioned. It is thus desirable to adopt a construction for the coil, which will increase L without increasing $\sqrt{R'}$ proportionally. The highest amplification will in general be obtained when L is as large as possible for the frequency in question; in other words, it will be possible to obtain a higher amplification when the tuning condenser in the tuned circuit amplifier is reduced to zero, giving a simple choke coil amplifier.

For a tuned circuit amplifier with an ordinary good inductance coil made of stranded wire and of an inductance of, for instance, 200 microhenries and a high-frequency resistance of about 5 ohms, the amplification at 800 kilocycles will not be higher than about 9 times, according to formula (6) (using the same kind of tubes as in the experiments above), while with a choke coil an amplification as much as 18 times was obtained. This means that in order to get high amplification, small coils made of fine, solid wire and with large inductance and small distributed capacity should be used, rather than large coils made of stranded wire and with smaller inductance, but with larger distributed capacity.

In practice, it is not important to go to extremes in order to reduce the distributed capacity by one or two $\mu\mu f$. because the coil will always be shunted by the tube capacities, which are of the order of 10 $\mu\mu f$. It may be mentioned that the distributed capacity of the coil used in the above experiment is 3.5 $\mu\mu f$. This means that the constructional details of such a coil are not very important, and the coil may be made as a single layer coil or as a coil wound in one or several sections of rectangular or square cross-sections, but in all cases it will be found that coils of the same inductance will have very closely the same resonance frequency provided that the same tubes and leads are used in all cases.

Some experiments made with a choke coil (or auto transformer) at about 50,000 cycles show that the formulas given above may be also used here.

The coil used in these experiments was wound on a core of iron dust and made with square cross-section. The total inductance of the coil was .33 henries and provisions were made so that the plate of the amplifier tube could be tapped across any part of the coil.

The circuit diagram was the same as that given in Fig. 3 with the exception that the condenser C was omitted. The maximum amplification curve for this coil, used as a choke coil, is given by Fig. 7, curve A. The step-up ratio necessary to obtain maximum amplification was 1:16; i.e., the plate was connected across 1/16 of the total number of turns.

The resistance of the coil is obtained from a resonance curve as before:

$$R' = 2\pi L(f_1 - f_2) = 2\pi \times .33 \times 1300 = 2700 \text{ ohms},$$

and inserting this in formula (6) gives:

$$K_{max} = \frac{6.1}{2} \; \frac{2\pi \times 54800 \times .33}{\sqrt{22{,}000 \times 2700}} = 45$$

while the experiment gave 44.5.

On Fig. 7 are also given the amplification curves B, C and D for a step-up of 1:4, 1:1 and 1:48, respectively.

In the two cases B and C, the selectivity of the circuit is determined almost entirely by R_p, the resistance of the circuit itself being negligible, while in case D the selectivity is practically determined by the resistance of the coil itself.

It is seen that the amplification curve *C* for a step-up ratio of 1:1 is extremely flat as compared with the amplification curve shown in Fig. 5 for a choke coil working at 850 kilocycles.

In connection with these experiments with tuned circuits and choke coils it may be mentioned that in order to separate the *DC* plate voltage from the *DC* grid voltage, it will often be found of advantage to replace the coil by a transformer with very close coupling. In all our experiments, we have found that the amplification curves obtained in the two cases are identical when the coupling coefficient for the two windings of the transformer is nearly unity.

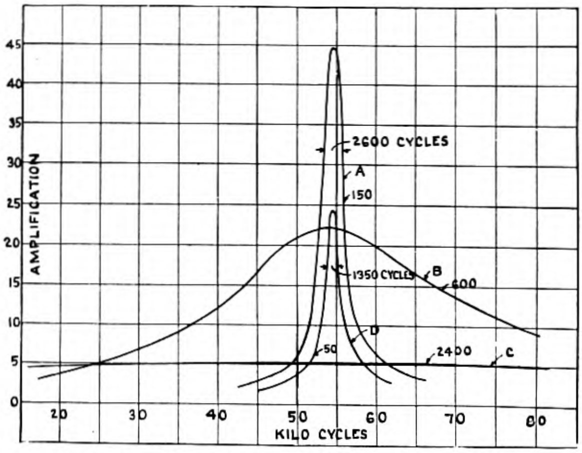


Fig. 7—Curve Showing the Effect of Ratio of Transformation in the Characteristic of a Choke Coil Coupled Amplifier

Loosely Coupled Transformer Amplifier

From the amplification curves obtained with choke coils, it will be seen that the frequency range obtainable with a choke coil amplifier is not as wide as might be desirable in some cases. This is especially true for higher frequencies between 300,000 and 1,000,000 cycles, and where a wide frequency band is desired these choke coils have, therefore, been replaced by transformers with a rather loose coupling, in which case the transformers will have the characteristics of two ordinary coupled circuits and give an amplification curve with two peaks.

It has been found by experiment that such transformers can actually be treated just as ordinary coupled circuits and the amplification curves can be computed by means of the well-known formulas for current and voltage conditions in two coupled circuits.

Before going into the details of these experiments, it is worth while to consider briefly the general relations involved as indicated by the curves obtained with two coupled circuits, each tuned to 52,000 cycles. These curves are shown in Fig. 8. The coils used

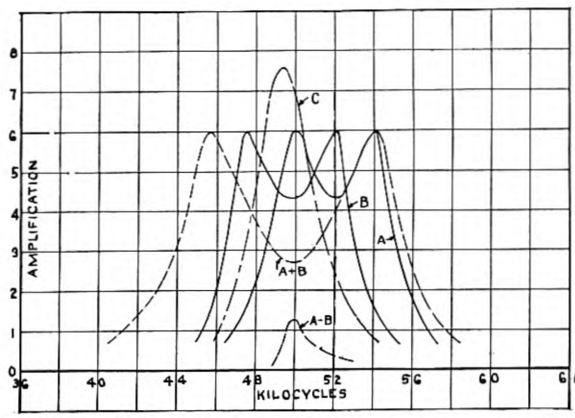


Fig. 8—Curves Showing the Effect of Coupling Inductive and Capacitive, on Amplification Characteristic of Coupled Tuned Circuits

had an inductance of 10 millihenries and were tuned by condensers. The circuit of the apparatus employed in obtaining the curves is given in Fig. 9.

Curve A gives the amplification for inductive coupling alone.

Curve B is for capacitive coupling alone.

Curve $A+B$ is for both capacitive and inductive coupling aiding each other, each coupling having the same value respectively as in curves A and B.

Curve $A-B$ is with the two couplings opposing each other and

Curve C is the same as $A-B$ but with different value of the inductive coupling.

The curves have the same shape as the well-known resonance curves for two coupled circuits with the oscillator input in series with the primary circuit, where the peak frequencies are given by the following approximate formulas:

Inductive coupling: $f' = \dfrac{f_o}{\sqrt{1-k}}, \ f'' = \dfrac{f_o}{\sqrt{1+k}},$

Capacitive coupling: $f' = f_o,\ f'' = f_o \sqrt{\dfrac{C}{C+2C'}}$,

where $f_o = \dfrac{1}{\sqrt{LC}}$, $k = \dfrac{M}{L}$ = coefficient of coupling.

Having thus demonstrated the general shape of the amplification curves for a two coupled tuned circuit amplifier, the action of a loosely coupled transformer amplifier for high frequencies will be treated.

The transformer used in this experiment was made up of two similar pancake coils, 2″ diameter, wound with 210 turns of solid wire. Fig. 9 shows the circuit diagram. Curves A and B in Fig. 10a

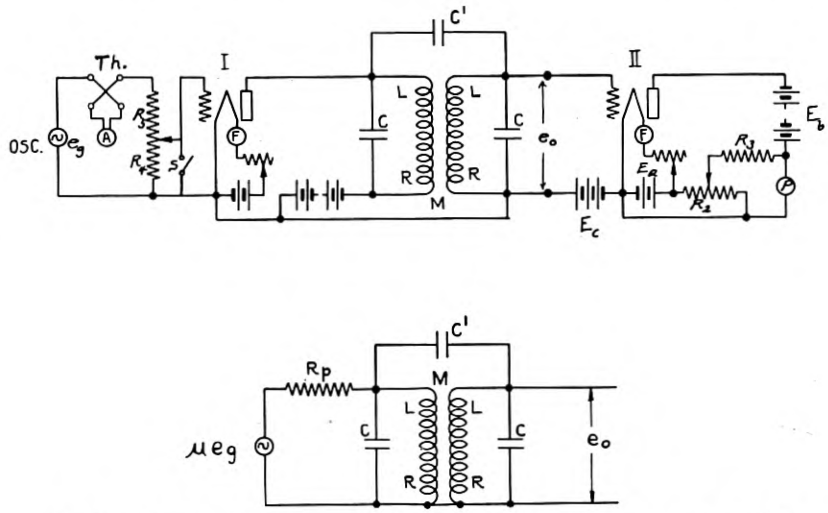


Fig. 9—Method of Measurement of Loosely Coupled Transformer Amplifier

show the measured amplification curves for a 3/8″ distance between the windings. The coupling condensers C' were omitted but even then there was some capacity coupling left due to the distributed capacity between the coils. The curves A and B correspond respectively, to an aiding and an opposing action of this capacitive coupling. Interchanging the leads to either coil changes the amplification curve from one type to the other.

The self inductance and mutual inductance of the coils were measured at low frequency and found to be:

$$L = 2.1 \times 10^{-3} \text{ henries}, \quad M = .95 \times 10^{-3} \text{ henries}.$$

The resistance of the coils was measured as described before by taking resonance curves at different frequencies. The distributed capacity of each coil was 14×10^{-12} farad (including tube capacity). By means of these values, the curve C was calculated.[2] The unknown capacity coupling makes it impossible to predict the exact shape of a transformer coupled amplifier from the constants of the circuits. However, the calculated curve C (calculated for inductive coupling only) will give a general idea of the shape of an experimental curve A.

Curves A and B, Fig. 10b, show the amplification curves for the case of capacity coupling alone. A is the experimental and B the calculated curve and they are seen to give fair agreement. The coupling capacity was 21×10^{-12} farad and the distributed capacity of the coils was 19.3×10^{-12} farad, the increase, as compared with the case of inductive coupling, being due to the ground capacities of the coupling condenser.

In connection with this type of amplifier it may be mentioned that a higher amplification naturally can be obtained if the plate of the amplifier tube is connected across a part of the primary circuit only, maximum amplification corresponding to the circuit impedance being equal to the plate impedance. However, the same effect will take place here as was shown for the tuned circuit amplifier, namely, that the band width will decrease with increase in amplification. Using transformers at their natural frequency instead of coupled tuned circuits with outside condensers will give broader bands or higher amplifications corresponding to the single tuned circuit amplifier.

[2] The following two formulas have been used for calculating the amplification of the circuit shown in Fig. 9.

Inductive Coupling:

$$\text{Amplification} = \frac{e_o}{e_g} = \frac{\omega M}{Z}\frac{1}{\omega C}\frac{1}{R_p(1-\omega^2 L^1 C)+R^1+j(\omega L^1+R_p R^1 \omega C)} \cdot \mu$$

$$\text{where } Z = \sqrt{R^2+\left(\omega L-\frac{1}{\omega C}\right)^2},\quad R^1 = R\left(1+\left(\frac{\omega M}{Z}\right)^2\right),\quad \omega L^1 = \omega L - \left(\frac{\omega M}{Z}\right)^2\left(\omega L - \frac{1}{\omega C}\right).$$

Capacitive Coupling:

$$\text{Amplification} = \mu\frac{R+j\omega L}{A+jB}$$

$$\text{where } A = R_p\left(2-2\omega^2 LC+2\frac{C}{C'}-\omega^2 LC\frac{C}{C'}-\frac{1}{R^2+\omega^2 L^2}\frac{\omega L}{\omega C'}\right)+R\left(1+\frac{C}{C'}\right),$$

$$B = RR_p\omega C\left(2+\frac{C}{C'}\right)-\frac{RR_p}{\omega C'(R^2+\omega^2 L^2)}+\omega L\left(1+\frac{C}{C'}\right)-\frac{1}{\omega C'}.$$

AMPLIFIERS WITH SEVERAL STAGES. "FEED BACK" ACTION

The experiments so far have shown, that with one stage of amplification and with the amplifier working into a detector tube without grid condenser and leak, it is always possible to calculate the amplification curve from the constants of the tubes and of the coils, regardless

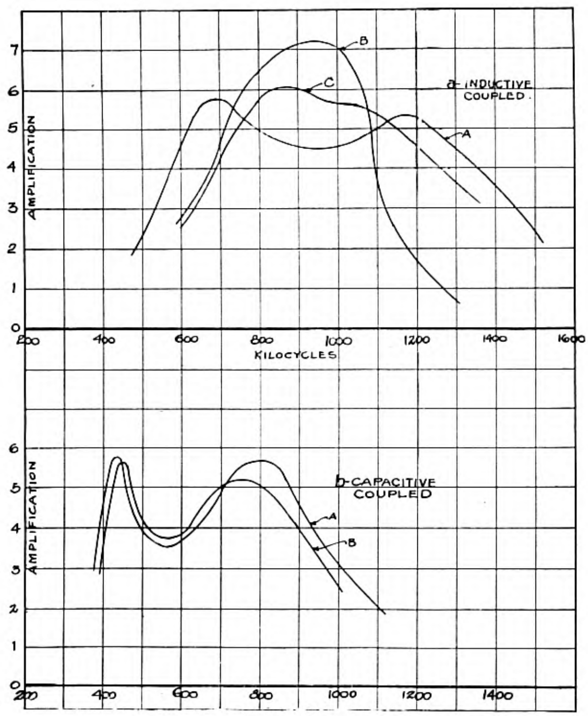


Fig. 10—Amplification Curves of a Loosely Coupled Transformer Amplifier Showing Effect of Coupling

of whether the connection between the amplifier and the detector consists of a simple tuned circuit, a choke coil, two coupled circuits or a loosely coupled transformer, the circuits being treated simply as ordinary tuned circuits.

Also the experiments have shown that a higher amplification can be obtained in the 50,000 cycles region than around 1,000,000 cycles, as might be expected from formula 6.

The next question is: What happens when more than one stage of amplification is used? If, for instance, the amplification for one stage is 10, will then the amplification for two stages be 100 or, in other words, in a multiple-stage amplifier is it possible to get the total amplification curve from the curve for the amplification per stage by multiplying them together?

The answer to this question is that the total amplification of a multi-stage amplifier will, in general, be lower than the value obtained by multiplying the amplification values per stage, and the reason for this is to be found in the input impedance of the tubes. So far, we have assumed the input impedance of the tube after the amplifier to be high as compared with the impedance of the tuned circuit (or transformer) and this is correct for a plate curvature detector, in which the impedance of the load in the plate circuit is negligible at the frequency of the amplified current but if the next tube is another amplifier it is only true at lower frequencies. It has been shown [3] that the input impedance of a vacuum tube can readily be calculated by means of the constants of the tube and the output impedance.

For the tubes used in the foregoing experiments we have the following approximate constants:

C_{g-p} = Grid to plate capacity $= 3\times 10^{-12}$ farad

C_g = Grid to filament capacity $= 5\times 10^{-12}$ farad

C_p = Plate filament capacity $= 5\times 10^{-12}$ farad

R_p = Plate impedance $= 20{,}000$ ohms

μ = Amplification constant $= 6$.

The output impedance including the plate-filament capacity will be assumed to be a resistance equal to the plate impedance.

If the input impedance is represented by an apparent resistance R'_g in parallel with an apparent capacity C'_g, we get for R'_g and C'_g the values given in Table IV.

TABLE IV

$$R'_g = \frac{\omega^2 R_p^2 C_{g-p}^2 + 4}{\omega^2 R_p C_{g-p}^2\,(\mu+2)} \qquad C'_g = C_g + 2C_{g-p}\,\frac{\mu+2}{\omega^2 R_p^2 C_{g-p}^2 + 4}$$

Frequency	R'_g	C'_g
10^3 cycles	7×10^{10} ohms	17×10^{-12} farad
10^4 "	7×10^{8} "	17×10^{-12} "
10^5 "	7×10^{6} "	17×10^{-12} "
10^6 "	7.5×10^{4} "	16×10^{-12} "

[3] H. W. Nichols, *Phys. Rev.*, Vol. 13, p. 405, 1919. John M. Miller, Bureau of Standards—Scien. Pap. No. 351, 1919.

From this table it is seen that the effect of the input impedance is negligible at frequencies up to about 100,000 cycles, but for frequencies in the broadcasting range, the input impedance will introduce an appreciable loss in the preceding circuit, which will result in a drop in amplification below the value obtained for a single stage amplifier. It is seen that the input impedance R'_g for broadcasting frequencies

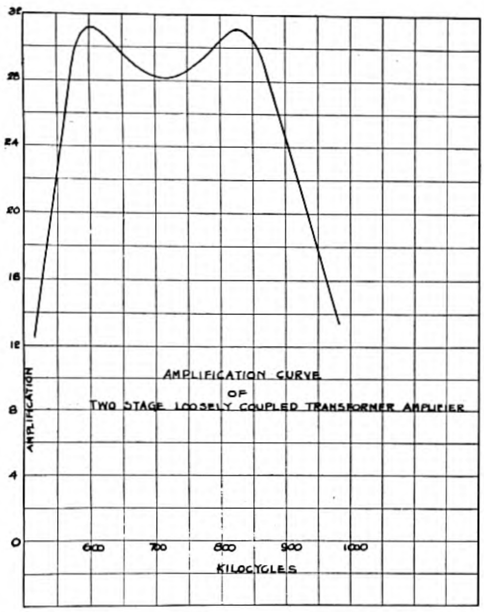


Fig. 11—Amplification Curve of Two Stage Loosely Coupled Transformer Amplifier

is of the same order of magnitude as the plate impedance R_p, which means that it will be of no advantage to use much step-up in choke coils or tuned circuits for an amplifier with more than one stage, since the amplification in no case will be much higher than μ per stage, except for the last stage, which is working into the detector.

The loosely coupled transformers of the type already discussed will, on the other hand, work very well in a two-stage amplifier, since there is no step-up used in these, and the amplification will be very nearly twice the amplification for a one-stage amplifier, as will be seen from the amplification curve shown in Fig. 11. The width of such an amplification curve can be increased by proper adjustment of the transformer inductances but the amplification will naturally drop correspondingly.

The values of R'_g and C'_g given in Table IV were calculated on the assumption of a pure resistance load R in the plate circuit. If the load in the plate circuit is an impedance $Z = R + j\,x$, it will be found that the sign of the apparent shunt resistance R'_g will depend upon the

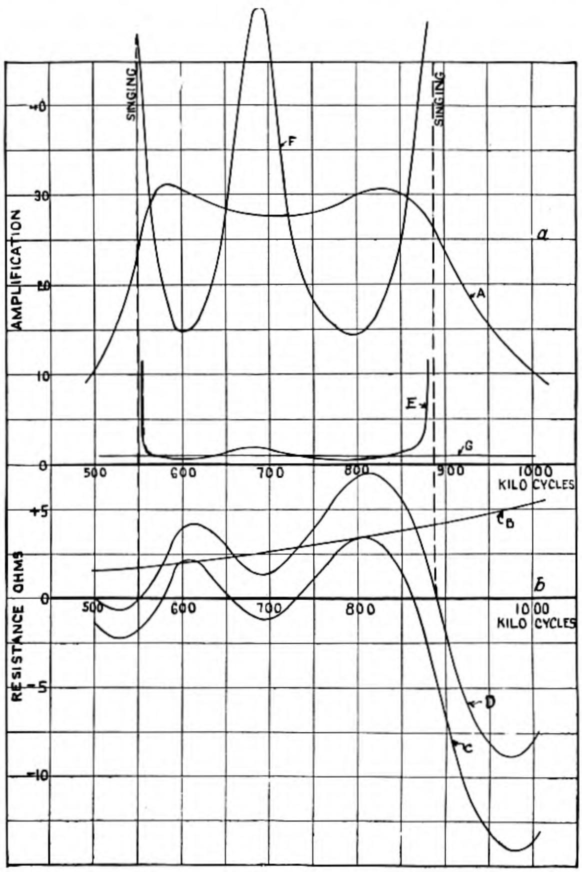


Fig. 12—Total Amplification of Transformer Coupled Receiver and Effect of "Feed Back" on Loop Resistance

sign of the reactance x. For a capacitive load, the resistance R'_g will always be positive, but for an inductive load, R'_g may in some cases become negative and we then have "feed back" or regeneration occurring through the tube. The negative resistance introduced in the circuit below the resonance frequency may in certain cases be so high that it more than neutralizes the positive resistance of this circuit which means that the set will start to oscillate or "sing."

As an illustration of the effect of this "feed back" action, there are given in Fig. 12 some curves obtained for a two-stage high frequency amplifier with loosely coupled transformer stages. The input circuit to the amplifier consisted of a loop antenna circuit tuned to the frequency of the induced signal.

Curve *A* shows the straight high frequency voltage amplification of the set, as measured with resistance input to the grid of the first high frequency amplifier. (Same as curve shown in Fig. 11.)

Curve *B* gives the actual resistance of the loop used with the set.

Curve *C* gives the resistance introduced in the loop due to "feed back" action from the first stage.

Curve *D* gives the resulting apparent resistance of the loop (Curve *B*+Curve *C*) and

Curve *E* shows the "feed back" amplification of the set. (Curve *B:* Curve *D*.)

Curve *F* shows the total amount of amplification obtained by the set which is the product of the ordinary voltage amplification (Curve

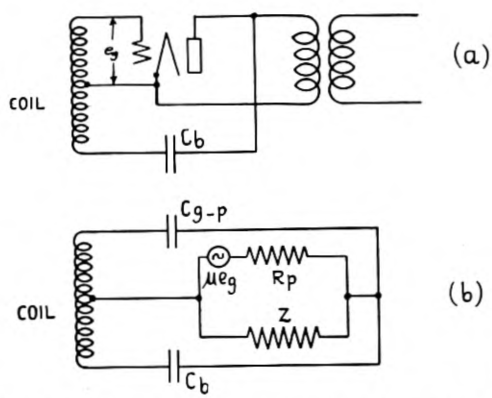


Fig. 13—Schematic of Balancing Condenser Action

A) and the "feed back" amplification (Curve *E*) and it is thus seen that the feed back action makes the total amplification vary irregularly in a very undesirable manner, and also makes the set "sing" at certain frequencies.

In order to avoid this, it is necessary to provide some means of balancing out the effect of the grid plate capacity of the tubes, and Fig. 13 (a) shows how this may be done.[4] The filament of the tube

[4] See Patent No. 1,183,875 issued to R. V. L. Hartley, and Patent No. 1,334,118 issued to C. W. Rice.

is connected to the middle of the coil, the grid to one end and the plate is connected through a small balancing condenser to the other end of the coil. In Fig. 13 (b) is given a schematic diagram of the circuit, which shows that the effect of C_b upon the coil circuit is just opposite the effect of C_{g-p}, so that the circuit can be regarded as an ordinary bridge circuit. It will, therefore, always be possible by proper adjustment of the condenser C_b to neutralize the effect of the feed-back action as shown by curve G in Fig. 12.

The same kind of an arrangement can be used between the different stages in a multi-stage high frequency amplifier, and it is thus seen that by proper use of such balancing condensers, it will be possible to obtain for a multi-stage amplifier a total amplification which is practically equal to the product of the amplifications per stage. This is true for a multi-stage tuned circuit coupled amplifier but for transformer coupled amplifiers, where it is more difficult to obtain a 180° phase difference of voltages, the advantage of the balancing condenser is not so great.

Of course, this favorable result presupposes that the wiring of the amplifier is properly done and the different stages shielded carefully from each other so that no external coupling exists between them.

Resumé

What has been said about amplifiers in the preceding sections can be summarized as follows:

With a given type of amplifier the same general shape of the amplification curve is obtained regardless of the frequency range at which the amplifier is designed to operate.

Thus, a low amplification over a wide frequency range will be obtained by using loosely coupled transformers or choke coils without any step-up, while a high amplification over a narrow range of frequencies can be obtained by using choke coils or tuned circuits with a proper step-up. In this last case, it will be necessary to use a small tuning condenser across the coils in order to make the frequency range of the amplifier wider, and the higher amount of amplification is, therefore, obtained only by a sacrifice of tuning facilities of the set. In a multi-stage amplifier it may, however, often be found of advantage to use a combination of low amplification stages and high amplification stages so that, for instance, one tuned circuit stage with high step-up and variable condenser is used in connection with one or several stages of choke coils or loosely coupled transformers with low amplification and a wide frequency range. The maximum amplification obtained with any kind of an amplifier will, in general, be higher at the lower frequencies, due to the lower loss and the higher ratio of L over C obtainable.

The width of the frequency band for a choke coil amplifier will be smaller, the higher the frequency due to the decrease in ωL with increasing frequency, and at broadcasting frequencies it will, therefore, in general be found advantageous to use loosely coupled transformers rather than choke coils, whenever a wide frequency band is desired. In addition to giving a wider frequency band, lower frequency amplifiers have the advantage of a smaller grid-plate feed back action.

Amplification Measurements at High Frequencies

In order to make a thorough study of radio frequency amplification, it is necessary to have a dependable method of measurement. Such a method developed in our laboratory and used very successfully will be described here.

In order to obtain an accurate comparison between different types of amplifiers, in which any type of resonant coupling is used, it is essential that these amplifiers be operated from a resistance input and not from an input containing a tuned circuit. With a tuned circuit it is not only very difficult to obtain an accurate measure of the voltage impressed upon the amplifier but considerable regeneration may occur between this input circuit and the output circuit of the first amplifier tube. There is, naturally, also a feed-back action in connection with a resistance input circuit, but its effect is negligible when the resistance is only a few hundred ohms. When the characteristic of a radio frequency amplifier with a resistance input has been accurately determined, its characteristic when used with a tuned circuit input may be determined as will be described later.

A schematic circuit diagram of the apparatus as used is shown in Fig. 3. To the left is shown the input apparatus which consists of an oscillator, a sensitive thermocouple and a potentiometer. The drop across the resistance R_4 of the potentiometer is used as the input to the amplifier stage I. The output of the amplifier stage is measured by the tube voltmeter II shown to the right in Fig. 3. The tube voltmeter II may be a low frequency detector in the case of amplification measurements of an actual receiver set.

It is necessary first to calibrate the tube voltmeter or detector II which is done by disconnecting it from the amplifier and connecting it directly across the potentiometer $R_4 - R_5$. R_4 is then adjusted to, say, 500 ohms and the current through it adjusted to some convenient value, such as 1 milliampere. This voltage of .5 volt will be sufficient with most tubes to give a change in the plate current of 30 to 40 microamperes.

The tube voltmeter is then reconnected to its normal place in the circuit and the resistance R_4 is connected to the input of the amplifier. Keeping the current constant at the value of 1 milliampere, the resistance R_4 is adjusted until the change in the detector plate current is the same as before. It is immediately apparent that the amplification will be the ratio of the known voltage on the grid of the detector, that is .5 volt, to the voltage on the input of the amplifier, as indicated by the product of the resistance R_4 and the current through it. The current having been kept constant, the amplification is the quotient of the 500 ohms used when calibrating the detector and the resistance value obtained with the amplifier included.

Considerable precaution must be observed to make sure that no energy is getting into the amplifier circuit except that which may be measured by the voltage drop across the resistance R_4. This necessitates the most careful shielding especially when the amplification is more than 50 times.

With the measuring apparatus described a dependable input voltage as small as 1 millivolt can be obtained. The maximum amplification which can be measured directly is, therefore, of the order of 500 times when the output voltage to the detector is of the order of one half of a volt.

For the measurement of higher amplification the following indirect method may be used.

The amplification is artificially decreased in some manner such as reducing the number of stages in the circuit and this reduced amplification is measured in the usual manner. The input current is then reduced and the input resistance increased keeping the plate current of the detector constant, the voltage impressed on its grid being determined by the previous calibration. The amplification is now increased to its normal value and the input resistance decreased until the detector plate current has its original value. The ratio of decrease in input resistance will thus give the increase in amplification and the total amplification will be the product of this and the smaller amplification as first measured.

The smaller current through the input resistance, which is obtained by this method and which will generally be less than can be determined by the most sensitive thermocouple, will reduce the pick-up to a sufficiently low value to give satisfactory results. In this connection it may be noted that an excellent test for the presence of undesirable pick-up is the closing of a switch (S) placed at the input of the amplifier. With this switch closed there should be no appreciable input to the detector.

Direct high frequency amplification measurements require input units made up of very carefully constructed attenuation boxes or potentiometers and well shielded oscillators. Such units have been developed in connection with field measurements and are described in a paper on "Radio Transmission Measurements," by Messrs. Bown, Englund and Friis and "Note on the Measurements of Radio Signals,"[5] by Englund.

On the right in Fig. 3 is shown, as mentioned before, the circuit diagram of a "tube voltmeter" such as is used in many high frequency measurements. The tube voltmeter is essentially a plate current curvature detector. The grid is made negative by means of the grid battery E_c, so that the normal plate current of the tube is very small (of the order of 50 microamperes or so), and this plate current is further balanced out by means of the potentiometer arrangement R_2, R_3, so that the plate current meter reads zero when the input to the tube voltmeter is short-circuited. This arrangement has the advantage of making it possible to utilize the entire scale of the meter and to obtain the measured voltage from a single reading instead of the difference of two readings. Such a tube voltmeter built with an "N" tube will give a deflection of 1 microampere for an input voltage of about 1/5 of a volt, and the calibration will stay remarkably constant for several months and is *independent* of the *frequency* at which it is calibrated. The values of the resistances in the resistance boxes used at high frequencies may, therefore, be checked by using the boxes for calibrating a tube voltmeter first at 60 cycles and afterwards at, for instance, 1,200 kilocycles. If the two calibration curves obtained are exactly identical, then the resistance has not changed appreciably within this frequency range.

In measuring the amount of "feed-back" amplification in a receiving set, it is not possible to use a method as direct as described above. The "feed-back" or regeneration in a set is, as already mentioned, due to the coupling between the grid circuit and the plate circuit of the tubes through the grid-plate capacity, and will depend upon both the load in the plate circuit and the nature of the input circuits. If, for instance, it is desirable to measure the amount of "feed-back" amplification due to the coupling between the loop circuit and the

[5] Proc. Inst. R. E., Vol. 11, No. 1, February, 1923. Proc. Inst. R. E., Vol. 11, No. 2, April, 1923.

plate circuit of the first amplifier in a high frequency amplifier set, it will not be possible to measure this with a resistance input to the amplifier since in this case the "feed-back" has no appreciable effect. In order to get the correct value for the "feed-back" amplification, the set must be connected up to the same loop with which it is going to be used and the measurements can then be made in the following way.

A resistance box is inserted in the middle of the loop and a tube voltmeter is connected across half of the loop in addition to the receiving set as shown in Fig. 14. With the filament circuit of the set open, a strong high frequency emf. is induced in the loop and the loop circuit is tuned until the tube voltmeter reads a maximum.

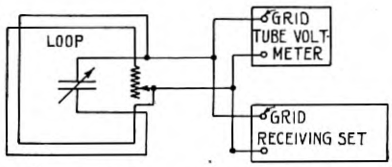


Fig. 14—Method of Measuring "Feed Back" Amplification

A "feed-back" action in the set will then produce a change in the tube voltmeter reading when the filament current is switched on. If the "feed-back" action is positive, i.e., if the resistance introduced in the loop is negative, then the tube voltmeter reading will increase, and in order to bring it back to its former value, the resistance of the loop is increased by an amount R' by means of the resistance box. If, on the other hand, the "feed-back" action is negative, the resistance of the loop must be decreased in order to obtain the former value of the tube voltmeter reading.

R' represents the equivalent series resistance introduced in the loop circuit by the "feed-back" action, and the apparent resistance of the loop is, therefore, $R-R'$, where R is the actual resistance of the loop. The voltage impressed upon the grid of the first tube is inversely proportional to the apparent resistance of the loop, and the amount of "feed-back" amplification is, therefore, defined as the ratio $K'=R/(R-R')$ where R' must be taken with the proper sign. This ratio is seen to be a direct measure of the increase (or decrease) in input voltage due to the "feed-back" action in the set and the total amount of amplification in a set at a certain frequency will then be given by the product of the ordinary voltage amplification factor K and the "feed-back" amplification factor K'.

In determining K', it is necessary to know the actual resistance R of the loop and this may be conveniently obtained by the reactance variation method using a tube voltmeter across half of the loop as the voltage indicating device. It has been found that the loss introduced by such a tube voltmeter is negligible, a fact which can be easily checked by connecting two similar tube voltmeters across the loop and determining the maximum reading of one of them. When the other one is then disconnected and the loop condenser slightly re-adjusted so as to again give maximum reading of the first tube voltmeter, it will be found, that the two readings obtained are exactly the same.

The discussion of the two types of amplification measurements of high frequency amplifiers may be summarized as follows:

The *ordinary voltage amplification* K is defined as the ratio of the amplified signal voltage impressed on the grid of the low frequency detector and the signal voltage impressed on the grid of the first amplifier tube. This amplification is measured by using a resistance input to the amplifier and includes the effect of "feed-back" action between the stages in the amplifier. This "feed-back" action between stages can naturally be analyzed by a method similar to the one used to determine the "feed-back" action between the amplifier and its tuned input circuit.

The *"feed-back" amplification* K' is defined as the increase (or decrease) of signal voltage due the "feed-back" action between amplifier and its tuned input circuit. The "feed-back" amplification depends upon the selectivity of the input circuit and will only vary slightly from unity when the resistance of this circuit is very large, while large variations, as shown in Fig. 12, may be found when a selective input circuit is used.

The total amplification is defined as the product of the ordinary amplification K and the "feed-back" amplification K'.

Design Characteristics of Electromagnets for Telephone Relays

By D. D. MILLER

NOTE: The electromagnets described are confined to relays, although the principles involved apply as well to selector magnets, clutch magnets and electromagnets in general. A treatment from the viewpoint of the telephone engineer is given of the important considerations which determine the design of the magnetic parts of relays and the economics of the winding dimensions. A knowledge of these factors as well as of the general considerations which are discussed is of great importance in the selection and application of relays to the telephone system. The operating and economic importance to the Bell System of the great number of relays required in the operation of the plant has been described in a previous paper.[1]

INTRODUCTION

ELECTROMAGNETS or relays as generally used in telephone switchboards are simply switches which are controlled electromagnetically. These switches may be required to open or close a number of separate and distinct circuits simultaneously or in a certain sequence. In many cases it is essential that the relay switch be opened or closed very quickly as this time may have a direct influence on the amount of apparatus required and consequently the first cost of the plant. The operating time of the relays also has a direct influence on the time required to establish a telephone connection. The above statements are particularly evident in automatic systems where selector apparatus is required to establish a connection between parties but is released during the conversation. It follows that the number of selector circuits and relays therein depends upon the amount of traffic and time required for the selectors to establish the connection.

To establish a telephone connection between two parties in certain automatic telephone systems, requires the opening and closing of about 2,000 electric switches of which 1,200 are operated by simpler types of electromagnetic relays. In a typical manually operated system a call is completed by the opening and closing of about 112 switches of which 70 are operated by relays. It is therefore evident that the relay switches must operate both quickly and reliably and maintain a high degree of stability throughout a long period of service.

In controlling the various circuits in telephone systems by relays, the character of the circuits determines the construction of the relay switches. If large currents are to be controlled the relay switch

[1] Relays in the Bell System, S. P. Shackleton and H. W. Purcell, *Bell System Tech. Journ.*, Vol. 3, p. 1, 1924.

construction differs materially in ruggedness from the construction where relatively small currents are to be controlled. In the operation of the relays larger amounts of power, of course, are required for those having the more rugged construction. It is also evident that more power is required for fast operation than for comparatively slow operation. Fast operation of relays is also dependent upon

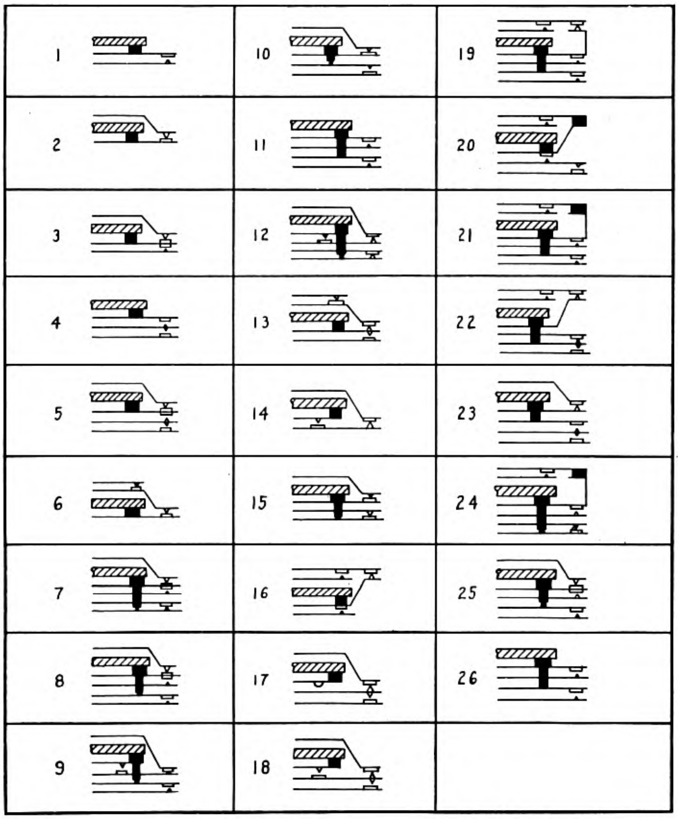


Fig. 1—Spring Combinations—Flat Type Relay

circuit arrangements which are effective in lowering the electrical "time constant" of the circuits in which the relays operate.

Electromagnets in telephone systems are designed and used for a great variety of conditions. The more common uses are for relay operation on direct current battery of 20 to 28 volts or 40 to 45 volts. Such relays perform a great number of switching functions, a few of which are shown in Figs. 1 and 12. Other designs are used for operation on alternating currents, ranging from 16 cycles ringing frequency to voice frequencies of 2,000 cycles per second. The load or work required of these relays and electromagnets varies from a fraction of a gram controlled through a few thousandths of an inch to 25 pounds controlled through a distance of $\frac{1}{4}$ of an inch. Some relays are operated where the annual power charges are negligible while in other designs annual power charges may be controlling. The technical considerations which determine the design features, therefore vary throughout a wide range as to the proportioning of the magnetic parts and the design of the windings. Other general design characteristics that must be carefully considered are as follows:

1. Operating capability of the structure—
   (a) Switching conditions or circuit control required of the relay.
   (b) Design of contacts required to safely carry the energy required by condition (a) throughout the estimated "life" requirements of the switchboard.
   (c) Capability of the structure with respect to the input power to satisfy condition (a).
2. Determination of winding best suited for the circuit.
3. Temperature limitation of the winding under extreme conditions.
4. Ease of adjustment.
5. Permanence of adjustment—
   (a) For a period of service operations representing the "life" of the relay in the switchboard.
   (b) Under extreme weather conditions.
6. Size and mounting facilities.
   (a) When used for additions to old equipment where it should mount in the same space as the apparatus it replaces.
   (b) Economy of space for new equipments.
   (c) Stability of mounting.

7. Terminals—arrangement and distribution for most advantageous electrical connections.

8. Insulating materials.
   (a) Windings.
   (b) Switch control of contacts.
9. Cover design.
   (a) Protection from dust.
   (b) Effect of cover on operation and protection from stray flux.
10. Speed of operation and release.
11. Transmission efficiency with respect to voice frequencies.

12. Mechanical design features with special reference to manufacture.

13. Electro-mechanical efficiency.

14. First cost and annual charges.

As it is not within the scope of the present paper to discuss in detail all of the above characteristics the following have been selected as perhaps the more important and the most interesting:

1. The design of the magnetic parts for various telephone switchboard requirements.

2. Methods of calculating windings and the determination of temperature characteristics.

3. Considerations which determine the spool dimensions.

4. Discussion of designs used extensively in the telephone plant.

Design of Magnetic Parts

The fundamental requirements of an electromagnet or relay are generally the load or pull, the distance through which the load must be moved and the time limits of operation. The last requirement, of course, is reflected in the load or pull requirement as an added pull or force of acceleration.

The fundamental constants of design are the flux leakage coefficient, the core flux density and the flux density in the pole face or area where the pull is exerted. If the designer is given data which fix these constants the remainder of the work is usually a comparatively simple matter of calculation.

The leakage coefficient has been determined experimentally throughout a range of designs where the load to be controlled varied from 1 gram to 5,000 grams. The results show that the leakage depends almost entirely upon the armature air-gap reluctance and the ratio of the core length to the core diameter. The leakage flux is defined as that percentage of the total core flux which does not cross the armature air-gap, and consequently can not be utilized for producing traction. The per cent useful flux is then the ratio of the flux crossing the armature air-gap to the total flux in the relay core. The curves in Fig. 2 for single spool electromagnets and Fig. 3 for double spool electromagnets give the per cent useful flux for various air-gaps and core lengths which are expressed in terms of the core diameter. In cases where the core is round and the pole face area is equal to the core section these data may be used directly. If, however, the pole face area differs from the core section, the air-gap used in looking up the leakage in Fig. 2 and Fig. 3 should be reduced to a value which,

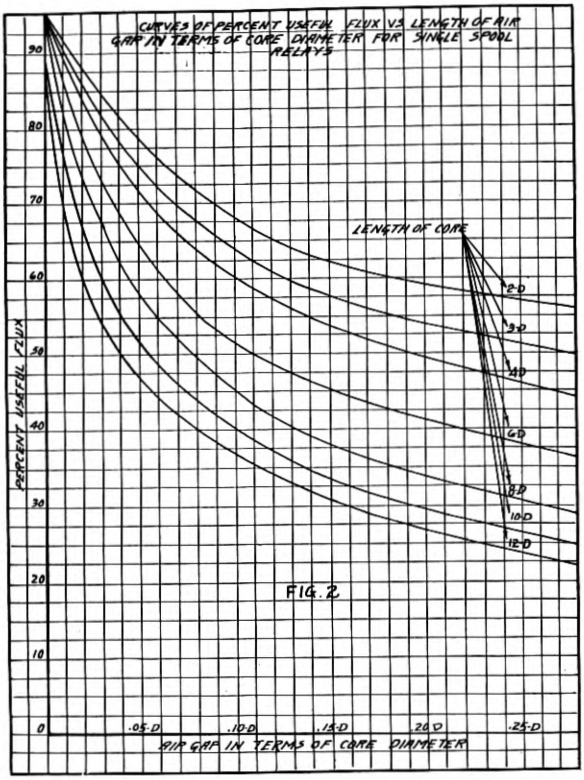


Fig. 2—Curves of Percentage Useful Flux vs. Length of Air Gap in Terms of Core Diameter for Single Spool Relays

with a pole face area equal to the core section, would give the same air-gap reluctance.

The core flux density and the pole face density depend largely upon the requirements of the particular design, but the considerations outlined in the next four paragraphs are of prime importance.

In some cases the annual power charges are relatively unimportant, there being plenty of power available during the short intervals of time required for operation. Obviously in this case efficiency of operation can be sacrificed, and consequently power, in order to obtain a low first cost. Referring to Maxwell's formula for traction or pull

$$P = \frac{B^2 S}{8\pi 980},$$

the pull P is proportional to the square of the armature air-gap flux density B, consequently the total flux required will be less the greater

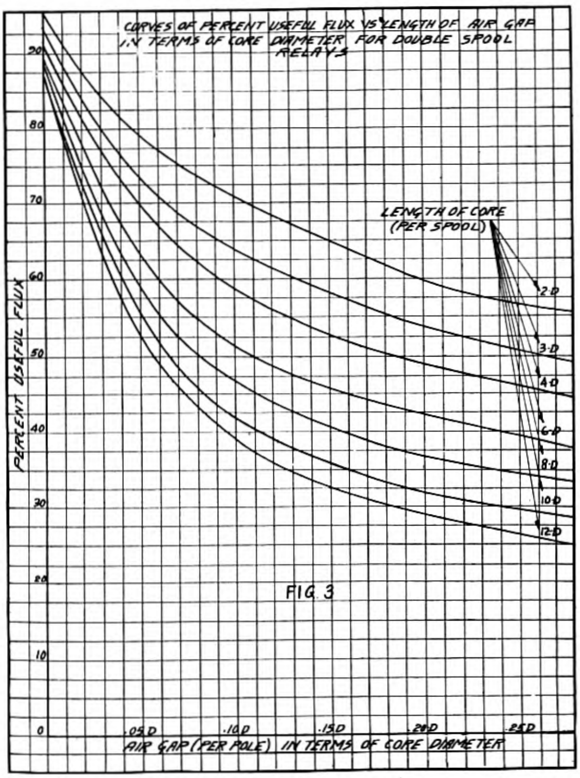


Fig. 3—Curves of Percentage Useful Flux vs. Length of Air Gap in Terms of Core Diameter for Double Spool Relays

the gap density. A high core flux density and pole face density gives a small core section and consequently a small and cheap magnet. The limit to the decrease in size is the allowable temperature limit of the winding.

Of course, there is a limit to the sacrifice of efficiency to obtain a low first cost. If the reasoning in the preceding paragraph is applied to a 5,000 gram electromagnet, the results will show three to four per cent of the total ampere turns required to saturate the core while on a relay which controls five grams the same assumptions show over 50 per cent of the total ampere-turns required to saturate the core. Where small forces such as five grams are involved, we are almost invariably concerned in maintaining a high efficiency

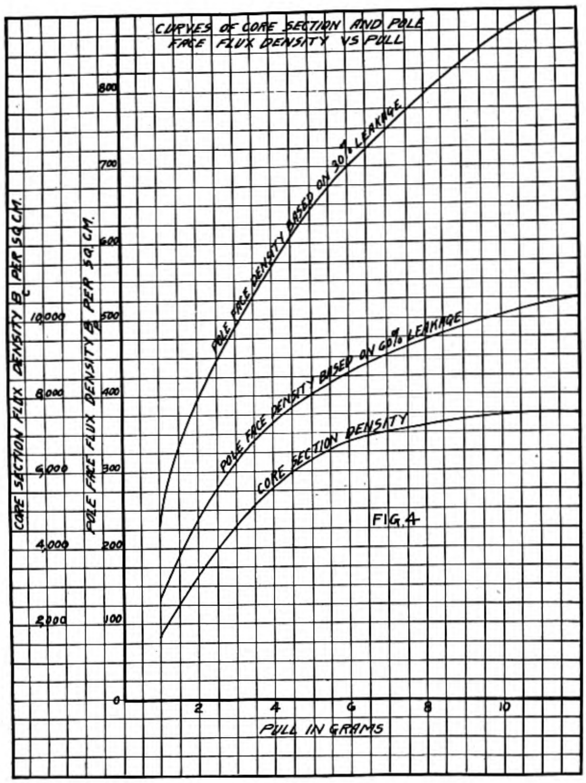


Fig. 4—Curves of Core Section and Pole Face Flux Density vs. Pull

whereas in designs for the heavier forces a few additional ampere-turns required in the core are relatively unimportant.

The work done by an electromagnet is $W = 980\ F\ L$ ergs where F is expressed in grams and L in centimeters. The energy in ergs required to magnetize the core is $W' = \frac{\phi NI}{20}$ where NI represents the ampere-turns required to force the flux ϕ through the core. The ratio of the core energy and the useful work may be taken as a criterion of the efficiency of the core design. Applying this reasoning to various designs it is found that the most efficient core design is obtained by choosing a core flux density at the maximum permeability of the core iron. If this reasoning is applied to a 5,000 gram relay a saving of approximately five per cent core energy results over working at a high density but the core section is increased in the

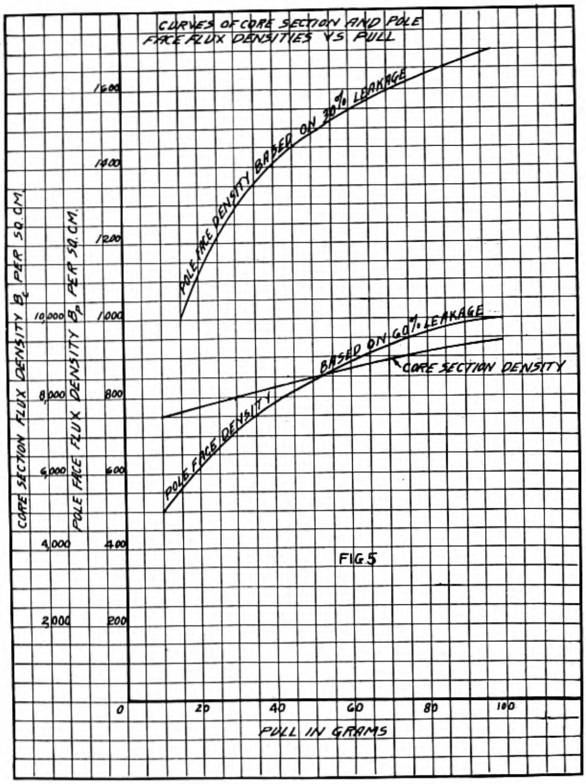


Fig. 5—Curves of Core Section and Pole Face Flux Density vs. Pull

ratio of six to one. Obviously the small improvement in efficiency results in an unreasonable increase in size and consequently first cost and is seldom if ever warranted by the requirements. Applying the same reasoning, however, to a five gram relay we obtain a reduction in core energy of approximately 20 per cent and although the core section has greatly increased this increase has practically no influence on the size or first cost of the magnet. Of course, a further consideration is mechanical strength as where light loads are encountered the core section, needed magnetically, may be entirely too small to give the requisite mechanical strength for winding or mounting. It may, therefore, be necessary to use a very low flux density in these instances in the core design.

The best flux density and area for the pole face as regards electromechanical efficiency is obtained by making the air-gap reluctance

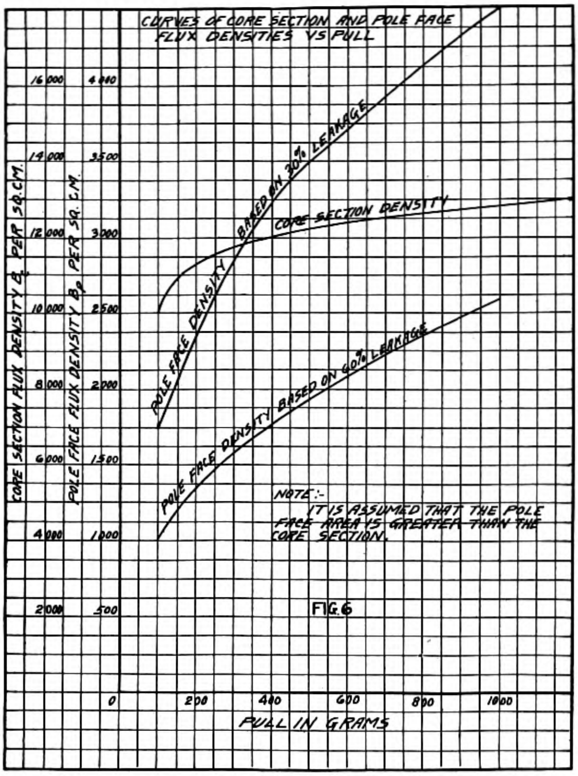


Fig. 6—Curves of Core Section and Pole Face Flux Density vs. Pull. These Curves Assume That the Pole Face Area is Greater than the Core Section

equal to the reluctance of the remainder of the magnetic circuit. Here again it is found that practical considerations must be carefully weighed, otherwise an unreasonable design results. If, for instance, the pole face density on a 5 gram relay is taken equal to the customary core density, a very small pole face area results. To make the air-gap reluctance, then, equal to the reluctance of the remainder of the magnetic circuit, it is found that an air-gap of possibly .001″ or less results. Such a small armature movement, of course, is generally of no practical value, and consequently, very low pole face densities are generally chosen.

As a result of the above considerations as well as the experience gained in designing a great number and variety of relays and electromagnets, the curves in Figs. 4, 5, 6 and 7 have been drawn which show

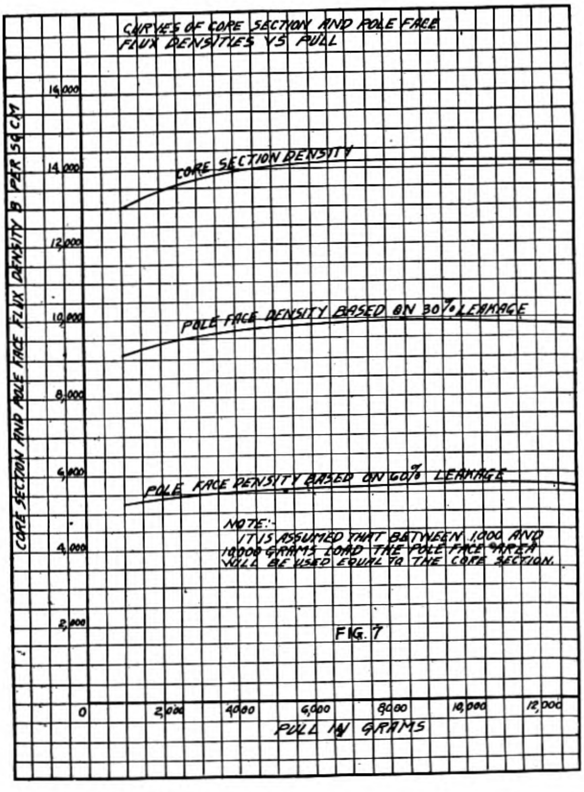


Fig. 7—Curves of Core Section and Pole Face Flux Density vs. Pull. These Curves Assume That Between Loads of 1,000 and 10,000 Grams the Pole Face Area Will be Used Equal to the Core Section

reasonable assumptions that may be made in working out new designs. These curves are to be employed, of course, with due consideration of the particular requirements in each case.

From the above discussion it is evident that magnetic irons which are capable of high flux densities are particularly desirable for the heavier magnets. The high densities permit of a small core section and consequently a small and low cost magnet. The magnets which control loads of a few grams, however, should be constructed of magnetic materials which have a high permeability and a low coercive force, but not necessarily capable of working at high densities. A relatively high permeability reduces the energy required to saturate the core although due to the reluctance of the air-gaps there is obviously a limit beyond which no practical gain results due to increased

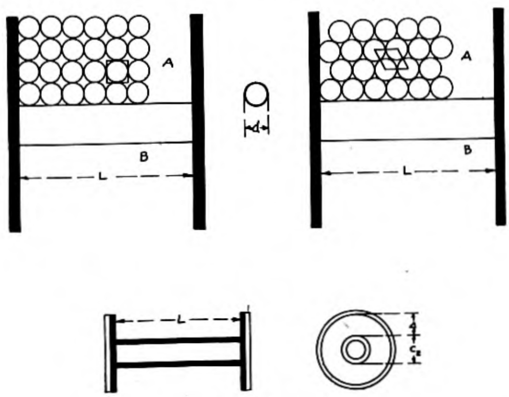


Fig. 8

permeabilities. The most important single requirement of a magnetic material for relays controlling light loads, is a low coercive force. A low coercive force reflects the ability of the magnetic parts to return to practically the same state of magnetization after repeated applications of magnetomotive forces. The effect of residual magnetism, if large, may cause sticking or holding forces of the same order of magnitude as the load requirements. Vacuum annealed silicon steels of comparatively high silicon content and certain nickel steel alloys which have low coercive forces are of great value for electromagnets which must control efficiently light loads of the order of one to fifty grams.

Winding Formulae

Before discussing the economics of the winding dimensions it is necessary to develop and carefully consider the winding formulae and the factors which determine the temperature characteristics.

Fig. 8 shows the one-half cylindrical section of a spool. Since a given wire occupies a similar space in both A and B we need only to consider winding space A. If d in Fig. No. 8 represents the diameter of the wire over the insulation, it is evident that each wire may occupy one of two positions with respect to adjacent wires. In the uniform layup each wire occupies an area d^2, and with the complete intermeshing of layers one wire occupies an area .866 d^2. In actual winding practice a combination of the two layups is obtained which gives .90 d^2 square inches as the space occupied by one wire. The area .90 d^2 may be taken as indicating perfect winding so that if the total winding space or area is represented by A and the total turns by N, we have under the best conditions

$$\frac{A}{N} = .90\,d^2.$$

The comparative merit or efficiency of any other winding may therefore be expressed as

$$\frac{.90\,d^2 \times 100}{\frac{A}{N}} = \text{per cent efficiency.}$$

As each size of wire and insulation winds with a somewhat different efficiency, the variation in the value A/N is generally determined experimentally for each gauge of wire. Thus

$$\frac{A}{N} = K = C_1\,d^2. \qquad (1)$$

The constant C_1 is often designated as a space factor constant and may include the insulating or interleaving paper used throughout the winding. The following are representative values of K for enamel and silk insulated wire of Western Electric Company manufacture.

VALUES OF K

B. & S. Gauge	Enamel Insulated Wire	Silk Insulated Wire
21	.000894	.000936
22	.000718	.000755
23	.000577	.000614
24	.000431	.000477
25	.000437	.0003825
26	.000280	.0003140
27	.000225	.0002615
28	.000183	.0002170
29	.000147	.000180
30	.000120	.0001510
31	.000096	.0001261
32	.0000781	.0001069
33	.0000628	.0000866
34	.0000516	.0000815
35	.0000410	.0000678
36	.0000338	.0000577
37	.0000269	.0000500
38	.0000222	.0000428

Referring to Fig. 8 the space or area available for winding is

$$A = L\Delta. \tag{2}$$

From equations 1 and 2 the total turns possible are

$$N = \frac{A}{K} = \frac{L\Delta}{K} \tag{3}$$

The total resistance of the winding is the product of the resistance of the mean turn R_m and the total turns N,

$$R_t = R_m N.$$

The length of the mean turn for a round core, Fig. 8, is

$$2\pi\left(\frac{C_2}{2} + \frac{\Delta}{2}\right) = \pi(C_2 + \Delta),$$

and if r is the resistance per unit length we have

$$R_m = \pi(C_2 + \Delta)r,$$

whence the total resistance is

$$R_t = \pi(C_2 + \Delta)rN;$$

or substituting the value of N from equation 3,

$$R_t = \frac{\pi(C_2 + \Delta)r\Delta L}{K}. \tag{4}$$

For a core of rectangular cross section equation 3 holds for the total number of turns and it will be found that the equation for total resistance is

$$R_t = \frac{\pi r L\Delta}{K}\left(\frac{p}{\pi} + \Delta\right), \tag{5}$$

where p represents the periphery of the core in inches.

Temperature Characteristics

The critical circuit conditions with respect to the relay winding specify either constant wattage, constant voltage or constant current. The constant voltage circuit is one in which a change in resistance of the relay winding materially affects the current flow. The constant current circuit is one in which a change in resistance of the relay winding does not materially affect the current flow. An approximate constant wattage condition is one in which a resistance such as a line in series with the relay is equal to the resistance of the relay and where the resistance external to the relay winding does not change appreciably with temperature variations.

The temperature formulae for the constant wattage condition are developed as follows:

Let Q be the quantity of heat in calories supplied to the winding per second, and $Q\,dt$ be the amount supplied in a small increment of time. Let S be the product of the specific heat and weight of the total wire on the spool expressed in calories. Let T be the temperature difference between the winding and the surrounding air. $S\,d\,T$ is then the amount of heat used in raising the temperature of the wire by the amount $d\,T$. Let ρ be the average dissipating constant throughout the temperature range. It depends upon the radiating surface of the winding, metal conducting parts of the structure and external convection of heat by the air. Given the constant ρ, $\rho\,T\,dt$ represents the calories dissipated during the interval dt.

The total heat supplied during the time dt is partially used in raising the temperature of the wire, and partially dissipated, consequently

$$Qdt = SdT + \rho Tdt. \tag{6}$$

If heat is continuously supplied the winding in the form of electrical energy, the rate of dissipation ultimately equals the rate of supply. This is true for temperatures that do not fuse the wire or permanently alter its resistance characteristic. Ultimately

$$SdT = 0$$

and

$$Qdt = \rho T_m dt.$$

If the final temperature reached is designated as T_m then

$$Q = \rho T_m \tag{7}$$

and from equations 6 and 7

$$\rho T_m dt = SdT + \rho Tdt,$$

$$-\frac{\rho}{S}dt = -\frac{dT}{T_m - T}$$

and integrating gives

$$-\frac{\rho}{S}t = \log(T_m - T) + C.$$

Observe that when $t=o$ the value of T is also zero and $C=-\log T_m$. Hence

$$T=T_m\left(1-\epsilon^{-\frac{\rho}{S}t}\right) \tag{8}$$

Equation 8 shows that the transient relation between temperature rise T and time is exponential and ultimately the temperature rise is $T=T_m$.

The final temperature T_m reached by the winding may be determined by writing equation 7 in the form

$$\rho T_m=\frac{EI}{4.186} \tag{21}$$

where $E\ I$ represents the constant wattage applied to the winding and 4.186 is the Joule equivalent. If the room temperature is T_r and the ultimate temperature rise T_m, it is evident that the final temperature of the winding is

$$T_f=T_m+T_r,$$

$$T_f=\frac{EI}{4.186\rho}+T_r.$$

By introducing a new constant K_1 which represents the ability of the structure to dissipate heat and also includes the factor $\frac{1}{4.186}$, we have [2]

$$T_f=\frac{EIK_1}{A_1}+T_r, \tag{9}$$

in which A_1 is the area of the winding but does not include the ends.

The value of K_1 can be readily determined by obtaining an experimental curve between $E\ I$ and T_m. This is obtained by gradually increasing $E\ I$ but holding the wattage constant for each value long enough for the final temperature rise to take place. The value of T_m is calculated by observing the change in resistance of the winding.

The constant current and constant voltage characteristics are determined in a similar manner with the important exception that the quantity of heat Q supplied per second is not constant but varies in accordance with the change in resistance with temperature. Thus for constant current conditions $4.186\ Q\ dt=I^2\ R\ dt$ and for constant voltage conditions $4.186\ Q\ dt=\frac{E^2}{R}dt$, where $R=\frac{R_0(234.5+T)}{234.5}$ for centigrade degrees and R_0 is taken at 0° C.

[2] For single spool relays $K_1=50$ to 60, and for double spool relays $K_1=35$ to 50.

The steps by which the solution fitting these conditions are obtained will not be given but the results, stated in practical units, are included

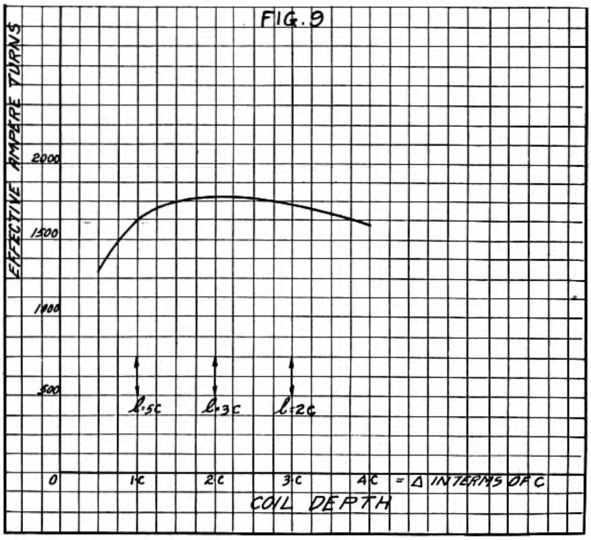


Fig. 9—Relation Between Coil Depth in Terms of Core Diameter and Effective Operating Ampere Turns

to bring out certain important facts relating to winding design. The final temperatures are

$$\text{Constant Wattage, } T_f = \frac{K_1 EI}{A_1} + 20°; \tag{10}$$

$$\text{Constant Current, } T_f = \frac{5090A_1 + 234.5I^2R_{20}K_1}{254.5A_1 - I^2R_{20}K_1} \tag{11}$$

$$\text{Constant Voltage, } T_f = -107 + \sqrt{16000 + \frac{254.5K_1E^2}{A_1R_{20}}} \tag{12}$$

The transient temperatures of constant wattage, voltage and current are all of the exponential form $T = T_m\,(1-\epsilon^{-ct})$, while the cooling of the winding after current is stopped is of the form $T = T_m\,\epsilon^{-ct}$. In these equations c is the constant pertaining to the particular condition considered.

An important observation in connection with these temperature characteristics is the great difference in temperature rise in the three cases with *like initial conditions* of energy input. Thus, it is important to note that an electromagnet which is correctly designed and worked to its temperature limit in a constant voltage circuit, would overheat in a constant wattage or constant current circuit. A relay properly designed to work at a safe temperature under a constant current condition, would be unnecessarily large and expensive in a constant voltage or constant wattage circuit. It is, therefore, evident that

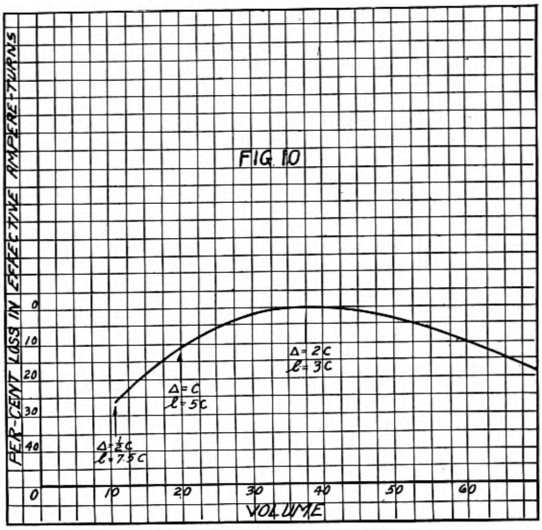


Fig. 10—Relation Between Copper Volume and Percentage Loss in Effective Ampere Terms

exact rules can not be given for the correct proportioning of spool and winding dimensions from a purely design standpoint without consideration of the circuit in which the electromagnet is to operate. Some general design features, however, can be indicated which will enable preliminary assumptions to be made that can be refined as the design is worked out for its particular operating conditions.

Spool Dimensions

Certain important facts regarding spool dimensions are indicated in Fig. 11. The spool dimensions for the winding may be investigated by assuming that a definite radiating surface must be used to dissipate the heat, and then determine the relative values of winding depth, length, and volume in terms of the core diameter. The volume of wire used in the spool is taken as a measure of the first cost and a variation in the length of the coil is reflected in the leakage flux which in turn may be taken as a measure of the effective ampere turns. The determination of the leakage flux involves reasonable assumptions from experience of the armature air-gap in terms of the core diameter.

If the electromagnet is to be operated on a definite voltage the assumption of a definite radiating surface to dissipate a certain input

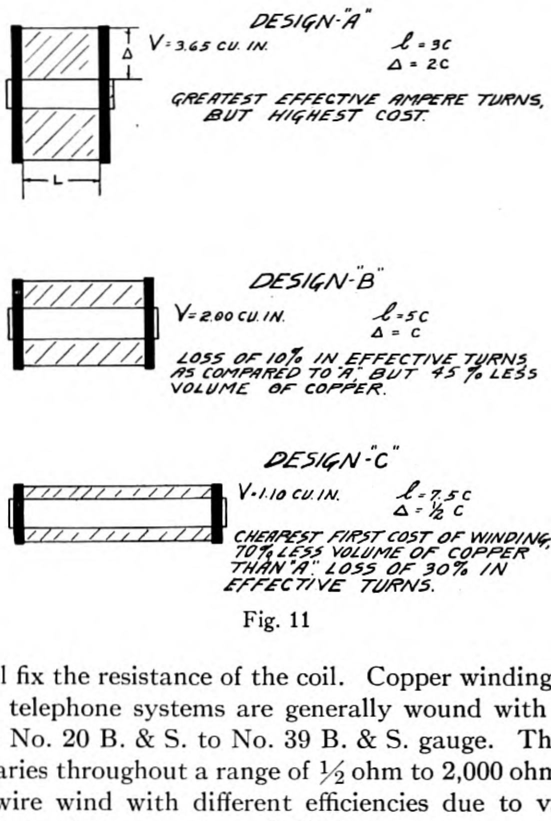


Fig. 11

wattage will fix the resistance of the coil. Copper windings of electromagnets in telephone systems are generally wound with wire which varies from No. 20 B. & S. to No. 39 B. & S. gauge. The resistance generally varies throughout a range of ½ ohm to 2,000 ohms. Various gauges of wire wind with different efficiencies due to variations in the space factor but a number of different gauges may be assumed and the calculations carried out which give the relation between the winding depth and the effective ampere-turns. With a constant radiating surface a variation in the winding depth causes a variation in the length which, of course, is reflected in the leakage flux. The results of a number of calculations on various windings are shown in Fig. 9. In Fig. 10 is shown the relation between the volume of wire on the spool and the per cent loss in efficiency due to a variation in the depth of winding which, with a constant [3] radiating area, causes

[3] The radiating area is taken as the surface only of the coil and the ability to dissipate through the ends and otherwise is reflected by the heating constant K_1.

a corresponding change in the length of the coil. Fig. 11 shows the relative dimensions of three designs of spools taken from Figs. 9 and 10.

Some very interesting information can be obtained from Fig. 11 in regard to the relation between the volume of wire, as reflecting the first cost, and the ampere turn operating efficiency. Design "A" contains a volume of copper of 3.65 cubic inches, while in design "B" the volume of copper has been reduced to 2.00 cubic inches although the loss in effective ampere-turns is only 10 per cent. In design "C" the volume has been reduced to 1.10 cubic inches with a loss in efficiency of 30 per cent.

Obviously the design "C" is the cheapest in first cost because of the small copper volume and will also give the lowest annual charge where the time of operation is very short and the charge for power relatively low. Where the magnet is required to operate very often and the price of power is high the design "A" will prove the most economical. Design "B" may be considered as intermediate between designs "A" and "C".

In the above considerations of spool dimensions the examples given should not be taken as an accurate generalization but simply as a method which, with a given set of requirements, should enable reasonable first approximations to be made. Thus, if annual power charges are controlling, a relatively short and deep spool will give the best results, although there may be exceptions where for instance, the operating current is reduced to a holding value and where the leakage is relatively small due to the fact that the armature is operated. In such a case and unless operating efficiency is also of prime importance the design "A" would be more expensive than necessary in first cost. Other cases often arise where the input wattage is very small but the operating requirements are very exacting so that the most efficient winding is required and the first cost is relatively unimportant. In this case a larger volume than "A" can be used to advantage. These examples may be used as a guide therefore, in determining spool dimensions which are later refined as the design is completely worked out. The illustrations of designs given in the latter part of this paper show how accurately certain final design dimensions can be worked out to give the minimum annual charge.

Discussion of Designs Used Extensively in the Telephone Plant

To any one familiar with telephone systems it is obvious that it is impracticable to design all the relays required at maximum efficiency and economy for each particular condition that arises. Such a procedure would involve endless equipment changes as well as the large and unnecessary manufacturing expense of making an excessive number of types of relays. Much of the relay engineering work of the past few years has therefore been directed toward the standardization of relay designs which would be flexible, reliable and economical

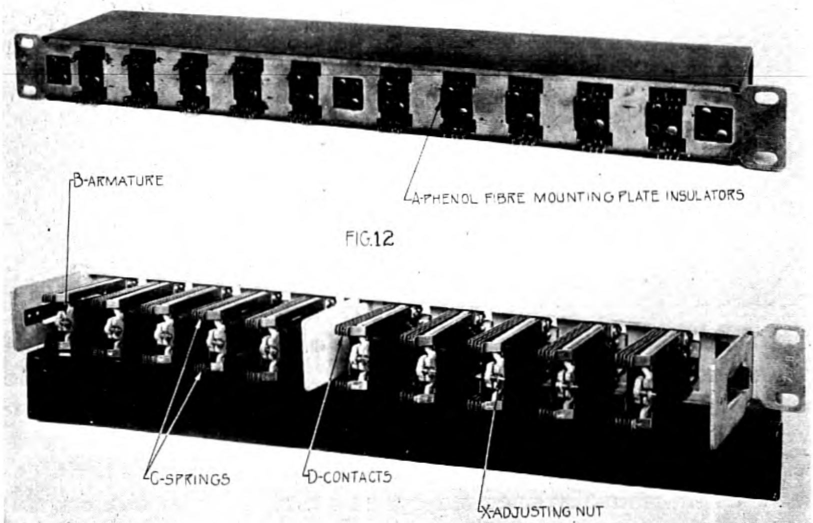


Fig. 12

as a whole in the telephone plant rather than the most efficient in all respects for any specific condition. The flat or punched type relay manufactured by the Western Electric Company represents largely the result of this effort.

The flat relay is essentially a punch press product manufactured yearly in large quantities and in about 3,000 varieties of windings and switching or contacting arrangements. The punch press method produces parts which are exact duplicates and therefore interchangeable which is particularly advantageous both for assembly and replacements or repairs. All the springs as well as the core and armature are punched and formed in bending fixtures to the required shapes. The mounting plates are also punched and designed to permit of uniform and economical mounting of the relays.

A number of these relays are shown on a punched mounting plate in Fig. 12. Referring to the figure it will be seen that the relays are insulated from the mounting plate by phenol fibre insulators "A," which are securely fastened to the mounting plate by means of metal eyelets. The armature "B" is hinged at the rear by the use of a thin, steel reed, securely riveted to the armature. The switching arrangements which the armature controls are in the form of nickel silver springs "C" with the contacts "D", at the front and in plain view.

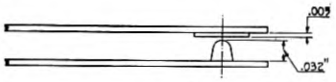


Fig. 13

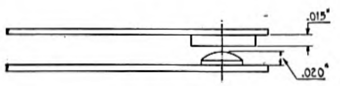


Fig. 14

The springs and contacts are mounted vertically which is particularly effective in keeping the contacts clean. The contact points are made from platinum or a recognized equivalent, and are designed in the form of points and discs to facilitate alignment and adjustment. Two designs of contacts have been standardized; one size being used for the customary electric currents and wear conditions encountered in manually operated systems and a larger size for the somewhat more severe conditions of wear frequently encountered in automatic systems. All contacts are electro-welded on their respective spring supports and the two sizes are shown in Fig. 13 and Fig. 14, respectively.

The springs and their associated contacts are designed in twenty-six switching arrangements as shown in Fig. 1. A single relay may be provided with one of these switching arrangements or any one of these twenty-six arrangements may be paired with any other arrangement. Thus on a single relay there may be chosen any one of 377 switching or contacting combinations. The 377 spring combinations provide a great flexibility in circuit design and permit of uniform and efficient equipment layouts.

In manufacturing the relays the spring assemblies are clamped together under high compression before tightening the screws which hold them together. This insures that the springs retain their position and adjustment throughout a long period of time. The arrangement of the springs is such that definite stops or supports are provided for each spring either on the front spool head or on the armature. In tensioning or adjusting the relay springs against their supports, sufficient tension is set up in the springs to insure a pressure of at least 15 grams between all contacts at the time of closure.

The amount of current and power required to operate each relay is dependent upon the tension and number of springs that must be moved and the distance through which this movement takes place. Relays or electromagnets operate most efficiently with the armature air-gaps set at the minimum required for the satisfactory opening and closing of the contacts. Consequently a method has been carefully worked out for these relays in which the armature travel is set in accordance with the requirements of the particular spring combination by the adjustment of the friction lock nut "X" shown in Fig. 12. This setting of the armature insures a normal separation of contacts of approximately .010 inch and at least .005 inch "follow" after closure of the contacts. The "follow" allows for a certain amount of contact wear as well as insuring a slight wiping action which gives a certainty of contact closure. The electrical operating current requirements are figured and specified on the basis of obtaining 20 grams pressure between all contacts; this margin being allowed so that no undue hardship will be experienced in maintaining the minimum requirement of 15 grams.

The insulating materials used throughout have been carefully studied and the best materials known to the present day art have been used. Thus the wire used in the winding is insulated with a high grade enamel and the insulating papers on the core are practically inert from an electrolytic corrosion standpoint. The coils are covered with a serving of cotton, treated with unbleached shellac which acts as a seal against moisture and protects the winding from abrasion. The phenol fibre used on the spool heads and spring insulators is much superior to hard rubber in regard to its ability to withstand a wide temperature range without appreciable expansion or contraction.

For this reason it is permissible to work these relays at higher temperatures without danger of fire hazard or deterioration of the insulation than relays insulated with hard rubber parts. These higher temperature limits permit a wider usefulness of the relays in circuits as well as economy in construction as the size of the coil often depends on the necessary area for radiation and this area is fixed by the permissible temperature range.

Where the relays are to remain operated a considerable length of time throughout the day the annual power charges become important and the design of the winding and in some cases the size of the spool must be altered to give the minimum annual charge. The group of curves in Fig. 15 show how nearly correct these relays have been designed for conditions where the operating ampere-turns are 260 and the relays remain operated from 60 to 600 minutes per day.

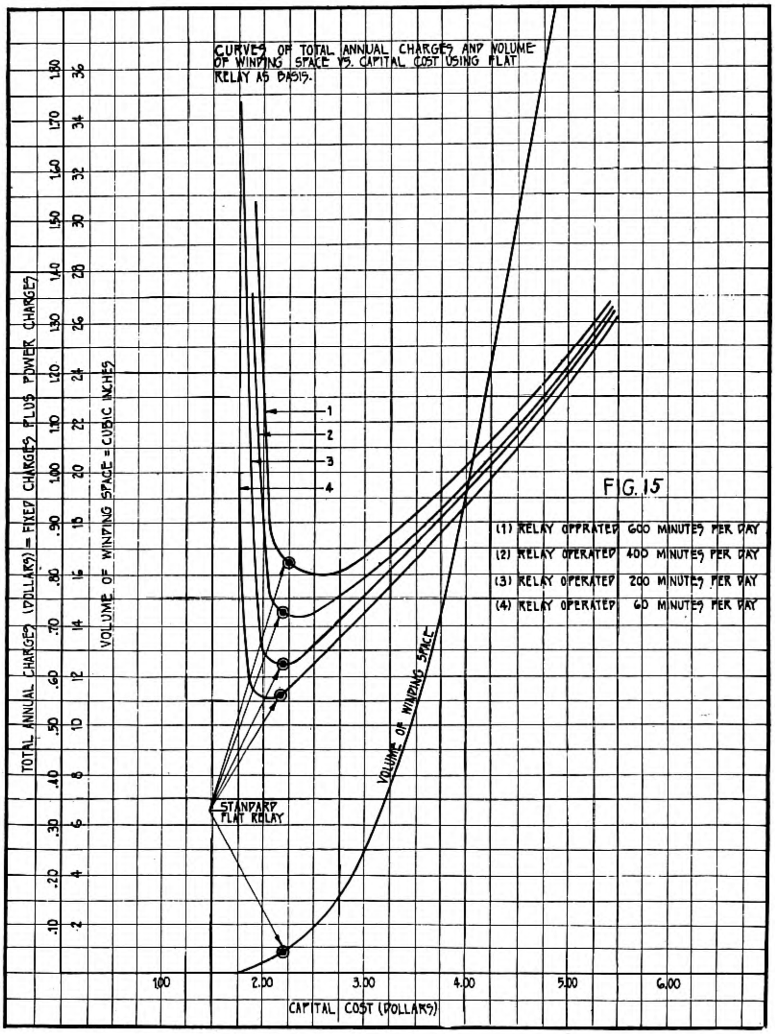


Fig. 15

The capital cost and annual charge figures should be taken as relative only as the correct values will vary with manufacturing conditions and with the cost of power for different localities.

Other designs of relays used extensively in the telephone plant are the relays that control the supervision of a telephone connection and the alternating current relays which operate on ringing currents of 16 to 20 cycles frequency.

Relays which are used for supervisory purposes and alternating current relays are generally constructed of silicon steel instead of the customary Norway or magnetic iron. The silicon steel is very

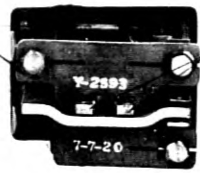


Fig. 16

satisfactory for these relays because of its comparatively high permeability, low coercive force and small hysteresis. The high permeability is advantageous for relays that are required to operate on a very small energy input and the low coercive force is very effective for obtaining a quick and positive release of the relay armature, particularly where a leak current exists due to faulty line insulation. A great improvement in many of these relays can be obtained by the use of certain nickel-iron alloys which have been recently developed and are known as "Permalloy."

A relay for use on ringing currents is shown in Fig. 16. The armature "A" of this relay is attracted to the bifurcated extensions of the core "B." One of these core extensions is completely surrounded by a part of the copper spool head "C." This arrangement is known as pole "shading" or phase splitting and is used to produce a substantially steady pull on the armature when the relay is energized by single phase alternating current.

Referring to Fig. 17 the theory of operation is shown by considering the vector diagram in connection with the schematic drawing of the relay core and armature. When an alternating current is applied to the winding we can assume that an alternating flux $2\ \phi_m$ is generated in the core. This flux divides into two approximately equal parts in the two bifurcated extensions of the core. If these two fluxes can be displaced in time phase it is evident that the armature will be attracted by one of the bifurcated extensions of the core, while the flux, and consequently the attraction of the other, is passing through zero. This may be explained by the vector diagram in which E_2 represents the induced voltage in the short circuited copper ring due to the alternating flux ϕ_m. The current in the copper ring I_2 lags behind the voltage E_2 as shown and the flux due to this current

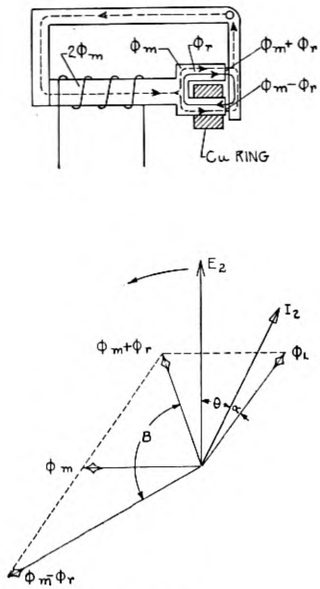


Fig. 17

is ϕ_r. This flux ϕ_r has a magnetic path through the bifurcated pole pieces and armature as shown by the arrows. Following out the arrows it will be seen that this flux adds to the flux ϕ_m in the upper part and subtracts in the lower part of the two core extensions.

The vector addition and subtraction of these two fluxes results in two vectors $\phi_m+\phi_r$ and $\phi_m-\phi_r$, each of which represents a flux that crosses an air-gap to attract the armature. These two fluxes differ in time phase as represented by the angle "B" so that a substantially constant attraction results on the armature. The operation of the relay under these conditions is very much the same as that of a direct current relay as no vibration or chatter of the armature or contacts occurs. The minimum effective alternating current ampere-turns required for operation are 70 to 100 ampere-turns.

Such a relay, of course, operates on direct current as well as on alternating current and in fact the direct current supervisory relays are quite similar to these relays in mechanical design.

Fig. 18 shows the design features for the supervisory and ringing frequency relays. In this figure the winding has been omitted so as to show clearly the unusually small core. This construction is especially efficient in circuits where the relay receives at times a

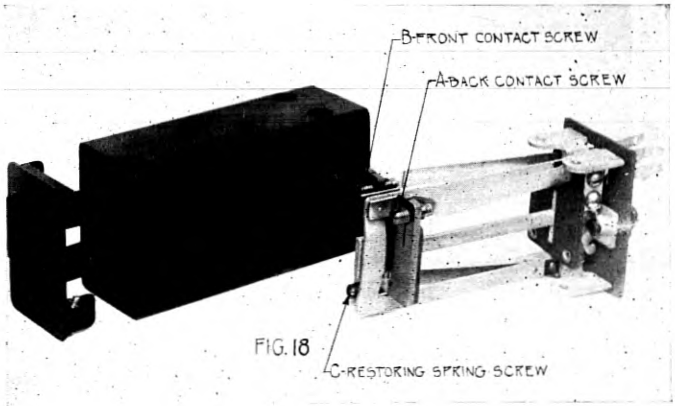


Fig. 18

very small amount of energy for operation and must also release reliably against a leak current immediately after operation by a comparatively large amount of energy. The small core saturates magnetically on a relatively small current or energy so that excessive energy does not store up additional magnetism which would retard or prevent the release of the relay.

Referring further to Fig. 18 the micrometer screws "A" and "B" are used to adjust the back and front contacts respectively, and to fix both the unoperated and operated positions of the armature. The screw "C" is used to control an armature restoring spring which is in the form of a flat spring riveted to the armature. These relays are generally provided with individual covers which are effective in preventing cross talk of telephone voice frequencies when used as supervisory relays in telephone switchboards.

A Dynamical Study of the Vowel Sounds

By I. B. CRANDALL and C. F. SACIA

Introduction

THE study of the vowel sounds presents a problem which has interested scientists and scholars in varied fields. A knowledge of their nature is of fundamental importance not only in communication engineering but also in acoustic science, phonetics and vocal music. From the earliest theories and the rough experiments of Willis (1829) and Helmholtz (1859) to the later measurements of D. C. Miller (1916) steady progress has been made toward the accurate determination of their characteristics.

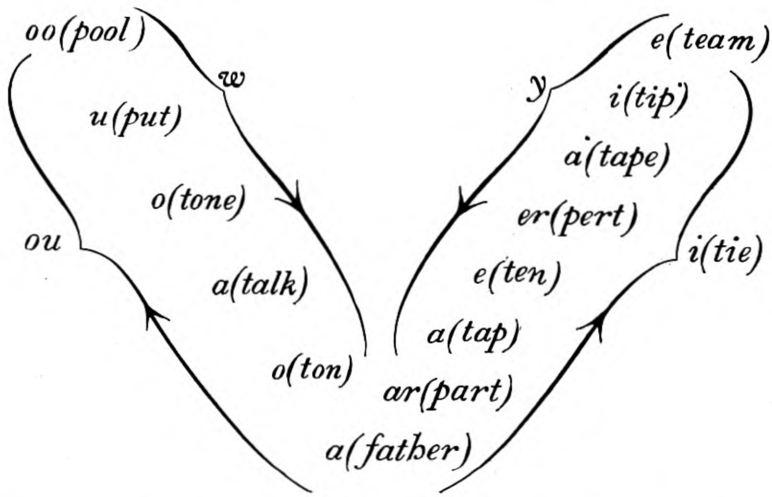
Further progress in this study has been made possible with improved facilities now available in the telephone research laboratory. It has been felt that there was need for more accurate records of the spoken sounds and the development of improved transmitters, amplifiers and other devices has made possible recording apparatus of greater accuracy, range and power than any heretofore used.

In this paper will be given the results of an analysis of spoken vowel sounds based on a set of accurate oscillographic records. The recording apparatus was designed to record the wave forms of the different speech sounds practically free from distortion over the frequency range from 100 to 5000 cycles. A brief description of this apparatus is given in the appendix. The emphasis in the present paper is placed on the composite frequency characteristics of the sounds as revealed by a particular method of analyzing the records so obtained.

Analysis of the Data

The thirteen vowel sounds investigated are shown arranged in a triangle in Fig. 1. The diphthongs *ou*, *w*, *y* and long *i* are not included. Eight records of each sound were taken, four by male and four by female speakers. In speaking these sounds the only constraint imposed on the speakers was that the sound should be completely uttered within an interval of one second. The recording mechanism was so arranged that the whole of the sound from beginning to end was recorded in one continuous graph. In practice the average duration of these sounds was about 0.30 second. Each record shows a sequence of growth and decay in amplitude somewhat as follows: first a period of rapid growth in amplitude lasting about .04 second during which all components are quickly produced and rise nearly to maximum amplitude; second a middle period in which the general amplitude is nearly constant but with varying phase relations between the different components and lasting about 0.17 second; and finally a period of gradual decay lasting about .09 second in which all the components disappear. A typical record so obtained is shown in Fig. 2.

A brief description of the method of mechanically analyzing such a record is given in the appendix. The essential point of the analysis is that the whole record from start to finish is taken as the unit for analysis and the data obtained are therefore the average characteristics of the sounds throughout their duration.



It is usual to exhibit the properties of a vowel sound in a spectrum diagram showing the amplitude of the component vibrations as a function of their pitches or frequencies. For each vowel sound there are, in addition to fundamental tones, certain characteristic regions of resonance which may be at high or low frequencies. It would be possible from the results of this analysis to present the sound spectra of each vowel showing the relative amplitudes for the different frequencies as present in the original air vibration[1] but this treatment has been modified to take into account the relative importance of the various pitches in hearing. Using the data available

[1] In previous publications (*Phys. Rev.* XIX, 1922, p. 228, Fig. 7, and *Bell System Technical Journal*, Vol. 1, No. 1, p. 124,) data have been given showing the actual distribution of energy in average speech. The tremendous concentration of energy in the lower frequencies is somewhat misleading unless account is also taken of the much reduced sensitivity of the ear in this region.

on the relative sensitivity of the ear at different frequencies[2] we have multiplied the acoustic amplitude at each frequency by the corresponding ear sensitivity factor and the results obtained are taken to be the effective amplitude frequency relations which are characteristic of these sounds.

The data from the four male records and from the four female records of each sound are separately composited and the resulting curves are shown in the diagram (Fig. 3). This compositing process was somewhat laborious because the analyses of the separate records were made not with reference to predetermined frequency settings, but rather for those critical frequencies which best determined the shapes of the spectrum curves. The individual curves were therefore plotted, and the average ordinates were then read off for small intervals of pitch. These ordinates were then averaged for each group of four analyses. These average ordinates (after being corrected for the calibration of the recording apparatus) were then multiplied by the ear sensitivity factors for the corresponding frequencies, and the curves so obtained were plotted on the musical pitch scale according to the usual practice. The final spectrum diagram thus shows the relative importance of the amplitudes of all the components of each vowel for male and female speakers.

The amplitude units are entirely arbitrary; it is only the shapes, not the sizes of these curves which have any significance. The order in which these curves are arranged is based upon the vowel triangle in Fig. 1.

Characteristics of the Vowel Sounds

The results of the analyses, as given in Fig. 3 show the essential dynamical properties of these sounds. Consider first the sounds numbered I to VI, which include those vowels usually designated as having single regions of resonance. Progressing through the sequence from I to VI this region of resonance rises in average frequency and becomes narrower in range. The rise in average frequency is of course a well known characteristic. There is also, at least with the male voices, a somewhat scattered and less well defined high frequency range of resonance, perhaps not essential in speech but more highly developed in well-trained singing voices.

The sound *a* (No. VI) is as it were the center of gravity of the vowel diagram and occupies the key position in the phonetics of

[2] See this Journal Vol. II, No. 4, October, 1923. The paper on audition, by H. Fletcher shows a cut of the "Threshold of Audibility" curve from which these data were obtained.

most languages. Now consider the sequence from this sound to No. XIII at the end of the diagram; these sounds include most of those which are known to have two characteristic regions of resonance. The main region of resonance now divides into two parts which gradually recede from each other as we follow the diagram downwards. (Sound X (*er*) is difficult to fit into the diagram in an exact position, but it is evident that it belongs in the series of doubly-resonant vowels.)

Contour lines (nearly vertical) have been drawn on the diagram to indicate the progressive changes in regions of resonance. Viewing the diagram as a whole it is important to consider not only the location of the resonant ranges but also their extent, and their relative separation from other resonant ranges in order to arrive at the essential characteristics of the vowel sound. In other words the individual vowel characteristic depends not only on the absolute pitch but on the relative pitches in case there is more than one region of resonance. It is only in this way that we can explain what is a matter of universal experience in using the phonograph; namely that moderate variations from normal speed in recording and reproducing speech leave the vowel sounds still intelligible.

It is expected to deal in a later publication with the semi-vowel sounds *l*, *ng*, *n*, *m* which seem to be related to the general diagram of the vowel sounds, and on which a preliminary report has already been made[3].

The more interesting features of the original records as such will also be dealt with in a subsequent publication.

APPENDIX

Recording and Analysis of Vowel Sounds

Recording Apparatus

The apparatus used in recording consisted of a condenser transmitter, an amplifier, and an oscillograph, in which important modifications were made. The vibrator was given great stiffness and damping so that the frequency response of the vibrator was nearly uniform up to 5000 cycles. Instead of the usual 12 inch film, special film 51 inches in length was used. This necessitated a much larger film drum. Furthermore the desired length of the record was about four times the circumference of the film drum, so the shutter was arranged to stay open during four revolutions while the vibrator was

[3] *Phys. Rev.* 23, 1924, p. 309—"Preliminary Analysis of Four Semi-Vowel Sounds."

given a slow uniform rotation about its vertical axis. With the film on the drum, the record thus had a helical form. In this way records of the requisite length were obtained.

The condenser transmitter was of the type developed by E. C. Wente, its characteristics combining with those of the amplifier and oscillograph vibrator in such a way that the combined amplitude response for the whole system was fairly uniform up to 5000 cycles, while the phase lag was approximately a linear function of frequency over the same range. This apparatus was therefore well adapted to the production of faithful records of the vowel sounds. The photographic equipment permitted the use of a time scale as great as six meters per second on the record (i.e. 2 inches = 0.01 sec.)

Transformation of Records for Analysis[4]

The oscillograms taken with the above apparatus were line records; in order to analyze these wave forms by the photo-mechanical method outlined below, it was necessary to transform the line record into a black profile. This was accomplished in the following steps:

(1) A positive print of the wave form on the original record was made on motion picture film.

(2) The emulsion of the positive print was then cut through to the base along the line of the wave by means of a stylus.

(3) The entire strip was blackened (on the emulsion side) with printer's ink.

(4) The emulsion on one side of the wave was stripped from the base, thus leaving the profile.

(5) The beginning and end were joined to form an endless belt.

Photo-Mechanical Analysis of the Prepared Records[4]

The principle of the photo-mechanical analysis is as follows: The motion of the strip past the image of an illuminated slit causes fluctuations in a beam of transmitted light which in turn, produce voltage fluctuations in the circuit containing a selenium or photo-electric cell. This voltage is then analyzed by means of a tuned circuit, an amplifier and a rectifier. The frequency of any component selected in this manner is determined by the tuning frequency divided by the ratio of speed transformation (analysis speed divided by the original speed of recording). The measured amplitude of the selected

[4] *Phys. Rev.* 23, 1924, p. 309. It is planned to publish a more detailed description of this apparatus later.

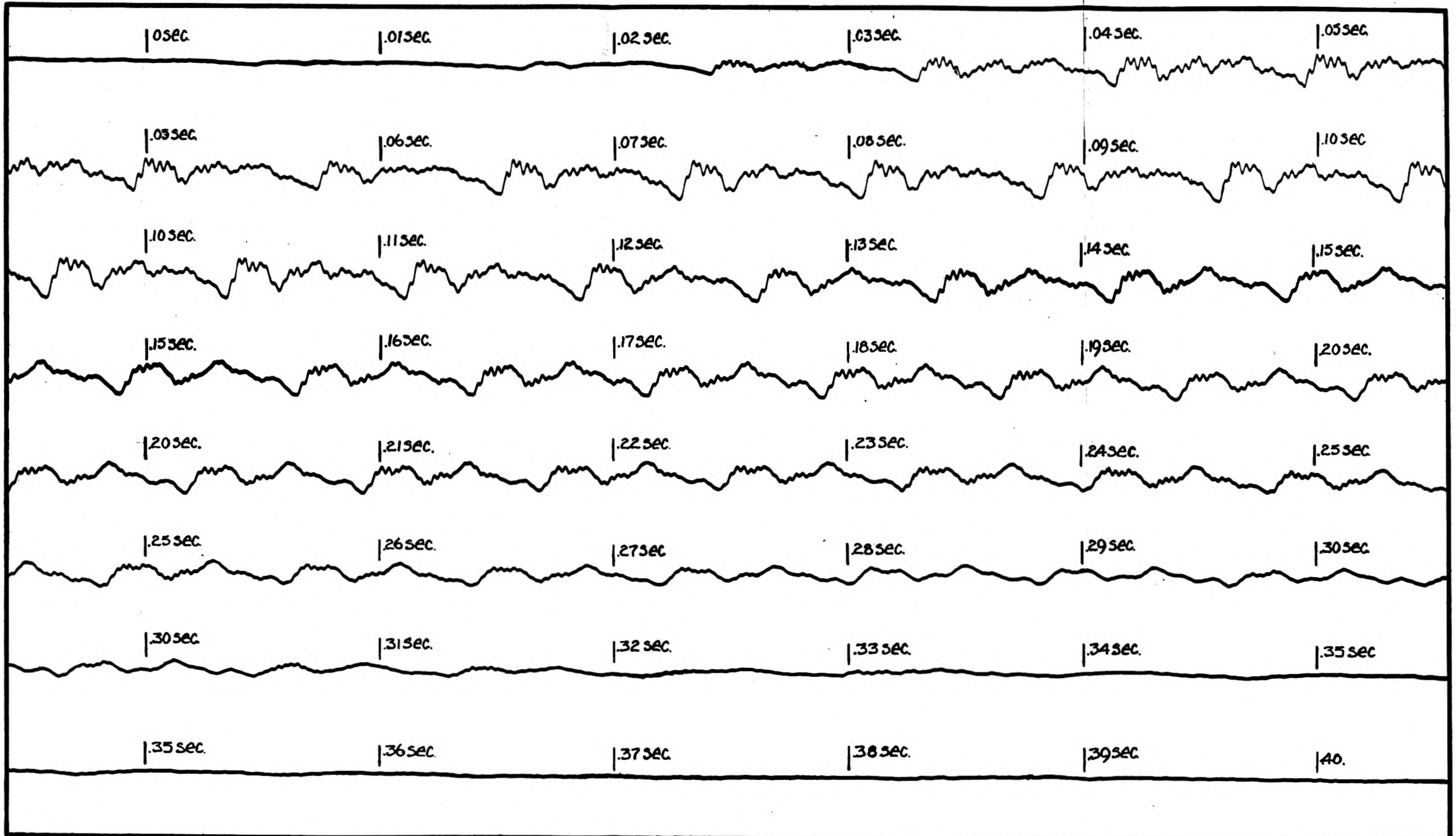


Fig. 2—*oo* as in p*oo*l; spoken by M. B.—male, low pitched. Plate No. 2: made from film record No. 157-A.

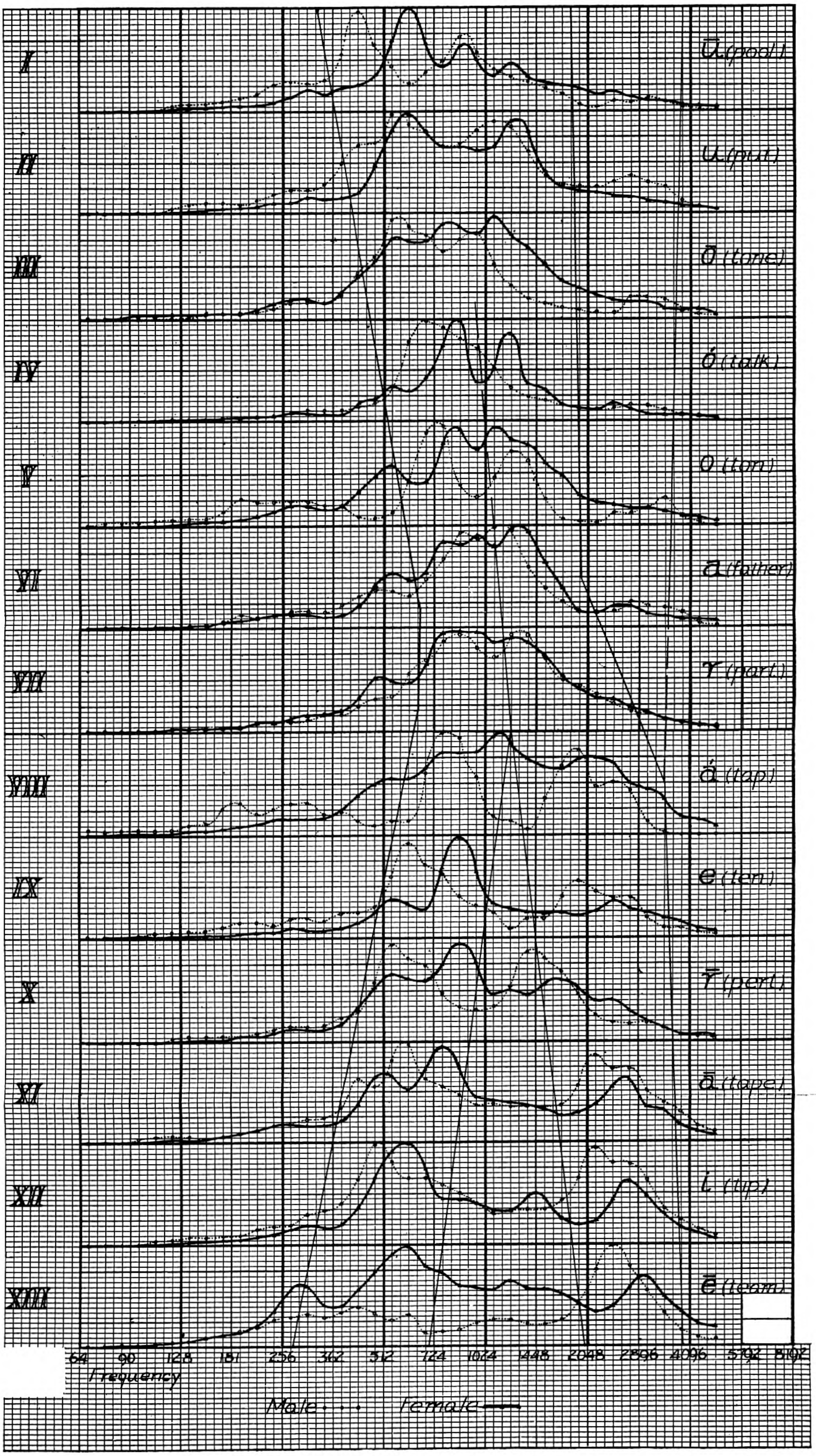

I
II
III
IV
V
VI
VII
VIII
IX
X
XI
XII
XIII
ū (pool)
u (put)
ō (tone)
ó (talk)
o (ton)
a (father)
r (part)
á (tap)
e (ten)
r̄ (pert)
ā (tape)
i (tip)
ē (team)
64
90
128
181
256
362
512
724
1024
1448
2048
2896
4096
5792
8192
Frequency
Male
Female

Fig. 3

component is determined by the rectifier output, the sensitivity factor of the selenium cell and the area of the frequency response curve of the tuning apparatus.

Since the wave form of a vowel sound is not a true periodic function, it is represented analytically by a Fourier Integral, not by a Fourier Series. The continued repetition of the motion of the wave past the slit, however, builds up a periodic function consisting of a fundamental and a series of harmonics. The magnitudes of these components bear a simple relation to those of the infinitesimal components of corresponding frequencies in the Fourier Integral. It is this series of harmonics which is measured by the above method, hence the problem of analyzing the aperiodic function represented in the record is solved by means of the related periodic function.

Humidity Recorders

By E. B. WHEELER

DURING recent years, the study of atmospheric conditions and their bearing on various industrial problems from the standpoint both of their effects on human efficiency and on manufacturing processes, is a matter that has received much attention, and the use of air conditioning systems, which have been developed in the last few years, has resulted in greatly improved working conditions, as well as in increased outputs of manufactured products of better quality than obtainable when air conditioning was not employed.

It is not so well appreciated, perhaps, that atmospheric conditions have a material effect upon the operation of intricate electrical and mechanical apparatus, such as those found in telephone systems.

Water vapor, and both gaseous and solid impurities in the air, hasten oxidation and corrosion of metals and also reduce the value of the insulation afforded by insulating materials. These effects usually are greatly accelerated if the temperature is high and if the materials are subjected to differences of electrical potential. Telephone apparatus and equipment consist of combinations of materials which are subject to both of these effects and, in general, the parts are small and the materials used in making them must be carefully chosen with regard to the necessary physical and electrical properties required for proper functioning of the apparatus. Therefore, the severe atmospheric conditions, which may be encountered in service, either must be eliminated by the use of air conditioning systems or the apparatus must be designed to withstand those conditions.

Accordingly, in order that the problem may be handled intelligently, accurate information must be available showing the character of the atmospheric conditions which exist in typical localities where telephone equipment is installed, so that the effects of these conditions on proposed designs may be studied under carefully controlled similar conditions in laboratory "humidity rooms." An outline of some of the work which has been done in an effort to obtain such information may therefore be of interest.

The first recourse would seem to be the data recorded by the various stations of the United States Weather Bureau. However, since these data usually represent periodic observations of outdoor conditions which are obtained primarily for meteorological purposes, it was found that while they indicate the general climatic conditions of different localities, they can not be taken to represent typical conditions in central office buildings, and therefore it has been necessary to devise methods by which we might secure such information.

The subject of hygrometry has long been one of the problems to which various investigators have given attention and the results of their work are a matter of record.

Thus it has been recognized [1] that, because of its ease of manipulation and its accuracy if suitable precautions are observed, the ventilated psychrometer is a suitable instrument for use in humidity measurements.

Consideration of the various types of hygrometers, commercially available, indicated however, that none would be suitable if reliable continuous records were to be secured. The use of simple wet bulb—dry bulb hygrometers would require practically constant attendance if frequent observations were made, and the results would not be accurate unless arrangements were made to circulate the air over the wet bulb. A pen recorder of the circular chart type to record wet and dry bulb temperatures had been used during one summer in a telephone central office where the humidity conditions were severe, but the results secured were not considered reliable because of the unsatisfactory method used to ventilate the wet bulb, as well as the sluggishness of the recorder due to pen friction on the chart.

Considerable experience in the laboratory with a recording hair hygrometer also had shown that, in addition to the inaccuracies to which hair hygrometers are commonly subject, the friction in the lever mechanism and between the pen and the chart made the instrument too erratic to be considered of possible use in the work being undertaken. Accordingly, a study was made to determine the possibility of developing apparatus which would overcome the troubles inherent in such recorders.

Development of a Recording Hygrometer

A promising method, developed by D. T. May of the Bell System Laboratories and operated successfully in the laboratory, consisted in the use of accurate and matched mercury thermometers, the stems of which were contained in a camera which would enable the heights of the mercury columns to be photographed upon a roll of sensitized paper. Arrangements were made for shifting the paper between

[1] U. S. Weather Bureau Psychrometric Tables for Obtaining the Vapor-Pressure, Relative Humidity and Temperature of the Dew-Point from Readings of the Wet and Dry Bulb Thermometers, by C. S. Marvin.

Proceedings of the Physical Society of London, Feb. 15, 1922. The Measurement of Atmospheric Humidity, by Sir Napier Shaw.

exposures, and a small exhaust blower was provided for circulating the air over the wet bulb. The whole apparatus was controlled electrically by a clock and was arranged to record the wet and dry bulb temperatures at any desired time interval. When the complete record roll had been exposed it was removed and upon development showed the thermometer readings from which the corresponding humidities could be found in the psychrometric tables. While this type of recorder would no doubt have enabled accurate information to be obtained, it had two inherent objections. These were, first, the bulkiness of the complete equipment which had to be placed at the location where the conditions were to be determined and,

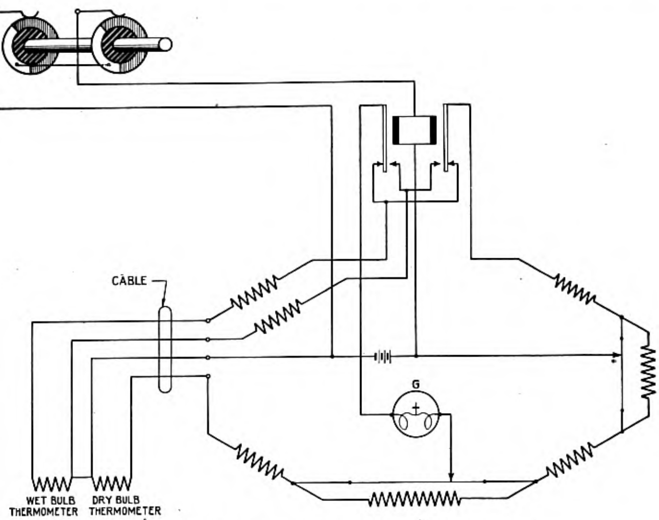


Fig. 1—Bridge Circuit of Difference Recorder

second, the thermometers could not be read because their stems were within the camera box, and therefore, the humidities and temperatures measured could not be ascertained until the record had been developed.

Accordingly, at this time, consideration was given to a type of mechanism which would produce a visible record upon a chart continuously available for observation by the operator. It was found that the Leeds & Northrup automatic recorder had been in commercial use for some time for the measurement of furnace temperatures, by means of thermocouples in conjunction with an automatically adjusted potentiometer circuit. The same type of recorder also had

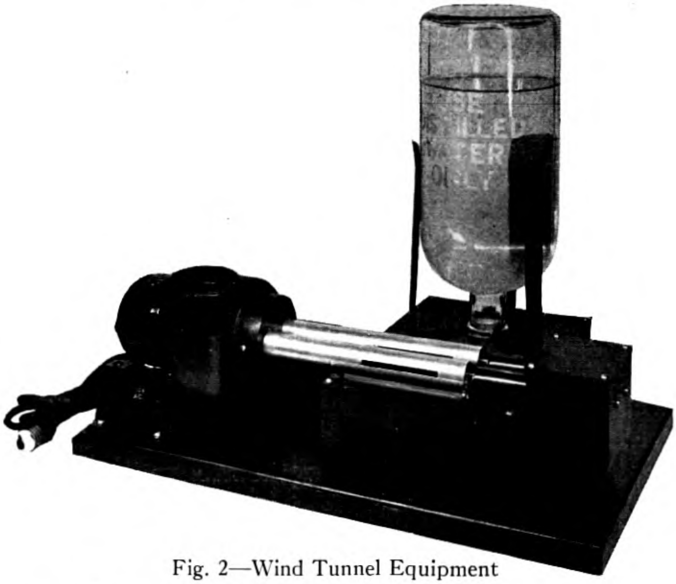


Fig. 2—Wind Tunnel Equipment

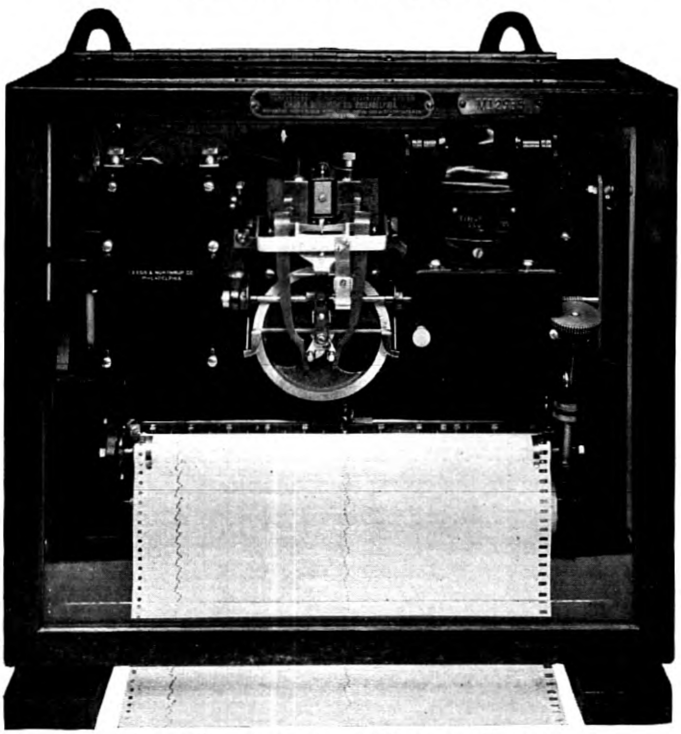


Fig. 3—Temperature and Difference Recorder

been used for recording temperatures and differences between two temperatures by means of resistance thermometers and a Wheatstone bridge arrangement. As it seemed feasible to adapt this instrument to meet our requirements, the double Wheatstone bridge circuit shown in Fig. 1 and the auxiliary wind tunnel equipment with resistance thermometers shown in Fig. 2 were developed. Fig. 3 is an illustration of the Leeds & Northrup recorder used.

This recorder was arranged to measure the resistance of the dry bulb thermometer and the difference between the resistances of the dry and wet thermometers, and to record these values upon a chart. Referring to the circuit diagram Fig. 1, it may be seen that, by means of a relay whose operation is controlled by the commutator on the recorder mechanism, the two Wheatstone bridges, one containing the dry bulb thermometer, and the other containing both the dry and wet bulb thermometers, may be balanced alternately by the recorder. After a sufficient interval has elapsed in each case for the bridge to become balanced the siphon pen is lowered into contact with the chart by a cam mechanism and the point of balance thus recorded. The record thus produced consists of dotted curves showing the successive indications of dry bulb temperature and difference between dry and wet bulb temperatures.

In order to secure the desired accuracy and sufficient sensitivity to follow the changes in temperature, the resistance thermometers used consist of platinum wire wound on mica cards and encased in flat nickel silver tubes with hard rubber ferrules. These are attached to a brass junction box in which is terminated the four conductor cable leading to the recorder mechanism.

The thermometers are enclosed in slotted brass tubes through which the air is drawn by a small blower driven by a universal motor. Mounted below these tubes is a shallow, covered water tank having a slot in the cover beneath the wet bulb thermometer through which the wick projects into the water. The desired water level in the tank is secured by an inverted water bottle, the neck of which projects into another opening in the cover of the tank.

The wind tunnel equipment[2] containing the resistance thermometers may be placed at any desired distance from the recorder mechanism, as the resistances of the thermometer leads have no effect upon the measurements provided they are equal. Leads consisting of a four conductor rubber insulated lead covered cable from 50 feet to 100 feet in length have been used.

[2]The wind tunnel and equipment is quite similar in operation to the "distance hygrometer," *Sci. Am.* June 6, 1914, p. 468.

As one of the difficulties encountered in the use of the wet bulb thermometer consists in the gradual clogging and drying up of the wick due to the accumulation of impurities left in it from the evaporation of the water, together with the dust which settles from the air which is drawn over it, special care must be taken to guard against trouble from this source. The cotton fabric used for the wicks which cover

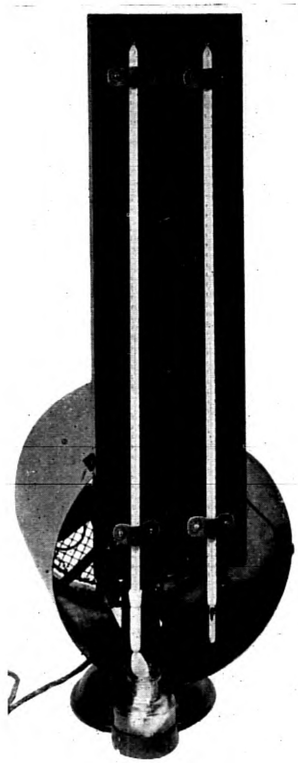


Fig. 4—Ventilated Psychrometer

the wet bulb must be treated to remove all traces of grease, with subsequent thorough washing to remove all traces of corrosive material. After this, the wicks should be handled only with thoroughly cleaned hands before they are placed on the thermometers. These wicks should be changed daily. Pure distilled water must be used in the tanks and they must be cleansed occasionally because they become contaminated by the impurities washed out of the air as it bubbles through the water. By rigid observance of such precautions no difficulty should be experienced in securing accurate records by means of this recorder.

Laboratory Tests

Several of these recorder mechanisms were built and after having been adjusted to operate satisfactorily, each wind tunnel equipment connected to its associated recorder was placed in a laboratory room controlled by air conditioning equipment, and given a run to test its operation under the range of conditions which might be expected to occur at the localities where the recorders were to be installed. During this test, the readings given by the recorder were compared with those obtained with a ventilated psychrometer, Fig. 4, equipped with accurate wet and dry bulb thermometers. Table I following gives a summary of the readings obtained in calibrating one of the recorders, while Fig. 5 shows a typical 12 hour record obtained in one of the laboratory rooms.

TABLE I

Ventilated Psychrometer			Leeds & Northrup Recorder			Per Cent Difference
Dry Bulb Temp. F°	Difference between Dry and Wet Bulb, Temp. F°	Relative Humidity Per Cent	Dry Bulb Temp. F°	Difference between Dry and Wet Bulb, Temp. F°	Relative Humidity Per Cent	
77.9	11.5	54.0	77.8	11.1	55.5	+2.8
77.2	10.2	58.0	77.0	9.9	59.5	+2.6
78.4	0.9	96.5	78.2	0.8	97.0	+0.5
84.4	0.7	97.0	84.4	0.6	97.5	+0.5
83.5	5.6	77.5	83.6	5.6	77.5	0.0
83.7	10.8	59.5	83.4	10.5	60.5	+0.1
98.3	10.5	65.5	98.2	10.5	65.5	0.0
98.3	13.4	57.0	98.3	13.4	57.0	0.0
97.4	1.3	95.0	97.8	1.5	94.5	−0.5
97.2	1.0	96.0	97.6	1.2	95.5	−0.5

Reference to these tabulated values of relative humidities obtained by the two methods indicates that the recorder is capable of giving reliable data particularly through the range of high humidities where the effects on materials or apparatus exposed to these conditions may be large. Difficulty was experienced in comparing the readings of the two instruments due to the sensitivity of the resistance thermometers to slight temperature changes, and also due to the slight differences in temperature between the two sets of thermometers which necessarily occurred because they were not in the same wind tunnel. This difficulty was encountered particularly when the "humidity

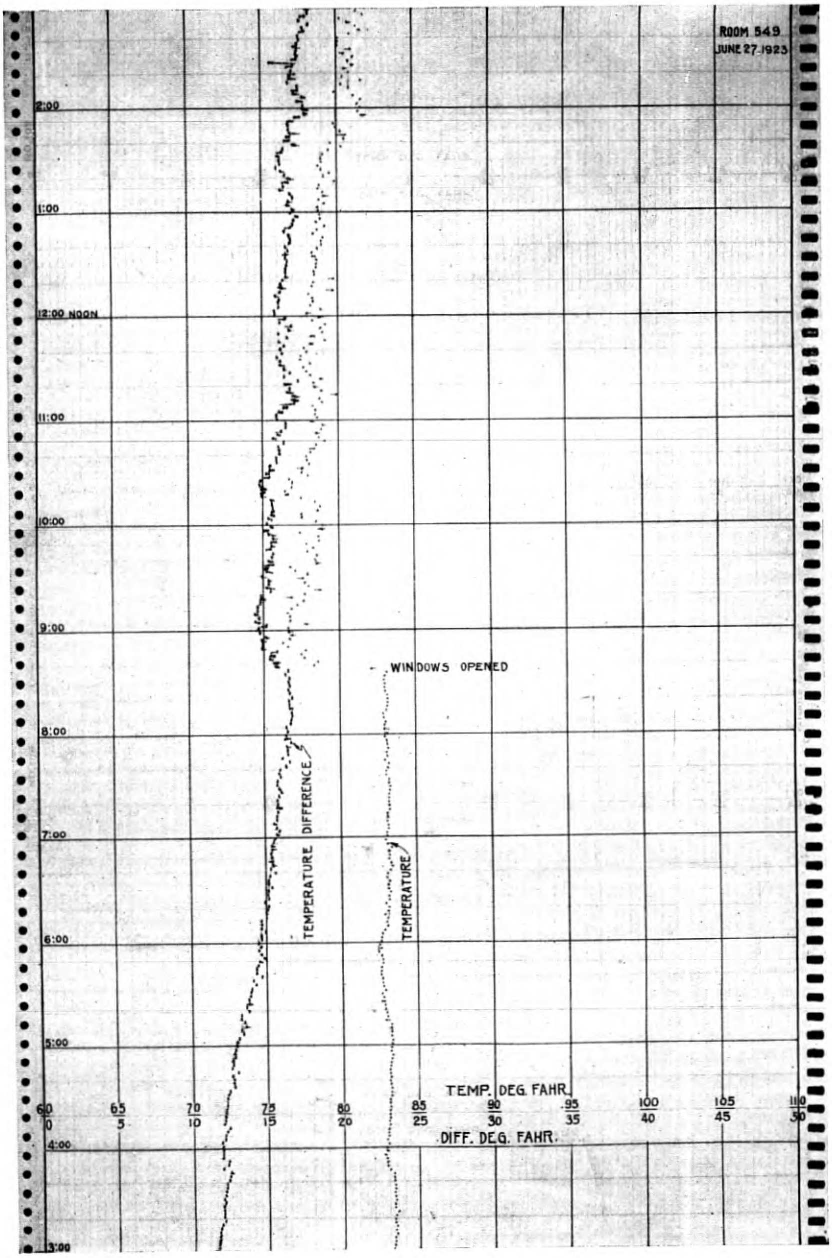


Fig. 5—Record from Temperature and Difference Recorder

room," in which the apparatus was located, was under a thermostatic control which allowed a temperature variation of approximately $\pm 0.5°$ F. However, the calibration of the resistance thermometers and the sensitivity of the bridges in which they are placed is such that temperatures and temperature differences are recorded with an accuracy of $\pm \frac{1}{4}°$ F.

Field Trials

In order to determine just what combinations of temperature and relative humidity prevail in widely separated localities of the United States, certain cities were selected in which moisture troubles with

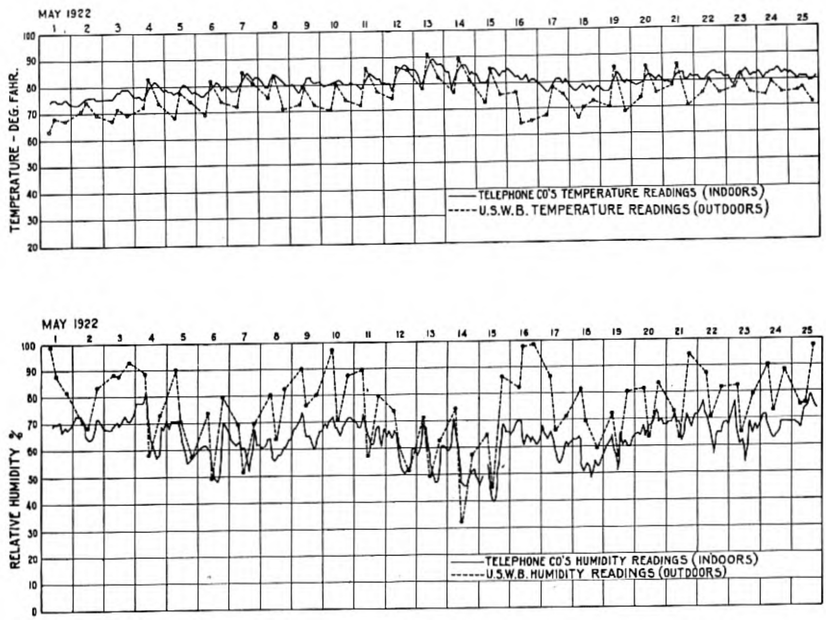


Fig. 6—Comparison of Indoor and Outdoor Temperatures and Relative Humidities at Savannah, Ga.

telephone equipment might be expected to occur, and at which local stations of the United States Weather Bureau were located, so that comparisons might be made between our records of indoor conditions and the observations of outdoor conditions.

Ten of these instruments were installed in central offices in New York (3), Boston, Savannah, New Orleans, Chicago, Minneapolis, Houston and Seattle, from which records have been obtained during the summer months of 1921 and 1922.

From the data accumulated in these cities, comprehensive information has been obtained as to the duration of conditions of average and maximum severity which occur during the humid months. It is of interest to compare the values of the central office conditions of temperature and relative humidity obtained from the recorders, with the corresponding Weather Bureau observations. The curves given in Fig. 6 show a typical comparison from data obtained at Savannah, Ga., during May, 1922. Study of these curves shows that the indoor temperature averaged somewhat higher than that out of doors, and that the indoor relative humidities were seldom higher than 75%, although the outdoor humidities often were higher than 85% for considerable lengths of time. The Weather Bureau data indicate very definitely when rain storms occurred and also periods of high humidity, due perhaps to foggy weather, although such periods are not well defined by the humidity curves showing the indoor conditions.

Since for a given absolute humidity, the relative humidity varies inversely with the change in temperature of the air, obviously it should be possible to keep the relative humidity in a central office building lower than that of the outside air by keeping the windows closed during periods of sudden temperature changes, and by the use of heat in switchboard sections. This latter remedy for humidity troubles has been successfully applied for several years to switchboards installed in some localities. Also the effects upon the indoor humidity and upon the performance of central office equipment, of closing the windows of central office rooms has been the subject of considerable investigation.

In the study of this method of reducing relative humidity, it is very desirable to have records which will show continuously the differences existing between indoor and outdoor temperatures and relative humidities, and in particular to study the effects on the indoor conditions when sudden changes in atmospheric conditions occur such as rain storms when the relative humidity outside reaches 100%. It was found that the automatic recorder described above would lend itself admirably to the study of this problem and that by the use of a simple relay switching mechanism on the recorder, two wind tunnel equipments could be operated with one recorder, enabling temperatures and differences between dry and wet bulb temperatures to be recorded alternately on the same chart for both indoor and outdoor conditions.

A recorder of this type was operated during the summer months of 1921 at the West Street laboratories of the Western Electric Co., Inc., to record the conditions in a well ventilated laboratory room

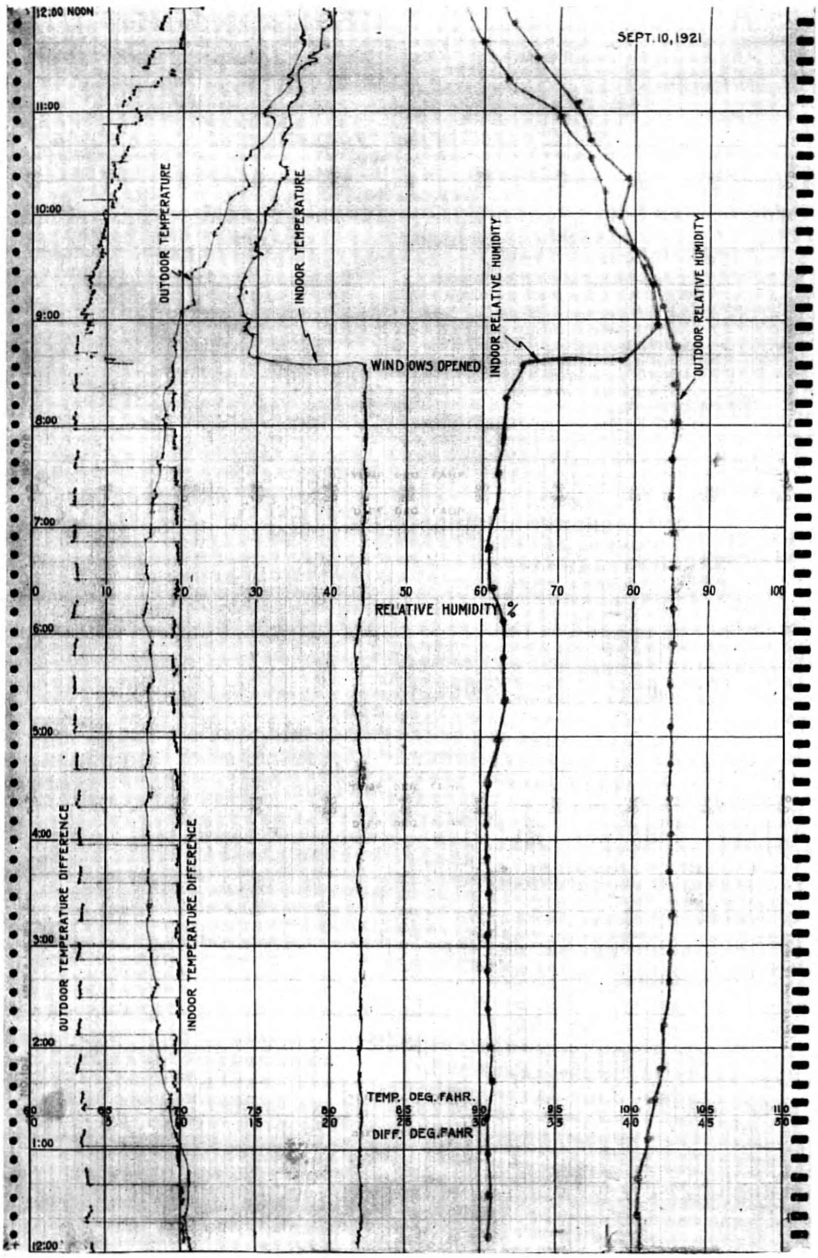


Fig. 7—Record from Double Recorder Comparing Indoor and Outdoor Conditions

about 25 feet x 27 feet and having two windows each in the east and south walls. The wind tunnel equipment was installed at a height of six feet upon a pillar in the center of the room. About ten people normally work in this room. The outdoor conditions were obtained by mounting a wind tunnel equipment in a standard Weather Bureau instrument shelter placed at the top of a tower, 14 feet high, which stands on the roof of a three story building far enough away from walls and other obstacles to permit free circulation of the air.

Figs. 7 and 8 show two typical 12 hour records upon which the indoor and outdoor relative humidities have been plotted from the curves of temperatures and temperature differences recorded by the instrument. A study of these records indicates that large differences often exist between the indoor and outdoor conditions and that the indoor conditions are much less severe than might be expected when the outdoor humidity is high. This difference is particularly noticeable when the windows are closed, but as soon as they are opened the indoor temperature decreases and the humidity generally increases to practically the same value as that of the outside air. Fig. 8 is of particular interest in showing the rapid decrease in the outdoor temperature and increase in relative humidity due to a thunderstorm.

The analysis of the records obtained from a number of recorders which record temperature and difference between dry and wet bulb temperature requires considerable labor in obtaining the corresponding relative humidities from the psychrometric tables and, obviously, periodic values only can be taken unless some rapid mechanical method of doing this is employed. Such methods have been developed and used successfully for this purpose.

A New Direct Reading Humidity Recorder

A much more satisfactory type of recorder is one which, in addition to tracing the temperature curve, traces a curve of the relative humidity. The only instrument of any prominence that has been used in this way is the recording hair hygrometer, the objectionable features of which have already been mentioned.

An improved type of direct reading humidity recorder which has been developed by E. B. Wood, of the Laboratories of the American Telephone and Telegraph Company and the Western Electric Company, employs the Leeds & Northrup automatic recorder mechanism, to which has been added an electrical mechanism which will be described, together with the principle upon which its operation is based.

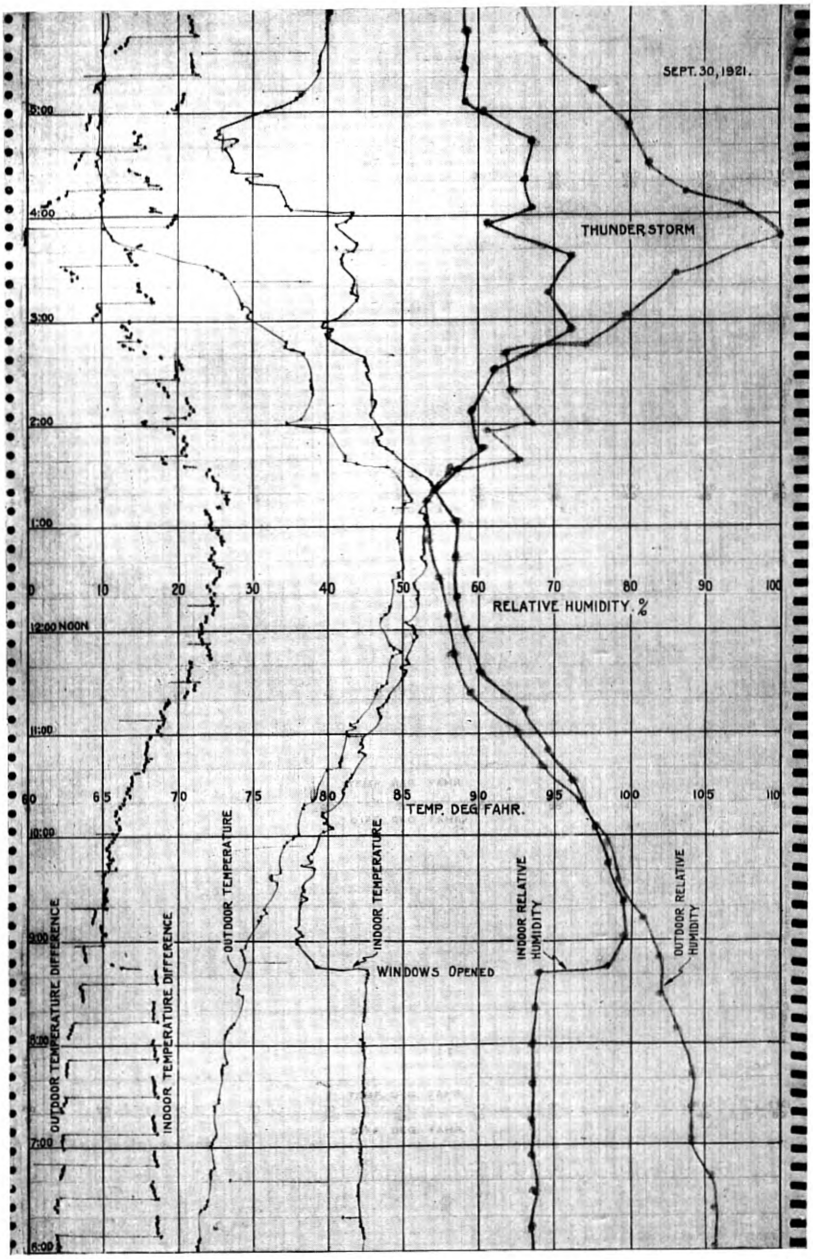


Fig. 8—Record from Double Recorder Comparing Indoor and Outdoor Conditions

This novel improvement depends, for its operation, on the approximate linearity and common intersection of the ordinary humidity curves as shown in Fig. 9.[3]

It is apparent that each of the humidity curves is in effect a straight line and that, with an accuracy sufficient for practical purposes, these curves, representing humidities of from 30% to 100%, converge at a point (a) whose coordinates are (b, c). Assuming that the humidity

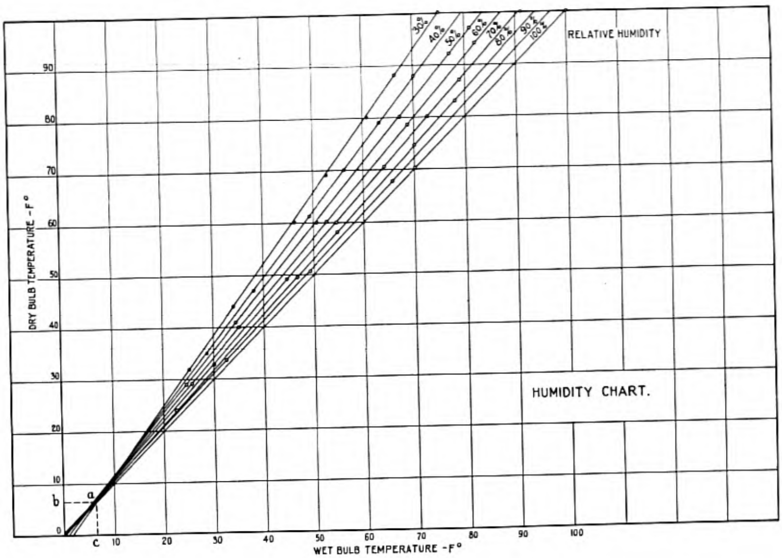


Fig. 9

curves are straight lines passing through point (a), it is apparent that the value of humidity is completely determined if the slope of the particular curve is known, since each curve represents only one value of humidity. It also is apparent that the slope is given by the ratio of dry bulb temperature minus the ordinate of point (a), to wet bulb temperatures minus the abscissa of point (a); or in other words, the relative humidity is completely determined, if the dry bulb and wet bulb temperatures are each known, above the datum coordinates (b, c) of point (a).

If then, a resistance is set off, proportional to the difference between the temperature of the dry bulb and temperature (b), and another resistance is set off proportional to the difference between the temperature of the wet bulb and temperature (c), the ratio between

[3] Bur. Stands. Cir. No. 55, p. 116.

these two resistances will indicate directly the relative humidity corresponding to the dry and wet bulb temperatures. The circuit arrangement by means of which this is accomplished is shown in Fig. 10, and the mechanism of the recorder employing it, is shown in Fig. 11.

The recorder circuit contains three Wheatstone bridges with one battery and galvanometer which are transferred in rotation from

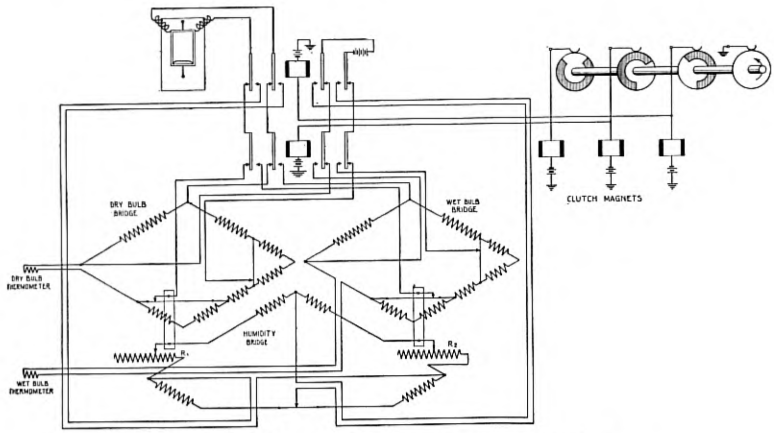


Fig. 10—Circuit of Direct Reading Recorder

each bridge to the next by the commutator and relays shown in the circuit. The three bridges are arranged so that they remain at their last positions of balance until mechanically connected to the balancing mechanism of the recorder by the electric clutch associated with each bridge whose operation also is controlled by the commutator. The first of these bridges, designated the "dry bulb bridge," contains the dry resistance thermometer and mechanically associated with its slide wire contact is a second slide wire contact operating upon a slide wire resistance arm in the third bridge, designated as the "humidity bridge." The second of these bridges, designated as the "wet bulb bridge," contains the wet resistance thermometer, and mechanically associated with its slide wire contact is a second slide wire contact operating upon a second slide wire resistance arm of the "humidity bridge."

The consecutive balancing of the "dry bulb bridge" and "wet bulb bridge" accordingly sets off resistances upon the two slide wire resistance arms of the "humidity bridge" proportional respectively to the temperature differences described in the second preceding paragraph.

The balancing of this bridge accordingly accomplishes the result already described of determining the ratio of the resistances R_1 and R_2 of these two slide wire arms, and consequently, the relative humidity corresponding to the dry and wet bulb temperatures previously

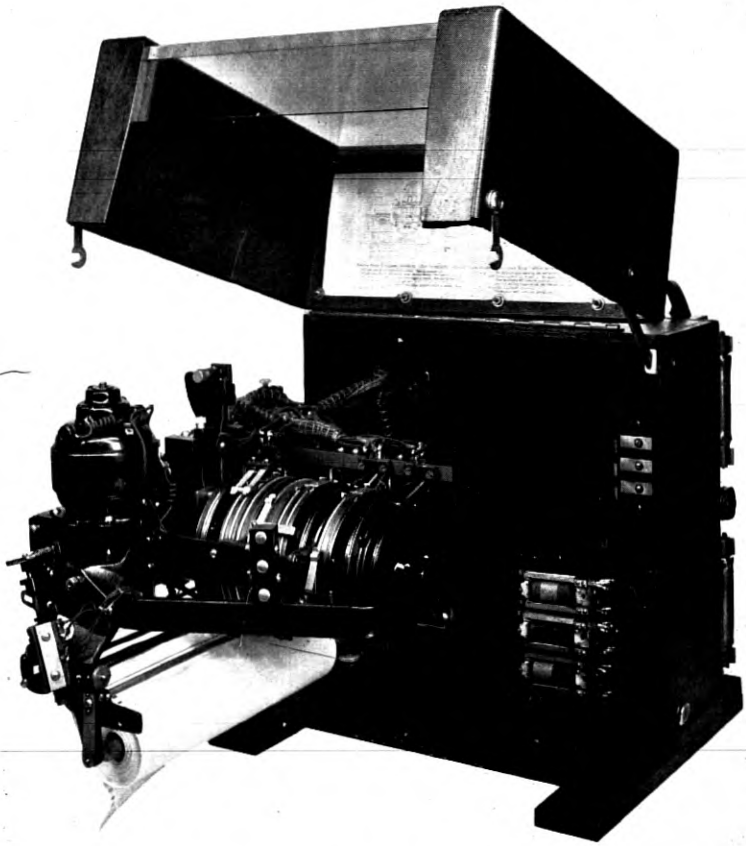


Fig. 11—Direct Reading Recorder

measured on their corresponding bridges. In the operation of the recorder, a period of about 20 seconds is allowed by the commutator to balance each bridge thus completing a cycle every 60 seconds.

The recorder is equipped with two pens one of which is associated with the slide wire of the "dry bulb bridge" thus recording the dry bulb temperature, while the other pen is associated with the "humidity

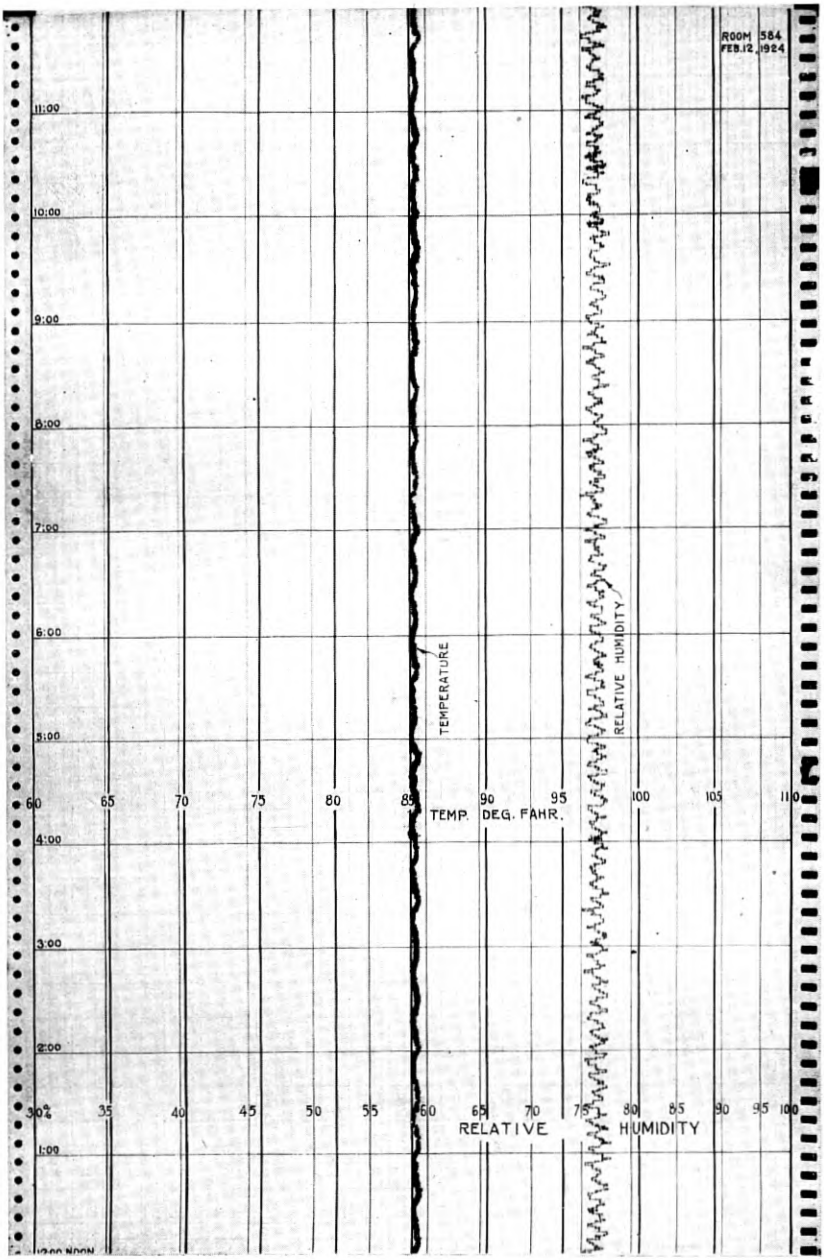


Fig. 12—Temperature and Relative Humidity in a Humidity Room

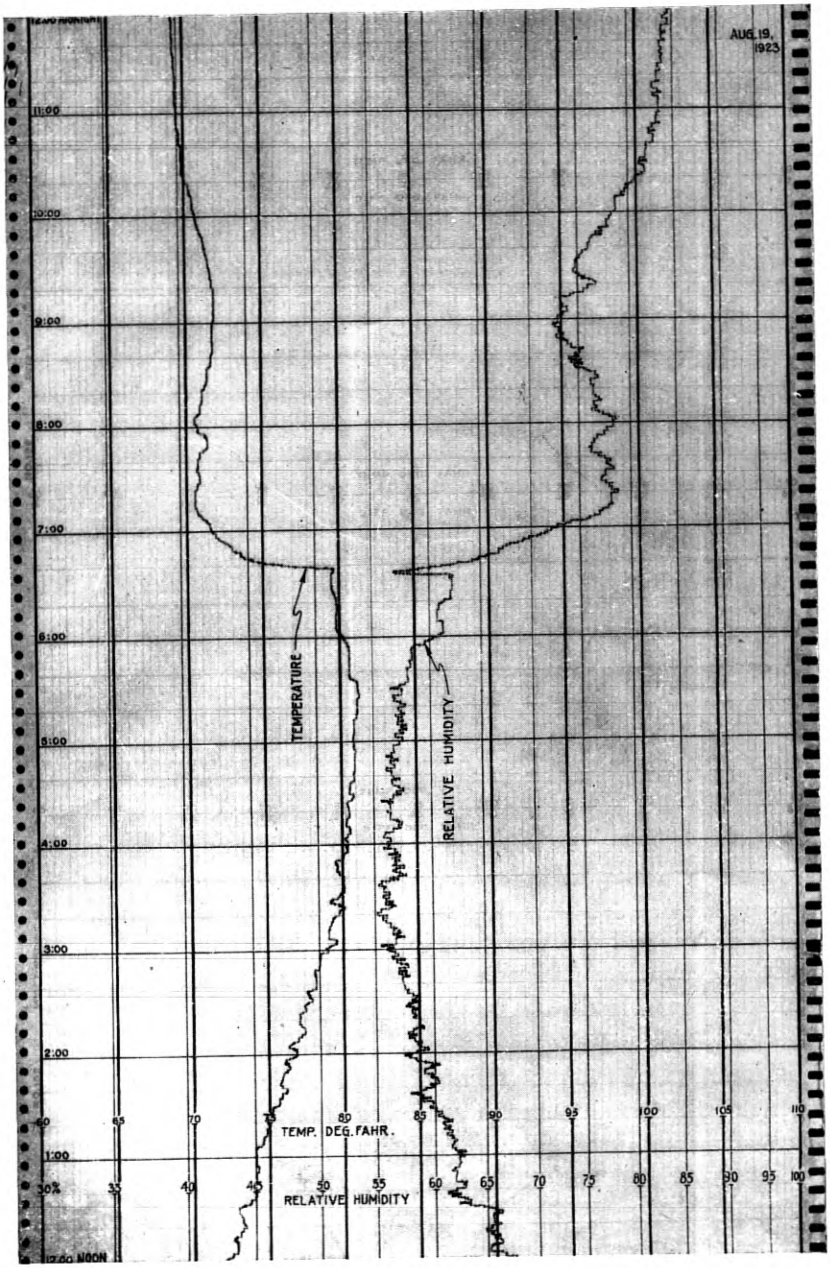


Fig. 13—Outdoor Temperature and Relative Humidity

bridge," thus recording the values of humidity directly. Inasmuch as successive operations of the recorder consist in the restoration of the balance of each bridge, if different from the last position of balance, it is evident that the pens will trace continuously the variations of temperature and relative humidity.

A recorder of this type with its associated wind tunnel mechanism has been used for some time to record the conditions in a laboratory "humidity room." The temperature record given by this recorder is accurate to $\pm\frac{1}{4}°$ F. as in the case of the difference recorder. The accuracy of the humidity record differs for various points on the scale, depending upon the values chosen for certain resistances in the recorder. When the recorder is adjusted for very close accuracy ($\pm\frac{1}{2}\%$ relative humidity) for relative humidities above 90%, the accuracy for lower values of humidity decreases until at 50% the maximum variation from the true value may be as much as $2\frac{1}{2}\%$ relative humidity. If desired, the adjustment may be made to transfer the point of greatest accuracy to any selected lower value of humidity. Experience with this model has suggested changes which should considerably improve this accuracy over the whole range of humidities. Fig. 12 shows a typical 12 hour record of conditions in the "humidity room" while under automatic control of an air conditioning equipment.

This recorder also was used during the summer months of 1923 to record outdoor conditions with the wind tunnel equipment installed in the Weather Bureau instrument shelter mentioned earlier. During this period of 4 months' operation, it required no attention save an occasional oiling of the mechanism and maintenance of the wet bulb equipment, and practically continuous records were secured. The records are of particular interest for observation of the variations of temperature and humidity which take place during changes in weather conditions such as rain storms. Figs. 13 and 14 are reproductions of typical consecutive 12 hour records obtained for outdoor conditions.

From consideration of the humidity recording apparatus which has been developed and the results which have been obtained with it, it may be stated that both the difference recorder and the direct reading recorder are satisfactory instruments with which accurate data may be obtained. However, they are instruments which, in common with other types of apparatus that have been developed to measure humidity, require careful attention of the wind tunnel equipment in order to secure reliable results; also the recorder mechanism itself requires the attention of an operator skilled in its maintenance.

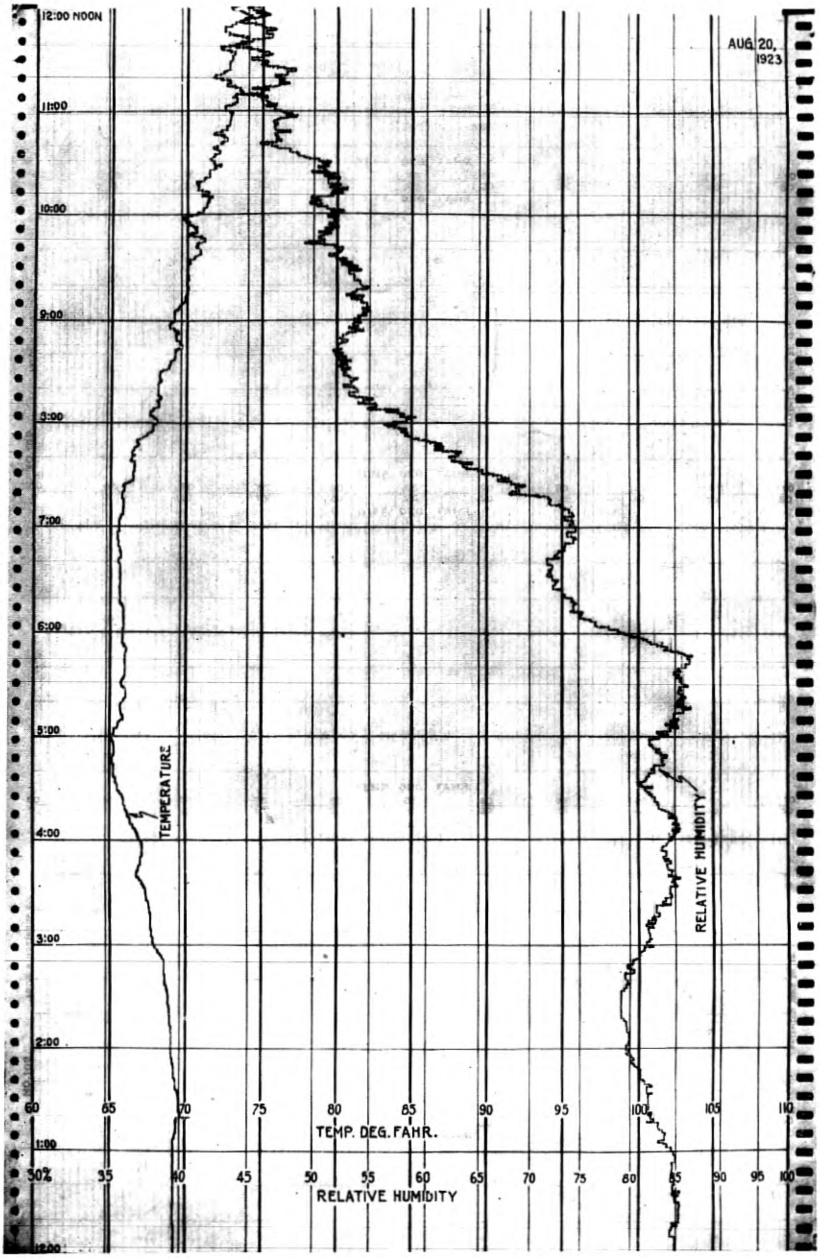


Fig. 14—Outdoor Temperature and Relative Humidity

While the mechanism of the direct reading recorder is more complicated than that of the difference recorder, it is a more useful instrument, both because the humidities may be read directly, thus saving the labor of interpretation of the records, and because the records are more significant. The direct reading recorder, furthermore, may be used to control the functioning of air conditioning apparatus at any desired conditions, at the same time that it is actually recording these conditions. Accordingly, it should prove particularly useful in maintaining proper humidity in apparatus and operating rooms.

A Reactance Theorem

By RONALD M. FOSTER

SYNOPSIS: The theorem gives the most general form of the driving-point impedance of any network composed of a finite number of self-inductances, mutual inductances, and capacities. This impedance is a pure reactance with a number of resonant and anti-resonant frequencies which alternate with each other. Any such impedance may be physically realized (provided resistances can be made negligibly small) by a network consisting of a number of simple resonant circuits (inductance and capacity in series) in parallel or a number of simple anti-resonant circuits (inductance and capacity in parallel) in series. Formulas are given for the design of such networks. The variation of the reactance with frequency for several simple circuits is shown by curves. The proof of the theorem is based upon the solution of the analogous dynamical problem of the small oscillations of a system about a position of equilibrium with no frictional forces acting.

AN important theorem[1] gives the driving-point impedance[2] of any network composed of a finite number of self-inductances, mutual inductances, and capacities; showing that it is a pure reactance with a number of resonant and anti-resonant frequencies which alternate with each other; and also showing how any such impedance may be physically realized by either a simple parallel-series or a simple series-parallel network of inductances and capacities, provided resistances can be made negligibly small. The object of this note is to give a full statement of the theorem, a brief discussion of its physical significance and its applications, and a mathematical proof.

THE THEOREM

The most general driving-point impedance S obtainable by means of a finite resistanceless network is a pure reactance which is an odd rational function of the frequency $p/2\pi$ and which is completely determined, except for a constant factor H, by assigning the resonant and anti-resonant frequencies, subject to the condition that they alternate and include both zero and infinity. Any such impedance may be physically

[1] The theorem was first stated, in an equivalent form and without his proof, by George A. Campbell, *Bell System Technical Journal*, November, 1922, pages 23, 26, and 30. By an oversight the theorem on page 26 was made to include unrestricted dissipation. Certain limitations, which are now being investigated, are necessary in the general case of dissipation. The theorem is correct as it stands when there is no dissipation, that is, when all the R's and G's vanish; this is the only case which is considered in the present paper.

A corollary of the theorem is the mutual equivalence of simple resonant components in parallel and simple anti-resonant components in series. This corollary had been previously and independently discovered by Otto J. Zobel as early as 1919, and was subsequently published by him, together with other reactance theorems, *Bell System Technical Journal*, January, 1923, pages 5–9.

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

constructed either by combining, in parallel, resonant circuits having impedances of the form $iLp+(iCp)^{-1}$, *or by combining, in series, anti-resonant circuits having impedances of the form* $[iCp+(iLp)^{-1}]^{-1}$. *In more precise form,*

$$S=-iH\frac{(p_1^2-p^2)\,(p_3^2-p^2)\,\ldots\,(p_{2n-1}^2-p^2)}{p(p_2^2-p^2)\,\ldots\,(p_{2n-2}^2-p^2)},\tag{1}$$

where $H\geq 0$ *and* $0=p_0\leq p_1\leq p_2\leq\ \ldots\ \leq p_{2n-1}\leq p_{2n}=\infty$.[3] *The inductances and capacities for the n resonant circuits are given by the formula,*

$$L_j=\frac{1}{C_jp_j^2}=\left(\frac{ipS}{p_j^2-p^2}\right)_{p=p_j}\quad(j=1,3,\ldots,2n-1),\tag{2}$$

and the inductances and capacities of the $n+1$ *anti-resonant circuits are given by the formula,*

$$C_j=\frac{1}{L_jp_j^2}=\left(\frac{ip}{S(p_j^2-p^2)}\right)_{p=p}\quad(j=0,2,4,\ldots,2n-2,2n),\tag{3}$$

which includes the limiting values,

$$C_0=\frac{p_2^2\,\ldots\,p_{2n-2}^2}{Hp_1^2p_3^2\,\ldots\,p_{2n-1}^2},\ L_0=\infty,\ C_{2n}=0,\ L_{2n}=H.$$

Formula (1) may be stated in several mutually equivalent forms.[4] This particular form is the driving-point impedance of the most general symmetrical network in which every branch contains an inductance and a capacity in series, with mutual inductance between each pair of branches. This includes as special cases the driving-point impedances of every other finite resistanceless network.

[3] Since the impedance S is an odd function of the frequency, resonance or anti-resonance for $p=P$ implies resonance or anti-resonance for $p=-P$. In enumerating the resonant and anti-resonant frequencies it is customary, however, to exclude negative values of the frequency. Thus, in the present case, we say that there are n resonant points (p_1, p_3, . . . , p_{2n-1}) and $n+1$ anti-resonant points ($p_0=0$, p_2, p_4, . . . , p_{2n-2}, $p_{2n}=\infty$).

[4] The expression for S given by formula (1) may be written in the mutually equivalent forms,

$$\left[-iH\frac{(p_1^2-p^2)\,(p_3^2-p^2)\,\ldots\,(p_{2n-1}^2-p^2)}{p\,(p_2^2-p^2)\,\ldots\,(p_{2n-2}^2-p^2)}\right]^{\pm1}\ \text{and}\ \left[iHp\frac{(p_2^2-p^2)\,\ldots\,(p_{2n-2}^2-p^2)}{(p_1^2-p^2)\,\ldots\,(p_{2n-3}^2-p^2)}\right]^{\pm1}$$

If the constant H and all the p_j's of these formulas are restricted to finite values greater than zero, the four cases, obtained by separating the plus and minus exponents, are mutually exclusive, but together they cover the entire field. If p_1 is allowed to be zero, either the first or the second pair covers the entire field. Finally, if in addition p_{2n-1} or p_{2n-2} is allowed to become infinite, while Hp_{2n-1}^2 or Hp_{2n-2}^2 is maintained finite, any one of the four expressions covers the entire field. Sometimes one, sometimes another way of covering the field is the more convenient. Formulas (2) and (3) apply to all of these expressions for S provided the p_j's include all the resonant points and all the anti-resonant points, respectively.

Physical Discussion

The variation of the reactance $X = S/i$ with frequency is illustrated by the curves of Fig. 1 in all the typical cases of formula (1) for $n=1$ and for $n=2$. For every curve the reactance increases with the frequency,[5] except for the discontinuities which carry it back from a positive infinite value to a negative infinite value at the anti-resonant points. Thus between every two resonant frequencies there is an anti-resonant frequency, no matter how close together the two resonant frequencies may be. The effect of increasing n by one unit is to add one resonant point, and thus to introduce one additional branch to the reactance curve, this branch increasing from a negative infinite value through zero to a positive infinite value.

That formula (1) includes several familiar circuits is seen by considering the most general network with one mesh, that is, an inductance and a capacity in series, with the impedance $iLp+(iCp)^{-1}$. This expression is given immediately by (1) upon setting $n=1$, $H=L$, and $p_1 = 1/\sqrt{LC}$. Since L and C are both positive these constants satisfy the conditions stipulated under (1), thus verifying the theorem for circuits of one mesh. This general one-mesh circuit includes as special cases a single inductance L by setting $H=L$ and $p_1=0$, and a single capacity C by setting $H=0$ and $p_1=\infty$ such that $Hp_1^2=1/C$.

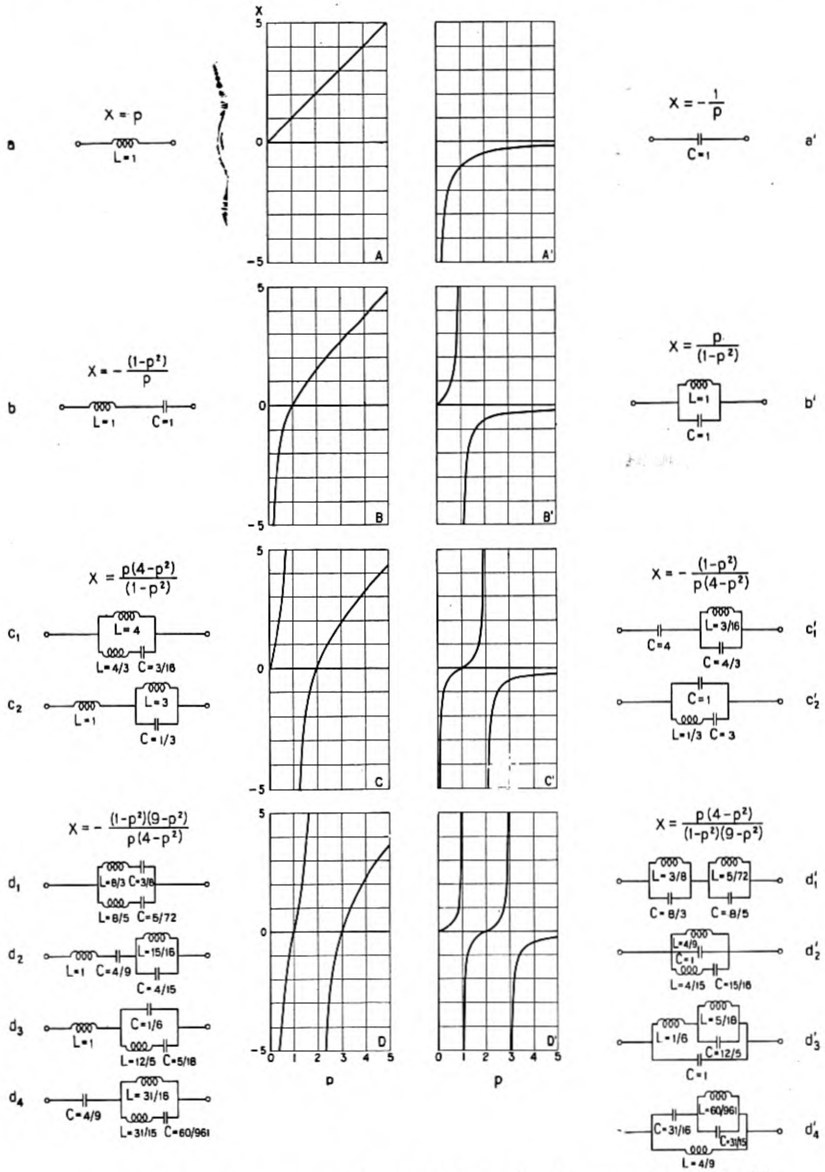
In Fig. 1 the reactances shown by the curves on the right are the negative reciprocals of those on the left. Fig. 1 also shows networks which give the several reactance curves, the networks being computed by means of formulas (2) and (3). The networks are arranged in pairs with reciprocal driving-point impedances and with the networks themselves reciprocally related, that is, the geometrical forms of the networks are conjugate,[6] and inductances correspond to capacities of the same numerical value and vice versa. This relation is a natural consequence of the reciprocal relation between an inductance and a capacity of the same numerical value, these being the elements from which the networks are constructed.

For $n=1$, formulas (2) and (3) give identical networks, as illustrated by the reactances A, B, A', and B' of Fig. 1, each of which is realized by a single network. For the reactances C and C' the two formulas give distinct networks, c_1 and c_2, c_1' and c_2', respectively, these

[5] This has been proved by Otto J. Zobel (loc. cit., pp. 5, 36), using the formula for the most general driving-point impedance given by George A. Campbell (loc. cit., p. 30).

[6] For a further treatment of conjugate or inverse networks, see P. A. MacMahon, *Electrician*, April 8, 1892, pages 601, 602, and Otto J. Zobel, loc. cit., pages 5, 36, and 37.

two being the only networks with the minimum number of elements which give the specified impedance. In general, however, there are four ways of realizing a given impedance when $n=2$, as illustrated by D and D' of Fig. 1; formulas (2) and (3) give only the first two



1—Reactance curves and networks for simple cases of formula (1).

networks, d_1 and d_2, d_1' and d_2', respectively. The total number of possible ways of realizing a given impedance increases very rapidly for values of n greater than 2; for $n=3$, there are, in general, 32 distinct networks giving a specified impedance.

Formulas (2) and (3) are to be used for determining the constants of the circuits which have certain specified characteristics, whereas most network formulas are for the determination of the characteristics of the circuit from the given constants of the circuit. The application of these formulas is illustrated by the following numerical problem:

To design a reactance network which shall be resonant at frequencies of 1000, 3000, 5000, and 7000 cycles; anti-resonant at 2000, 4000, and 6000 cycles, as well as at zero and infinite frequencies; and have a reactance of 2500 ohms at a frequency of 10,000 cycles.

By formula (1) the reactance of such a network must be

$$X=-H\frac{(p_1^2-p^2)\,(p_3^2-p^2)\,(p_5^2-p^2)\,(p_7^2-p^2)}{p\ (p_2^2-p^2)\,(p_4^2-p^2)\,(p_6^2-p^2)}, \tag{4}$$

where p_1, p_3, p_5, and p_7 are determined by the resonant frequencies to be $1000\times 2\pi$, $3000\times 2\pi$, $5000\times 2\pi$, and $7000\times 2\pi$, respectively; p_2, p_4, and p_6 are determined by the anti-resonant frequencies to be $2000\times 2\pi$, $4000\times 2\pi$, and $6000\times 2\pi$, respectively; and H must be made equal to 0.0596 in order that the reactance at $p=10{,}000\times 2\pi$ may be 2500. The variation of the reactance with the frequency is shown by the curve of Fig. 2.

A network having this reactance may be constructed by combining $n=4$ simple resonant circuits in parallel, or $n+1=5$ simple anti-resonant circuits in series. These two networks are shown by Fig. 2. The numerical values of the elements are determined as follows: Applying formula (2) we have

$$L_1=\frac{1}{C_1p_1^2}=H\frac{(p_3^2-p_1^2)\,(p_5^2-p_1^2)\,(p_7^2-p_1^2)}{(p_2^2-p_1^2)\,(p_4^2-p_1^2)\,(p_6^2-p_1^2)}=0.349,$$

$$L_3=\frac{1}{C_3p_3^2}=H\frac{(p_1^2-p_3^2)\,(p_5^2-p_3^2)\,(p_7^2-p_3^2)}{(p_2^2-p_3^2)\,(p_4^2-p_3^2)\,(p_6^2-p_3^2)}=0.323,$$

$$L_5=\frac{1}{C_5p_5^2}=H\frac{(p_1^2-p_5^2)\,(p_3^2-p_5^2)\,(p_7^2-p_5^2)}{(p_2^2-p_5^2)\,(p_4^2-p_5^2)\,(p_6^2-p_5^2)}=0.264,$$

$$L_7=\frac{1}{C_7p_7^2}=H\frac{(p_1^2-p_7^2)\,(p_3^2-p_7^2)\,(p_5^2-p_7^2)}{(p_2^2-p_7^2)\,(p_4^2-p_7^2)\,(p_6^2-p_7^2)}=0.142;$$

and applying formula (3) we have

$$C_0 = \frac{p_2^2 p_4^2 p_6^2}{H p_1^2 p_3^2 p_5^2 p_7^2} = 0.0888 \times 10^{-6},\ L_0 = \infty,$$

$$C_2 = \frac{1}{L_2 p_2^2} = \frac{-p_2^2(p_4^2-p_2^2)(p_6^2-p_2^2)}{H(p_1^2-p_2^2)(p_3^2-p_2^2)(p_5^2-p_2^2)(p_7^2-p_2^2)} = 0.0461 \times 10^{-6},$$

$$C_4 = \frac{1}{L_4 p_4^2} = \frac{-p_4^2(p_2^2-p_4^2)(p_6^2-p_4^2)}{H(p_1^2-p_4^2)(p_3^2-p_4^2)(p_5^2-p_4^2)(p_7^2-p_4^2)} = 0.0523 \times 10^{-6},$$

$$C_6 = \frac{1}{L_6 p_6^2} = \frac{-p_6^2(p_2^2-p_6^2)(p_4^2-p_6^2)}{H(p_1^2-p_6^2)(p_3^2-p_6^2)(p_5^2-p_6^2)(p_7^2-p_6^2)} = 0.0725 \times 10^{-6},$$

$$C_8 = 0, \qquad L_8 = H = 0.0596.$$

These formulas give the numerical values of the inductances in henries and the capacities in farads. The entire set of numerical values is shown in Fig. 2. It is to be noted that the anti-resonant circuit corresponding to $p_0 = 0$ consists of a simple capacity since the inductance is infinite and thus does not appear in the network, whereas for $p_8 = \infty$ the anti-resonant circuit consists of a simple inductance, the capacity being zero and thus not appearing in the network.

Mathematical Proof

We shall first prove that the driving-point impedance S, as given by (1), may be physically realized by either a simple parallel-series or a simple series-parallel network of inductances and capacities, provided resistances can be made negligibly small.

The rational function $1/S$ can be expanded in partial fractions,

$$\frac{1}{S} = \frac{iH_1 p}{p_1^2 - p^2} + \frac{iH_3 p}{p_3^2 - p^2} + \dots + \frac{iH_{2n-1} p}{p_{2n-1}^2 - p^2},$$

where $$H_j = \left(\frac{p_j^2 - p^2}{ipS}\right)_{p=p_j} \quad (j = 1, 3, \dots, 2n-1).$$

Hence S is equal to the impedance of the parallel combination of the n circuits having the impedances $(p_j^2 - p^2)/(iH_j p) = iH_j^{-1} p + [i(H_j p_j^{-2})p]^{-1}$, that is, n simple resonant circuits in parallel, each circuit consisting of an inductance and a capacity in series, with the numerical values given by (2). Furthermore, these numerical values of the inductances and capacities given by (2) are all positive, an even number of negative factors being obtained upon substituting $p = p_j$, since in every case $p_j \leq p_{j+1}$. Hence the network defined by (2) has the impedance S as given by (1) and is physically realizable.

Likewise, by expanding S in partial fractions, it can be shown that the network defined by (3) has the impedance S as given by (1) and is physically realizable.

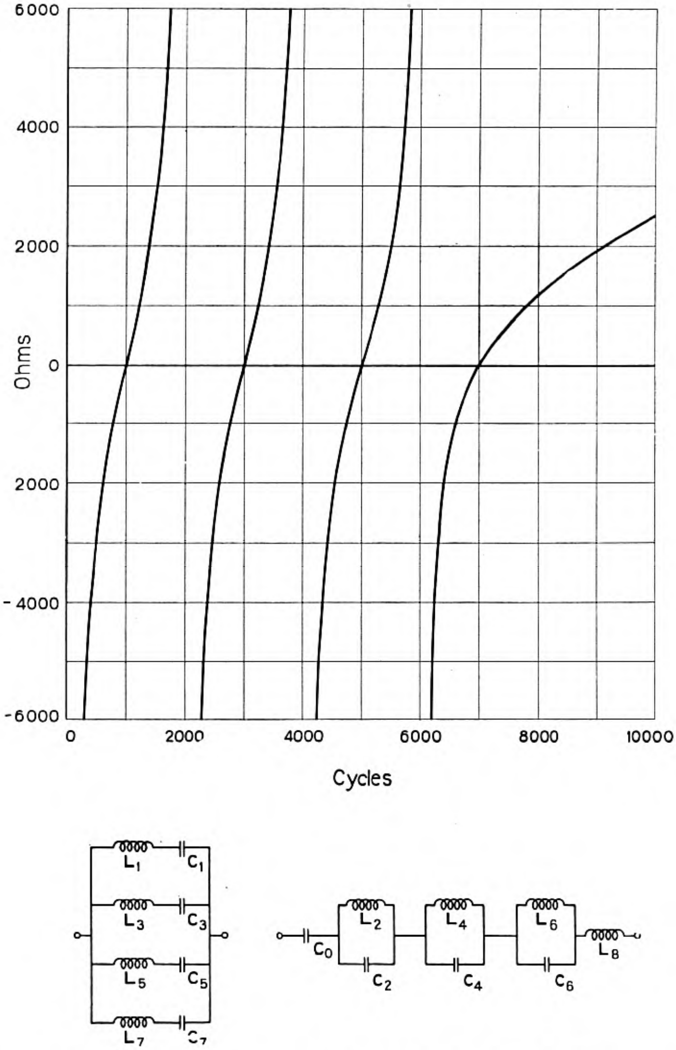


Fig. 2—Reactance curve and networks for formula (4).

The values of the inductances and capacities are (in henries and microfarads):

$L_1=0.349$	$C_1=0.0726$	$L_2=0.137$	$C_0=0.0888$
$L_3=0.323$	$C_3=0.00872$	$L_4=0.0302$	$C_2=0.0461$
$L_5=0.264$	$C_5=0.00384$	$L_6=0.00971$	$C_4=0.0523$
$L_7=0.142$	$C_7=0.00363$	$L_8=0.0596$	$C_6=0.0725$

The electrical problem of the free oscillations of a resistanceless network is formally the same as the dynamical problem of the small oscillations of a system about a position of equilibrium with no frictional forces acting. The proof of formula (1) may be derived from the treatment of this dynamical problem as given, for example, by Routh.[7]

In any network the driving-point impedance in the qth mesh, S_q, is equal to the ratio A/A_q, where A is the determinant[8] of the network and A_q the principal minor of this determinant obtained by striking out the qth row and the qth column. The determinant of a network has 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 network.

Hence the determinant A has the element $Z_{jk}=iL_{jk}p+(iC_{jk}p)^{-1}$, where L_{jk} is the total inductance and C_{jk} the total capacity common to the meshes j and k. Upon taking the factor $(ip)^{-1}$ from each row and substituting $-p^2=x$, the expression for A may be put in the form $A=(ip)^{-n}D$, where D is a determinant with $L_{jk}x+1/C_{jk}$ as the element in the jth row and the kth column. This is of exactly the same form as the determinant given by Routh[9] for the solution of the dynamical problem; it is proved there that this determinant, regarded as a polynomial, has n negative real roots which are separated by the $n-1$ negative real roots of every first principal minor of the determinant.

Hence, we may write $D=E(x_1+x)(x_3+x)\ldots(x_{2n-1}+x)$, where $x_1, x_3, \ldots, x_{2n-1}$ are all positive and arranged in increasing order of magnitude, and where E is also positive since D must be positive for $x=0$. The determinant D_q may be expressed in similar manner since it is of the same form as D but of lower order.

[7] E. J. Routh, "Advanced Rigid Dynamics," sixth edition, 1905, pages 44–55. In the notation of the dynamical problem as presented here, the coefficients A_{jk} correspond to the inductances, $1/C_{jk}$ to the capacities, $p/(i2\pi)$ to the frequency, and θ', ϕ', etc., to the branch currents in the electrical problem.

A complete proof of formula (1) has been worked out for the electrical problem, without depending in any way upon the solution of the corresponding dynamical problem. This proof has not been published here in view of the great simplification made by using the results already worked out for the dynamical problem.

[8] A complete discussion of the solution of networks by means of determinants has been given by G. A. Campbell, Transactions of the A. I. E. E., 30, 1911, pages 873–909.

[9] The determinant given by Routh (loc. cit., p. 49) has the element $A_{jk}p^2+C_{jk}$.

The driving-point impedance is given by

$$S_q = \frac{A}{A_q} = (ip)^{-1}\frac{D}{D_q} = (ip)^{-1}\frac{E(x_1+x)\,(x_3+x)\,\ldots\,(x_{2n-1}+x)}{E_q(x_2+x)\,\ldots\,(x_{2n-2}+x)},$$

where $0 \leq x_1 \leq x_2 \leq x_3 \leq \ldots \leq x_{2n-2} \leq x_{2n-1}$, since the roots of D are separated by the roots of D_q. Upon substituting $x = -p^2$ and introducing the notation $H = E/E_q$ and $p_1^2, p_2^2, \ldots, p_{2n-1}^2 = x_1, x_2, \ldots, x_{2n-1}$, respectively, we see that formula (1) is completely verified as the most general driving-point impedance obtainable by means of a finite resistanceless network.

Some Contemporary Advances in Physics—III

By KARL K. DARROW

ELECTROMAGNETIC waves of every frequency from 10^4 to 10^{20} exist; they can be generated and perceived; their frequencies in nearly every instance can be measured; their actions and reactions with matter can be studied. This brief statement is the synthesis of a great multitude of inventions, experiments and observations upon phenomena of extraordinary diversity and variety. When Herschel in 1800 carried a thermometer across the fan-shaped beam of colored light into which a sunbeam was resolved by a prism, and observed that the effect of the sunbeam on the mercury column did not cease when it passed beyond the red edge of the fan, he proved that the boundary of the spectrum beyond the red is imposed by the limitations of the eye and not by a deficiency of rays. Almost at the same time Ritter found that the power of the violet rays to affect salts of silver was shared by invisible rays beyond the violet edge of the beam. Maxwell developed the notion of electromagnetic waves from his theory of electricity and magnetism, and described some of the properties they should have; and the light-waves and the infra-red and ultra-violet rays were found to have some of these properties, while the outstanding discordances were explained away by Maxwell's successors. Hertz and many others built apparatus for producing Maxwell's waves with frequencies far below those of light, and apparatus for detecting them, with consequences known to everyone. Years after X-rays and gamma-rays were discovered emanating from discharge-tubes and disintegrating atoms, Laue proved that these too are waves, lying beyond the visible spectrum in the range of high frequencies. Radiations emerging from collapsing atoms and radiations diverging from wireless towers; waves conveying the solar heat and waves carrying the voice; rays which disrupt atoms by extracting their electrons, rays which alter atoms by rearranging their electrons, rays which almost ignore atoms altogether, were successively discovered or created; and all these radiations were brought into one class, and identified with light.

This enormously extended electromagnetic spectrum was interrupted until lately by two regions unexplored. They were known as the gap between the X-rays and the ultra-violet, and the gap between the infra-red and the Hertzian waves, according to the names by which the various explored regions of the spectrum commonly go; but to understand why they remained unclosed for so long, and what kinds of rays are being found within them, it is necessary to consider how certain properties of the waves vary along the spectrum. Enough is known about the origin of electromagnetic waves to justify using it as a basis of classification. Classifying the rays, therefore, by *mode of production*, we can distinguish at least four sharply-contrasted types: first, rays emitted from atomic nuclei in process of disintegration; second, rays emitted from atomic electron-systems in process of rearrangement; third, rays due to atoms vibrating to and fro about their positions of equilibrium as constituents of molecular groups or of space-lattices; and finally, waves generated by oscillating electrical circuits.[1] For each of these classes there is a region of the spectrum which is particularly, although not exclusively, its own.

The rays emitted from disintegrating nuclei lie at the topmost end of the frequency-scale; they overlap the rays of the second class, but do not approach either of the gaps. The rays resulting from rearrangements of the electron-systems surrounding atom-nuclei extend over an enormous range. The minimum wave-length of this range is .1075A, the K-frequency of the uranium atom; it is and will almost certainly remain the definitive limit, unless someone should succeed in discovering a substance further up the periodic table than uranium, or in removing some of the deepest electrons from the electron-system of some heavy[2] atom. As maximum wave-length we might take that of a line 40500A lately recognized by Brackett as belonging to atomic hydrogen; but this is certainly not the definitive limit. Emission-bands due to atoms vibrating within molecular groups are found in and beyond the "near infra-red" (and indeed in the ultra-violet around 3000A, if we include bands of "compound" origin, resulting from processes occurring together which if happening separately would produce rays of the second and third types, respectively); while the "residual rays," which are ascribed to atoms vibrating within the gigantic molecular group which is a crystal lattice, extend as far as 0.152 mm. (residual rays of thallium iodide). Between 0.1 mm. and 0.4 mm. rays have been discovered emanating from the mercury

[1] This classification is obviously not an exhaustive one. Continuous spectra have been omitted—thermal emission spectra of solids, and continuous X-ray spectra, which may be ascribed to random accelerations of free electrons. The continuous bands in gas spectra, of which one has just been explained by Gerlach (*ZS. f. Phys.*, 18, pp. 239–248; 1923) and others by Bohr (*Phil. Mag.* 26, p. 17; 1913), can be included in the second class by a slight generalization; and so, probably, can some fluorescence and phosphorescence spectra, at least if we extend "atomic electron-systems" to include "electron-systems of grouped atoms." There is also the possibility of rays due to changes in rate of rotation of molecules, not compounded with changes in oscillation or electron-arrangement.

[2] Meaning an atom with a large nuclear charge, which would have heaviness, or more properly massiveness, as a secondary characteristic. A short and simple adjective to describe where an atom stands in the scale of nuclear charge, i.e. in the periodic table, would be very welcome.

arc, which probably belong to the second or third class, but it is not certain which. If we gather all these classes together into a single great class of *natural* rays, extending from .02A or 2.10^{-10} cm. to 4,000,000A or 0.04 cm., they may be contrasted with the *artificial* rays generated by man-made electrical circuits, lying entirely beyond the long-wave limit of their range.[3]

One of the two lacunae in the spectrum, extending from 0.4 mm. to 7 mm., separated the range of natural rays from the range of artificial rays. To close this gap it was necessary literally to invent new rays, by designing oscillating electrical circuits which would generate frequencies which perhaps had never existed before in nature. The other lacuna, extending from 13A to 1200A, lay by contrast in the very centre of the range of natural rays, and precisely where we expect to find the frequencies resulting from certain peculiarly interesting and important processes in the electron-systems of atoms. These processes, it appears, are not in all cases easy to incite by the usual methods of stimulating atoms to radiate; but this difficulty is only one, and probably the least serious one, of the three hindrances which combined to delay the exploration of this region. A second impediment comes from the limitations of our devices for measuring wave-length, every one of which is unavailable over a certain sector of the region, extending roughly from 13A to 150A (limits which may later be forced somewhat closer together); but the most conspicuous obstacle is the extraordinary obstructiveness and opacity of every kind of matter to these rays.

The ability of electromagnetic waves to penetrate matter varies enormously from one part of the spectrum to another. At the uppermost end of the frequency-scale, the rays penetrate every sort of matter with astonishing ease. A layer of lead 8 mm. thick is required to remove half of the energy of a ray of wavelength .025A; and even this, it is probable, is not absorbed in the strict sense of being converted from radiant energy into another form, being merely deflected or *scattered* out of its original direction of motion.[4] With rays of greater wave-length, a true absorption is superposed upon the scattering, and increases very rapidly, about as the third power of the wave-length. The absorbed energy is used in extracting electrons from

[3] The distinction between natural and artificial rays is striking, but I fear not quite exact, since lightning-discharges and the causes of "static" offer instances of natural sources of radio frequencies. Also the selective absorptions of certain substances in the Hertzian range strongly suggest natural emission-frequencies. Still the distinction is not yet unsound enough to be dangerous.

[4] If A. H. Compton's theory of X-ray scattering is eventually triumphant, it will be necessary to admit that some radiant energy is transformed into kinetic energy of moving masses when scattering occurs.

the deeper levels of atomic electron-systems, as I described in the second of these articles. The absorption in any particular substance does not increase with an uninterrupted upward sweep; there are occasional setbacks, each of which occurs at a critical frequency where the radiation ceases to be able to extract electrons from a particular level. But though the lower-frequency rays cannot extract the deeper electrons of the atoms, they more than make up for it by expelling the outer electrons in greater and greater abundance; and when wave-length 13A is reached, they can remove only the outermost electrons or shift them from one orbit to another,[5] but they perform these actions so often that the beam is rapidly absorbed (even at 0.5A, 0.01 mm. of lead is sufficient to abstract half its energy).

Beyond 13A there is a region of well-nigh total eclipse. All we know about it is derived from a few measurements by Holweck. According to him, rays of wave-length 40A lose half their energy in traversing half a millimetre of air at atmospheric density; at 100A, the same proportion is consumed in a twentieth of a millimetre of air, or in a quarter of a millimetre of hydrogen, the most tenuous of all substances; and even these are not the most absorbable rays. A sheet of celluloid, .0001 mm. thick, which absorbs only 8% of the energy of a beam of wave-length 40A and 36% at 100A, abstracts 94% of the energy at 250A. It actually absorbs 97.3% of a ray of wave-length 308A; but this may be the least penetrating radiation of the entire scale, for the transmission apparently is a little greater at 400A (although Holweck seems to distrust the reliability of the last result). It must be admitted that the various beams of radiation on which these measurements were made are not monochromatic, but comprise each a continuous range of wave-lengths extending down to the quoted value, which is the minimum. Since the beam is in every case filtered through as many absorbing layers as possible before the final measurement of transmission through the celluloid sheet is made, and those remove preferentially the longer waves, it is probable that each datum refers to a finite, yet comparatively narrow, band of wave-lengths with its lower end at the specified value.[6]

[5] Some of the absorbed energy may be utilized in other ways, but there is no known alternative mechanism.

[6] The curve of Fig. 1, taken from Holweck's article, shows his data for the absorbing power of celluloid plotted (logarithmically) against wave-length. All the points refer to wave-lengths between 40A and 400A except the one marked "a," which refers to the rays emitted by gaseous hydrogen bombarded by electrons of energy between 13 and 38 volts (the transmission is the same for every bombarding-voltage within this range). It is probably a sort of "weighted-mean" value for the various radiations of the Lyman series and possibly the secondary spectrum of hydrogen, and the value 1140A which Holweck assigns as its effective wave-length is probably as good as any. The straight line on the left relates to nitrogen.

Whether or not Holweck's measurements are accurate enough to fix the point of greatest opacity, it is certain that somewhere between 300A and 1200A the eclipse begins to pass off. Fluorite commences to transmit at about 1200A, quartz and gelatine at about 1800A (each

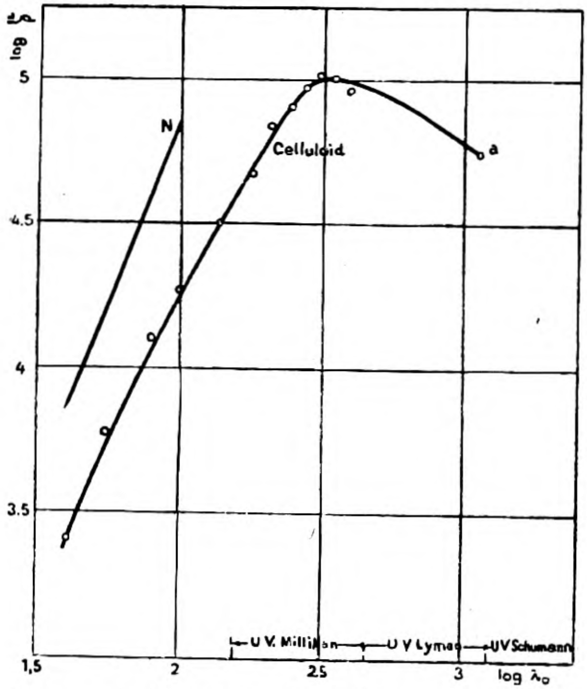


Fig. 1—Absorbing power of celluloid and of nitrogen plotted versus wave-length in the region of greatest opacity. (*Annales de Physique.*)

of these was, for reasons of experimental technique, long the limit of the explored region). Air begins to let through the light at about 1800A; the atmosphere indeed arrests the rays of the sun and stars as far along as 2900, but this is ascribed to ozone in the upper strata. Henceforward the absorption of radiant energy in gases consists mainly in shifting the valence-electrons of the atoms from one level to another, or in altering the amplitude of vibration of atoms built into molecular groups. The characteristics of individual atoms become steadily less influential; the groupings of the atoms into molecules, crystals, liquid or solid continua determine the amount of absorption. The question whether a particular solid is a conductor or an insulator, entirely irrelevant at high frequencies, eventually becomes the only question that matters; and radio-frequencies penetrate great thicknesses of rock or brick more readily than the thinnest sheet of metal foil.

To explore the region of the spectrum in which the absorbing-power of matter is at its greatest, it is necessary to make a high vacuum over the entire path of the rays from their source to the receiver (photographic plate, ionization-chamber, or electrode for photoelectric emission). This necessity can be escaped only if the obligation of measuring wave-lengths is evaded, for then the path may be very short; the receiver may be brought quite close to the piece of solid substance or the stratum of gas in which the rays are excited. If the wave-lengths are measured, it must be done with a ruled or crystalline diffraction-grating, which enforces a lengthy path (often as much as two metres). No solid windows can be interposed in it to confine a diffusing gas to the region where the rays are excited (the only exceptions yet developed are Holweck's .0001-mm. celluloid windows, which when stretched over and sustained by a fine-meshed gauze are said to be able to support a 5-cm. pressure-difference between their two faces). The excitation must therefore take place, whenever possible, in vacuo. This is simple enough when dealing with the rays excited from solids by electron-bombardment, and originating from displacements of electrons deeper down in the atomic system than the valence-electron; for the bombardment can be carried on in vacuo. But the arcs and sparks which are commonly used to displace the valence-electrons of free atoms or molecules, and so produce the frequencies for which these are responsible, are usually operated in an atmosphere composed of a comparatively few of the atoms being studied, mingled with a large amount of air or some other permanent gas. Yet it has been found possible to operate both arc and spark discharges "in vacuo," that is, without the atmosphere of permanent gas; though they differ in various ways from the like-named and familiar discharges in air, and do not display quite the same spectra.

Vacuum arcs, when once ignited, can be maintained with a moderate voltage between electrodes of various metals; the mercury vapor lamp is the familiar example, but arcs of such metals as magnesium, aluminium, and lead were developed as early as 1905. The name "vacuum arc" is, of course, a misnomer; the discharge occurs in an atmosphere of the vapor of the metal, but this congeals as soon as it starts to diffuse away from the discharge, and does not impair the vacuum in the light-path. The condition for an easily-maintained vacuum arc is that the vapor-pressure of the metal involved be comparatively high. Yet arcs between carbon electrodes in vacuo seem to be easy to maintain, though the vapor pressure of carbon is immeasurably small; one is led to suspect the gases inevitably occluded in this element.[7] Saunders produced waves as short as 978A with an arc in calcium vapor, and Simeon waves down to 375A with a "carbon vacuum arc."

These vacuum arcs are started either by heating the electrodes to produce a momentary high vapor-density, and applying a transient high voltage between them; or by touching them together and drawing them apart while the moderate voltage is applied. If the latter method is tried when the voltage is too low to maintain an arc, there is a transitory flash, the *breakspark*; its spectrum in the visible region has been noticed by von Welsbach, who finds the relative intensities of certain lines strangely altered from what they are in the ordinary spark; but according to McLennan and Lang, it yields no rays of wave-length inferior to 2000A.

The *vacuum spark* or *hot spark* employed by Millikan and his associates is an altogether different affair; it is a brilliant spark which occurs between electrodes a millimetre or so apart (the limits 0.1 mm. and 2 mm. have been assigned) in an extremely high vacuum, when a transient potential-difference of the order of several hundreds of thousands of volts is laid across them. This is a mysterious phenomenon, which has been studied by several scientists, without satisfactory conclusions. Whatever the vacuum spark really is, there is no doubt that it exists, and that wave-lengths are found in its spectrum which are shorter than any hitherto observed in any spectrum of arc or spark; and it is likely that these high-frequency rays are not excited at all in the ordinary electrical discharges of relatively low voltage, so that the high vacuum provides the conditions for stimulating as well as for transmitting them. The least wave-length yet measured with an optical method (ruled grating), which is 136A, occurs in the spectra of some of these sparks.

Most difficult of all is obviously the problem of detecting the rays emitted by the atoms or molecules of a permanent gas, which must of necessity occupy the entire path of the light from the place where it is excited to the place where it is received, unless intercepted by a solid partition which would intercept the desired waves also. If the discharge-tube containing the luminous gas communicates only by a narrow slit with the chamber containing the diffracting and receiving apparatus, it is practicable to connect a powerful pump to

[7] The minimum maintaining voltage for arcs in vacuo is given by Simeon as follows, for electrodes of the following materials: *C* 30 to 40 volts, *Na* 30 to 40, *Al* 80 to 100, *Si* 95 to 105. The distance between the electrodes is described as "slight," the degree of vacuum before arcing is not stated.

a branch-tube opening near the slit into the latter chamber, and so maintain in it a considerably lower density of gas than is required in the discharge. Hopfield has succeeded in maintaining an atmosphere of one kind of gas in the discharge-tube, and an atmosphere of another and a more transparent kind of gas in the chamber; the two gases are prevented from mingling by the same pumping-arrangement.

As for the measurement of wave-lengths from 1200A down to about 100A, it must be made with a concave diffraction-grating, which separates rays of different wave-lengths and itself focusses them at different places; for the rays cannot penetrate the prism of a prism spectograph, or the lens which is commonly used[8] to focus the beams diffracted by a plane grating. Rowland of Johns Hopkins, the first great master of the art of making diffraction-gratings, ruled them both upon plane and upon concave surfaces. The plane grating was so much the more easily ruled, that the concave grating fell into desuetude; but it became invaluable as soon as Lyman began to work in the region where the lenses extinguish the light. One might have anticipated that it would refuse to diffract rays the wave-lengths of which are only one-twentieth, one-fiftieth, even one one-hundredth of the spacing between its lines; but as Lyman and Millikan advanced farther and farther beyond the earlier limit of the ultra-violet, the concave grating proved itself competent to an extent which would probably have astonished its inventor. In one of Millikan's articles we may read an account of the ruling of new gratings by Pearson of Chicago; the spacing of the lines was by no means unusually small (about 500 per mm.) but they were ruled "with a very light touch so as to leave a portion of the original surface functioning in the production of spectra"—partly so that successive rulings might be nearly alike, but chiefly because if just half the original surface could be left intact, a large proportion of the total radiant energy would be diffracted into the first-order spectrum (this is the only usable one, because the higher-order images formed by the small wave-length rays encroach on the first-order images of the rays of greater wave-lengths). The arrangement of apparatus in experiments with the concave grating has varied little from the form which Lyman originally gave it. In Fig. 2 (from an article by McLennan) one sees the cross-section of a large tubular air-tight chamber, containing the grating at L (it is mounted on a carriage Q sliding on rails O, P), the slit at S and the photographic

[8] There is no apparent reason against using concave mirrors instead of lenses, unless the multiple reflections consume too much of the light. Luckiesh mentions an instrument designed with focussing mirrors of nickel (Houston, Proc. Roy. Soc. Edinb., 1912), which, however, were found inferior to quartz lenses in the range in which it was tested.

plate at C. The rays are excited at the centre of a tube V communicating by the slit with the grating-chamber. In this instance the source of light was a vacuum-spark between the electrodes sketched; had it been a vacuum arc or a glow-discharge in a permanent gas, the tube might have been different in appearance, but would have been sealed onto

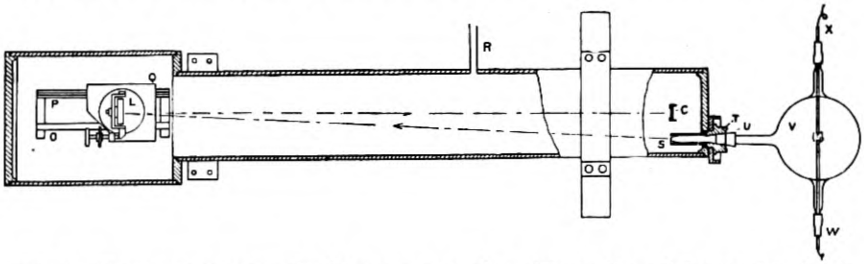


Fig. 2—Vacuum spectrograph with concave grating. (*Proceedings of the Royal Society.*)

the chamber at the slit in the same manner. The distance SL and LC are each one metre, and the sum of them constitutes the major part of the light-path (Lyman has reduced the sum to 40 cm. by using a more curved grating).

The extension of the explored or explorable region of the spectrum from 1200A onward to 136A does not entirely close the lacuna; but it brings into the accessible range every one of a certain very important class of rays—the rays emitted by a free atom when its valence-electron has been displaced and is returning towards or to its normal position. The reason for distinguishing one electron of the atomic electron-system above the others as the *valence electron* (the name is chosen rather for its meaninglessness than for its meaning) lies in the existence of line-series in the spectra. Magnificently regular series of rays are observed in the spectra of the atoms of hydrogen and of ionized helium, each of which has an electron-system consisting of a single electron in the inverse-square field surrounding the atom-nucleus.[9] Series which resemble these, though they are not arranged according to so elegantly simple a numerical law, are found in the spectra of the elements of the first column of the periodic table (Fig. 3) and suggest forcibly that one of the electrons of the atom of lithium, or sodium, or potassium lies so much farther out than all the others that it moves by itself in a field which is almost identical with the inverse-square field of a nucleus of charge e (the resultant of the fields of the nucleus and the inner electrons approaches such a

[9] This inverse-square field seems to be assured by the experiments on deflections of alpha and beta particles by atom-nuclei, quite apart from the successes of Bohr's special assumptions about atomic structure and radiation.

field as the distance from them increases). The same argument applies to elements of the second, third, and fourth columns, though with diminishing force, for the series become more difficult to trace and depart greatly from the archetype. In the crowded and complicated spectra of elements such as neon, argon, and iron, it is very

	I	II	III	IV	V	VI	VII	VIII	0
1	1 H 13.54								2 He 24.6^{x}
2	3 Li 5.36	4 Be	5 B	6 C	7 N 11^{x}?	8 O 13.56	9 F		10 Ne 21.5^{x}
3	11 Na 5.12	12 Mg 7.61	13 Al 5.9	14 Si 7.2 ?	15 P 13.3^{x}?	16 S 10.31	17 Cl 8.2		18 A 15.3^{x}
4	19 K 4.32	20 Ca 6.09	21 Sc 5.7	22 Ti 6.0	23 V	24 Cr 6.8	25 Mn 7.38	26 Fe 27 Co 28 Ni	
	29 Cu 7.69	30 Zn 9.35	31 Ga 6.0	32 Ge	33 As 11.5^{x}	34 Se	35 Br 10.0		36 Kr 12.7^{x}
5	37 Rb 4.16	38 Sr 5.67	39 Yt	40 Zr	41 Nb	42 Mo 7.0	43—	44 Ru 45 Rh 46 Pd	
	47 Ag 7.54	48 Cd 8.95	49 In 5.6	50 Sn	51 Sb 8^{x}	52 Te	53 I 9		54 Xe 10.9^{x}
6	55 Cs 3.88	56 Ba 5.19	RARE EARTHS	72 Hf	73 Ta	74 W	75—	76 Os 77 Ir 78 Pt	
	79 Au 8.7	80 Hg 10.39	81 Tl 6.1	82 Pb 7.38	83 Bi 8	84 Po	85—		86 Nt
7	87—	88 Ra 5.3	89 Ac	90 Th	91 Pa	92 U			

Fig. 3—Periodic table of the elements showing their atomic numbers and ionizing-potentials. (Cf. footnote 15.)

difficult, though apparently not impossible, to arrange frequencies into series, and this is in accord with the belief (founded on evidence of other kinds) that in these atoms there is no single outer electron far beyond all the others, but rather an outer shell of several similarly-placed electrons. Any one of these might imitate the behavior of a valence-electron, however, when removed to an unusually large distance from the nucleus and from the rest. It is to be observed also that when atoms are brought close together in the liquid or solid state, the line series can no longer be excited.

Wherever, therefore, there are discernible line-series, one infers an electron far enough beyond all the others to have a behavior and deserve a title of its own. Generalizing Bohr's wonderfully successful model of the atoms of hydrogen and ionized helium, we imagine that this electron enjoys a particular set of orbits, in the narrowest and deepest-lying of which it normally abides, while in any one of the others it can make only a transient halt.[10]

[10] It may not be superfluous to complete the description of Bohr's model by saying that when the electron goes from one orbit to another, the difference ΔU between the values of the energy of the atom in the two states is radiated in a ray of frequency $\Delta U/h$.

Now all the line-series observed in the spectra of excited atoms and all which there is any reason to imagine as existent but undiscovered, lie entirely at wave-lengths greater than 136A; indeed most of them lie in the already-accessible region beyond 1200A, but a few of the most important are in the newly-opened range. Hydrogen is entitled to first mention, being the leader of the procession of elements as well as the most completely understood of them. The visible spectrum of (atomic) hydrogen consists of the archetype of all line-series, the Balmer series, extending from 6563A to 3650A, the frequencies of its lines being equal to the numbers of the series

$$\text{(A)} \qquad R\left(\frac{1}{2^2}-\frac{1}{3^2}\right),\ R\left(\frac{1}{2^2}-\frac{1}{4^2}\right),\ R\left(\frac{1}{2^2}-\frac{1}{5^2}\right),$$

and so forth, in which R is a certain constant ($R=3.29 \cdot 10^{15}$). According to Bohr's theory, this means that the energy-values[11] of the consecutive orbits of the valence-electron (in this case the only electron) are given by the numbers of the succession

$$\text{(B)} \qquad -Rh\left(\frac{1}{2^2}\right),\ -Rh\left(\frac{1}{3^2}\right),\ -Rh\left(\frac{1}{4^2}\right),\ -Rh\left(\frac{1}{5^2}\right),$$

and so forth, and the consecutive rays of the series are emitted when the electron drops into the first of these orbits from the second, third, fourth and consecutive orbits. Most people, on looking at the succession of numbers (B), would instinctively complete it by adding a term $-Rh$ at the beginning; and if there is truly an orbit of which the energy-value is $-Rh$ there must be an additional line-series,[12] the frequencies of its lines being equal to the numbers of the series

$$\text{(C)} \qquad R\left(1-\frac{1}{2^2}\right),\ R\left(1-\frac{1}{3^2}\right),\ R\left(1-\frac{1}{4^2}\right),\ \text{and so forth.}$$

The first three lines of this series should lie at 1216A, 1026A and 972A. They were discovered by Lyman in 1913, and the series bears his name.

[11] The energy-value of an orbit is the energy of the atom when the valence-electron is in this orbit; the energy of the atom being set equal to zero, when the valence-electron is removed to infinity. It follows from this last convention that the energy-value of an orbit, with sign reversed, is equal to the energy which must be imparted to the atom to remove the valence-electron completely from the atom when it is initially in the orbit in question. Thus the energy-value of the orbit which the valence-electron normally inhabits is equal to the ionizing-potential of the atom, when it is expressed in appropriate units and its sign reversed. The practical advantages of this convention are so great that we endure its annoying and confusing consequence of making all the energy-values of non-ionized atoms negative.

[12] The existence of this series was anticipated long before Bohr's interpretation of the Balmer series, being suggested by the form of the series itself.

Helium follows hydrogen in the procession of elements. Its spectrum includes several line-series. The frequencies of the first four members of one of these series, the principal series of the singlet or parhelium spectrum, are as follows (all the numbers in the successions D, E, F, G, and H should be multiplied by 10^{14}):

(D) 1.457, 5.981, 7.567, 8.300

Subtracting each from the frequency of the series-limit, which is 9.609, we obtain the succession of numbers

(E) $(9.609-8.152)$, $(9.609-3.628)$, $(9.609-2.042)$, $(9.609-1.309)$

which suggests a succession of orbits, having the following consecutive energy values[13]:

(F) $-9.609h$, $-8.152h$, $-3.628h$, $-2.042h$, $-1.309h$.

The consecutive frequencies of this series are emitted when the valence-electron falls from the second, third and consecutive orbits of this succession into the first one. One would suppose that the valence-electron normally abides in this first orbit. But if this were so the energy required to ionize the atom would be $9.609h \cdot 10^{14}$, equivalent to 3.96 volts; and waves of the frequencies given by (D) could displace the electron and be absorbed thereby. But the ionizing-potential of the atom is about 25 volts and the frequencies (E) do not appear as dark lines in the absorption-spectrum of helium. Therefore there must be still another orbit much deeper down, with a much higher (negative) energy-value, than any listed under (F). In 1921–22 Lyman discovered (with his highly-curved grating and shortened light-path, and pumping arrangement for keeping the pressure low) a new series of lines of wave-lengths 584.4A, 537.1A, 522.3A and 515.7A. Their frequencies are

(G) 51.34, 55.85, 57.44, 58.18

which may be written as the succession of numbers

(H) $(59.49-8.15)$, $(59.49-3.64)$, $(59.49-2.05)$, $(59.49-1.31)$.

Comparing these with the succession (E) we recognize the same set of subtrahends,[14] and accordingly identify the common quantity $59.49 \cdot 10^{14}$

[13] It is customary to designate the orbits by their energy-values divided by hc, or $19.68 \cdot 10^{-17}$.

[14] It would not be necessary to call attention to this if we could calculate the frequency of the series-limit, which would give the energy-value of the new orbit immediately; but the four discovered lines are hardly sufficient for such an extrapolation (there are fourteen of the other series to use for calculating its limit).

as $-1/h$ times the energy-value of an additional orbit. If this orbit is the permanent home of the valence-electron, the energy required to ionize the atom must be $+59.49h \cdot 10^{14}$, equivalent to 24.5 volts. When the new lines were discovered, the accepted value was 25.3 volts, largely because of a certain measurement by Franck. After the publication of Lyman's discovery, Franck re-examined his method and data and found them compatible with the value 24.5 volts; and very recently C. A. Mackay has ascertained that the ionizing-potential of helium is 14.1 volts greater than that of mercury, which is quite definitely known to be 10.4. One could hardly desire a better illustration of the confluence of measured values of the energies of atoms and measured values of their radiation-frequencies, when both are interpreted according to the contemporary theory of radiation.

These newly-discovered waves must be the shortest in the spectrum of helium; the atom cannot emit a ray of wave-length less than the series-limit 504A, calculated by the equation

$$h\nu = hc/\lambda = \text{energy-value of the deep-lying orbit} = 59.49 \cdot 10^{14}h.$$

They are much shorter than the waves of the Lyman series of hydrogen, which Bohr's theory, together with the observed value of ionizing-potential of atomic hydrogen, justify us in declaring to be the shortest waves emitted by that atom. Furthermore, it is almost certain that they are shorter than any waves for which the valence-electron of any atom is responsible; for the ionizing-potential of helium is greater than any other measured ionizing-potential, and there is no reason to believe that any of the yet unmeasured ones exceed it. Its nearest rivals are the ionizing-potentials of the inert gases which share the last column of the periodic table. The experimentalists have not agreed very well in their estimates of these, although all agree that the values are comparatively high. Hertz, the latest to make measurements upon neon and argon, gives 21.5 volts for the first and 15.3 volts for the second. Both, therefore, should emit some rays lying below 1200A, but above 575A and 800A, respectively, and resulting from transitions of the valence-electron. Dejardin gives 12.7 volts for the ionizing-potential of krypton and very lately 10.9 volts for that of xenon. In the other columns of the periodic table, the values of ionizing-potential are prevailingly lower than in the column of inert gases. The value 10.4 volts (for mercury) is the highest among the metals; several of the non-metallic elements appear to have ionizing-potentials between 12 and 17 volts, but for some of these it is difficult to tell whether the observed value pertains to the atom or to a molecule. The experimental material is abundant[15] enough to give practical certainty that "valence-electron rays" below 1000A occur in the spectra of only a few elements, and below 500A in none.

Nevertheless, Millikan and Bowen, photographing the spectra of all of the first twenty elements (neon and argon excluded, and chromium and copper added) down to the extremity of the region accessible with the concave grating, discovered great numbers of lines, of which they attribute dozens or scores to particular elements (for example, some forty lines ascribed to potassium, though its ionizing-potential of four volts corresponds to a minimum wave-length exceeding 2500A). Some of these may be lines of compound origin, resulting from two simultaneous changes in the electron-system of the atom, one being a transition of the valence-electron and the other a rearrangement of the other electrons. (Saunders mentions such lines in the spectra of elements of the second column of the table.) Others are due to rearrangements of internal electrons following upon a displacement of one of these. Many others are attributed to displacements of the valence-electrons of ionized atoms. Of this new field of research, the spectroscopy of ionized atoms, I wrote briefly in the first article of this series. In the more easily accessible regions of the spectrum, Paschen had discovered rays of doubly-ionized aluminium and Fowler rays of trebly-ionized silicon.[16] Millikan and Bowen go a step further by identifying certain rays of quadruply-ionized phosphorus; indeed they believe that, under the violent excitation provided by their vacuum-spark, the waves emitted by atoms which have lost all but one of the electrons from their outermost electron-shells (the three just specified are in this state) are especially abundant and intense.

[15] In the periodic table of Fig. 3 the ionizing-potentials of the elements are given along with their atomic numbers. Overlined figures are values calculated from series-limits and confirmed by direct experiment; starred values are data of experiment, for elements of which the series have not been worked out; the remaining values are calculated from series-limits and have not been verified. The data are from the cited sources, from Foote and Mohler, and from Saunders' graphic tabulation (*Science*, volume *59*, pp. 50–51, January 18, 1924;). Interesting variations with atomic number are observed which yield bases for estimating the values for still other elements by interpolations.

[16] "It might be mentioned that the spectrum of silicon was first selected for investigation on account of its astrophysical interest. The lines representing successive stages of ionization of this element appear in stars which there is every reason to believe are at successively higher temperatures. The complete series-data for the four spectra [trebly,- doubly-, once- and non-ionized atoms] and the ionization-potentials deduced from them, may be expected to find an important application in fixing the scale of stellar temperatures . . . All the series predicted in [these four] spectra of silicon; in the spectra of doubly-, once-, and non-ionized aluminum, of once-ionized and neutral magnesium, and of neutral sodium, have been actually produced and have been found to have the character and constants expected."—A. Fowler.

The "valence-electron" rays emitted by ionized atoms should lie at lesser wave-lengths (roughly ¼ as great) than the valence-electron rays of neutral atoms, and therefore should be particularly at home in the region newly opened to exploration. The highest frequencies emitted by the ionized-helium atom are perfectly calculable from Bohr's theory; they are the frequencies of the Lyman series, quadrupled, and the wave-lengths therefore lie between 304A and 230A. They have not been reported, but Lyman in 1919 observed two lines in the spectrum of violently-excited helium, near the positions 1214.9 and 1640.1A calculated for the first and third lines of the next helium series (having the frequencies of the Balmer series, quadrupled). The place of the second member of the series was obscured by an alien line.

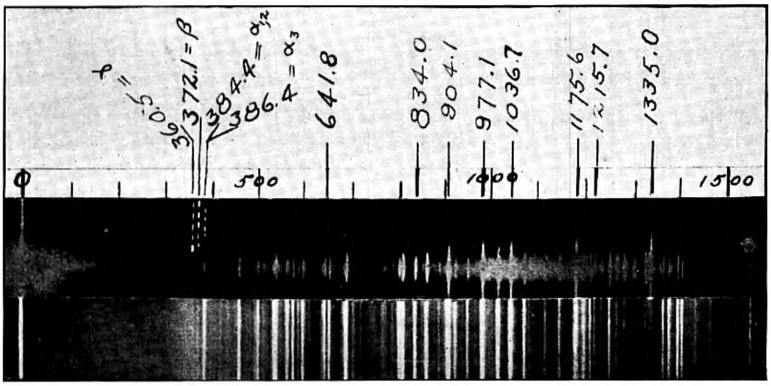


Fig. 4—Spectrum of a vacuum spark between carbon electrodes. (*Astrophysical Journal.*)

The once-ionized lithium atom, judging from the example of neutral helium, should display higher frequencies than any other once-ionized atom, and they should be arranged in recognizable series, somewhere near the extreme limit of the explorable range as it stands at this moment. They have not, however, been reported; Millikan says that his plates show no lithium lines of any sort from 1700A down at least to 370A, if not farther.

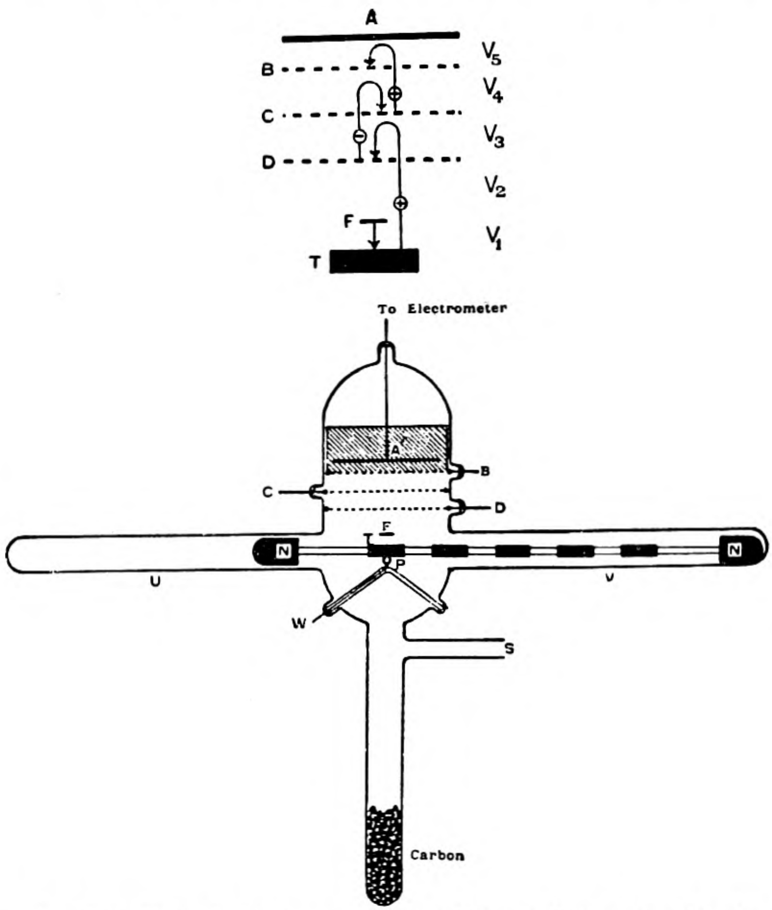
As an example of a spectrum extending far into the newly-conquered field, a plate representing the spectrum of a vacuum spark between carbon electrodes is reproduced from one of Millikan's articles as Fig. 4. The actual spectrum is in the middle; it is drawn out for better intelligibility, at the side. Most of the marked lines, including the extreme line at 360.5A and the strongest line at 1335A, are attributed to carbon; some to other elements, particularly the one at 1215.7 which is the first line of the Lyman series of hydrogen. The interpretation of spectra like this is not a simple matter of putting electrodes of the desired substance into the tube and ascribing to it all the lines which come out on the plate. It appears that impurities, even when present in what might be considered small proportions, contribute their own rays to the spectrum in great abundance and intensity. Millikan found that all the lines present in the spectrum of the vacuum spark between magnesium electrodes were also present when aluminium electrodes were used, and vice versa, and finally assigned them all to oxygen. Lyman found it extremely difficult to decide which lines belong to hydrogen and which to helium, since the spectra of glow-discharges in these gases have so many lines in common. Helium has a pronounced habit of encouraging excitation of the rays of whatever other gases are mixed with it, since the helium atoms require so much energy to displace their valence-electrons that free electrons shot into helium gas are liable to bounce harmlessly from one helium atom to another until they strike and excite an atom of another variety. Even if one can be sure that all the rays in a spectrum belong to a single element there remains the problem of assigning them to neutral or variously-ionized atoms. It is clear that the completion of the spectroscopist's task is deferred by this extension of it to what the Germans call the unforeseeable time.

We return to the consideration of the lacuna in the spectrum, which extends from 13A up to a boundary which by the use of high vacua, concave gratings, and violent excitations, has been forced from 1200A down to 136A. This wave-length 136A stands for the moment as the lowest which has ever been actually measured with the ruled grating; and in spite of the unexpected and fortunate adequacy of the instrument down even to this point, little more can be demanded from it. The reason is, that the substance on which the rulings are made must eventually cease to reflect the rays on account of its own looseness of texture. Being a congeries of atoms themselves separated by finite distances, the metal will not behave as a continuum towards waves of a length not very large compared to its own atomic spacing. Below 13A waves are not reflected. Little is known of the rate at which the reflecting-power dwindles away to zero between 136A and 13A and this little we owe again to Holweck. He directed a beam of radiation (it was a mixed beam, as was previously made clear, and the wave-length-value is merely the minimum wave-length in it) against a polished bronze mirror at the very oblique incidence of 73.°8; the reflected beam had one-third the intensity of the incident beam at wave-length 123A (practically the extreme wave-length of Millikan's experiments), but only 10% at 60A and only 3% at 40A. The performance was much better at a still more oblique incidence; on the other hand Holweck thinks that it gets rapidly worse as the incidence is made more nearly normal, and if this is true the outlook for the concave grating, with its condition of almost normally-incident light, is most unpromising.

Below the boundary 13A, the atomic constitution of solid substances turns from a hindrance into an advantage, and crystals serve as natural diffraction-gratings of incomparable fineness—too fine, indeed, for our convenience in this part of the spectrum, since the boundary is fixed by the smallness of the distance d between successive layers of atoms in the diffraction-grating. Rocksalt, one of the standard crystals, for which $d=2.814$A, has been used successfully up at least to 4A (by Fricke) and the rest of the way to 13A has been explored with crystals of gypsum ($d=7.58$A) or sugar ($d=10.56$A); in this region it was necessary to evacuate the light-path, precisely as in the region beyond 136A. The only possibility of a new advance depends on the utilization of crystals of still greater inter-atomic spacings. Holweck mentions a crystal with a formidable name, for which $d=19$A, and de Broglie and Friedel found that the oleates of sodium, potassium and ammonium presented spacings of the order of 40A between consecutive molecule-layers. If these substances can be adapted for use in crystal spectographs, the boundary of the explored region may be pushed far beyond its present place. It may be found, however, that the crystal absorbs the rays before they go deeply enough to be diffracted.

As for the region between 13A and 136A, no one has ever measured the wave-length of a radiation lying within it; but there is a method which indicates the existence and something about the wave-lengths of rays which almost certainly belong in it. In applying this method the photographic plate is replaced by a metal electrode (usually of platinum) which, when irradiated by rays of any wave-length (less than a certain critical one which always lies far above this range) emits electrons. For this reason it is often known as the *photoelectric method*, although the substitution of one kind of receiver for another is not its most distinctive characteristic. A target made of the substance to be studied is sealed into a tube, opposite a source of electrons (generally a filament for thermionic emission); the photosensitive electrode is placed somewhere in the tube where whatever rays are excited at the surface of the target will fall directly upon it. It is all-important to protect this electrode from electrons and ions, negative or positive, proceeding from target, filament, or anywhere else.

Usually the electrode is screened by a family of gauzes, with their potentials adjusted as is indicated in Fig. 5 (with the paths of intruding ions of both signs, including those excited from the outer gauzes themselves, mapped out to show how they are rebuffed). Naturally



Figs. 5 and 6—Horton's apparatus for determining excitation-potentials by the "photoelectric method." (*Philosophical Magazine.*)

the arrangement of potentials in front of the electrode must be such that the emitted electrons are all drawn away from it, not driven back onto it. The rate of emission of electrons, the *photoelectric current*, may be measured with an electrometer connected either to the sensitive electrode or to a gauze so placed as to gather in all the electrons emitted from it. Figs. 5 and 6, the latter of which shows a completely equipped tube with filament at F, a row of targets which can be moved consecutively into place at T, and the photosensitive electrode at A with its gauze shields in front of it, come from the work of Horton, Andrewes and Davies. A tube designed and used by E. H. Kurth is shown in Fig. 7; the filament is seen in perspective at C, the target at T, and the sensitive disc at D; the family of diverging straight lines represents a set of metal laminae, which being charged

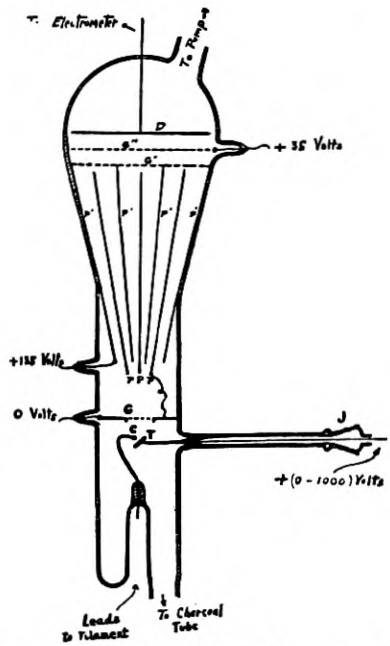


Fig. 7—Kurth's apparatus for determining excitation-potentials. (*Physical Review.*)

alternately to potentials 0 and +135 volts gather in any ions which start up towards the disc. The method can also be adapted to gases, and this application has an interesting and important history; but as nearly all the data respecting gases refer to wave-lengths superior to 1200A, they fall out of the province of this discourse. Foote and Mohler, however, penetrated to 26A with the apparatus of Fig. 8, filled with oxygen. The filament is at A; the electron-accelerating voltage V is applied between A and the gauze B, so that the target is essentially a thin layer of gas enveloping B; the photosensitive electrode is the gauze C, the photoelectric current from which is gathered in by the plate D (screened against positive ions by its high potential).

The art of detecting radiations by this method consists in giving various values to the "bombarding voltage" V between target and filament, which is the measure of the energy of the electrons impinging on the target; measuring the photoelectric current i, which is the measure of the intensity of the rays; plotting i (or better the ratio of i to the current of bombarding electrons) versus V; and examining the curve to see whether it displays sudden changes of slope. If it does,

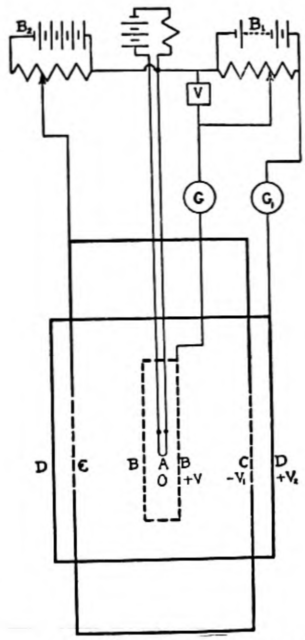


Fig. 8—Mohler and Foote's apparatus for determining excitation-potentials of gases and vapours. (*Bulletin of the Bureau of Standards.*)

one infers that at the corresponding voltages new radiations suddenly burst forth. The method therefore consists in finding critical bombarding-voltages, that is, critical electron energies which just suffice to excite particular sorts of radiation; it is a method for discovering *excitation-potentials*. Three excellent instances of such abrupt changes in slope, or *breaks* as they are frequently called, appear in the (i, V) curve determined with an aluminium target by Horton and his associates (Fig. 9). Very many such curves appear in the literature, with more or less conspicuous breaks; some are as striking as these in the figures, some require a good deal of care and experience to locate them properly, and some, one is driven to conclude, are visible only to the eye of faith. But it is hardly possible to doubt that such a corner as the three here reproduced marks the *entrée en scène* of a new ray or set of rays.[17]

But a determination of an excitation-potential is not a measurement

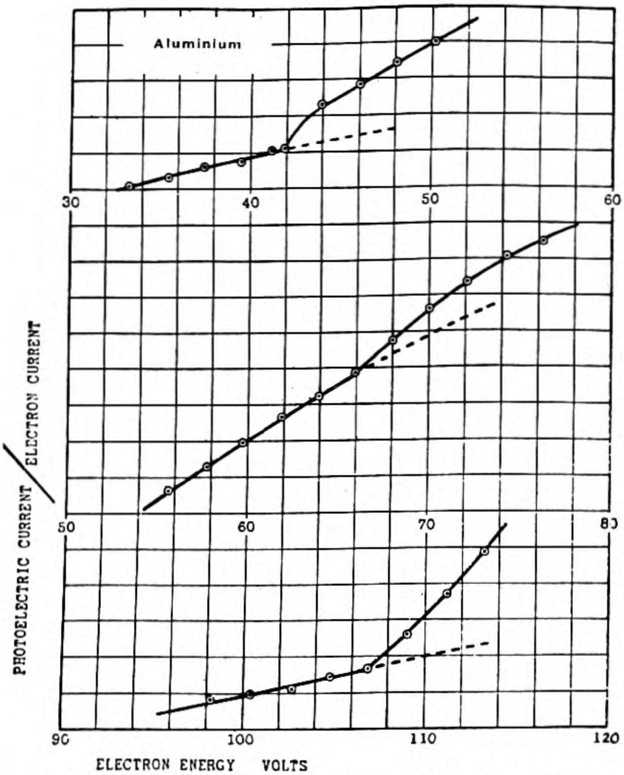


Fig. 9—Breaks in a photoelectric-current curve indicating excitation-potentials. (*Philosophical Magazine.*)

of the wave-lengths of the excited rays; and while it is supposed that excitation-potentials between 1000 volts and 100 volts are associated with rays of wave-lengths between 12A and 123A, this is merely a supposition.[18] We require a theoretical relation between excitation-

[17] It is clear from this account that the photosensitive disc might be replaced by a photographic plate, on which the opacity due to the rays produced by consecutive values of V could be measured; or by an ionization-chamber, in which the ionization-currents could be measured. It is equally clear that neither method would be so suitable for detecting slight discontinuities in rate of increase of radiant energy with increase of V. However, both methods are used at higher frequencies, where by dispersion of the waves a discontinuity in the intensity at a single wave-length is made more conspicuous.

[18] This is the best place to remark that electrons of voltage V bombarding a solid, in addition to exciting (if V is high enough) rays characteristic of the bombarded atoms, excite also a continuous spectrum of rays of all frequencies up to a maximum

potentials and excited frequencies. The question is of high importance, not simply because we are interested to know whether some of the excited rays really lie in the hitherto unpenetrated range, but primarily because excitation and emission are among the fundamental qualities of atoms. Excitation-potentials exceeding 1000 volts generally produce rays of which the wave-lengths are less than 12A and can be measured with the crystal spectrograph, so that a rule or law can be deduced from the two sets of measurements. Excitation-potentials inferior to 25 volts generally produce rays of which the wave-lengths are greater than 500A and can be measured with optical apparatus, and again a law can be deduced from the two sets of data. But the law is not the same in the two cases; this is because excitation, in the former case, consists in displacing a deep-lying electron, while in the latter case it consists in a displacement of the valence-electron. We are forced to the disconcerting conclusions that excitation-potentials between 1000 volts and 25 volts involve electrons of an intermediate type, and that the still-unverifiable law connecting them with the frequencies of their excited rays is not identical with either of the laws in the accessible regions of the spectrum.

The law for excitation-potentials involving displacements of the valence-electron is twofold. Each atom has at least two such excitation-potentials. One of them is its ionizing-potential. When the accelerating-voltage of an electron-stream playing against a multitude of free atoms forming a gas is raised just past the value V_i at which an individual electron has just enough energy to remove the valence-electron of an atom, there is an outburst of radiation. This comprises rays of many frequencies—probably all those which we have called valence-electron rays—and they are emitted as the valence-electrons descend step-by-step along their ladders of orbits. All these frequencies conform to the relation

$$h\nu < eV_i. \tag{1}$$

The other excitation-potential is the *resonance* potential of the atom (there may be more than one of these [19]). When the accelerating-

equal to eV/h. The heterogeneous beams used by Holweck in the experiments previously cited consisted chiefly, if not entirely, of this continuous spectrum. All the excitation-potentials mentioned in these pages, however, relate to individual rays or groups of individual rays characteristic of atoms.

[19] This question is still incompletely solved, in spite of much labor. At one time it was supposed that the valence-electron could be raised either altogether out of the atom, or else to the deepest-lying of the transient-sojourn orbits (or to either of the two deepest-lying orbits, if there are two complete families of orbits such as the mercury atom possesses); but not to any of the other transient-sojourn or "virtual" orbits. This restriction would apply only to displacements caused by impinging electrons; quanta of appropriate frequencies can lift the valence-electron to any of

voltage of the electron-stream is raised just past the value V_r at which the individual electron has just enough energy to raise the valence-electron from its normal to one of its transient-sojourn orbits, there is an outburst of radiation. This comprises rays of a single frequency, emitted when the valence-electrons return in single leaps from the orbits to which they were momentarily raised to the orbits of their normal habitation. This frequency conforms to the relation:

$$h\nu = eV_r. \tag{2}$$

The law for excitation-potentials involving displacements of deep-lying electrons bears a certain resemblance to the first of the foregoing laws. When the accelerating-voltage of an electron-stream playing against a multitude of atoms assembled in a solid or liquid is raised just past the value V_e at which an individual electron has just enough energy to extract a certain deep-lying electron, say a K-electron, there is an outburst of radiation comprising many frequencies, all conforming to a relation resembling (1), to wit:

$$h\nu < eV_e. \tag{3}$$

But it would be misleading to assume that the processes resulting in (1) and in (3) are identical. In the first place, it is not certain that the deep-lying electron need be completely extracted. Suppose it possessed a set of transient-sojourn orbits in the outskirts of the atom, their energy-values differing from one another and from that of the "orbit at infinity" (the state in which the electron is quite detached) by amounts less than the 25 volts which is the maximum difference between the energy-values of any valence-electron. Then there might be several excitation-potentials, differing from one another by 25 volts at most; but this difference would be so inconsiderable a fraction of the value of the extraction-potential V_e, which ranges from more than 100,000 volts for the K-electrons of uranium to about 1100 volts for those of neon, that they would be difficult to distinguish. Indications of multiple excitation-potentials have, how-

---

an immense number of orbits of a certain set, but not to transient-sojourn orbits of certain other sets. Lately it has been affirmed that impinging electrons of the right energy can lift the valence-electron to any one at all of its transient-sojourn orbits, even those to which it cannot be lifted by quanta; but this rule, if it is the true one, has not yet been illustrated by any extensive set of experimental data, though Hertz has lately intimated in a brief note that he has assembled such a set by experiments on helium. Franck and Knipping detected excitation-potentials corresponding to the lifting of the valence-electron of helium from its normal orbit to several distinct P-orbits; but I gather from a later paper by Franck that nobody has been able to reproduce the result. Olmstead and Compton discerned excitation-potentials corresponding to the lifting of the electron of hydrogen from its normal orbit to each of the next six transient-sojourn orbits.

ever, been discerned in the "fine structure" of the K absorption-edges of the lighter elements (notably the elements from sodium to potassium). In the second place, the process of emission is different in the two cases described by equations (1) and (3). In the former case, the rays were emitted as the valence-electron (or another replacing it, which comes to the same thing) redescended its ladder of orbits; but when a deep-lying electron is extracted, the resulting rays are emitted because of rearrangements of the other internal electrons of the atomic electron-system, which occur irrespective of whether the departed electron quickly returns to the atom, or remains a long time away.

I will now risk the making of a distinction which may eventually turn out not to be the most natural or practical, by reserving the name *deep-lying electrons* for those electrons which lie entirely within at least one completed electron-shell of an atom, and designating the others (exclusive of the valence-electron, which has already been set apart from the rest) as the *shallow-lying electrons.* It follows from this definition that the first nine atoms of the periodic table, up to fluorine (inclusive) possess only shallow-lying electrons; the next eight (*Ne* to *Cl*) have one set of deep-lying electrons, the K set; the next eighteen (*A* to *Br*) have at least four sets of deep-lying electrons, the K set and three L-sets (the last three can be grouped as one). It follows also that every instance in which an excitation-potential has been measured, and the wave-lengths of the excited rays have also separately been measured, is an instance in which a deep-lying electron is involved. For example, the excitation-potentials involving extraction of the K-electrons have been measured from the top of the periodic table down to the twelfth element (*Mg*), over which range they decline from 115,000 volts to 1100 volts; the excited waves have been measured over the same range and down to the eleventh element (*Na*), over which range they rise (for the principal ray) from .10A to 11.88A. At this point, and just before the K-electrons pass over into the category of shallow-lying electrons at the ninth element, the wave-lengths enter into the inaccessible range. The wave-lengths of the rays excited when one of the L electrons is displaced have been measured from the top of the table down to the twenty-ninth element (*Cu*) where, arriving at 13.3A, they too pass into the immeasurable class.

The general consequence of all this is, that the excitation-potentials involving shallow-lying electrons must be below 1000 volts; that, conversely, the excitation-potentials observed between 25 volts and 1000 volts are chiefly those of excitations which consist in displacements of shallow-lying electrons; and finally, that the wave-lengths of the excited rays lie below 13A, many of them in the inaccessible range, some in the range newly opened to exploration. This is a most unfortunate coincidence, for instead of being able to apply laws which prevail in other ranges to compensate for our inability to measure wave-lengths in this range, we have to expect distinct laws within it. Must shallow lying electrons be extracted altogether from the atom if they are to be displaced at all or have they certain transient sojourn orbits to some or all of which they may be raised by electron-impacts? Do the emitted rays result from a step-by-step return of the displaced electron? or from a return in a single leap? or from a rearrangement of the remaining electrons? or from a compounding of changes of the two latter types? So long as the emitted wave-lengths are not measured, these questions cannot be answered with confidence.

Some little can be inferred from numerical relations among excitation-potentials. McLennan and Clark, for example, observed three excitation-potentials of lithium, at 37.0, 31.8 and 12.0 volts. The first two of these voltages stand nearly in the ratio of the first two frequencies of the Lyman series in the hydrogen spectrum, which suggested to the discoverers that the processes involved in the excitations were the raising of a K-electron to the first and second of a pair of transient-sojourn orbits, standing in the same relation to the normal orbit of the K-electron as the orbits of energy-values $-Rh/4$ and $-Rh/9$ stand to the normal orbit of energy-value $-Rh$ in the hydrogen atom. That is to say, they conceive these excitation-potentials to be comparable to resonance-potentials, and the K-electron of lithium to behave like a valence-electron. They also found excitation-potentials of beryllium at 20.3 and 16.0, and of boron at 27.92 and 23.45. The ratio of each pair of numbers is about equal to the ratio of the first two frequencies of the Balmer-series, suggesting that these are resonance-potentials of an L-electron; the details of the analogy may be left to the reader to work out. Each of the latter elements displayed additional higher potentials, to be associated with the K-electrons. Rollefson lately discovered seven excitation-potentials of iron in the range between 160 and 264 volts, expressible by a formula $(a-b/n^2)$ if the integer values 5, 6, 7, 8, 9, 10 and 12 are successively given to n. If these seven potentials correspond to elevations of a certain shallow-lying electron to seven transient-sojourn orbits, the extraction-potential for this electron can be calculated by an extrapolation (so also in the cases cited from McLennan and Clark). Rollefson interprets certain other excitation-potentials as corresponding to elevations of certain deep-lying electrons to transient-sojourn orbits.

Some assistance in identifying the excitation-potentials of the light atoms can be obtained by plotting the recognized excitation-potentials of the heavier atoms, and also the frequencies of the rays excited; plotting curves representing them as functions of atomic number; and extrapolating the curves into the range of low atomic numbers. The best procedure is to plot the square roots of the excitation-potentials and the emission-frequencies, as then the curves are nearly straight lines (Moseley's law). Some of these lines are shown in

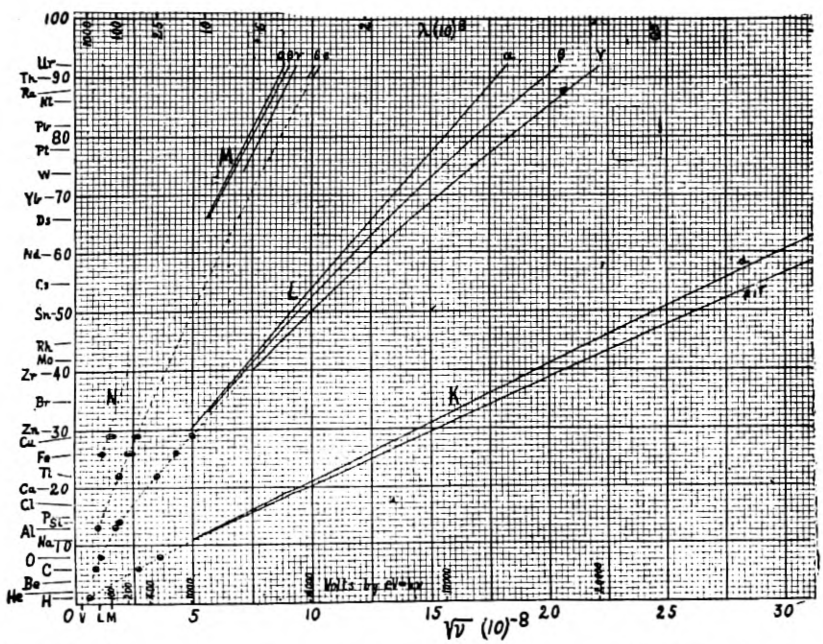


Fig. 10—Curves representing square roots of emission-frequencies of heavier atoms as functions of atomic number. (*Physical Review.*)

Fig. 10 (from Kurth). Since the atomic numbers are laid off (contrary to usage) along the axis of ordinates, the lowest-lying line represents the highest recognized emission-frequencies (the $K\beta$ and $K\gamma$ frequencies, which actually are slightly different, but are not indicated separately upon the graph). The next line, marked $K\alpha$, represents another particular emission-frequency. Excitation is the same for every ray of this group, and consists in extracting one of the deepest-lying or K electrons of the atom; and the excitation-potential for the entire group, the K excitation-potential, is also represented by a straight line, the K line, which may be taken as coincident with the lowest-lying line in the graph, provided that we translate frequencies into potentials by the relation $V=h\nu/e$ (both frequencies and

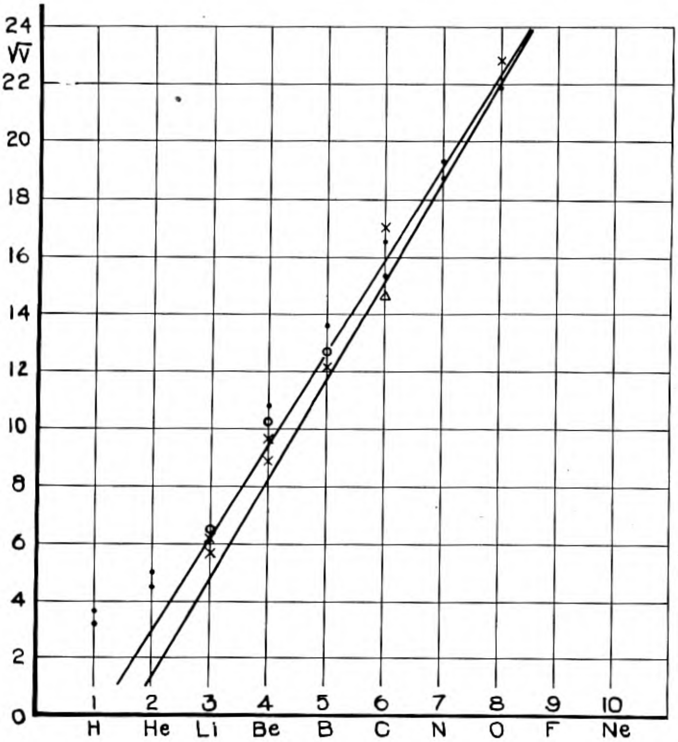


Fig. 11—Excitation-potentials of light elements, correlated with displacements of K electrons. (Cf. footnote 20.)

potentials are laid off along the axis of abscissae). This K line, it must be realized, extends the whole way from atomic number 92 to atomic number 12.

The circles upon the graph represent excitation-potentials inferior to 1000 volts, observed by Kurth. Three of these lie very close to the downward prolongation of the K line; the almost inevitable inference is, that in these three cases the excitation consists in the extraction of one of the electrons nearest the nucleus. The others lie so much above the extended K-line that they must belong to a distinct class. Many additional measurements have been made since this graph was published, and in Fig. 11 I have set down all the experimental values known to me which have been given for excitation-potentials of the first eight elements, omitting those which are so small that they obviously do not belong to the K class.[20] The

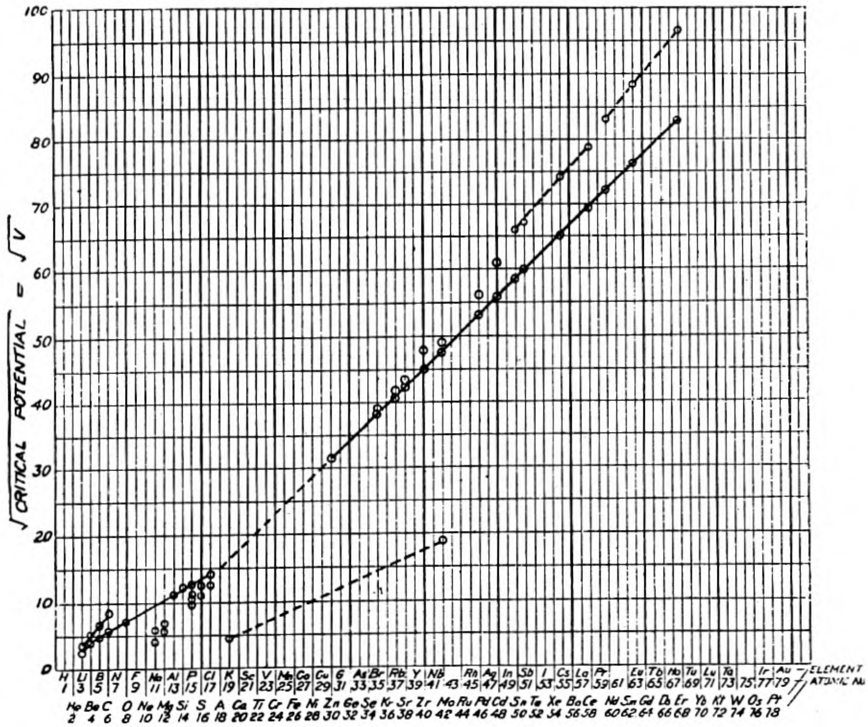


Fig. 12—Emission-frequencies of heavier and excitation-potentials of lighter elements, correlated with L electrons. (*Proceedings of the Royal Society.*)

[20] The data are from various sources, as follows. The dots for hydrogen and helium represent the observed ionization and resonance potentials of these atoms. The dots for *Be*, *B*, *C*, *N* and *O* are at values of excitation-potentials given by Mohler and Foote from experiments on gaseous compounds of these atoms. All the other data except Holweck's are values of excitation-potentials for solids. The crosses for *Li* and *Be* stand for the excitation-potentials observed by McLennan, the circles for the extraction-potentials of the K electrons which they infer from these data. The cross for B represents three values lying so close together as to be indistinguishable (from McLennan, Hughes, and Holtsmark) and the cross for C also three coincident values (Kurth, Richardson and Bazzoni, Holweck). The circle for B is at the potential corresponding to a discontinuity in absorption, observed by Holweck. The triangle for C is a value observed by Hughes, and the cross for O a value from Kurth (obtained with oxidized copper). No data for F or Ne are available. At Na measurements on the wave-length of $K\alpha$ and at Mg measurements on the K absorption-edge commence.

lower of the continuous straight lines coming downward from the right is the prolongation of the K line from the heavier elements downward; the upper is the prolongation of the $K\alpha$ line. The fact that these intersect proves that linear extrapolation from the range of heavier atoms is unjustifiable.

Reverting to the graph in Fig. 10, the problem of properly extra-

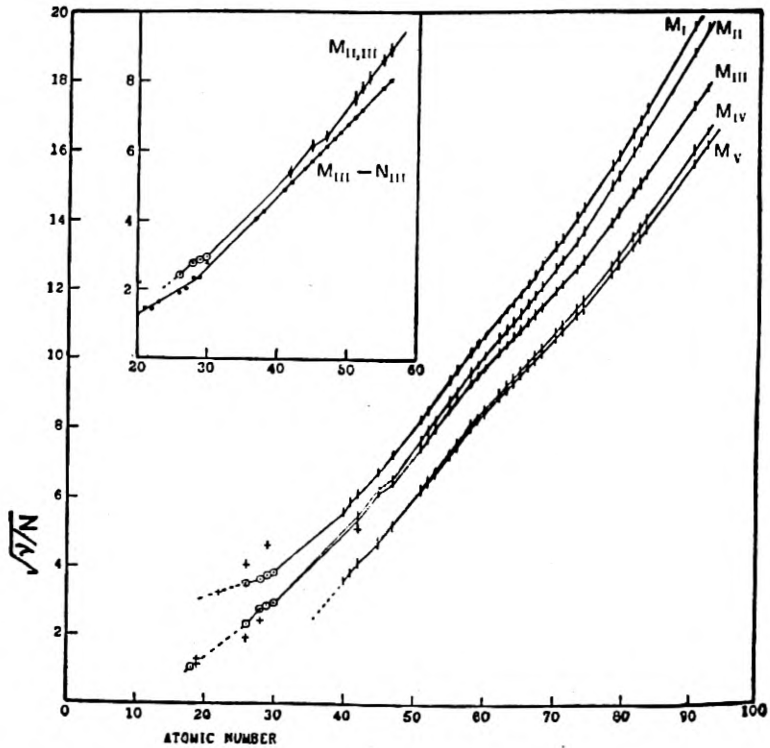


Fig. 13—Excitation-potentials correlated with M electrons. (*Philosophical Magazine.*)

polating the L and M curves is clearly not so simple as it was for the K curve; since they extend over shorter segments and do not come so far down into the range of light elements. In Fig. 12 (from McLennan and Clark) the circles for the elements from number 3 to number 17 represent observed excitation-potentials which they attribute to displacements of L electrons; those for the elements from number 30 to number 69 represent the highest and the lowest recorded emission-frequencies of the L-series for these elements, frequency being translated into equivalent voltage by the same relation as above. As for the excitation-potentials of the heavier elements, few measurements on potentials of the L class have been made, and very few indeed upon potentials of the M class—not nearly enough for an extrapolation. The deficiency is partially compensated by calculating the L and M excitation-potentials from the K, L and M emission-frequencies—an elaborate process, requiring a good deal of care in measuring and properly interpreting the various emitted rays. In this manner the potentials for the group of five M levels have been estimated for the various elements from the ninety-second down to the fortieth, and Horton has attempted to link onto them certain excitation-potentials which he and others observed when bombarding elements between number 20 and number 30 with electrons (Fig. 13). The curves must be supposed to bend, somewhere between the thirtieth and the fortieth elements; it is in this region that the M electrons pass from the status of deep-lying to the status of shallow-lying electrons. The excitations and emissions involving the shallow-lying electrons of the heavy atoms form a complicated system, of which the study has scarcely been begun, and will certainly prove perplexing. When research in this field is completed, each of the excitation-potentials and each of the emission-frequencies of every kind of atom will be entered upon curves, each of the curves corresponding to a definite and definitely-pictured process of rearrangement in the atomic electron-system, and extending over all the atoms of the periodic table which can be theatres of that process. This achievement may be reserved for a later generation.

LITERATURE

F. S. Brackett: Phys. Rev. 20, pp. 111–112; 1922.

British Association symposium "Spectra of the Lighter Elements"; Nature 112, pp. 217–224; 1924.

M. de Broglie and G. Friedel: C.R. 176, pp. 738–740; 1923.

G. Dejardin: C. R. 176, pp. 894–897; 1923. ibid. 178, pp. 1069–1071; 1924.

P. D. Foote and F. L. Mohler: *Origin of Spectra* (Chemical Catalog Co., 1922).

J. Franck: ZS. f. Phys. 11, pp. 155–160; 1922.

H. Fricke: Phys. Rev. 16, pp. 202–215; 1920.

G. Hertz: ZS. f. Phys. 18, pp. 307–316; 1923. Naturwiss. 11, pp. 778–779; 1923.

J. Holtsmark: Phys. ZS. 23, pp. 252–255; 1922.

F. Holweck: Annales de Physique, (9) 17, pp. 5-53 (1922). C. R. 173, pp. 709–712; 1922. (C. R. 176, pp. 570–573; 1923 (reflection of X-rays).

J. J. Hopfield: Phys. Rev. 18, p. 327 (1921), 20, pp. 573–588 (1922).

F. Horton, U. Andrewes, A. C. Davies: Phil. Mag. 46, pp. 721–741; 1923.

A. L. Hughes: Phil. Mag. 43, pp. 145-161; 1922.

E. H. Kurth: Phys. Rev. 18, pp. 461-476; 1921.

T. Lyman: *The Spectroscopy of the Extreme Ultra-violet* (Longmans, Green & Co., 1914). Astroph. Jl. 43, pp. 87–102; 1916. Nature 104, p. 314; 1922 (lines of ionized helium). Phil. Mag. 41, pp. 814–817 (1921) and Science, 56, pp. 167–168 (1922) (lines of neutral helium; with H. Fricke,) J. Opt. Soc. 7, pp. 495–499; 1923. (Vacuum grating spectrograph.)

C. A. Mackay: Phys. Rev. 21, p. 717; 1923.

J. C. McLennan: Proc. Roy. Soc. 95*A*, pp. 258–273; 1919 (with R. J. Lang). ibid. 98*A*, pp. 114–123; 1920. (Vacuum grating spectrograph). ibid. 102*A*, pp. 389–410; 1923 (with M. L. Clark; excitation potentials).

R. A. Millikan: Astroph. Jl. 52, pp. 47–64; 1920. ibid. *53*, pp. 150–160; 1921 (with R. A. Sawyer, I. Bowen). Phys. Rev. 23, 1–34; 1924 (with I. Bowen). Cf. also R. A. Sawyer.

F. L. Mohler and P. D. Foote: Bull. Bur. Standards 17, pp 471–496; 1923.

P. S. Olmstead and K. T. Compton: Phys. Rev. 22, pp. 559–565; 1923.

O. W. Richardson and C. B. Bazzoni: Phil. Mag. 42, pp. 1015–1019; 1921.

G. K. Rollefson: Phys. Rev. 23, pp. 35–45; 1924.

F. A. Saunders: Astroph. Jl. 40, pp. 377–384; 1914. (Vacuum arc spectra.) Science 59, pp. 50–51; 1924. (Ionizing potentials.)

R. A. Sawyer: Astroph. Jl. 52, pp. 286–300; 1920.

F. Simeon: Phil. Mag. 46, pp. 816–819; 1923. Proc. Roy. Soc. 102*A*, pp. 484–496; 1923. ibid. *104A*, pp. 368–375; 1923.

A. von Welsbach: Ann. d. Physik 71, pp. 7–11; 1923.

An Electrical Frequency Analyzer[1]

By R. L. WEGEL and C. R. MOORE

SYNOPSIS: An apparatus has been developed by means of which it is possible to measure and obtain a permanent record of the frequency components of an electric current wave. The device has two frequency ranges: 20 to 1250 cycles and 80 to 5000 cycles; the amount of power required does not in general exceed 500 microwatts; and the time necessary for making a record is about 5 minutes. An attachment is provided which permits of the making of simultaneous harmonic analyses of two complex waves in the same length of time.

In principle, the process consists in feeding the complex wave to be analyzed into a selective network, the essential feature of which is a sharply tuned circuit whose frequency of tuning is controlled by varying the capacitance in small steps with a pneumatic apparatus similar to that in a player piano. A maximum of response of the circuit occurs at each frequency of tuning which coincides with a component of the complex wave. An automatic photographic recorder of the response to each frequency of tuning is provided by means of which the frequency and magnitude of each component of the complex wave may be obtained. For convenience of operation, an automatic control apparatus is provided, so that it is only necessary to connect the complex source or sources to be analyzed and press a starting button. The completed record of the analysis is delivered after the machine has passed through the entire range of frequencies.

The application has so far been principally to problems in the communication field such as the analysis of performance and distortion at audio frequencies of vacuum tube and mechanical oscillators and amplifiers, analysis of complex telephone waves and speech sounds, and the effect on a complex wave of transmission through electrical and acoustic apparatus. In the power field many applications are obvious, such as for example, quantitative comparison as to frequency content of the voltage and current supplied to and delivered by transformers, voltage and magnetic flux studies in generators and motors, commutation, and the effect of wave-shapes in power transmission line problems and control apparatus.

INTRODUCTION

THE harmonic analyzer described in this paper consists of a variable tuned circuit into which the complex current wave to be analyzed is introduced, and an automatic recording apparatus to register its response as the frequency of tuning is changed.

The first recorded use of a tuned circuit as an analyzer was by Pupin in 1894.[2] He analyzed power waves by measuring the response of circuits tuned to each of the harmonic frequencies. It has been the practise for a number of years to determine the frequency characteristics of currents and voltages on power circuits and noise on telephone lines by means of a variable resonant circuit which includes a telephone receiver for listening.

[1] Presented at the Midwinter Convention of the A. I. E. E., Philadelphia, Pa., February 4-8, 1924.

[2] Resonance Analysis of Alternating and Polyphase Currents, Trans. A. I. E. E., Vol. XI, p. 523.

During the recent war a rapid automatic method was developed for varying the tuning of a circuit in such an analyzer in connection with the analysis of sounds radiated by submarines. The analyzer described in this paper is in principle the same as this apparatus but includes such improvements as were found desirable by experience to increase the speed, dependability and convenience of use. The present apparatus is capable of recording the frequency and magnitude of each component in a complex wave between 20 and 1250 cycles or 80 and 5000 cycles in about five minutes. This analyzer does not measure the phase of the various components but has the advantage that the frequencies need not be simple multiples of the fundamental as is the case with graphical analyzers. With this apparatus it is possible to measure quite accurately component frequencies as close together as about fifteen cycles at the lower end of the range and about 200 cycles at the upper end of the range, and to detect components as close together as three to five cycles at the lower end and fifty cycles at the upper end of the range.

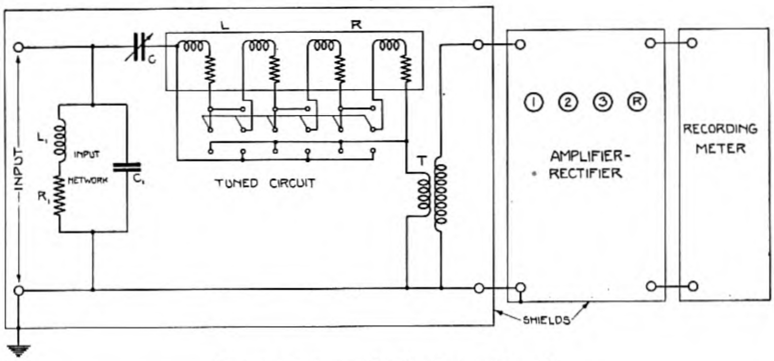


Fig. 1—Schematic Analyzer Circuit

Principles of Operation of the Analyzer

Fig. 1 is a schematic diagram of the essential elements of the analyzer circuit. The wave to be analyzed is introduced at the input terminals from which it passes to an input equalizing network and to the variable tuned circuit. The tuned circuit consists of a variable condenser of capacitance C and a coil whose inductance is L and resistance R. The value of the capacitance C is varied in small steps by an automatic device to be described in the next section. The inductance L consists of four identical windings on a toroidal core which, by means of a switch, may be thrown in series or in parallel, thereby changing the value of the inductance in the ratio of 16 to 1. With the same range of capacitance values this change in inductance gives the two frequency ranges of tuning, 20-1250 cycles and 80-5000 cycles. By means of the high-ratio transformer T the response of this circuit is applied to a vacuum tube amplifier-rectifier and registered by means of the recording meter.

This circuit arrangement will analyze a complex wave by virtue of the selective shunting of current by the tuned circuit from the input network. The impedance of the source of the complex wave is in practise maintained high in value at all frequencies compared to that of the input network so that the input wave-shape is independent of the small changes in impedance of the analyzer due to the varying of condenser C. The current fed into the analyzer traverses two paths, the input network and the tuned circuit. The impedances of these paths are respectively,

$$Z_1 = \frac{(R_1 + j\omega L_1)/j\omega C_1}{R_1 + j\omega L_1 + 1/j\omega C_1}$$

and

$$Z = R + j\omega L + 1/j\omega C.$$

The transformer T introduces into the tuned circuit a small resistance and inductance, both of which are negligible. The input network impedance Z_1 varies gradually from 0.4 ohms for direct current to about 10 ohms at 5000 cycles. The values of the elements are: $R_1 = 0.4$ ohms, $L_1 = 0.075$ milhenries, $C_1 =$ about 15 microfarads. Impedance Z of the tuned circuit depends on the setting of the variable condenser C. The resistance R of the iron-core coil, varies with frequency; its values for the parallel connection are 0.7 ohms for direct current, 1.5 ohms at 2500 cycles and 4.2 ohms at 5000 cycles The value of the inductance L for the parallel connection is 23.4 milhenries and is practically constant with change of frequency. For the series connection both R and L are sixteen times as great. The capacitance is varied from about 200 microfarads to about 0.05 microfarads. It will be seen that for each capacitance value there is a frequency, $f_r = 1/(2\pi\sqrt{LC})$, for which the tuned circuit impedance, Z, is R. For other frequencies Z is much greater due to the reactance. An incoming current of frequency f_r is, therefore, largely shunted through the tuned circuit while current of any other frequency passes through the input network. In this way if the capacitance C is varied gradually the tuned circuit will shunt selectively from the input network the successive components of the complex wave.

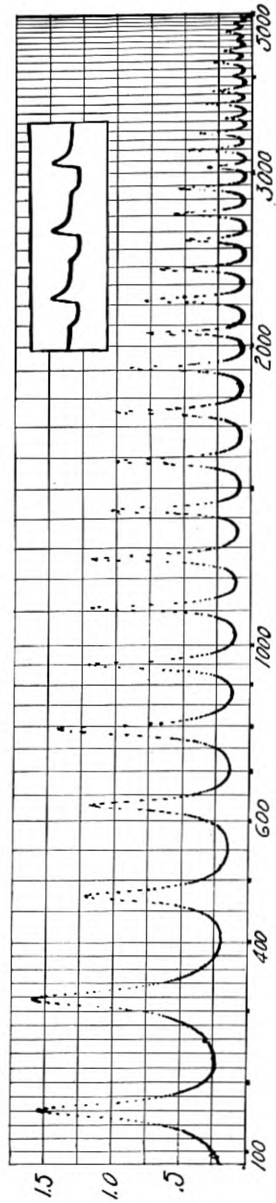


Fig. 2—Record of 160-cycle Buzzer Output

The special features of design of this analyzer circuit can be better explained by reference to a typical record made by the apparatus. Fig. 2 is the record of analysis of the current from a buzzer which vibrates with a frequency slightly under 160 per second and gives an irregularly shaped wave which is shown in the accompanying oscillogram. In taking this record the windings of the tuning inductance were in parallel so as to give the frequency range 80-5000 cycles. The vertical scale gives approximately the r. m. s. current in milliamperes at each frequency (as read on the horizontal scale) at which a peak occurs. It will be seen that a peak occurs at each multiple of the frequency of the buzzer. The r. m. s. values of input current at the corresponding frequencies as read from the peaks on the record are: 160, 1.6 milliamperes; 320, 1.6 milliamperes; 480, 1.25 milliamperes; 640, 1.2 milliamperes; 800, 1.45 milliamperes; 960, 1.25 milliamperes; 1120, 1.2 milliamperes; 1280, 1.1 milliamperes; 1440, 1.05 milliamperes; 1600, 1.0 milliamperes; 1760, 1.0 milliamperes; etc. The root square sum of all components shows that 4.7 milliamperes was the effective value of the complex current fed into the analyzer.

The fact that the 80-5000 cycle records read directly the current at each frequency component is due to the special design of the input network. A small correction is still necessary but can be neglected except where maximum obtainable accuracy is desired. If the input network were a pure resistance the higher frequency components would produce relatively lower peaks because of the falling off of efficiency with frequency of the amplifier-rectifier circuit and the increase in resistance of the tuning coil. The input network was designed empirically so that with constant input current the voltage drop across the input terminals increases with frequency in such a way as to compensate for these high-frequency losses. The tests to determine this were made by taking records of single frequencies of known amounts.

It will be seen that the frequency scale is gradually contracted as the upper end of the record is approached. Owing to the increase in resistance of the coil with frequency, the sharpness of tuning of the analyzing circuit decreases with frequency. Each peak on the record corresponding to a single frequency is a plot of the resonance curve of the variable tuned circuit. The sizes of the capacitance steps are so adjusted that a sufficient number of points, necessary to trace a resonance peak at all frequencies, is recorded. The length of the record and the time required for an analysis are determined by the number of points needed.

When peaks on the record are so close together as to overlap greatly, the reading on the scale is untrustworthy. If, instead of a rectifier and direct-current meter, an alternating-current meter giving deflections proportional to total r. m. s. values, were used, it would be theoretically possible to determine the component frequencies and amplitudes making up any composite peak, provided the number of frequencies could be determined. This procedure, however, would be impracticable. An examination of the theory of the rectifier shows that the problem of separation of the components of a composite peak is in general indeterminate. The rectifier however resolves adjacent peaks somewhat better than an alternating-current meter.

The analyzer has been most used in the analysis of audio-frequency currents for which the higher frequency range, 80-5000 cycles, is more useful. For the investigation of power problems the lower range would ordinarily be more suitable. In order to simplify the change from one frequency range to the other the tuning inductance only, is changed, leaving the mechanism for varying the capacitance in steps the same for both ranges. Since the inductance change in going from the high to the low-frequency range is in the ratio 1:16 and the change in the frequency range 4:1, the abscissas on the low-frequency records have one-fourth the value of those on the high-frequency records.

Since the smallest frequency divisions at the lower end of the high-frequency records are 20 cycles, these divisions on the low-frequency records are 5 cycles. There are, therefore, four times as many steps of tuning in the same frequency interval on the low as on the high-frequency record. The low-frequency record is therefore not of minimum practicable length. Since the same input network is used with the 20-1250 range as with the 80-5000 range, the low-frequency records are not direct reading in input current, but must be used with a calibration. Our use of the low-frequency range, however, has been so limited as not to justify the preparation of additional equipment for this use of the analyzer.

The apparatus is equipped with a device which permits of making simultaneous analyses of two complex waves. The principal reason for making such double records is to reduce errors in comparing two sources which may vary with time. The device may also be used simply to save time. It operates by connecting alternately to the analyzer the two complex waves in such a way that the record for each wave is traced by points representing alternate tuning condenser settings.

Description

The mechanism of the analyzer is so designed that to take a record it is only necessary, after starting the amplifier and connecting to a 110-volt power source, to attach the leads from the source or sources to be analyzed and press a starting button. The completed record is then automatically delivered in about 5 minutes after which the apparatus returns to the starting condition ready to repeat the

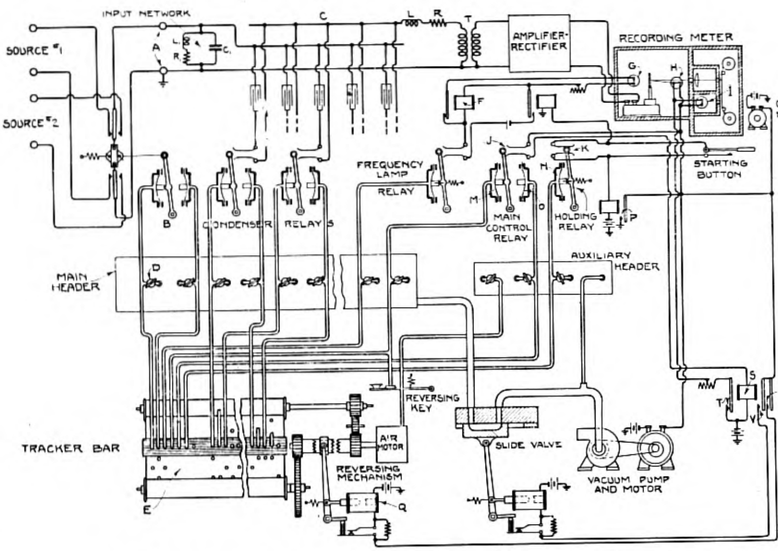


Fig. 3—Arrangement of Pneumatic and Electrical Apparatus

operation. This is accomplished by means of pneumatic apparatus operating in conjunction with a photographic recording device.

The pneumatic arrangement is a modification of a piano player mechanism in which a paper roll of standard dimensions is used. By proper perforation of the roll special pneumatic relays are operated in proper sequence to switch the condensers of the tuned circuit, flash frequency lines on the record, stop the mechanism after a record has been completed, rewind the piano roll, and perform other functions necessary to leave the analyzer in the starting condition. Electrical relays for switching the tuning condensers were not found practicable on account of the disturbances induced into the analyzer circuit.

The photographic recording apparatus consists of the camera motor for moving the sensitized record paper at a constant rate proper arrangement of lenses and lamps for illuminating the mirror galvanometer and tracing the scale and frequency lines, and suitable baths for developing and fixing the record. The record is drawn through the mechanism by means of two motor-driven rubber rollers, which also serve to remove excess solution.

The development of the pneumatic switching apparatus was carried out with a view to making use of as many standard piano player parts as possible. However, it was found necessary to make some

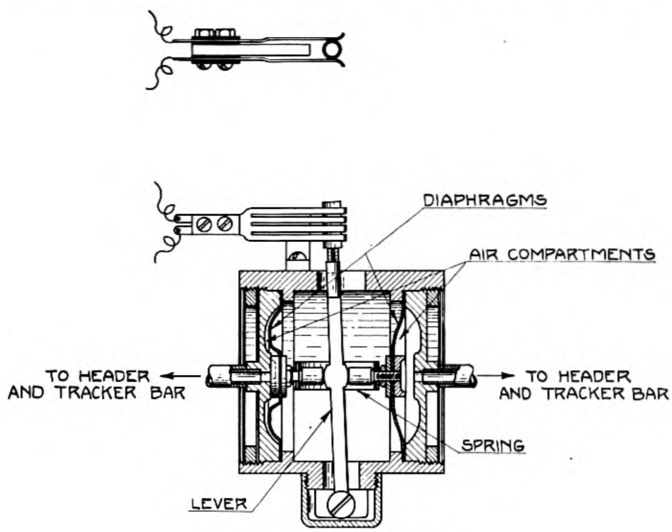


Fig. 4—Pneumatic Relay

modifications in method and apparatus; in particular a new pneumatic motor element (air relay) for switching the condensers at the requisite speed had to be developed.

Fig. 3 is a schematic drawing showing the principal features of the analyzer. In this drawing the vacuum pump is shown driven by an electric motor, and connected by means of pipes to the auxiliary and main headers and relays. This pump maintains in the headers an absolute pressure of about 4 or 5 lb. per square inch. The player piano roll E operates the entire mechanism by passing over the tracker bar in the usual manner. The air motor and tracker bar equipment are substantially as supplied by the manufacturers except that the reversing mechanism is arranged to be operated electrically instead of by hand.

The essential features of the air relay which was developed for this analyzer may be better understood by reference to Fig. 4. A cylindrical casting is arranged to mount two flexible diaphragms and two end plates in such a way as to form at each end of the cylinder, compartments, one side of each of which is a diaphragm. When assembled the two diaphragms face each other and are connected together by a circular spring made of steel strip. In use the two end compartments are partially evacuated thus causing the diaphragms to pull apart, straining the spring. When distended the diaphragms lie against the inner faces of the end plates which are shaped as shown. Obviously if air be allowed to enter either of the compartments the diaphragm belonging thereto will be pulled toward the other diaphragm by the spring. Passing through the circular spring is a lever pivoted at one end and carrying on the other end an insulated metallic sleeve. This lever is not attached in any way to either diaphragm and will of itself remain in position where last placed. Switch points are mounted in such a way that the sleeve may be forced in or out between them by the action of the diaphragms. This relay has proved very satisfactory in service and is particularly fast in its operation.

Connections between the tracker bar, main header, and the pneumatic relays are made by means of rubber tubing. As shown in Fig. 3 each of these relays requires two rubber tubes leading to the main header and two from the header to the tracker bar. These tubes are connected to the header by means of stop cocks D so connected that the direct passage of air from tracker bar to relay is practically unobstructed but the passage leading from the junction to the header may be made as small as desired by turning the finger valve. As adjusted, the opening to the header is small compared to the size of the tubes so that if air be permitted to enter one of the tube lines (as at the tracker bar), the diaphragm of the relay associated therewith is immediately released. When the tube is closed again, the entrapped air is soon removed through the small opening leading to the header thus restoring the diaphragm to its original position. The relay lever, however, does not follow the diaphragm.

This arrangement possesses the advantage that small openings only are necessary in the player piano roll, and that the opening which connects a condenser into the circuit is not in line on the roll with the opening which disconnects this condenser. Also at the beginning of an analysis by suitable perforations in the roll all air relays can be set simultaneously in the off position (condensers disconnected), thus making sure of the initial conditions. The apparatus is so designed that all the openings causing condenser circuits to close are on one side of the roll and those causing them to open are on the other side.

As before mentioned the tuned circuit is made up of inductance L, having some resistance R, and a bank of condensers designated by C. The function of the "Condenser Relays" is to connect into the tuned circuit any one or any combination of the 25 fixed condensers, thus tuning the circuit in small steps over a wide range of frequencies. The input is fed into this circuit as shown at A, and the degree of resonance, that is the response of the circuit at any

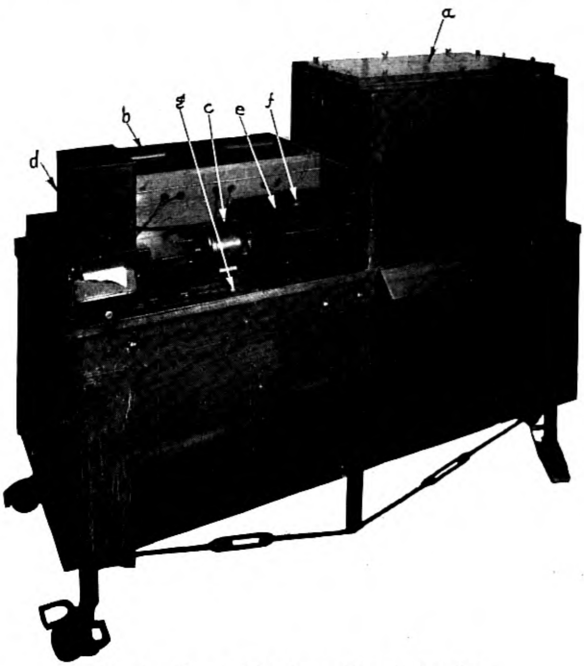


Fig. 5—View of Analyzer Ready for Use

a—Input and tuned circuits
b—Amplifier-rectifier
c—Camera motor
d—Recording meter
e—Control box
f—Starting button
g—Reversing key

particular frequency of tuning, is measured by means of the small transformer T, the amplifier-rectifier and the recording meter.

In addition to operating the tuned circuit a few of the air relays are used to operate the control circuits, mark frequency lines on the chart, etc., uses which required slight modification as indicated schematically in Fig. 3. In two of these control relays only one diaphragm is used, and the switch lever and diaphragm are fastened together by means of a flexible link. It has already been noted that the analyzer is equipped to trace two curves simultaneously on a single record. This is accomplished by means of air relay *B* which is so arranged as to connect two sources of input alternately to the analyzer. These input connections are alternated rapidly and are effected by appropriate punching of the roll.

The above covers the essential features of the analyzer but there

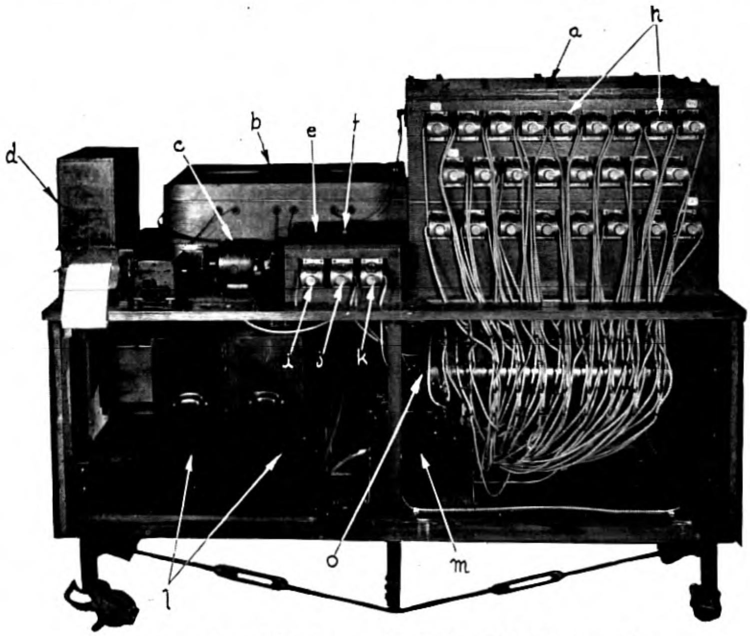


Fig. 6—View of Analyzer with Relay Side Uncovered

a—Input and tuned circuits
b—Amplifier-rectifier
c—Camera motor
d—Recording meter
e—Control box
f—Starting button
h—Condenser relays
i—Holding relay
j—Frequency lamp relay
k—Main control relay
l—Plate battery
m—Air motor
o—Main header

remain a few details having to do with assembly, control, etc., that may be of interest.

Fig. 5 shows the analyzer as completed and ready to operate. The apparatus is assembled on a two-deck, structural-steel table equipped with castors for convenience in handling. Much of the equipment is inclosed for protection against moisture and dust. The recording meter, camera motor, amplifier-rectifier, control relays, and input and tuned circuits are placed on the top. Below are mounted the batteries, the vacuum pump with its motor, and the tracker bar with paper roll mechanism. The arrangement is clearly shown in Figs. 6 and 7 taken with the protecting panels removed. In Fig. 6, *a* is a moisture-proof box containing the input and tuned circuits. The inductance coil is placed in the center of the upper half of this box together with the switch for connecting the windings

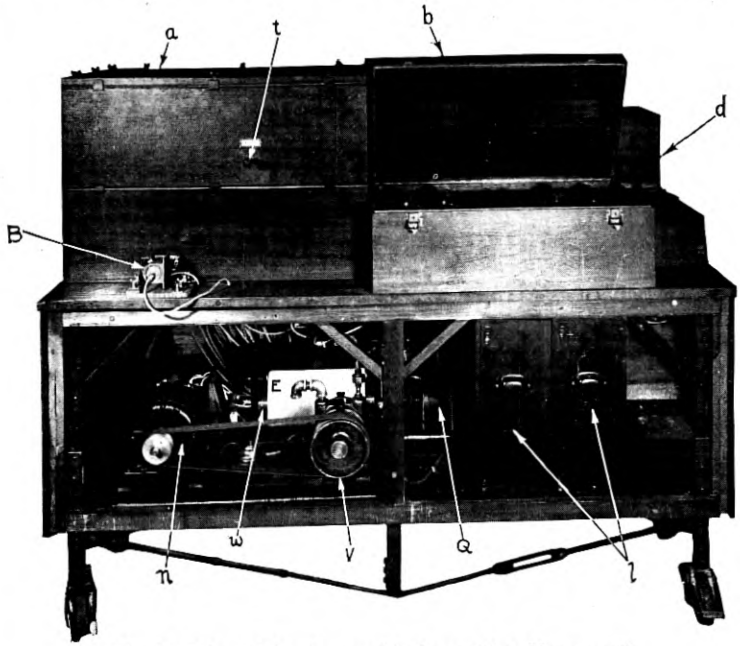


Fig. 7—View of Analyzer with Pump Side Uncovered

a—Input and tuned circuits
B—Amplifier-rectifier
c—Source-alternating relay
d—Recording meter
E—Paper roll
l —Plate battery
n—Pump motor
Q—Reversing solenoid
t —Series-parallel coil switch
v —Vacuum pump
w—Tracker bar

in series or parallel. The smaller capacitances are of mica and are arranged around the coil and switch assembly in such a way that they may be connected with a minimum length of lead to the air relays which are located on one side of the box. The larger capacitance units are made up of paper condensers and are placed in the lower half of the box. The metal lined box *b* contains the amplifier-rectifier, and at *d* is shown the recording meter. Box *e* contains the control circuits with the necessary relays. The method of mounting the air relays, main vacuum header (attached to underside of table top), air motor, etc., is also clearly shown in this figure. Each air relay is equipped with two rubber tubes leading to adjustable cocks on the header which in turn are connected to the tracker bar. The three-control relays are also shown in Fig. 6. The vacuum pump is shown at v in Fig. 7. The piano roll E moves over the tracker bar w and is reversed by means of solenoid Q. In boxes l are placed the plate batteries for the amplifier-rectifier.

The control apparatus by means of which the analyzer becomes practically an automatic machine will now be described. Referring again to Fig. 3 it will be seen that there is provided an auxiliary header and an electrically operated slide valve. The functions of these devices will be discussed presently.

The machine is started by pressing the starting button which should be kept closed for a few seconds while normal vacuum is being established in the headers. The air motor then starts and the paper roll E begins to travel across the tracker bar. Perforations in the roll are so made that when the roll is in its initial position an opening allows air to enter chamber N of the holding relay. As soon as the paper starts, however, this opening is closed, chamber N is exhausted, and contacts K close. This short-circuits the starting button which the operator may now release, and the machine is in full operation. It will be noted that the closing of the contacts of the starting button or contacts K puts into operation motors which drive the vacuum pump and the camera apparatus. Simultaneously recording meter lamp H and scale-line lamp I are lighted. The latter illuminates the record through small holes in an opaque scale strip thus marking horizontal lines due to the motion of the record.

As the roll E traverses the tracker bar, appropriate perforations control the condenser relays so as to switch the proper condensers into and out of the tuned circuit. Proper perforations also control the frequency lamp relay which flashes frequency lines on the record by means of Lamp G. Relay F is inserted in order to make the flash of short duration.

The tracker bar-paper-roll apparatus was received as a unit from the manufacturer and was installed after making modification in the reversing mechanism as mentioned above. This was done in the interest of automatic control. The paper roll is kept in its proper course over the tracker bar by means of an automatic adjusting device such as used in practically all high grade player pianos.

As the paper progresses over the tracker bar a point is finally reached where the last condenser connections are made and it becomes necessary to rewind the roll and to restore the entire mechanism

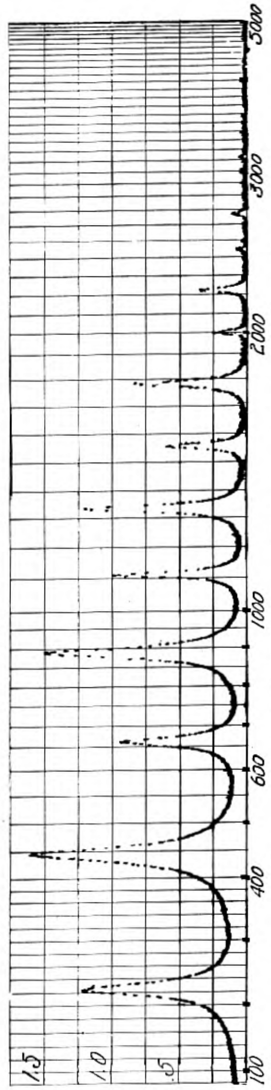


Fig. 8—Output of Carbon Button Driven at an Excessive Amplitude

to its starting condition. This is accomplished by means of a perforation at the end of the record which admits air to chamber *M* of the main control relay, thus closing contacts *J*. Relay *S* then operates since its circuit to ground is completed through contacts *P*. Operation of relay *S* opens contacts *T* thus disconnecting lamps *H* and *I*, and closes contacts *U* and *V*. It will be seen that the closing of contacts, *U* operates the reversing mechanism, and rewinding of the roll begins immediately. The closing of contacts *V* operates the slide valve thus releasing the vacuum on the main header, allowing the roll to be rewound with minimum mechanical drag.

It may be noted that means are also provided for rewinding the roll from any point in its forward travel by admitting air manually at the reversing key. This will cause the main control relay to operate so that rewinding will begin. Vacuum is kept on the auxiliary header during the rewind so that control of the analyzer may be maintained to the end of the operating cycle.

When the paper has been completely rewound perforations allow air to enter simultaneously chamber *O* of the main control relay and chamber *N* of the holding relay. This action opens contacts *J* and *K*, thus bringing the entire mechanism to rest in its initial starting condition.

Applications

To show the variety of problems in which the analyzer is a useful means of investigation, a few illustrative records have been made and will be discussed. These records were taken in each case to illustrate the use of the analyzer and are not parts of investigations to which they are related. They cannot, therefore, be taken as representative of the performance of the apparatus tested.

One of the uses of the analyzer has been in the study of the performance of microphone buttons. Fig. 8, for example, illustrates the character of the distortion in a button when driven at an excessive amplitude. The button was mounted so that its movable electrode could be driven at a single frequency by a very heavy reed at its natural frequency so that the motion was very nearly sinusoidal. The frequency of the motion was a little less than 450 cycles corresponding to the second peak on the record. The amplitude of motion was 0.001 centimeters or 0.0004 inches which is of course much greater than normally obtains in a transmitter. The circuit consisted simply of the button and a battery in series with the analyzer so that the record is an analysis of the current fluctuations in the button. The record shows two series of frequencies generated

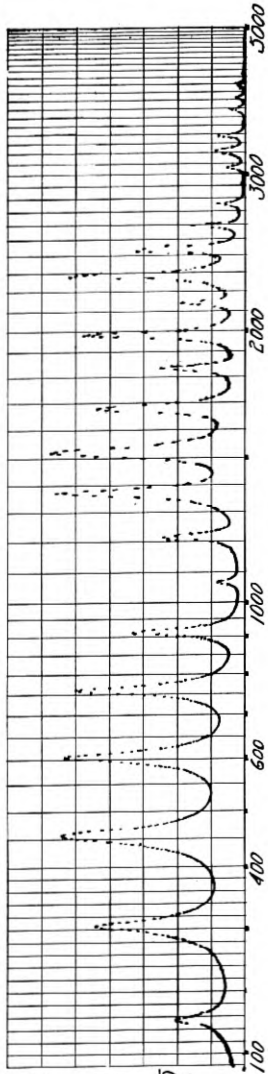


Fig. 9—Noise in Room as Picked up by Condenser Transmitter

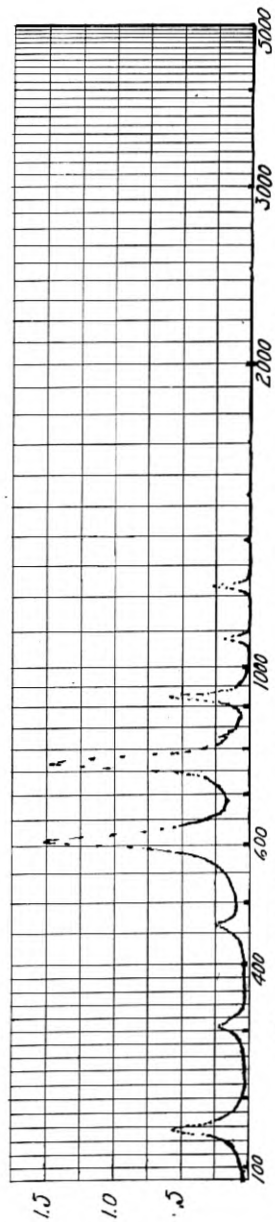


Fig. 10—Noise in Room as Picked up by Telephone Receiver Used as a Transmitter

by the button; a primary series having for its fundamental the driving frequency, 450 cycles, and a subsidiary series, having for its fundamental half the driving frequency or 225 cycles. The even harmonic components of the secondary series coincide, of course, with the frequencies of the primary series. The primary series can be accounted for by the fact that with such large amplitudes the changes in resistance are not a linear function of the amplitude of

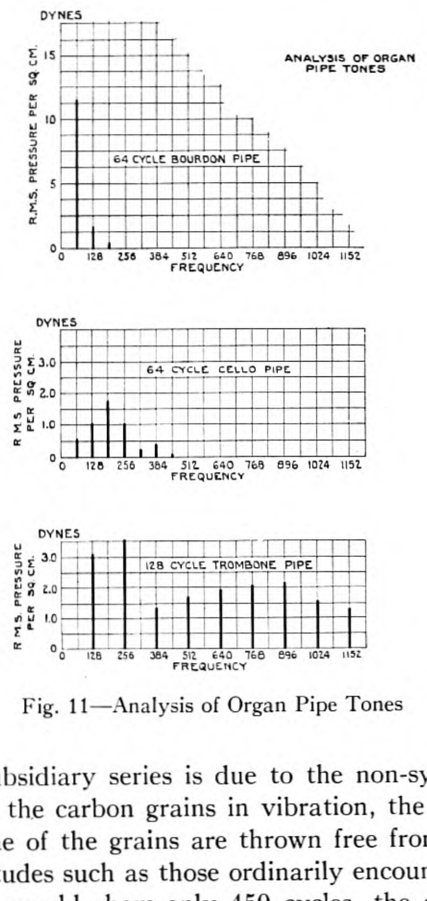


Fig. 11—Analysis of Organ Pipe Tones

motion. The subsidiary series is due to the non-symmetrical effect of the inertia of the carbon grains in vibration, the motion being so violent that some of the grains are thrown free from their contacts. For small amplitudes such as those ordinarily encountered in a transmitter, a record would show only 450 cycles, the other frequencies occurring in negligible amount; for intermediate amplitudes the primary series only occurs.

The analyzer has been used in connection with the study of sustained sounds and of the performance of acoustical apparatus. Fig. 9 is a record of the noise in a room originating from a buzzer as

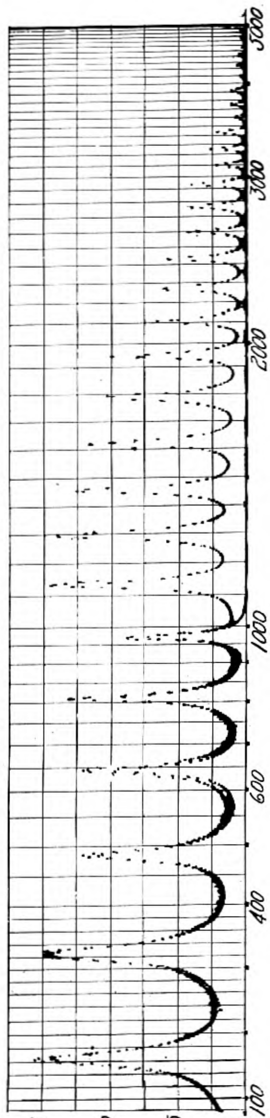


Fig. 12—Record Showing Action of Low Pass Filter

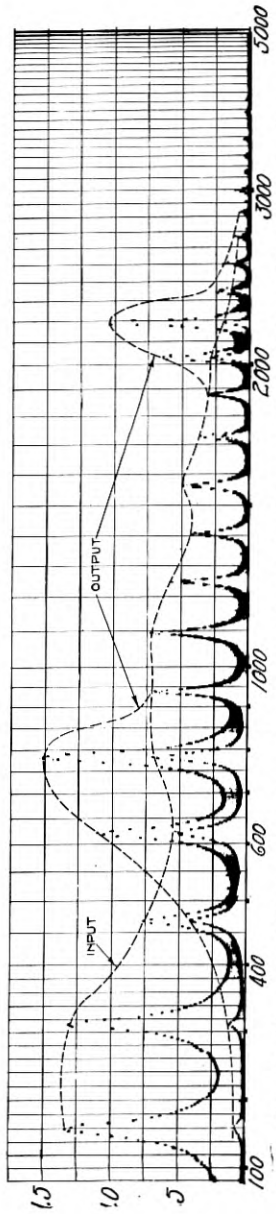


Fig. 13—Records of Electrical Input and Acoustic Output of a Common Type of Loud Speaking Receiver

picked up by a condenser transmitter.[3] The reverberation in the room probably had a large effect on the character of this record. With such a source of frequency the analyzer may be used to study the acoustics of rooms. Fig. 10 is a record of the same noise as in Fig. 9 but as picked up by a common type of telephone receiver placed in the same position as the condenser transmitter. A comparison of Figs. 9 and 10 will show the inadaptability of such a receiver for use as a transmitter. The receiver, owing to the resonance of its diaphragm, is seen to be relatively sensitive in the region of 600 to 800 cycles and insensitive at most other frequencies. When this instrument is placed against the ear, as when used as a receiver, the diaphragm resonance is damped so as to give more nearly uniform response.

By means of the calibration of the condenser transmitter and its amplifier, it is possible to make an analysis of the absolute intensity of a sustained sound in the air. This method has been used to study the frequency characteristics of musical instruments. Fig. 11 shows the analyses of three low-frequency organ pipes. These are plots of r. m. s. pressure change in the sound wave as obtained from the analyzer records. Each vertical line corresponds to a peak on the original record. The upper chart shows the almost pure tone given by a 64-cycle Bourdon pipe. In the case of the cello pipe, also having a fundamental of 64 cycles, the third harmonic is seen to be more prominent than the fundamental or second harmonic. The third chart is for a 128-cycle trombone pipe which was found to be rich in harmonics. The pressure in the single components of the cello and trombone pipes is less than in the case of the Bourdon pipe, and a larger scale of ordinates is therefore used.

To illustrate the use of the attachment which permits the making of two simultaneous analyses, a few double records will be presented. An electric wave filter which has been used in the study of telephone quality was connected to the buzzer source whose output is shown in Fig. 2. Simultaneous analyses of the current delivered to and transmitted through the filter are shown in Fig. 12. This filter is designed to pass all frequencies below 1000 cycles and to suppress all others. The input is represented by a more or less continuous series of peaks along the entire length of the record. The peaks corresponding to the output coincide rather closely with the input

[3] "A Condenser Transmitter as a Uniformly Sensitive Instrument for the Absolute Measurement of Sound Intensity." E. C. Wente, *Physical Review*, July 1917.

"The Sensitivity and Precision of the Electrostatic Transmitter for Measuring Sound Intensities." E. C. Wente, *Physical Review*, May 1922.

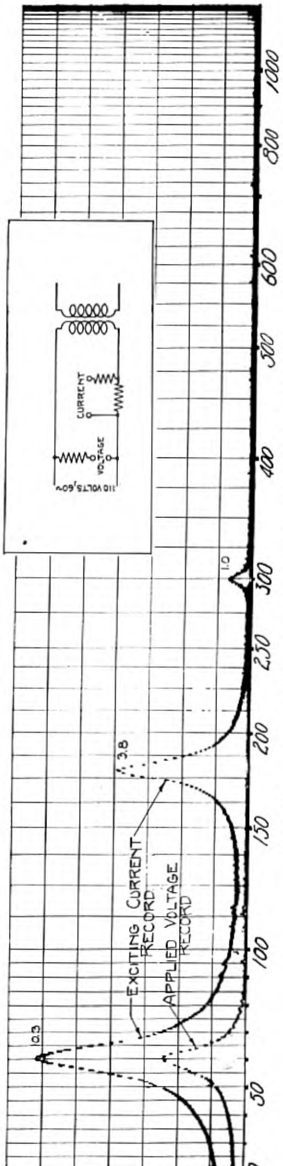


Fig. 14—Record Taken on Transformer at No Load

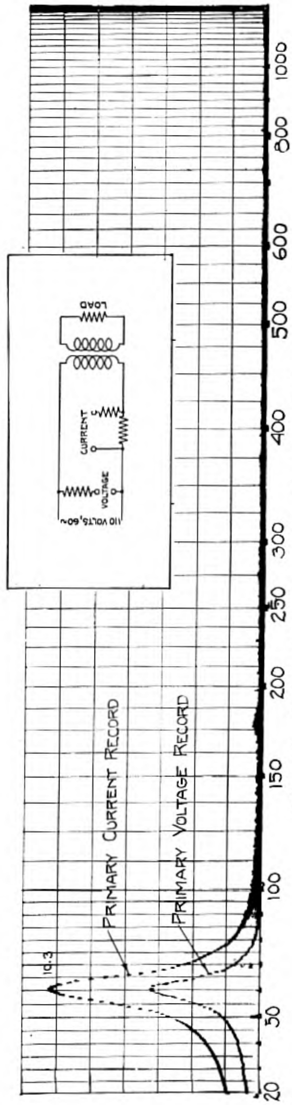


Fig. 15—Record Taken on Transformer Under Load

peaks for all frequencies below 1000 cycles and are not detectable for the higher frequencies.

Fig. 13 is a double record showing the analysis of the wave from a buzzer as fed into a common type of loud speaking receiver and the acoustic output as picked up by a condenser transmitter placed in front of it at a distance of about 15 inches. The analysis of the input current wave to the loud speaker is shown by the comparatively continuously decreasing series of peaks. The acoustic output is represented by the series having maxima in the neighborhood of 800 cycles and 2200 cycles. This record cannot be taken as an adequate analysis of this loud speaker because of probable reverberation effects in the room.

The analyzer has thus far not been used in the study of power problems. A few illustrative records have been taken, however, on transformers and generators and will be shown as suggestive of the use of this method of attack in such problems.

Fig. 14 is a double record showing applied voltage and exciting current of a small 110-volt, 60-cycle transformer operating at normal voltage and frequency under the no-load condition. The presence of the well known third and fifth harmonics in the exciting current is clearly shown. Because of the rise in the calibration curve of the analyzer at the low end of the lower frequency range, a scale of ordinates is not shown on this record. Instead, the values of the analyzer current at each frequency are noted on the record. The circuit used in making this record is drawn on the figure. A computation of the components of the exciting current from the record and constants of the circuit shows that at 60 cycles the current was 175 milliamperes, at 180 cycles, 65 milliamperes and at 300 cycles, 17 milliamperes. The total r. m. s. exciting current was therefore 187 milliamperes.

The operation of this transformer under full load is shown in Fig. 15, where, as before, the primary voltage and current are analyzed. The transformer load consisted of a pure resistance. It will be noted that the third and fifth harmonics have become very small compared with the fundamental. The analyzer currents at each frequency are again noted on the record. In obtaining the analysis of the current it was necessary to further shunt the analyzer. The primary current was 310 milliamperes.

Problems relating to commutation may also be conveniently studied qualitatively and quantitatively by means of the analyzer. The use of an apparatus which will indicate the source and measure the extent of parasitic frequencies is obvious. Information has

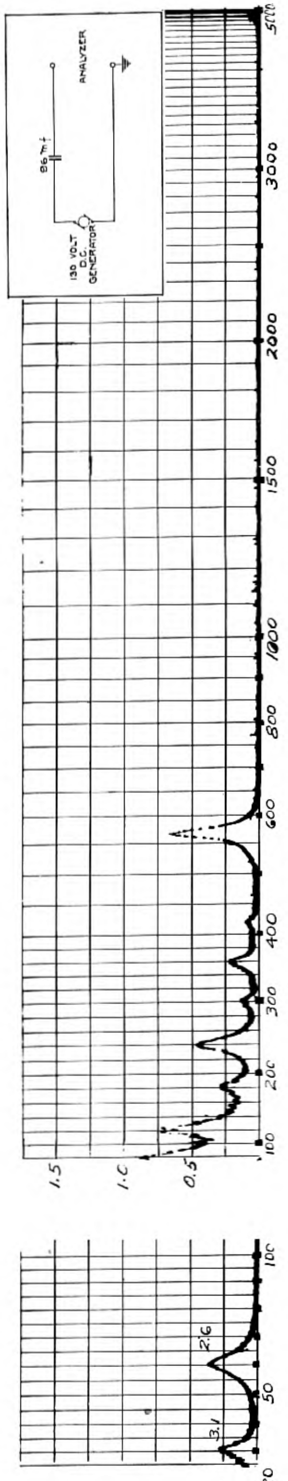


Fig. 16—Record Taken on D. C. Generator at No Load

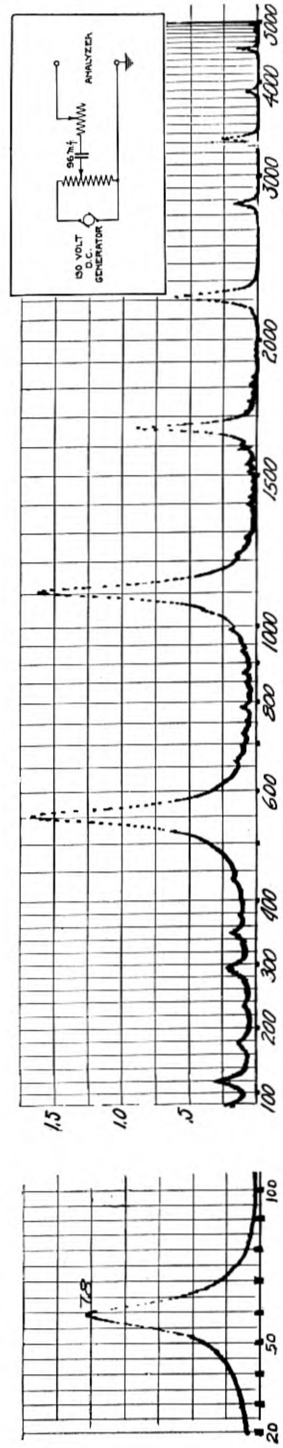


Fig. 17—Record Taken on D. C. Generator Under Load

been obtained on a small machine direct-driven by a ½-h. p., 60-cycle single-phase motor. Data of importance relating to the generator tested are as follows:

Capacity of Generator	¼ kw.
Number of Poles	2
Speed	1725-1800 r. p. m.
Voltage	125
Field	Shunt-connected
Diameter of Commutator	2.75 in.
Number of Commutator Bars	38
Number of Armature Slots	19
Size of Brush	⅜ in. square
Yoke	Ring type

Records obtained from this machine when operating under no-load and half-load conditions are shown in Figs. 16 and 17, respectively. The corresponding speeds are approximately 1800 and 1750 r. p. m. In order to show what frequencies the machine gives out over the entire range 20 to 5000 cycles each figure is made up of two parts: a portion of a 20-1250 record and a complete record over the range 80-5000 cycles. On each figure is drawn the circuit connecting the d-c. generator to the analyzer. It will be noted that a large condenser is inserted to prevent the passage of heavy direct current through the analyzer.

The consideration of these records leads to the conclusion that there are at least three independent major causes of alternating voltage operating in this d-c. machine. The fundamental frequencies due to these causes are 30, 60 and 570 cycles. It will be noted that the 30-cycle peak occurs only on the no-load record under which condition the average speed is practically 30 revolutions per second. Sixty cycles and a series of its harmonic overtones are seen to be present under both conditions of load. Under load the 60 cycles is augmented whereas its harmonics are reduced. No harmonic overtones of 30 cycles except such as might coincide with the harmonics of 60 cycles are found in either case. This indicates the existence of independent causes of the 30 and 60-cycle frequencies, that the 30-cycle cause produces an almost sinusoidal voltage, and that the 60-cycle cause under no load produces an irregular wave which becomes smoother as the machine is loaded.

The no-load record, Fig. 16, shows 570 cycles with no harmonics while the load record, Fig. 17, shows 570 cycles with a complete series of harmonics. This indicates that at no load the cause of 570 cycles feeds a relatively smooth wave to the line while under load this cause feeds an irregular wave to the line. The fact that 1140 cycles is about as strong as the fundamental and that its harmonics are stronger than alternate ones which are overtones of 570 only, suggests the likelihood of a fourth cause having a frequency of 1140 cycles. Small irregularities at frequencies other than those already mentioned occur in the record. These are more prominent under load than at no load and indicate the presence of small, more or less irregular pulses, which increase with load. All of the above frequencies may be accounted for by a consideration of the construction and operating condition of the machine.

The generator was driven by a single-phase, 4-pole, 60-cycle motor which may give rise to torque fluctuations once per revolution, or 30 times per second. Under no load this may produce considerable corresponding fluctuations in speed while under load conditions the generator acts as a damper, eliminating these oscillations.

The 60-cycle peak may be due to any one or some combination of a number of causes, *e. g.*, eccentricity of generator armature, non-uniform winding, non-uniform thickness of mica separators in commutator, high mica between one or more pairs of segments, etc. The records show that for this particualr machine in its present condition (new) at normal speed the 60-cycle voltage developed increases considerably with load indicating strongly that the cause is largely influenced by an IR drop somewhere in the machine. The most likely causes therefore appear to be commutator eccentricity, irregular spacing of the segments, or high mica.

The peak at 570 cycles may be accounted for by cyclic variation of flux entering the armature core as the teeth pass the pole faces. At no load the speed is approximately 1800 r. p. m. The number of teeth being 19, it is obvious that there will be 570 fluctuations of air-gap reluctance per second. Under no-load conditions the record shows a comparatively pure wave form for this cause. This is to be expected because of the comparatively uniform distribution of flux under the pole faces at no load. As the machine is loaded, however, the field is distorted and shifted giving rise to an irregular wave form of voltage which is responsible for at least a part of the large harmonic content shown by the load record.

The presence of 1140-cycle peak which is present only under the load condition may be due to the cyclic variation of voltage produced by the commutator bars leaving the brushes. Inasmuch as the speed is roughly about 29 revolutions per second the frequency with which bars leave brushes is about 1100 cycles. This frequency is present under the load condition only, thus indicating that it is due to an *I R* drop at the brush contacts or to an e. m. f. developed in the short-circuited coil with the brush off the magnetic neutral.

The very small irregularities on the record shown particularly between peaks above 550 cycles on the load record are probably due to slight chattering of the brushes.

It is of interest to note that the so called frequency of commutation does not appear in either of the records. For this machine this frequency at no load is approximately 346 cycles per second.

From these records it is possible to determine the r. m. s. value of the alternating voltage at any frequency of interest. This is computed from a knowledge of the circuit constants and analyzer impedance. We thus obtain for the 550-cycle peak (Fig. 17) a value of 0.8 volts and for the 60-cycle peak a value of 1.1 volts.

In general the records taken by means of the analyzer on this commutating machine, confirm quantitatively the well known fact that such machines may give rise to frequencies in the audible range. Consideration of the records indicates that these frequencies may be divided into two classes: First, those pertaining to and controlled by design, and second, those caused and controlled by the physical condition of the machine at any particular time. It is also interesting to note that the driving motor may produce an appreciable effect, particularly under the no-load condition.

Summary

In the above paper there has been given a short statement of the theory and construction of an automatic, recording, electrical frequency analyzer, together with illustrations showing its use and limitations in various fields.

This apparatus has been found very useful in the laboratory in the investigation of many different types of problems chiefly because of the speed with which records can be made and harmonic analyses obtained without computation.

In conclusion the authors wish to express their appreciation to Mr. C. E. Lane and Mr. C. E. Dean, of the Western Electric Company, Inc., for their assistance in the building of this machine and the preparation of this paper.

Certain Factors Affecting Telegraph Speed[1]

By H. NYQUIST

Synopsis: This paper considers two fundamental factors entering into the maximum speed of transmission of intelligence by telegraph. These factors are signal shaping and choice of codes. The first is concerned with the best wave shape to be impressed on the transmitting medium so as to permit of greater speed without undue interference either in the circuit under consideration or in those adjacent, while the latter deals with the choice of codes which will permit of transmitting a maximum amount of intelligence with a given number of signal elements.

It is shown that the wave shape depends somewhat on the type of circuit over which intelligence is to be transmitted and that for most cases the optimum wave is neither rectangular nor a half cycle sine wave as is frequently used but a wave of special form produced by sending a simple rectangular wave through a suitable network. The impedances usually associated with telegraph circuits are such as to produce a fair degree of signal shaping when a rectangular voltage wave is impressed.

Consideration of the choice of codes show that while it is desirable to use those involving more than two current values, there are limitations which prevent a large number of current values being used. A table of comparisons shows the relative speed efficiencies of various codes proposed. It is shown that no advantages result from the use of a sine wave for telegraph transmission as proposed by Squier and others[2] and that their arguments are based on erroneous assumptions.

Signal Shaping

SEVERAL different wave shapes will be assumed and comparison will be made between them as to:

1. Excellence of signals delivered at the distant end of the circuit, and

2. Interfering properties of the signals.

Consideration will first be given to the case where direct-current impulses are transmitted over a distortionless line, using a limited range of frequencies. Transmission over radio and carrier circuits will next be considered. It will be shown that these cases are closely related to the preceding one because of the fact that the transmitting medium in the case of either radio or carrier circuits closely approximates a distortionless line. Telegraphy over ordinary land lines

[1] Presented at the Midwinter Convention of the A. I. E. E., Philadelphia, Pa. February 4-8, 1924, and reprinted from the Journal of the A. I. E. E. Vol. 43, p. 124, 1924.

[2] A. C. Crehore and G. O. Squier. "A Practical Transmitter Using the Sine Wave for Cable Telegraphy; and Measurements with Alternating Currents upon an Atlantic Cable." A. I. E. E. Trans., Vol. XVII, 1900, p. 385.

G. O. Squier. "On An Unbroken Alternating Current for Cable Telegraphy." *Proc. Phys. Soc.*, Vol. XXVII, p. 540.

G. O. Squier. "A Method of Transmitting the Telegraph Alphabet Applicable for Radio, Land Lines, and Submarine Cables." *Franklin Inst., Jl.*, Vol. 195, May 1923, p. 633.

employing direct currents will next be considered. This will be followed by a consideration of the more complicated case of transmission over long submarine cables.

It will be shown that the waves produced by sending rectangular signal elements through suitable electrical networks which round them off before they are impressed on the transmitting medium are probably best in most cases. Comparsion will be made between waves shaped by sending rectangular signal elements through suitable networks and waves made up of half cycles of a sine wave, bringing out the inferiority of the latter.

Direct-Current Telegraph Transmission Over a Distortionless Line

Before proceeding with this discussion two terms, which will be used in this paper, and which are considered to be of fundamental importance, will be defined—"signal element" and "line speed." It is usually possible, especially when sending is done mechanically, to divide the time into short intervals of approximately *equal* duration, such that each is characterized by a definite, not necessarily constant, voltage impressed at the sending end. The part of the signal which occupies one such unit of time will be called a "signal element." For example, the letter *a* in ordinary land telegraphy will be said to be made up of five signal elements, the first constituting a dot, the second a space and the next three a dash. The "line speed," as used in this paper, equals the number of signal elements per second divided by two. In ordinary land telegraphy the line speed is equal to the dot frequency when a series of dots separated by unit spaces is transmitted.

The discussion will first be limited to the case of direct-current telegraphy over a distortionless line. This case is the simplest, and in addition the results will aid in understanding the more complex cases. It may aid in obtaining an understanding of this case to assume that the distortionless line is made up simply of series and shunt resistances.

A distortionless line, such as the one which has been assumed, will transmit all frequencies with equal efficiency from zero upward. In considering applying direct-current telegraph to this line, it will be assumed that the telegraph circuit will have assigned to it only a limited range of frequencies from zero upward, the remaining frequency range being assigned to some other uses, such as ordinary telephone and carrier telephone and telegraph. It will also be assumed that the direct-current telegraph circuit is worked at as high a speed as the frequency range assigned to it will permit.

A number of different wave forms which might be employed to make up the telegraph signal elements will next be examined, consideration being given first to the waves which will be received at the distant end when the different wave forms are impressed at the transmitting end and second to the interference which will be produced in the higher range of frequencies which has been assigned to other uses.

Three forms of voltage waves which will be considered are shown in Fig. 1. *A* in that figure shows the simplest form of voltage wave,

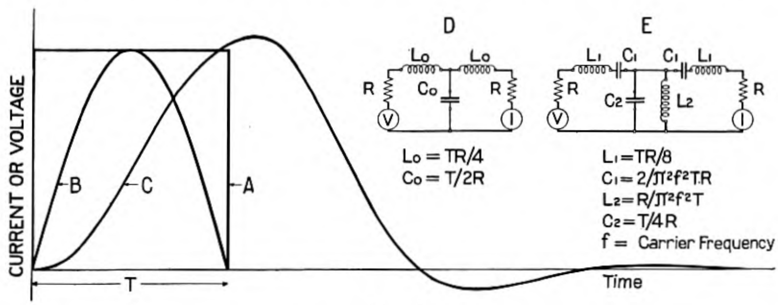


Fig. 1

A—Rectangular Voltage Wave
B—Half Cycle of Sinusoidal Voltage Wave
C—Rectangular Voltage Wave Modified by Being Passed through Network Shown at *D* or *E*.

namely, the ractangular form which is produced by applying a battery for a given interval of time and then substituting a short circuit for it. *C* in the figure is the wave produced by transmitting the rectangular voltage wave *A* through an electrical network which is the one indicated by the letter *D* in the figure. (Other forms of networks might also be selected which would produce similar results.) *B* in the figure is a wave which has the shape of a half cycle of a sine wave. In what follows this wave will be referred to as the "half-cycle sine wave."

In considering the waves which will be received when the above waves are applied at the transmitting end, use will be made of the following general principles, which have been stated by Malcolm,[3] for the case of a submarine cable circuit and discussed for the general case in Appendix A.

[3] H. W. Malcolm. "Theory of the Submarine Telegraph and Telephone Cable." The Electrician Printing & Publishing Co., London, March 1917.

When a telegraph circuit is worked at a line speed as high as will be permitted by the available frequency range, the shape of the received signal will be practically independent of the shape of the transmitted signal, and further, the magnitude of the received signal will be approximately directly proportional to the area included within the impressed voltage wave.

The area included within the impressed voltage wave being of principal importance so far as the wave received at the distant end is concerned, the areas under the three voltage waves shown in Fig. 1 will next be examined. The areas under waves A and C will be found to be substantially equal while the area under the wave B is only about 0.6 as great. Consequently, it should be expected that waves A and C will be about equally good from the standpoint of the received signals, while wave B will be poorer, producing received signals only about 0.6 as great in magnitude. If the maximum voltage (or power) impressed at the sending end is limited to some given value, the rectangular wave is seen to be the optimum, since this wave has the maximum area. While the area shown under curve C is approximately equal to that under the rectangular wave, the effect produced when a number of signal elements of the same polarity and magnitude are sent in succession is such that the maximum voltage transmitted will exceed slightly the corresponding voltage for the case of the unmodified rectangular wave due to overlapping of adjacent signal elements.

The above comparison of the three waves of Fig. 1 from the standpoint of received signals holds not only for signal elements, but also for complex waves comprising a number of elements. Since for the speeds under consideration the received currents for different shapes of signals applied at the sending end are substantially of the same form, differing, at most, in magnitude, it follows from the principle of superposition that any complex signal, whether built up of elements of one shape or another at the sending end, will produce substantially the same wave form at the receiving end, the differences in the shapes of the elements at the sending end producing differences principally in magnitude of the received waves.

Consideration will next be given to the relative interference which the different wave forms of Fig. 1 will produce in the frequency range assigned to other circuits. Since interference into other circuits results from having the telegraph signal elements contain frequencies which spread into the ranges assigned to other circuits, it is evident that the wave will be the best from the standpoint of interference which contains the least amount of these outside frequencies. By making use of a method which is discussed in Appendix C, the frequency components of the three waves illustrated in Fig. 1 have been computed and are shown in Fig. 2. The frequency marked 1/2 T in the drawing equals the line speed. T in this connection has the same value as in Fig. 1. The letters in this figure refer to the corre-

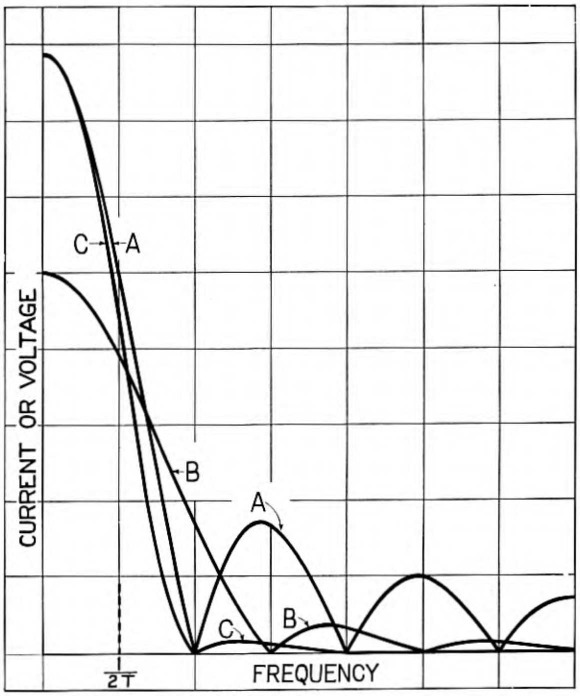


Fig. 2

A—Frequency Components of a Single Dot, Rectangular Wave
B—Frequency Components of a Single Half Cycle of a Sine Wave
C—Frequency Components of a Single Dot, Rectangular Wave Passed through Network Shown in Fig. 1

sponding waves in Fig. 1, A being the components of an isolated rectangular wave, B the corresponding components for the half-cycle sine wave, and C those for the rectangular wave after it has been transmitted through the network D in Fig. 1. It is seen from Fig. 2 that the rectangular wave form A contains the greatest amount of currents of higher frequencies and is, therefore, the poorest from the standpoint of interference. The half-cycle sine wave contains less of these higher frequencies although, as will be seen, the high-frequency components are far from negligible. The wave C is the best from the standpoint of interference, since it contains the least amount of these higher frequencies.

From the preceding it is concluded that for the case under consideration, the wave form C in Fig. 1 produced by sending a rectangular shaped signal element through a suitable network is the most suitable. This wave form is almost the optimum from the standpoint of the received signals while from the standpoint of interference into other circuits it leaves little to be desired.

Carrier and Radio

The results for the distortionless line are particularly applicable to the cases of radio and carrier telegraphy because in these cases we have a transmitting medium which is substantially distortionless. We may again make use of Fig. 1 to illustrate three possible voltages, it being understood that these curves represent the envelope or outline of the transmitted currents which are in reality of a frequency considerably higher than the signaling frequency. If now we limit consideration to the case where the carrier frequency is located in the middle of the transmitted frequency band, then, this case becomes very similar to the direct-current case and what has been said about the received wave shape being independent of the transmitted one and its magnitude being directly proportional to the area under the transmitted voltage curve still holds. One important difference is that, whereas in the direct-current case the network shown at D, Fig. 1, is used in the alternating-current case having the carrier located in the middle of the free transmitted range, the network shown at E, Fig. 1, is used. A further difference is that in the case of radio where very high frequencies are involved, it may not be practicable to construct the required networks. In that case, however, it is practicable to produce the corresponding direct-current wave and utilize it to modulate the radio wave.

What was said about interference from the circuit in question into other circuits in the direct-current case above also holds for the case of radio and carrier with the difference that whereas Fig. 2 shows a band of frequencies extending from zero up, the corresponding curve in the case of radio and carrier consists of two such bands. The complete curve for radio and carrier is substantially symmetrical with respect to the ordinate corresponding to the carrier frequency, and the right-hand portion is similar to the curve shown in Fig. 2. It will be obvious that the rectangular wave and the half-cycle sine wave are both objectionable, as voltage waves to be applied to the transmitting medium, because they contain frequency components which may easily extend into the range allotted to neighboring carrier bands. For this reason it is customary in carrier telegraph practise to make use of a transmitting filter to cut off these interfering frequencies. The voltage impressed on this filter is substantially rectangular in outline but after passing the filter it has a shape which is approximately similar to curve C in Fig. 1, and which, therefore, produces less interference than a half-cycle sine wave.

Land Lines

The case of land lines is somewhat different from the case discussed previously because it is not economically desirable to utilize the full frequency range available. In other words, the great expenditure for terminal apparatus that may be proper in the case of submarine cables and long distance radio circuits is not warranted. In land circuits the highest frequencies transmitted are considerably greater than the required line speed. When this is the case, it is usually possible and desirable to make use of the available range to increase the steepness of the received wave. A steep wave front results in prompt operation of the receiving relay and this in turn results in minimum distortion. If a half-cycle sine wave were to be employed instead of the usual rectangular wave or if a network were to be employed which were to round off the wave to the extent indicated in Fig. 1, the received wave would necessarily lose a great part of its steepness and as a consequence the response of the receiving relay would be less positive and the signals would be distorted. It will, of course, be understood that by means of suitably proportioned networks the wave can be rounded just enough to meet the interference requirement, still retaining sufficient steepness to insure prompt operation of the receiving relay. Therefore, rounding by means of networks is preferable.

If it should be desirable and practicable to utilize the frequency range to its fullest, what has been said above about a distortionless line holds without any substantial modification and it would, in that case also, be more advantageous to use a wave rounded by means of suitable networks than to impress on the line a wave of the half-cycle sine form.

Submarine Cables

In the case of submarine-cable telegraphy, there is a limitation on voltage which has not been emphasized in the simple direct-current case discussed above. The voltage which may be impressed on the cable is limited to a definite value. Moreover, for certain reasons, the cable has an impedance associated with it at the sending end which may make the voltage on the cable differ from the voltage applied to the sending-end apparatus. Inasmuch as the limitation in this case is voltage limitation at the cable, the ideal wave is one which applies a rectangular wave to the cable rather than to the apparatus, because it insures that the area under the curve should be the maximum consistent with the imposed limitations. It would be possible to make the transmitting-end impedance approximately proportional to the cable impedance throughout most of the important range. This would insure that the wave applied to the cable would have approximately the same shape as the wave applied to the apparatus. It would probably be desirable for practical reasons to make this impedance infinite for direct current.

In connection with the submarine cable a special kind of interference is particularly important, namely, that due to imperfect duplex balance. For a given degree of unbalance, the interference due to this source may be reduced by putting networks either in the path of the outgoing current or in the path of the incoming current. These facts, together with the frequency distributions deduced above for each of the several impressed waves as exhibited in Fig. 2, make it apparent that the beneficial reaction on the effect of duplex unbalance, which can be obtained by the use of a half-cycle sine wave instead of a rectangular wave, can be obtained more effectively by the use of a simple network, either in the path of the outgoing or in the path of the incoming currents. Either of these locations is equally effective in reducing interferences from duplex unbalance, but the location of the network in the path of the outgoing current has the advantage that it decreases the interference into other circuits, whereas the location in the path of the incoming current has the effect of reducing the interference from other circuits.

Before leaving the matter of submarine telegraphy, it may be well to point out that it is common in practise to shorten the period during which the battery is applied so as to make it less than the total period allotted to the signal element in question. For instance, if it is desired to transmit an *e* the battery may be applied for, say, 75 per cent. of the time allotted to that *e* and during the remaining 25 per cent. the circuit is grounded. The resulting voltage is shown in Fig. 3F. From the foregoing, it is concluded that this method is less advantageous than the application of the voltage for the whole period, because while the shape of the received signal is substantially the same in the two cases, the magnitude, being proportional to the area under the voltage curve, will be less. A cursory examination of the literature does not disclose that anything has been published on the experimental side either to confirm or to oppose this result.

Choice of Codes

A formula will first be derived by means of which the speed of transmitting intelligence, using codes employing different numbers of current values, can be compared for a given line speed, *i.e.*, rate of sending of signal elements. Using this formula, it will then be shown that if the line speed can be kept constant and the number of current values increased, the rate of transmission of intelligence can be materially increased.

Comparison will then be made between the theoretical possibilities indicated by the formula and the results obtained by various codes in common use, including the Continental and American Morse codes as applied to land lines, radio and carrier circuits, and the Continental Morse code as applied to submarine cables. It will be shown that the Continental and American Morse codes applied to circuits using two current values are materially slower than the code which it is theoretically possible to obtain because of the fact that these codes are arranged so as to be readily deciphered by the ear. On the other hand, the Continental Morse code, as applied to submarine cables, or other circuits where three current values are employed, will be shown to produce results substantially on par with the ideal. Taking the above factors into account, it will be shown that if a given telegraph circuit using Continental Morse code with two current values were rearranged so as to make possible the use of a code employing three current values, it would be possible to transmit over the rearranged circuit about 2.2 times as much intelligence with a given number of signal elements.

It will then be pointed out why it is not feasible on all telegraph circuits to replace the codes employing two current values with others employing more than two current values, so as to increase the rate of transmitting intelligence. The circuits, for which the possibilities of thus securing increases in speed appear greatest, are pointed out, as well as those for which the possibilities appear least.

Theoretical Possibilities Using Codes with Different Numbers of Current Values

The speed at which intelligence can be transmitted over a telegraph circuit with a given line speed, *i.e.*, a given rate of sending of signal elements, may be determined approximately by the following formula, the derivation of which is given in Appendix B.

$$W = K \log m$$

Where W is the speed of transmission of intelligence,
m is the number of current values,
and, K is a constant.

By the speed of transmission of intelligence is meant the number of characters, representing different letters, figures, etc., which can be transmitted in a given length of time assuming that the circuit transmits a given number of signal elements per unit time.

Substituting numerical values in this formula gives the following table which indicates the possibilities of speeding up the transmission of intelligence by increasing the number of current values.

Number of Current Values Employed	Relative Amount of Intelligence which can be Transmitted with a Given Number of Signal Elements
2	100
3	158
4	200
5	230
8	300
16	400

This table indicates that there is considerable advantage to be secured in going to more than two current values where the circuits are such as to permit it and where the line speed is not lowered as a result. The limitations will be outlined below. It should also be noted that whereas there is considerable advantage in a moderate increase in the number of current values, there is little advantage in going to a large number.

Codes Now in Common Use—Comparison with Ideal

In the case of printer codes, the theoretical results derived correspond closely to practise, as will be obvious from the method of deriving the formula.

In order to compare the theoretical possibilities indicated by the formula with the results which are obtained when non-printer codes are constructed, several codes were assumed, and for each one the number of signal elements required to produce an average letter was deduced. The method of doing this is set forth in Appendix D. This work resulted in the following table:

	Signal Elements per Letter	Relative Number of Letters for a Given Number of Signal Elements
American Morse (two current values)	8.26	74
Continental Morse (two current values)	8.45	73
Ideal (two current values)	6.14	100
Continental Morse (three current values)	3.77	163
Ideal (three current values)	3.63	169

The column in the above table headed "Relative Number of Letters for a Given Number of Signal Elements" makes possible direct comparison with the results predicted from the formula as given in the table which preceded. It will be noted that the ideal three-current-value code gives an increase in the number of letters for a given number of signal elements as compared with the ideal two-current-value code which is in fair agreement with the theoretical ratio of 1.58:1. It will also be noted that the Continental three-current-value code which is actually in use in the case of submarine cables appears to come quite close to the ideal. In the case of the Continental and American Morse codes, however, where only two current values are used, the results fall short of the ideal, the ratio between the results actually obtained and the ideal being approximately 1.4:1. The reason for this is that a certain proportion of the possible speed is sacrificed in order to make it possible to read the signals by means of a sounder instead of recording them. For instance, the dash has been assumed to be approximately three times as long as the dot. If the signals were mechanically formed at the sending end and recorded at the receiving end, it would be possible to make use of markings 1, 2, 3, etc., signal elements long, as well as corresponding spacings. The ideal codes were so constructed.

It will be seen that the figures deduced for the Continental Morse and the American Morse are substantially identical for two current values. This result probably does not correspond with practise; it is thought that the difference in speed between these two codes is considerably greater, say on the order of 10 or 15 per cent. in favor of the American Morse. The discrepancy is due partly to the fact that no account has been taken of figures and punctuation marks in the present computations and partly to the fact that the assumptions as to relative lengths of space is not strictly in accordance with practise.

From the foregoing, it is seen that there is a two-fold gain in changing from the two-current-value American or Continental Morse codes to the three-current-value Continental code. In the first place, there is a theoretical increase in the ratio of 1.6:1 which accompanies the change from the two-current-value to the three-current-value code. In the second place, there is an incidental increase in the ratio of 1.4:1, due to the fact that the present two-current-value codes are longer than would be necessary, if receiving were done by means other than the ear. The total increase in going from the two-current-value Continental or American Morse codes to the three-current-value Continental code is, therefore, in the ratio of $1.6 \times 1.4:1$ or 2.2:1, provided the line speed is the same. In this connection it should be noted that in the case of the American Morse, the ratio is probably somewhat less than this for the reasons pointed out above.

Limitations in Applying Codes with More than Two Current Values

Certain inherent limitations which have to do with how much the number of current values can be advantageously increased are as follows:

1. Fluctuations in transmission efficiency of the circuit,
2. Interference,
3. Limitations on the power or voltage which it is permissible to employ.

In addition it may be stated that, in general, whenever more than two current values are employed it is necessary to make the sending and receiving means more complicated and expensive. There may be nothing to gain, therefore, in using codes other than those made up of two current values where the telegraph circuits are cheap.

Considering now the features which limit the number of current values which can be employed, it is believed that the importance of the first factor will be obvious. If the line is subject to fluctuations so that the stronger currents at certain times become less in magnitude than the weaker currents at other times, it will be impossible to discriminate between the different current strengths making up the code, particularly if the fluctuations are rapid.

In connection with interfering currents, it is evident that these may be of such polarity as to add to or subtract from the signaling currents and it is consequently necessary to separate the various current values employed sufficiently so that one current value with the interference added may be distinguished from the next larger current value with the interference subtracted.

The spacing between the current values being determined by the interference and fluctuations in transmission efficiency, it will be seen that the maximum number of current values which can be employed is determined by the maximum power which it is permissible to use.

In the case of land line telegraph circuits operated with direct currents, it is well known that quadruplex circuits are much more seriously affected by fluctuations and interference than are circuits employing only two current values. (A quadruplex telegraph circuit employs four current values for transmission in one direction.) In general, it may be said that the possibilities of improving ordinary direct-current operated telegraph circuits in this manner do not appear particularly promising.

In the case of wireless transmission over great distances all three of the above factors are important in limiting the number of current values which can be effectively employed. In the first place, as is well known, large variations take place in the efficiency of the transmitting medium so that the received signals vary considerably in magnitude from time to time. Secondly, the interference, at least at certain seasons, is great enough to make it difficult to distinguish between the current values even when the usual method which employs only two current values is employed. Thirdly, the received power is limited because of the great attenuation suffered by the wireless waves.

In the case of carrier transmission, it may be that there will be a field for the use of more than two current values. The relative cheapness of the line circuits, however, will tend to limit the amount by which it will be economical to increase the cost and complexity of the receiving apparatus. Moreover, it should be borne in mind that no allowance has been made for the effect on the line speed of increasing the number of current values, this being considered outside the scope of the present paper.

Changing an existing network of telegraph circuits so as to employ a code with three instead of two current values would require new types of telegraph repeaters as well as new sending and receiving apparatus, and new operating methods. It is considered to be outside of the scope of this paper to go into a discussion of the details of this matter.

"Sine Wave" Systems

Considerable interest and discussion has been created by suggestions which have been made to use so-called "sine wave" systems of telegraphy. In view of this, a brief discussion of these systems is given below.

A brief analysis of what are the fundamental features of these systems will be given and, based on the results which have been developed in the preceding discussion, comparison will be made of these systems with systems based on other principles. A particular effort will be made to clear up what appears to be fundamentally incorrect assumptions which underlie the arguments which have been advanced in favor of these "sine wave" systems.

Crehore-Squier System. The use of a sine wave envelope to improve the characteristics of telegraph signals was advocated by Crehore and Squier.[4] The words "United States" formed by means of a wave of this type are shown in Fig. 3*d*. The code employed is the same as the ordinary Continental Morse, the only difference being that the signal elements consist of half-cycle sine waves.

In what has preceded, it has been shown that a half-cycle sine wave has a smaller area than a rectangular wave rounded off by passing through an electrical network and, consequently, the sine wave is inferior to the latter from the standpoint of the received signals. From the standpoint of interference into other circuits, it has also been pointed out that the half-cycle sine waves contain more high-frequency components than properly rounded off rectangular waves. Consequently more interference into other circuits will be produced with the wave made up of signal elements consisting of half-cycle sine waves.

Squier System Applied to Submarine Cables. A more recent suggestion of Squier[5] gives the wave shown in Fig. 3*a*. This wave resembles the one advocated by Crehore and Squier in that each signal element consists of a half-cycle sine wave. As has been pointed out, there is no advantage gained by this.

The difference between the two systems lies in the fact that the wave in Fig. 3*a* uses three absolute values and crosses the axis once every half cycle. The code is the same as the Continental, a space being indicated by a half-cycle sine wave of one unit amplitude, a dot by a half-cycle sine wave of two units amplitude and a dash by a half-cycle sine wave of three units amplitude.

By referring to the figure, it will be seen that the resulting wave resembles a continuous sine wave, except for the fact that successive half cycles differ in magnitude. For this reason, the code may be termed an "unbroken-reversals" code.

In considering the application of this code to submarine cable telegraphy, it is convenient to make use of an analysis which is carried

[4] Crehore and Squier, loc. cit.

[5] Squier, loc. cit. *Proc. Phys. Soc.*

out in Fig. 3. Fig. 3*a* shows the words "United States" written in the code advocated by Squier. Fig. 3*b* shows a constant sine wave whose amplitude is equal to the amplitude of a dot in Fig. 3*a*. Fig. 3*c* shows the result obtained by subtracting the wave of Fig. 3*b* from the wave of Fig. 3*a*. On comparing this last wave with the wave

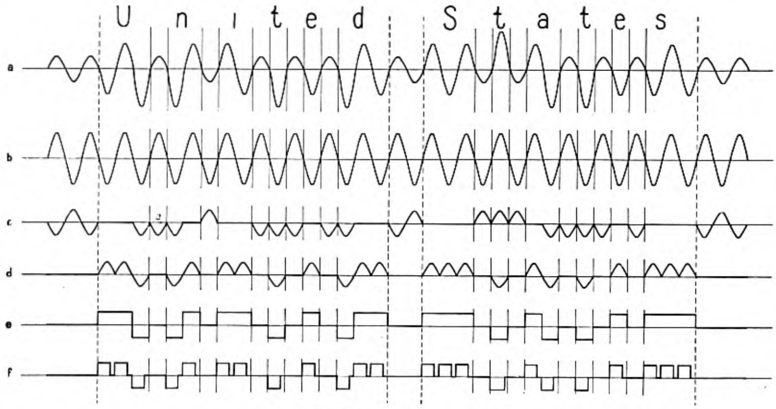


Fig. 3

a—Unbroken reversals code (space = 1 unit, dot = 2 units, dash = 3 units)
b—Constant sine wave, 2 units
c—Wave resulting when substracting *b* from *a*
d—Sine Wave code: note similarity between *c* and *d*
e—Rectangular wave, unmodified
f—Rectangular wave, modified by grounding apex one fourth of the marking time in addition to the spacing time

shown in Fig. 3*d*, it will be seen that the two waves are electrically equivalent. They differ only in having the signal elements permuted.

It is thus evident that the wave shown in Fig. 3*a* is made up of two components; one being the inert component shown in Fig. 3*b* which transmits no intelligence, and the other the intelligence carrying component illustrated in Fig. 3*c*.

The fact that the component shown in Fig. 3*b* does not carry intelligence from the sending station to the receiving station is made clear when we consider that its value at any moment is predictable and that the component can in fact be produced locally.

The net effect of this component is to reduce the voltage available for intelligence transmission to one-third of the total voltage. For example, if it is permissible to apply 60 volts to a particular cable, 40 volts out of these would be used up in transmitting the inert alternating-current wave and only the remaining 20 volts would be useful for the transmission of intelligence.

Radio and Carrier Telegraphy. Squier has also advocated [6] that the combination of sine wave envelopes, unbroken reversals and a three-current-value code be applied to radio and carrier telegraphy.

The advantages and limitations in applying codes with more than two current values have been fully discussed above, and do not need to be gone into further here. It will be evident that the combining with these of sine wave envelopes and unbroken reversals does no good.

The matter of using sine wave envelopes was discussed above, the discussion pointing out that waves with sine-wave envelopes are inferior to waves produced by sending rectangular shaped signals through suitable networks, both from the standpoint of the received signals, and from the standpoint of interference into other circuits.

The "unbroken reversals" bring in again the use of an inert component. Due to the fundamental difference between cable telegraphy on the one hand, and radio and carrier as usually practised on the other, the inert component in the latter case is somewhat smaller than in the former. In the code advocated by Squier, the current which may be subtracted without greatly affecting the intelligence-carrying capacity of the signals, is about one unit in value, which is the current corresponding to a space. When this current has been subtracted, the space current is reduced from one unit to zero, the dot current from two units to one, and the dash current from three units to two. This subtraction having been carried out, it is seen that the maximum intelligence-carrying component is approximately two-thirds of the maximum current actually employed. (This figure of two-thirds compares with the figure of one-third for the submarine cable.)

In the case of radio, the amount of power which must be radiated from the transmitting station is of particular importance. Since with the system advocated by Squier about two-thirds of the maximum voltage which is radiated is effective in transmitting intelligence, it is evident that about twice as much power must be radiated as would be required if the inert component were not transmitted.

Incorrect Assumptions. Two incorrect assumptions are made in the papers referred to and underlie a considerable portion of the arguments advanced in favor of the systems advocated by Squier.

One of these is that a wave, whose elements are half-cycle sine waves, lends itself to tuning. It is true that in the case of the "unbroken-reversals" code a certain amount of tuning can be secured, but this tuning applies only to the inert unvarying component in the wave, which carries no intelligence. The fact, shown in Fig. 2, that

[6] Squier, loc. cit., *Franklin Inst., Jl.*

the intelligence-carrying component contains no outstanding narrow range of frequencies to which tuning can be applied should make obvious the error in this assumption.

The other assumption is that a wave, which is ideal for the transmission of power, is also ideal for the transmission of intelligence. As a matter of fact, the transmission of intelligence inherently involves rapid and unpredictable changes in the current, whereas the transmission of power is best brought about by steady current, either direct or alternating. These two conditions are, of course, incompatible.

APPENDIX A

Use has been made of the following two principles:

1. In a telegraph circuit in which the line speed is near the maximum, the shape of the received dot is substantially independent of the shape of the impressed dot, and

2. The magnitude of the received current is approximately proportional to the area under the transmitted voltage curve.

The following general discussion of these principles has been furnished by J. R. Carson.

Let the arrival curve, due to suddenly impressed unit battery be denoted by $A\ (t)$; then the received signal $S\ (t)$, due to the elementary dot impressed signal $f\ (t)$ is given by [7]

$$S(t) = \int_0 f(x) A'(t-x)\,dx \tag{1}$$

the upper limit of integration being t for $t<T$ and T for $t \geqq T$. The latter case will alone be considered since the conclusions arrived at in this case are conservative.

Expanding $A'\ (t-x)$ in (1), we get

$$S(t) = \left[A'\ (t) - \frac{h_2 T}{2!} A''(t) + \frac{h_3 T^2}{3!} A'''(t) \ldots \right] \int_0^T f(x)\,dx \tag{2}$$

where
$$h_2 = \frac{\int_0^T x f(x)\,dx}{\frac{T}{2!}\int_0^T f(x)\,dx},$$

$$h_3 = \frac{\int_0^T \frac{x^2}{2!} f(x)\,dx}{\frac{T^2}{3!}\int_0^T f(x)\,dx}, \text{ etc.}$$

[7] J. R. Carson. "Theory of the Transient Oscillations of Electrical Networks and Transmission Systems." A. I. E. E. Trans., Vol. XXXVIII, 1919, p. 345.

It follows at once that, provided

$$\int_0^T f(x)dx \neq 0$$

and provided the duration T of the signal is sufficiently short, the arrival dot is given approximately by the leading term

$$A'(t)\int_0^T f(x)dx$$

and that this approximation becomes increasingly close as the speed of signaling is increased, *i.e.*, as the duration T of the dot is decreased.

The conclusions from the foregoing may be stated in the following propositions:

I. If the speed of signaling is sufficiently high the arrival signal representing the elementary dot is independent in shape of the form of the impressed signal, and is proportional in amplitude to the time integral or "area" of the impressed signal.

It will be evident, however, that if no restrictions are imposed on $A'(t)$ and $f(t)$, the foregoing proposition requires, in general, that the duration T of the dot shall be so small as to make the series expansion rapidly convergent from the start. This, however, requires a speed of signaling very considerably greater than that actually necessary in practise in order that the foregoing proposition shall hold to a good degree of approximation, at least for the types of impressed dot signals specially considered in the present paper. To show this, it is necessary to establish two less general propositions, valid for the types of impressed signals under consideration.

II. If the impressed signal $f(t)$ is everywhere of the same sign, then a value τ exists, such that $0</\tau/<T/2$, and such that

$$S(t+T/2)=A'(t+\tau)\int_0^T f(x)dx \tag{3}$$

This proposition follows from the mean value theorem.

III. If $f(t)$ is everywhere of the same sign, and if further it satisfies the conditions of symmetry,

$$f(x)=f(T-x),\ (x\leq T/2)$$

then a value τ exists, such that $0<\tau<T/2$ and such that

$$S(t+T/2)=1/2[A'(t+\tau)+A'(t-\tau)]\int_0^T f(x)dx \tag{4}$$

This last equation also follows from the mean value theorem. Furthermore, the conditions stated in proposition III are satisfied by the rectangular wave, the half-cycle sine wave, and the rectangular wave extending through part of the dot provided the reference time $t=0$ is properly chosen.

Returning to proposition II, let us write

$$S_j(t+T/2)=A'(t+\tau_0+\tau_j)\int_0^T f_j(x)\,dx,$$

the subscript $_j$ indicating the particular type of impressed dot signal, and τ_0 the value of τ for any type of signal, taken as reference. Then

$$S_j(t+T/2)=\left[A'(t+\tau_0)+\frac{\tau_j}{1!}A''\,(t+\tau_0)+\ .\ .\ .\right]\int_0^T f_j(x)dx \quad \text{(2a)}$$

Now, the condition that proposition I shall hold to a good degree of approximation is that the expansion (2a) shall converge rapidly. Since the maximum possible value of τ_j is $T/2$ and since in practise it is much smaller than $T/2$, the required convergence obtains for much larger values of T, that is, slower speeds of signaling than that required in the expansion (1). Furthermore, for the three types of signals specifically under consideration τ_1, τ_2 and τ_3 differ from one another by quantities very much smaller than $T/2$ in all actual transmission systems.

If the conditions of proposition III are introduced, the approximation is still closer and proposition I is valid for still lower signaling speeds.

In order to arrive at quantitative ideas of the minimum signaling speeds at which the foregoing proposition is valid, it is necessary, of course, to specify the arrival curve of the transmission system under consideration. An application of the foregoing analysis to representative transmission systems both with and without a "cut-off" frequency has shown that it is valid to a very good degree of approximation for speeds considerably lower than the highest attainable under practical conditions.

APPENDIX B

Use has been made of the formula

$$W=K\log m$$

where $W=$ the speed of transmission of intelligence

$K=$ a constant

and $m=$ the number of current values employed.

The assumptions which underlie this formula and its derivation will now be given.

Let us assume a code whose characters are all of the same duration. This is usually the case in printer codes. If n is the number of signal elements per character, then the total number of characters which can be construed equals m^n. In order that two such systems should be equivalent, the total number of characters that can be distinguished should be the same. In other words,

$$m^n = \text{const.} \tag{1}$$

This equation may also be written

$$n \log m = \text{const.} \tag{2}$$

The speed with which intelligence can be transmitted over a circuit is directly proportional to the line speed and inversely proportional to the number of signal elements per character provided that the relations above are satisfied. Hence, we may write

$$W = s/n \tag{3}$$

where s is the line speed. Substituting the value of n derived from the equation above, this equation becomes

$$W = \frac{s \log m}{\text{const.}} \tag{4}$$

which may also be written

$$W = K \log m \tag{5}$$

In applying this formula to practical cases it will be found impossible to comply strictly with the condition expressed by equation (1). As an example, consider the comparison between a three-current-value code where each character is made up of three signal elements, and a two-current-value code where each element is made up of five signal elements. It is obvious that the speed with which *characters* can be transmitted in the former case is five-thirds the speed in the latter case for a given line speed. In other words the ratio is 1.67:1 whereas the formula gives the ratio 1.58:1. It should be noted, however, that the former code possesses only 27 characters whereas the latter possesses 32. In other words one *character* of the latter code represents the transmission of more *intelligence* than one *character* of the former. Thus the figure 1.67 for the relative speeds of transmission of *characters* and the figure 1.58 for the relative speeds of transmission of *intelligence* are not incompatible.

It will be noted that the formula has been deduced for codes having characters of uniform duration and that it should not be expected to be anything but an approximation for codes whose characters are of non-uniform duration. To establish the formula for the latter case it would be necessary to make an assumption as to the relative frequencies of the various characters. It seems reasonable to suppose that the formula will give a fair approximation to the facts in this case also, but it should not be expected to be accurate.

APPENDIX C

The deduction of the curves given in Fig. 2 from the curves given in Fig. 1 requires some explanation. Looked at casually, it would seem as if an isolated dot would not possess any frequency characteristics whatsoever. Nevertheless, if a voltage, such as any of those represented in Fig. 1, is applied to a network capable of being thrown into oscillation, the network will respond to the voltage by oscillating. Suppose, for simplicity, that the network consists of an inductance, a capacity and a very small resistance in series, the response of the network to the application of any of the voltages illustrated is that it oscillates at constant frequency and gradually decreasing amplitude. Further, the response varies when the natural period of the circuit is varied.

There are two ways of looking at this phenomenon. We may say, on the one hand, that the oscillations of the frequency in question are manufactured by the network out of the voltage applied and that the frequency does not exist in the original voltage. On the other hand, we may say that the original voltage contains components at or near the resonant frequency and that the circuit responds to these components, because it offers them a small impedance, while it does not respond to other components because it offers them a large impedance. Either of these views is permissible, but it is convenient for the purposes of this paper to use the nomenclature of the second view and to consider the applied voltages to be made up of an indefinitely large number of frequencies. The problem of determining the response of oscillating networks is then solved by deducing the frequency characteristic or the response characteristic of the impressed voltage. This characteristic may be determined by means of the Fourier integral, whose computation is described in any standard textbook on the subject. The following is intended to outline the considerations, from a physical standpoint, which lead to establishing this integral.

To deduce the frequency characteristic of an isolated dot, it is simplest to start with a long series of dots which are uniformly spaced. If such a series of dots is considered to extend indefinitely, it is possible to analyze the resultant wave into a Fourier series by well known methods. Now, suppose that such a Fourier series has been obtained for a given spacing of the dots. The next step is to increase the spacing between the dots. The result of this is to increase the number of Fourier components in a given frequency range and to decrease the magnitude of each. If this process of increasing the space between the dots is continued indefinitely, we approach the condition of an isolated dot. Moreover, as we approach this condition, the number of components in a given frequency range increases indefinitely and the magnitude of each decreases indefinitely. This limiting result is known as the Fourier integral for the wave in question.

APPENDIX D

A table has been given in the paper in which the relative efficiency of various codes in transmitting intelligence is listed. The derivation of that table will now be given.

The comparison will include the following codes based on two current values: American Morse, Continental Morse, and the so-called "ideal" two-current-value code. It will also include the following codes based on three current values: Continental Morse and an "ideal" three-current-value code.

The assumption is made that the text is made up of five-letter-words, no allowance being made for punctuation. The following table gives the length of the spaces assumed in terms of signal elements.

	Ordinary Spaces Within Letters	Special Spaces in "Spaced" Letters	Spaces Between Letters	Spaces Between Words
American Morse (two current values).....	1	2	3	4
Continental Morse (two current values)...	1	—	2	3
Continental Morse (three current values)..	—	—	1	2

It is assumed that the dashes in the two-current-value codes are of three signal elements duration, except for the letter *l* in American Morse which is assumed to occupy five signal elements. It may be that in practice, the dashes are somewhat shorter than has been assumed but the resulting error is not great. In connection with the relative spacings between letters and words assumed for the Continental and American Morse codes, it is also questionable whether they accord strictly with practise. It may be that these spacings are on the average more nearly equal than the table indicates. However, this assumption affects only the relative speeds obtainable with the American Morse and the Continental Morse and does not materially affect the comparison between codes based on two current values on the one hand and codes based on three current values on the other.

The term "ideal" has been applied to two codes which will next be explained. These codes are constructed on the same principles as the Continental and American Morse codes with an effort to make them as brief as possible without making the reading too difficult. It is thought that the two ideal codes chosen are comparable in the matter of ease of reading. In constructing the two-element code, two steps are involved. In the first place it is assumed that the markings and spacings of any integral number of signal elements' duration can be used so that in addition to the values for markings and spacings assumed above, there may be dashes of two, four, etc., units duration. With these assumptions the 26 shortest characters that can be constructed are next made up. It is found that one character is of 1 unit duration, 1 of 2 units, 2 of 3 units, 3 of 4 units, 5 of 5 units and 9 of 6 units duration. The remaining 5 characters are taken of 7 units duration each. The second step is to ascribe the 26 letters of the alphabet to these characters in such an order that the most frequent letters correspond to the shortest characters. It is most efficient to use the same spacing as was assumed above for the Continental two-current-value code, with the addition that spaces of longer duration than three units may be employed within a letter.

The matter of constructing the ideal three-current-value code is similar. First, the 26 shortest characters are constructed. Two characters can be constructed having a duration of 1 unit, four characters having a duration of 2 units and eight characters having a duration of 3 units. The remaining twelve characters are taken 4 units in duration. Next, the most frequent letters are assigned to these characters in the order of their duration. It is best in this case to use the same assumptions as to spacings between letters and words as was used above in connection with the three-current-value Continental code. The use of spaces within letters is not economical in this case.

A frequency table given by Hitt [8] was used to determine the relative frequency of the various letters. The average duration per letter was computed from this table and corrected for spaces between words and letters. The resultant average duration is as follows:

Code	Signal Elements per Letter
American Morse (two current values)	8.26
Continental Morse (two current values)	8.45
Ideal (two current values)	6.14
Continental Morse (three current values)	3.77
Ideal (three current values)	3.63

[8] Parker Hitt. "Manual for the Solution of Military Ciphers." Army Service Schools Press, Fort Leavenworth, Kansas. Second edition, p. 7.

Abstracts of Bell System Technical Papers Not Appearing in the Bell System Technical Journal

The Auditory Masking of One Pure Tone By Another and Its Probable Relation to the Dynamics of the Inner Ear.[1] R. L. WEGEL and C. E. LANE. The authors used an air damped telephone receiver supplied with variable currents of two frequencies and determined the amount of masking by tones of frequency 200 to 3500 for frequencies from 150 to 5000. Except when the frequencies are so close together as to produce beats the masking is greatest for tones nearly alike. When the masking tone is loud it masks tones of higher frequency better than those of frequency lower than itself. If the masking tone is introduced into the opposite ear the effect occurs only by virtue of conduction through the bones of the head.

It is shown that combinational tones result when two tones of sufficient intensity are introduced simultaneously, these combinational tones being due to a non-linear response of the ear.

A dynamical theory of the cochlea is given which ascribes pitch discrimination to a passing of vibrations along the basilar membrane and a shunting through narrow regions of the membrane at points depending on the frequency. This view of the action of the ear offers an explanation of the masking effects.

Distribution of Radio Waves from Broadcasting Stations over City Districts.[2] RALPH BOWN and G. D. GILLETT. This is a description and analysis of the results obtained in a radio transmission survey of the cities of New York and Washington, D. C., and contiguous territory. Measurements of the field strength of radio signals from stations WCAP at Washington and WEAF at New York were made at a large number of points. Based on these data, curves are drawn showing how different kinds of territory cause different attenuations and showing radio shadows caused by mountains and by large masses of steel buildings. In order to visualize the phenomena, the data have also been plotted on maps, contour lines of equal signal strength being drawn. These contour maps illustrate graphically the non-uniformity of transmission in city areas and show the nature and extent of the "dead spots" and shadows.

[1] *Physical Review*, II, Vol. XXIII, p. 266, 1924.

[2] Presented to the Institute of Radio Engineers, January 16, 1924, at New York.

Measuring Methods for Maintaining the Transmission Efficiency of Telephone Circuits.[1] F. H. BEST. The circuits involved in the transmission of speech in a modern telephone plant, particularly those designed for long distance operation, necessarily involve a considerable amount of complexity. The use of telephone repeaters the development of long toll cables, the application of carrier systems and other developments associated with these, while increasing the efficiency and economy of telephone toll circuits have also increased their complexity and have required the development of more effective means of insuring that the circuits are maintained at all times in good condition and adjustment.

Maintenance of the transmission efficiency of the telephone plant is conducted by a special force, using methods and apparatus that have been developed for this purpose. This paper gives a brief description of the transmission characteristics of some of the common type of telephone circuits, outlines a general method for measuring their transmission efficiency and describes several of the most modern types of transmission measuring sets, together with a brief, mention of the oscillators which supply the power for testing.

A Primary Standard of Light Following the Proposal of Waidner and Burgess.[2] HERBERT E. IVES. The primary standard of light proposed in this paper consists of a black body constructed of platinum; the light from which, at its melting point, constitutes the photometric fixed point desired. The platinum black body consists of a cylinder of highly polished platinum with a narrow slit for observing the interior. Studies of the optical properties of reflecting cylindrical enclosures show that at certain angles of observation the interior is practically "black." The platinum cylinders are heated electrically and the light from the interior is observed by throwing an image of the slit on to a photometer field. Two series of observations were made, one by a visual photometric method, the other by a photoelectric cell giving a photographic record by means of a string electrometer. The two methods of observation gave practically identical results, yielding a final value for the brightness of the black body at the melting point of platinum of 55.4 candle power per square centimeter. The advantages of this proposed standard over the present unsatisfactory flame standards are discussed.

High Quality Transmission and Reproduction of Speech and Music,[3] W. H. MARTIN and HARVEY FLETCHER. Radio broadcasting has

[1] Journ. A. I. E. E. Vol. XLIII, p. 136, 1924.
[2] *Journal Franklin Institute,* Vol. 197, p. 147, p. 359, 1924.
[3] Journ. A. I. E. E. Vol. XLIII, p. 230, 1924.

drawn attention to the problems involved in obtaining high quality in systems for the electrical transmission and reproduction of sound. This paper gives the general requirements for such systems, discusses briefly the factors to be considered in design and operation and indicates to what extent the desired results can be obtained with the means now available.

It was pointed out in this paper that broadcasting stations and connecting lines can be made practically perfect but that most of the loud speaking apparatus now extensively used for reproduction, causes distortion. At the time of reading this paper the authors demonstrated a laboratory model of a new loud speaker of unusual design. This apparatus reproduces all frequencies from the lowest to the highest of the audible range with approximately equal facility. This results in reproduced music which the ear can scarcely distinguish from the original.

Telephone Transformers.[1] W. L. CASPER. After outlining the varied sets of conditions which different types of telephone transformers must meet, this paper discusses the design and construction of transformers to handle efficiently the range of frequencies ordinarily present in speech. Two winding transformers only are dealt with, and the three most common impedance combinations of the two circuits connected by the transformer are considered; namely, both circuits comprised of resistances, one circuit a resistance, and the other a positive reactance, and one circuit a resistance and the other a negative reactance.

The efficiency with which energy is transmitted is measured by comparison with an ideal transformer, and the transformer is studied by supposing it replaced by an equivalent T network. The variation of transformer losses with frequency is discussed and characteristic curves are shown for transformers of different mutual impedances. Characteristics are also given showing the operation of the in-put transformer associated with the vacuum tube.

The mechanical construction of the common battery repeating coil, telephone induction coil, and of certain types of transformers for vacuum tube circuits, are shown. These transformers are all constructed so as to give the desired accuracy of speech transmission under their respective circuit conditions.

Radio Telephone Signaling—Low Frequency System.[2] C. S. DEMAREST, M. L. ALMQUIST and L. M. CLEMENT. The system described

[1] Journal of the American Institute of Electrical Engineers, Vol. XLIII, p. 197, 1924.

[2] Journ. A. I. E. E. Vol. 43, p. 210, 1924.

provides a means whereby any one of about seventy-five radio stations, operating on the same wave length, may be called without signaling the remaining. Obviously this is an important improvement in the radio art for in many cases it permits a radio station operator to pursue other duties which would be impossible if he were required to listen in at all times.

The engineering problem presented, being remarkably similar to many telephone problems, was solved in a very similar manner. When it is desired to signal a station, an alternating current of a very definite frequency is impressed on the transmitter. This modulates the power radiated similar to the way the undulations of the voice modulate the power when speech is transmitted. The station to be signaled is determined by the code transmitted. This code consists of a definite grouping of dots and spaces and dashes.

At the receiving station this modulated power is detected in the usual manner and results in an alternating current identical in nature to that used in transmitting the code. A special alternating current relay of high selectivity and sensitivity, in conjuction with a more common direct current relay system, converts the code into a series of direct current impulses. These impulses pass into a selector like that used in common train dispatching circuits. The mechanism of this selector will be unlocked and a local ringing circuit closed if the code is that for which it has been set. Thus it is seen that the code is received by all stations but only one selector of the system will operate to ring its local annunciator bell. The number of stations which can operate in the same system is determined by the number of possible combinations on the selector. At present this is set at seventy-eight but this may be readily extended to include more than two hundred.

Because of the high selectivity of the alternating current relay and its associated direct current relay system, the apparatus is particularly free from interference such as the operation of nearby spark or I.C.W. Stations. In fact, tests show that the signaling system will continue to function satisfactorily long after interference is so bad as to make conversation impossible. As designed, the signaling system may be made an integral part of a standard radio system without altering the apparatus already in use.

Contributors to this Issue

H. R. Friis, E.E., Royal Technical College in Copenhagen, 1916; Columbia University, 1919–1920. Research Department, Western Electric Company, 1920—. Mr. Friis' work has been largely in connection with radio reception methods and measurements. He has published papers on vacuum tubes as generators, radio transmission measurements and static interference.

A. G. Jensen, E.E., Royal Technical College of Copenhagen, 1920. Research Assistant to Professor P. O. Pedersen, 1920–21. Columbia University, 1921–22. Research Department, Western Electric Company, 1922—. Mr. Jensen has been mainly engaged in work relating to radio reception methods and measurements.

D. D. Miller, B.S. in electrical engineering, Tennessee, 1909. Installation Department, Western Electric Company, Hawthorne, 1909–1910. Physics Laboratory, Engineering Department, New York, 1910–1917. Apparatus Development, 1917—. Mr. Miller is in charge of the design of relays and has contributed much to the development of the modern flat types of relays which combine cheapness of manufacture with improved operating characteristics.

I. B. Crandall, A.B., Wisconsin, 1909; A.M., Princeton, 1910; Ph.D., 1916; Professor of Physics and Chemistry, Chekiang Provincial College, 1911–12; Engineering Department, Western Electric Company, 1913—. Dr. Crandall has published papers on infra-red optical properties, condenser transmitter, thermophone, etc. More recently he has been associated with studies on the nature and analysis of speech which have been in progress in the Laboratory.

C. F. Sacia, B.E.E., University of Michigan, 1916; Engineering Department of the Western Electric Company, 1916—. Mr. Sacia has been engaged upon methods for recording and analysis of speech.

E. B. Wheeler, B.S., University of Illinois, 1905. Engineering Department, Western Electric Company, Chicago, 1905–1907. Engineering Department, Western Electric Company, New York. Physical Laboratory, 1907–1921. General Development Laboratory, 1921—. Mr Wheeler has been actively connected with the development of improved types of switchboard and telephone cords, dry cells, condensers, and other types of telephone equipment; and with the investigation of the effects of atmospheric conditions upon the performance of telephone apparatus.

RONALD M. FOSTER, S.B., Harvard, 1917; American Telephone and Telegraph Company, Engineering Department, 1917–19; Department of Development and Research, 1919—.

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.

R. L. WEGEL, A.B., Ripon College, 1910; assistant in physics, University of Wisconsin, 1910–12; physicist with T. A. Edison, 1912–13; Engineering Department of Western Electric Company, 1914—. Mr. Wegel has been closely associated with the development of telephone transmitters and receivers, and has made important contributions to the theory of receivers.

CHARLES R. MOORE, B.S. in Mechanical and Electrical Engineering Purdue, 1907; E.E., Purdue, 1910; Instructor and Assistant Professor Electrical Engineering, Purdue, 1907–13; Manager of LaFayette Electric and Mfg. Co., 1913–14; Associate in Electrical Engineering, University of Illinois, 1914–16; Engineering Department of the Western Electric Co., 1916—. At the Western Electric, Mr. Moore, for several years, has had charge of transmitter development work and has contributed important inventions relating to telephone instruments and acoustic devices.

HARRY NYQUIST, B.S. in electrical engineering, North Dakota, 1914; M.S., North Dakota, 1915; Ph.D., Yale, 1917; Engineering Department, American Telephone and Telegraph Company, 1917–1919; Department of Development and Research, 1919—. Mr. Nyquist has been engaged in work on both direct current and carrier telegraph systems as well as problems in line compositing.