

JOURNAL OF The British Institution of Radio Engineers

(FOUNDED IN 1925 - INCORPORATED IN 1932)

"To promote the general advancement of and to facilitate the exchange of information and ideas on Radio Science."

Vol. VII (New Series) No. 2

MARCH-APRIL 1947

THE CONVENTION

Three years ago, the Council agreed to the holding of a post-war Convention and it should be quite clear as to what we, as a professional institution, understand by a Convention. Literally, of course, the word means a "coming together," and if this were the sole meaning, we could have had a convention in London by arranging a series of papers on successive days, starting, say, with an official luncheon and ending with a banquet. Such a "convention" would, obviously, have been far easier to arrange than the one now to take place.

The Council intended, however, that the Convention should be more than a series of technical papers, delivered in a group, and arranged. From the outset the Convention should not only be a "coming together," but also a "getting away from" everyday surroundings. To those delegates who, it is regretted, cannot be accommodated in the main convention hotel, we would explain that it was necessary to book the hotel over a year ago, at which time it was impossible to do more than estimate roughly the accommodation likely to be required. Until recently, the estimate had been very accurate, but in the last few weeks belated reservations have necessitated some members being accommodated in a nearby, but similarly attractive, hotel. Those members will, however, be able to take their main meals in the convention hotel with the other members.

As far as the business of the Convention is concerned, the response to the Council's request for papers has been excellent, and not only is there a very full programme of papers to be delivered, but a number of others have been approved for eventual publication in the *Journal*. In selecting the papers, many of which are by eminent radio physicists and engineers, the Papers Committee has had in mind the covering of radio and electronics in the broadest possible sense. The Convention has thus not been limited to one aspect of radio engineering, and as a result every member, whatever his own particular work, should find papers of interest to him.

For those who find that a long succession of papers without a break is apt to encourage mental indigestion, the Papers Committee has provided alternative interests for each afternoon. Instead of hearing a paper, small parties of members will be able to pay visits to Government establishments, airports, etc., to see some of the equipment being used and the work which is going on.

Only difficulties of travel and the general uncertainty of affairs at the present time have prevented the number of overseas delegates from being much greater. During the course of the Convention, we hope to establish radio contact with India, South Africa and Canada, and to have the opportunity to send best wishes from the Convention to our members in those countries.

Our greatest regret is that our President will not be with us in person. He has been one of the most enthusiastic advocates of this new departure on the part of the Institution and would have attended had he been in Great Britain. We hope that, at a future Convention, we shall be able to welcome him in person. Meanwhile, Admiral the Viscount Mountbatten of Burma has asked Air Vice-Marshal Aitken, Vice-President, to act on his behalf.

It would not be right to conclude these notes without paying tribute to the small group of members on whom most of the preliminary work has fallen. No one who has not experienced the many problems that arise can have any idea of what is involved in organizing a Convention, and the committees and individuals concerned have done everything possible to ensure the success of this Convention.

It is hoped that those who will not be with us at Bournemouth this year will attend the 1948 Convention. To the 1947 delegates the Council extends best wishes for an informative, enjoyable and refreshing break in the normal routine.

W. E. M.

QUARTZ OSCILLATORS*

by

P. Vigoureux, D.Sc., A.C.G.I.

*Read before the Scottish Section of the Institution on October 31st, 1945,
and the London Section on December 19th, 1946.*

1. Quartz Crystals

Quartz crystals become electrically polarized when subjected to stress in certain directions. Figure 1 shows two regular crystals of quartz. If the crystal is compressed along X, it becomes electrically polarized along X, i.e., one side (the left-hand side in the right-hand crystal of Fig. 1) becomes positively electrified, the other side negatively. Also, if the same crystal were compressed in a direction perpendicular to X, i.e., along Y, it would again become polarized along X, but this time in the opposite sense. The other crystal shown in Fig. 1 is similarly polarized by pressure, but in the opposite sense. Finally, no polarization is produced by strain along Z.

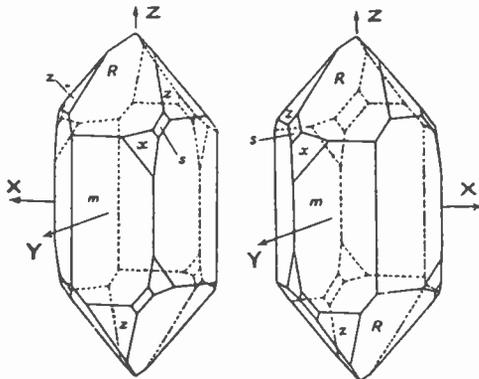


Fig.1. —Right-handed and left-handed quartz crystals.

There are three directions like X along which the crystal can become electrically polarized if suitably stressed: these directions are called electric axes. The direction Z is called optic axis because when polarized light is transmitted through the crystal in that direction, the plane of polarization of the light is rotated. There are two types of crystal, right-handed and left-handed, so called because the plane of polarization is rotated in opposite senses by the two varieties. These two varieties can also be distinguished by the positions of the small facets *s*, *x*,

of Fig. 1. But needless to say, a small proportion only of the quartz crystals found so abundantly in nature are so perfect as to exhibit all the edges and faces drawn in these diagrams. In most cases one only of the pyramids is present, and occasionally so little of the crystal is available that it becomes difficult to identify its various faces.

But all crystals are endowed with the property, that however badly shaped they may be, yet the angles between the various faces do not vary from crystal to crystal of any particular type. For instance, in the prismatic portion of the crystal shown in Fig. 1, the six straight edges are parallel to the Z axis, and if a slab be cut from the crystal in a direction perpendicular to these edges, it will be hexagonal in shape. The hexagon is always equiangular, but usually it is not regular, i.e., the sides are not all equal. The axial directions can be determined much more accurately by X-ray methods.

Quartz crystals are subject to twinning, which is said to occur when the minute crystals which make up the whole block are not all of the same variety, or are not all oriented in the same direction. Since left-handed and right-handed crystals become electrically polarized in opposite senses by pressures, it is important, for piezo-electric purposes, to select pieces which are made of only one variety. Thus, after the slab has been cut, it must be examined for that kind of twinning, which is called optical twinning. This examination is easily carried out with the help of two sheets of polaroid. The sheets are placed one behind the other a short distance apart, and one of them is rotated until no light, or minimum light, comes through; the second sheet is then intercepting the polarized light coming through the first. If the slab is now inserted between the sheets, the plane of polarization of the light coming through the first sheet is rotated by the slab before it reaches the second sheet, and some of it can therefore pass through the second sheet, with the result that there is some illumination. Wherever the slab contains only one variety of crystal, this illumination is uniform, except that if ordinary sunlight be used,

* Revised manuscript received 16th March, 1946.
U.D.C. No. 621.396.611.21: 537.228.1.

the colours of the rainbow appear because the plane of polarization is not rotated to the same extent by light of all colours. But all portions containing adjacent left and right-handed crystals produce a discontinuous illumination, as shown along three edges of the slab of Fig. 2.



Fig. 2.—Quartz slab illuminated by polarized light viewed through a Nicol prism. The twinned portions lie close to three edges.

Once a slab, or part of a slab, free from optical twins has been secured, slices like A B F D, Fig. 3, are cut off it. Here the electric axis is along A D. If the depth of the slab across the plane of the diagram be comparable with the length AB, and large compared with the thickness AD, the slice is called a plate. If desired circular plates or *discs* can be cut off the slice by means of swiftly rotating brass tubes fed with carborundum powder. Finally, *rods* can be obtained by cutting the slice parallel to the face A B F D into narrow portions of length AB and width comparable with the thickness AD.

2. Direct Piezo-Electric Effect

When the slice is strained to an extent ξ in either of the directions AD or AB, an electric polarization P is set up along AD, of value 0.16ξ coulomb per square metre. The production of electrical polarization by strain is called the *direct* piezo-electric effect. This effect can be demonstrated by compressing the specimen and measuring the charge with an electrometer. If the compression is applied locally by two small spheres, the slice can be moved about and examined all over its surface. The slices are good when they give the same deflection throughout. With some slices the deflections vary widely, and in some cases even change sign. Such

irregularities are due to a second type of twinning in which, although the minute constituent crystals are all right-handed or all left-handed, the electric axes of some are opposite in sense to the electric axes of the others. From the piezo-electric point of view, this type of twinning also is a defect, and plates exhibiting it to an appreciable extent must be rejected.

3. Converse Piezo-Electric Effect

Soon after the discovery of the direct piezo-electric effect (J. and P. Curie, 1880) Lippmann deduced from energy considerations that if the quartz plate were subjected to electric force along the electric axis, i.e., along AD, Fig. 3, its dimensions AB and AD would change. From the known elastic constants it can be predicted that the strain ξ resulting from a potential gradient E_x volts per metre is $2.15 \times 10^{-12} E_x$. This production of strain by electric force is called the *converse* piezo-electric effect.

To obtain an idea of the magnitude of the phenomenon in a practical case, suppose AB, Fig. 3, is 10 cms. and AD is 1 mm. Let the faces perpendicular to the electric axis be coated with sheets of tinfoil, called *electrodes*, and apply 10 kilovolts to them; the length of the rod will change by two thousandths of a millimetre. Thus, even

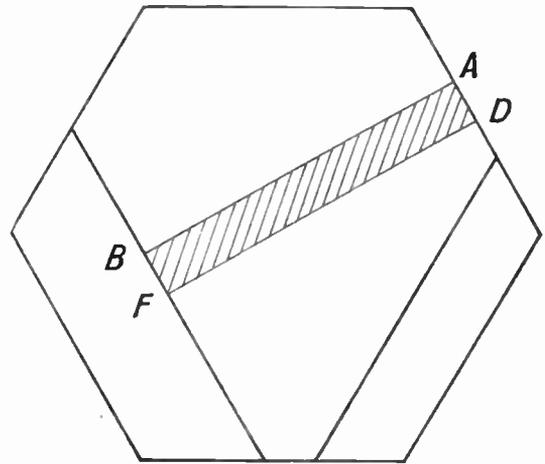


Fig. 3.—Diagram showing how quartz plates are obtained from a hexagonal slab.

high voltages, when they are steady, produce but infinitesimal deformations. Such, however, is not necessarily the case with alternating voltages. Since the strain is reversed when the p.d. is reversed, an alternating p.d. causes the rod to vibrate

at a frequency equal to that of the alternating p.d. These vibrations are *forced*, except near a few frequencies, and, according to the example given above, they are very small. But if the frequency of the alternating p.d. be varied until it coincides with the natural frequency of the rod, resonance comes into play, and the amplitude of vibration is increased hundreds of times. It is limited only by the damping, i.e., by the losses connected with the motion of the rod.



Fig. 4.—Luminous quartz resonator for wavemeter calibration.

4. Luminous Resonators

By virtue of the direct piezo-electric effect, this considerable motion produces at the electrodes large alternating quantities of electricity. A beautiful method of demonstrating the presence of those charges is to enclose the quartz rod in a vessel containing helium or some other gas at low pressure, and excite the rod by tiny electrodes placed a small distance away from each face, as in Fig. 4. For then the potential gradients produced by the charges in their neighbourhood ionize the gas, and a glow appears at the points of maximum mechanical stress, viz., at the nodes. Fig. 5 shows the glow produced by vibrations of the 3rd and 9th order respectively, and Fig. 6 shows the glow for vibra-

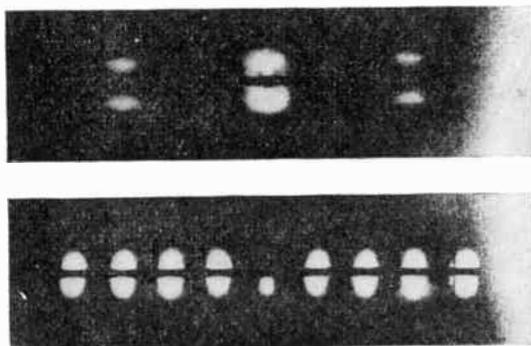


Fig. 5.—Vibrations of the third and ninth order.

tions of the 15th and 33rd order. Since resonance is very sharp, since quartz is a material of great permanency, since the thermal coefficient of frequency

of these rods is only a few parts in a million per °C., and since as many as 14 or 15 different frequencies can be obtained from one rod, these luminous resonators are useful for calibration of wavemeters.

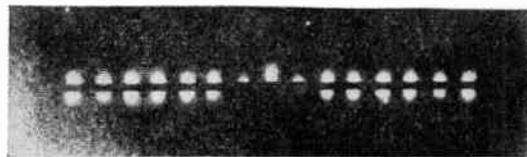
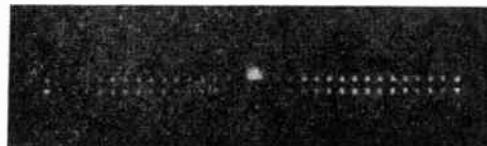


Fig. 6.—Vibrations of the fifteenth and thirty-third order.

They were developed in Germany, and have not been used to any extent in Britain, probably because the valve-maintained quartz oscillator described by G. W. Pierce in America has proved still more useful, more simple and cheaper as well.

The natural frequency of rods and plates can be predicted from the elastic constants and the density, and it can also be measured, e.g., by the glow method. The results of calculation and experiment are in good agreement, and yield $2,720/l$ kc/s for a rod l mms. along X or Y, and $2,870/x$ kc/s for a thin plate of thickness x .

5. Dip Method of observing Resonance.

When the quartz vibrates in resonance, the large alternating charges produced at the electrodes react on the electric circuit employed for establishing the p.d.; and that reaction can be used to detect

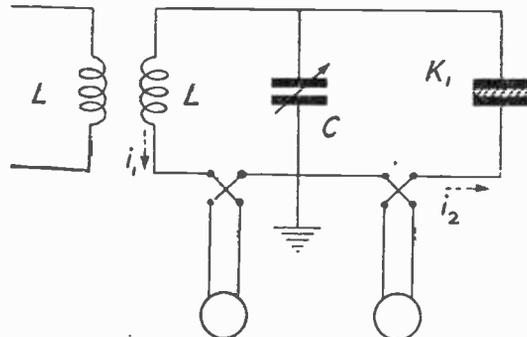


Fig. 7.—Circuit for study of quartz resonator.

resonance and study the behaviour of the vibrator. The electrodes are connected to the terminals of the capacitor of an oscillatory circuit L, C, Fig. 7, and provision is made for measuring the current in L by means of a thermo-junction and galvanometer. An alternating e.m.f. is induced in L by means of a valve generator of adjustable frequency. Let C be adjusted so that the natural frequency

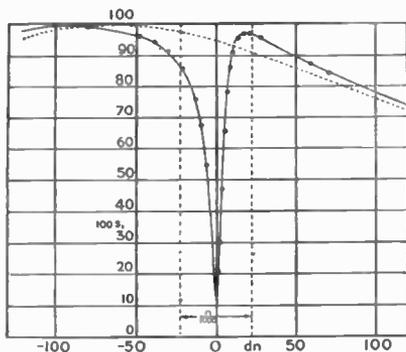


Fig. 8.—*Experimental response curve for a slightly detuned circuit (Dye).*

frequency of the coil and capacitor, viz., $1/2\pi\sqrt{LC}$, is approximately equal to the natural frequency of vibration of the quartz. At first suppose the quartz disconnected; as the frequency of the generator is varied from a low value, the current in the coil L increases, reaches a maximum and decreases to

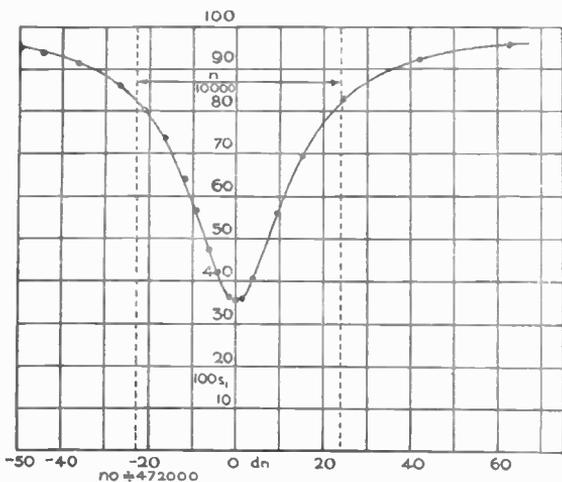


Fig. 9.—*Experimental response curve for a tuned circuit.*

nearly zero. This curve is the ordinary resonance curve of a coil and capacitor. Now suppose the quartz connected across the capacitor, and the

variation of frequency repeated; at first the current increases as before and reaches very nearly the same maximum, but then the vibrations of the quartz become large owing to resonance, and the quartz acts as a shunt across the capacitor, the lower the damping, the larger the vibrations and the lower the equivalent shunt resistance. In consequence, the overall damping of the electrical circuit is increased and there is a sharp dip in the resonance curve. In other words, a sharp resonance curve, that of the quartz, is added upside down to the relatively flat resonance curve of the coil and capacitor. As the frequency is further increased the quartz vibrations die-down, and the current resumes the course of the resonance curve of the coil and capacitor. Fig. 8 shows the result of an actual experiment, the dotted curve is the top of the resonance curve of the coil and capacitor, and the full curve shows the variation of current when the quartz is in circuit. By trial and error adjustment of the capacitor, a symmetrical crevasse can be obtained, as in Fig. 9.

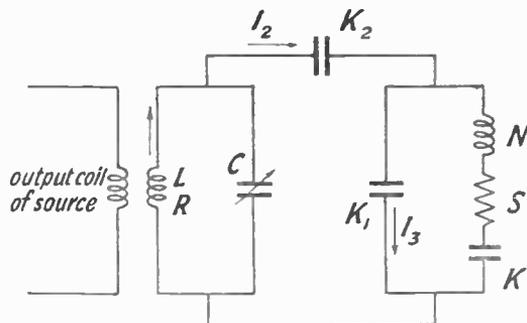


Fig. 10.—*Electrical equivalent network of quartz resonator.*

6. Electrical Equivalent Circuit

Since the vibration of the quartz is both produced and detected electrically, it is convenient to represent the resonator by an electrical network, as in Fig. 10, where N is an inductor representing the mass of the quartz and K is a capacitor representing the restoring constant; since there is damping, although only to a slight degree, a resistor S is added. The capacitor K_1 stands for the dielectric capacitance, which is much larger than K. The decrement of a properly mounted resonator is much smaller than that of the best electrical oscillatory circuits, and that is why quartz gives a much sharper resonance curve than electric

circuits. That is also one of the reasons why quartz vibrators are so useful for filters and as precision standards of frequency.

The electrodes are not always sheets of tinfoil coated to the faces; for many purposes it is more convenient to use brass plates a little distance away from the faces. The gap between the quartz and electrodes then constitutes an additional capacitance in series with the network of the quartz proper, and this capacitance affects the natural frequency to a slight extent. If, therefore, a quartz resonator is to be employed for frequency standardization, the holder must be of a construction which ensures constancy of air gap.

7. Effect of Air Gap on Damping and Frequency

An interesting phenomenon can be observed when the gap above a plate is varied gradually; the vibration of the quartz causes a forced vibration of the short column of air in the gap, and when the gap is of such a length that the column of air

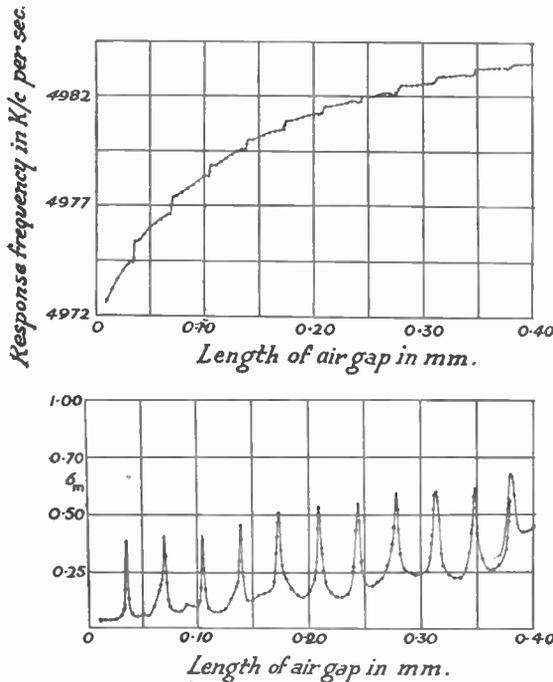


Fig. 11.—Influence of air gap on minimum of crevasse and on frequency. Illustrating effect of resonance of air column in gap.

resonates, the overall damping of the resonator increases, and therefore the crevasse of Fig. 9 is not so deep as it was before. The lower curve

of Fig. 11 is a graph of the minimum of the dip against length of air gap; in other words, the distance from the top represents the depth of the crevasse. Whenever the column of air is an integral number of half wavelengths it resonates, and there is a sharp decrease in the depth of the dip. This experiment yields the velocity of supersonic waves in air; in this case the result is 344 metres per second. In that experiment the response frequency also was measured, and plotted as the upper graph in Fig. 11. The rapid change of frequency each time the air resonates is to be noted. If the measurements had been conducted *in vacuo* both curves would have been quite smooth.

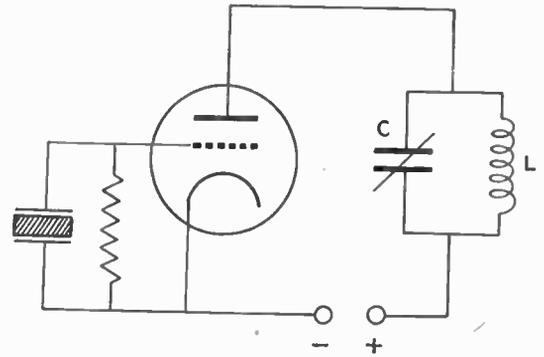


Fig. 12.—Pierce circuit for self-maintained oscillations.

8. Self-maintained Oscillations

In all the cases considered above, alternating voltage has been applied to the quartz, and the frequency has been adjusted for resonance. But it is possible to connect the quartz to a valve in such a way that self-maintained oscillations are produced. Fig. 12 shows a simple circuit for the purpose. As long as the capacitance C is less than the value which would make the natural frequency of the coil and capacitor equal to that of the quartz, the feed-back through the grid-plate capacitance is in the right direction, and oscillations build up and are maintained at the natural frequency of the quartz.

These oscillators open up fresh possibilities, for the quartz is made to furnish power at its natural frequency, instead of being used merely to measure a frequency controlled by other means. For calibrations of wavemeters, for instance, a valve-maintained quartz oscillator is more useful than a luminous resonator, because a large number of calibration points can be obtained by the method

of harmonics. A good procedure is to set an auxiliary valve generator by means of the quartz, and take the wavemeter reading corresponding to the setting of the auxiliary generator. If f_0 is the frequency of the quartz, the auxiliary generator can be set at all frequencies Lf_0/M , where L and M are integers.

Quartz oscillators are also used to drive radio transmitters, and the frequency can be stepped up two, four, or six times if necessary by tuning the anode circuit of one amplifying stage to a harmonic of the output of the previous stage.

9. Quartz Vibrators with Low Thermal Coefficient of Frequency

Most of the work done on vibrators during the few years preceding the war was devoted to the production of vibrators of negligible thermal coefficient of frequency. It may be of some interest to review the methods by which this result has been achieved. For this purpose it is convenient to divide vibrators into two broad classes: plates which vibrate at a frequency controlled by their thickness, and plates which vibrate at a frequency controlled by one of the other dimensions. This second class includes rods vibrating at a frequency determined by their length.

The thickness vibrations will be considered first. Plates cut as described above, i.e., with their normal directed along the X axis, have a thermal coefficient of frequency of about -20 parts in a million per $^{\circ}\text{C}$., and are therefore not very suitable for accurate frequency control.

A method of obtaining a negligible thermal coefficient at some particular temperature is described by Marrison and Lack. If the width of a plate be adjusted so that the frequency of an overtone vibration along the width is nearly equal to the frequency of the pure thickness vibration, there is coupling between the two types of vibration, and the resulting vibration has a frequency in the neighbourhood of the other two. The thermal coefficient of this frequency depends on the coefficient of the overtone mode as well as on that of the pure thickness mode, and if these two coefficients are of opposite sign, the resultant coefficient can be made to vanish at a particular temperature by adjustment of the width of the plate.

An X plate cannot be used for this purpose, because the thermal coefficients of the two modes independently, and of the coupling factor, are not of the right order to yield a zero resulting coefficient. But a Y plate, i.e., a plate

with its normal directed along Y, can also be excited piezo-electrically at a frequency dependent on the thickness y , but the vibrations are of shear type instead of being longitudinal, as with X plates. They exhibit, however, a positive thermal coefficient of frequency of about 75 parts in a million per $^{\circ}\text{C}$. At the same time, the shear strains give rise to vibrations governed by the width x of the plate. As the width x of the plate is decreased gradually, the frequency of these vibrations in-

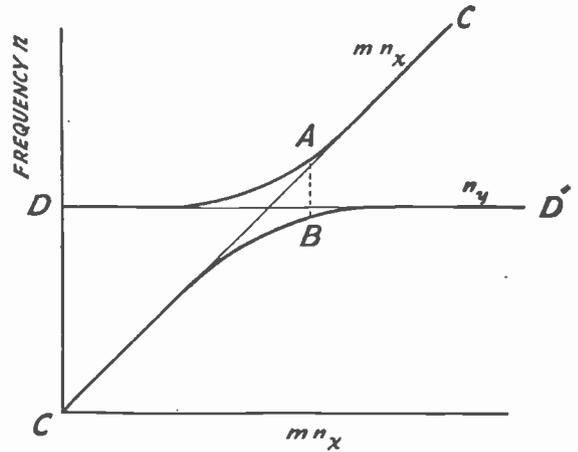


Fig. 13.—Natural frequency of two coupled systems as function of natural frequency of one of them alone.

creases. Denote by n_x the frequency of the fundamental mode, and concentrate attention on the vibration of m th order, whose frequency, approximately equal to mn_x , is near the frequency n_y of the thickness vibration. Taking mn_x as abscissa and frequencies in general as ordinates, and assuming for a moment that the two modes of vibration do not influence each other, the graph of the overtone frequency is a straight line OC, Fig. 13, at 45° to the co-ordinate axes, and the graph of the frequency of the thickness vibration is a straight line DD' parallel to the horizontal axis. But in reality there is some mechanical coupling between the two modes of vibration, with the result that when mn_x is nearly equal to n_y , the frequency of vibration of the plate is equal to neither, but has either of two values, one below and the other above n_y . Thus, if the crystal is ground down so that the width x , originally larger than the value for which mn_x is equal to n_y , is decreased gradually, the frequency of the thickness vibration increases slightly, following the curve DA, until at some point A there is an abrupt decrease to a frequency

corresponding to the point B, lower than n_y . With further reduction of the width of the plate the frequency of the thickness vibration increases gradually along the curve BD' until it reaches the normal value n_y .

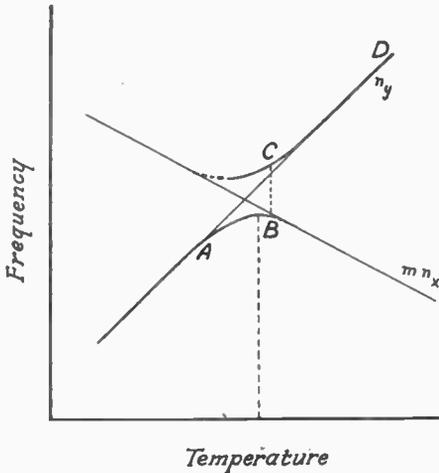


Fig. 14.—Effect of temperature on the natural frequency of two coupled systems when the thermal coefficients of the individual systems have opposite signs.

The thermal coefficient of frequency of the resulting vibration is also a function of the individual thermal coefficients. In Fig. 14 the two intersecting straight lines represent the variations with temperature which would occur if the two modes were not

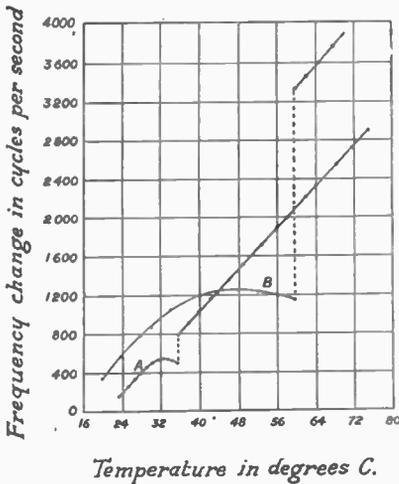


Fig. 15.—Frequency change with temperature of two Y plates of different widths (Lack)

Curve A: $x = 32$ mm. $m = 8$
 Curve B: $x = 19.35$ mm. $m = 5$
 Both curves: $y = 2.76$ mm. $z = 47$ mm.

influenced by each other. The abscissa of the point of intersection can be adjusted to any desired temperature by variation of the width of the plate. If then the temperature is increased from a low temperature, the frequency follows the curve AB until just past the region of zero coefficient, after which there is a sudden increase of frequency to C, and with further increase of temperature, the frequency increases along the curve CD. Figure 15 shows two experimental curves. They indicate that the smaller the value of m , i.e., the smaller the width/thickness ratio, the larger is the range of temperature for which the coefficient is negligible.

This necessity for a small width/thickness ratio is a drawback for high frequencies, because then it is necessary to use thin plates, and they would have to be very narrow for resonance of, say, the 5th order width vibration with the thickness vibration. A greater drawback still is the possibility of vibration at an alternative frequency at any one temperature. For instance, if after the temperature has been increased past the points C, B, Fig. 14, it is decreased gradually to the most favourable temperature, the frequency may retain a value corresponding to that part of the upper curve which is to the left of C, instead of resuming

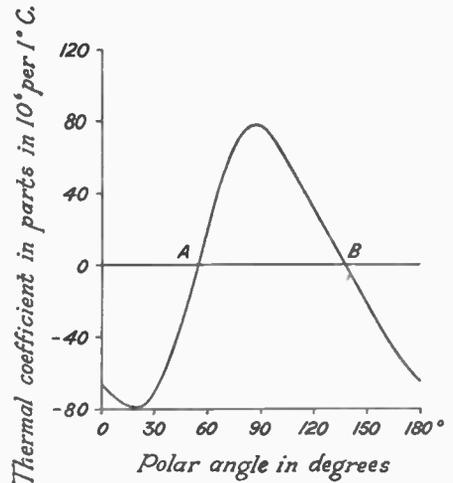


Fig. 16.—Thermal coefficient of frequency of plate rotated about X (Bechmann).

the value corresponding to the curve AB. In fixed installations where the temperature is carefully controlled, this defect may be overcome, and Y plates have been used with success by Marrison in the construction of highly accurate quartz oscillator clocks.

For other purposes, however, it becomes desirable to explore other means of securing a negligible thermal coefficient. Bechmann, Koga, Lack, Willard and Fair and others have shown that the thermal coefficient of frequency of rectangular plates having one side parallel to X, the other in the X plane, at an angle with Y, varies with that angle, being zero for two angles, as shown at A, B, Fig. 16 (p. 52). The orientation of the plates in the quartz crystal is illustrated in Fig. 17, reproduced from Mason, who distinguishes the two cuts by the letters AT and BT. The angles measured counter-clockwise from Y are 55° and 138.5° . It is also found that the frequency varies considerably with angle, as shown by Fig. 18, and it has been pointed out that the angles 59° and 150°

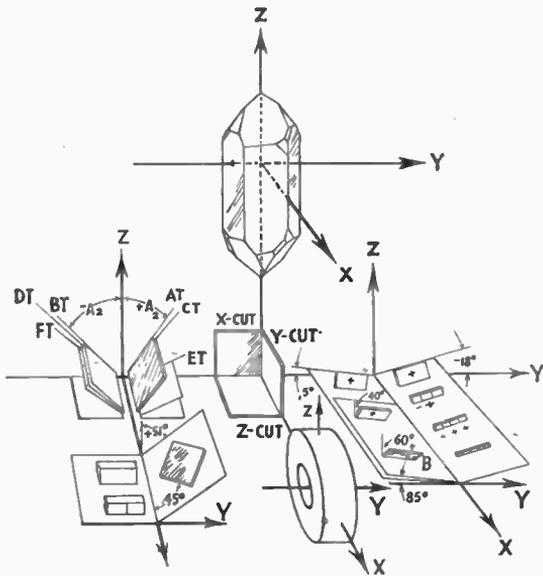


Fig. 17.—Oriented quartz crystal cuts in relation to the natural crystal (Mason)

of minimum and maximum frequency are also the orientations for which the thickness shear mode has zero coupling with the troublesome low-frequency overtones. In view of the small difference between the angles 55° of zero thermal coefficient and 59° of zero coupling, the AT cut is very advantageous, and free from the sudden frequency changes which constitute one of the great drawbacks of Y plates adjusted for negligible thermal coefficient; Fig. 19 illustrates this point.

For lower frequencies, e.g., in filters, it becomes necessary to explore the second class of vibrators, in which the frequency is controlled by the length or width of the plate.

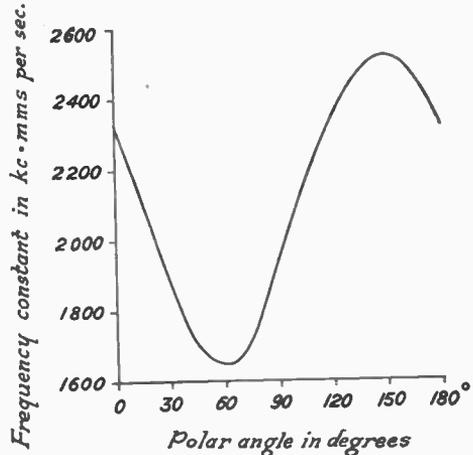


Fig. 18.—Frequency constant of plate rotated about X (Bechmann)

Mason has shown how to cut such vibrators for negligible thermal coefficient. A plate cut at an angle to the Y plane, as in Fig. 17, can be made to vibrate in a low frequency pure shear mode when excited by electrodes applied to the faces, and that mode is found to have, in general, a positive thermal coefficient. Suppose a smaller plate to be cut from

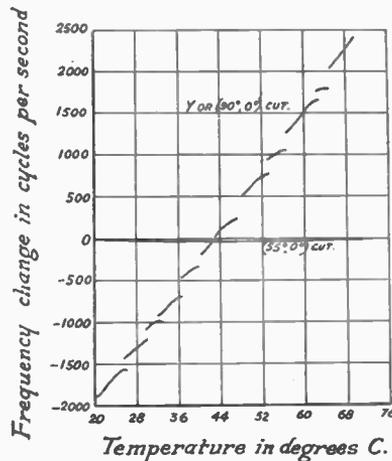


Fig. 19.—Frequency-Temperature characteristic of Y and of 55° plates of 1 Mc/s (Lack, Willard & Fair)

Y Plate thickness 1.970 mm.
 55° plate thickness 1.675 mm.
 Both plates 38×38 mm.

the first with its sides at 45° to the original sides, as in Fig. 20. The shear mode of vibration is seen to be equivalent to two coupled pure longitudinal

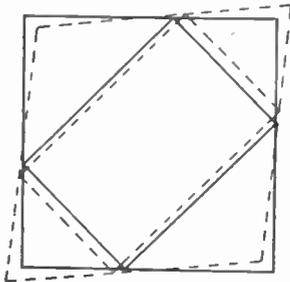


Fig. 20.—Relation between a face shear mode and two coupled longitudinal modes (Mason)

modes, with a resulting positive thermal coefficient. But since a pure longitudinal vibration not coupled with any other mode has, in general, a negative thermal coefficient, it follows that a reduction of the coupling, effected by an alteration in the ratio of the sides of the plate, tends to reduce the thermal coefficient. There is, therefore, as Mason points out, a particular ratio of sides for which the coefficient vanishes at a particular temperature. If t denotes the temperature deviations from that temperature, and f the frequency deviations from the frequency at that temperature, the relation between f and t is of the form

$$f = a_2 t^2 + a_3 t^3 + \dots$$

This formula indicates that if some way could be found of making a_2 vanish, the frequency variations with temperature would be still further

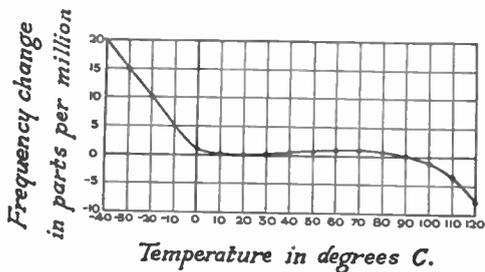


Fig. 21.—Temperature-Frequency characteristic for GT plate (Mason)

reduced, because the origin would be at a point of inflexion instead of being at a maximum or minimum. Mason has shown that for a plate at 51° 30' (see Fig. 17), the large square term disappears, and the frequency is constant to one part in a million over a 100° C. range centred at 50°, as shown by Fig. 21. Further experiments showed

that this flat range can be shifted by slight simultaneous changes in the angle of cut and the ratio of axes. Thus, the so-called GT plate, cut at 51° 7.5' with an axial ratio of 0.859 has negligible thermal coefficient over the range - 25° C. to + 75° C., and has been extensively used in frequency standards and filters.

Long thin rods are also useful for low-frequency filters, and are more economical than plates. Here again, it is possible to secure zero thermal coefficient by proper orientation of cut, for longitudinal vibrations can be excited piezo-electrically in rods oriented otherwise than along Y. For example, in the case of rods lying in the X plane, the component of strain along the length of the rod is sufficient to excite vibrations in all rods except those which are oriented along Z itself. When experimenting with rods in a search for orientations of negligible thermal coefficient, it is essential to keep the lateral dimensions small com-

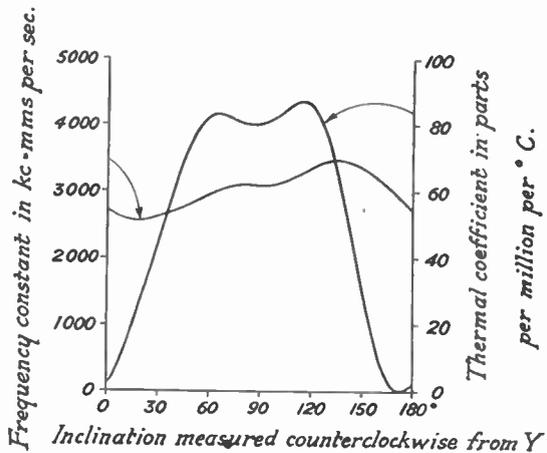


Fig. 22.—Frequency constant and thermal coefficient of rods in X plane

pared with the length, otherwise the result may be vitiated by coupling effects. Matsumura and Kanzaki (1932) and Bechmann (1934) observed the frequency constant and the thermal coefficient of a number of rods in the X plane at various inclinations to the Y axis, and found that the thermal coefficient is exceedingly small for an inclination of 5° with the Y axis. This orientation is also shown in Mason's diagram, Fig. 17. As it differs by only 5° from that for which the piezo-electric effect is a maximum, excitation is easy, and the performance is good. Figure 22 shows the frequency constant and the thermal coefficient for all orientations.

Lastly, mention should be made of Essen's ring, for longitudinal vibrations of the third order along the circumference. The ring is cut with its axis along Z, and the electrode arrangement (shown in Fig. 23) is extremely simple, consisting merely of two metal rings coaxial with the vibrator, one

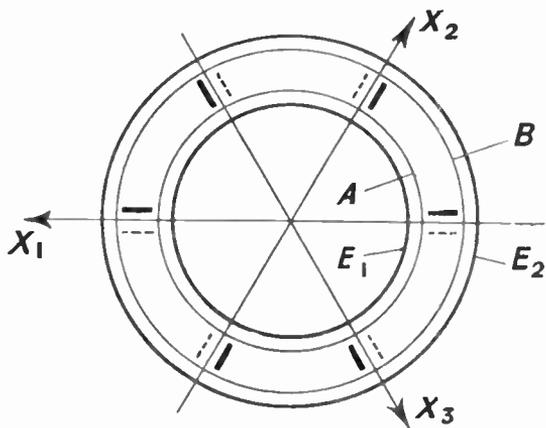


Fig. 23.—Ring vibration (Essen)

inside, the other outside. The diagram indicates that piezo-electric excitation will produce maximum stress at each of the six points where the electric axes cut the circumference, but since at adjacent points the axes cut the circumference in opposite sense, the sense of the stress will alternate. Thus, at any time when a segment equal to one-sixth of the circumference is in compression, the adjacent segments are in tension, but the total length of the circumference remains unchanged, with the result that air damping is minimized, and from that point of view there is no necessity for a vacuum. Also there is a node at each of the six points above referred to, and the ring can be clamped at three of them without disturbing the vibrations. The thermal coefficient of frequency of a thin ring in the Z plane is only -3 parts in a million per $^{\circ}\text{C}.$, as can be seen from Fig. 22. But this coefficient can be reduced to zero by an increase of the lateral dimensions to secure some coupling of an overtone of the circumferential mode with a lateral mode. Final adjustment of the frequency to a given value, and simultaneous adjustment of the thermal coefficient to vanish at some particular temperature are easily effected with rings, because grinding inside and outside by equal amounts alters the thermal coefficient but not the frequency, whereas grinding one side only alters the frequency. The

curve relating frequency with temperature is shown in Fig. 24. While not so independent of temperature as the GT plate, the ring is seen to be perfectly well suited for permanent installations, like standard clocks, which admit of fairly precise thermostatic control.

10. Quartz Clocks

The development of vibrators having a negligible thermal coefficient of frequency has facilitated the construction of quartz clocks which not only constitute frequency standards of higher accuracy than any made before, but also keep better time over long periods than the best pendulum clocks available. In fact, sudden changes in the rate of rotation of the earth have been detected by quartz clocks.

A GT plate or a ring vibrating at 100 kc/s forms a suitable vibrator. The holder must be designed for great constancy of frequency, for it is essential

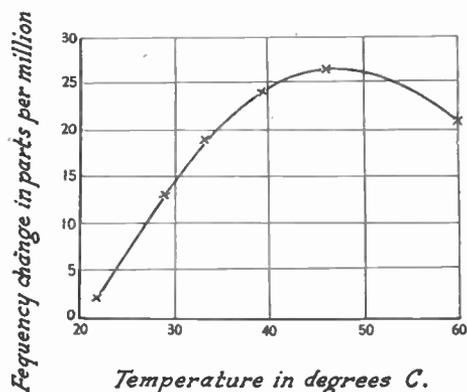


Fig. 24.—Variation of frequency with temperature for ring vibrator (Essen)

to rule out secular changes as completely as possible. Whatever the type of vibrator chosen, the holder is kept in a temperature-regulated enclosure at constant (usually low) pressure, in order to eliminate the effect of a residual thermal coefficient, and of humidity and variation of atmospheric pressure.

The amplifier coupled to the Pierce or other circuit by which the vibrations are maintained provides voltage alternating at the frequency of the vibrator. The next problem is to obtain hours, minutes and seconds from that voltage. For that purpose, multivibrators can be used as frequency dividers. A multivibrator, Fig. 25 (p. 56), consists of two valves intercoupled through resistors and capacitors. It gives a relaxation oscillation rich in

harmonics, at a frequency depending on the values of the resistors and capacitors. The capacitors are adjusted so that the frequency of the relaxation

The output at 1 kc obtained by two or more divisions, Fig. 26, is amplified and applied to a synchronous motor, which either runs synchro-

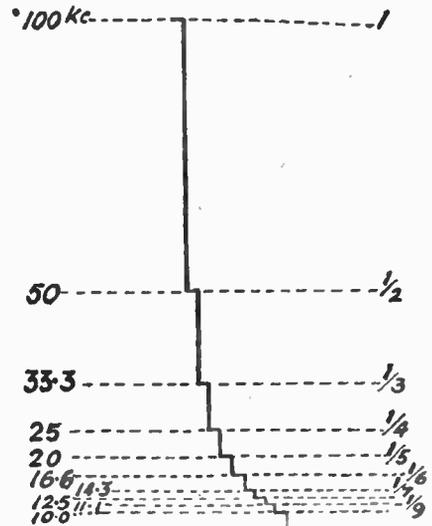
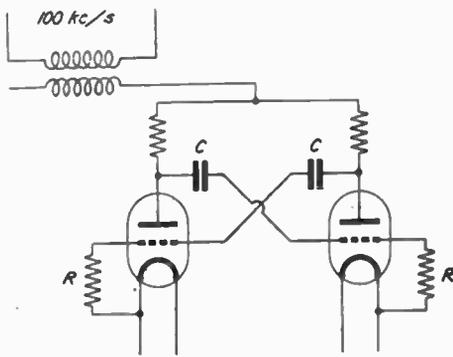
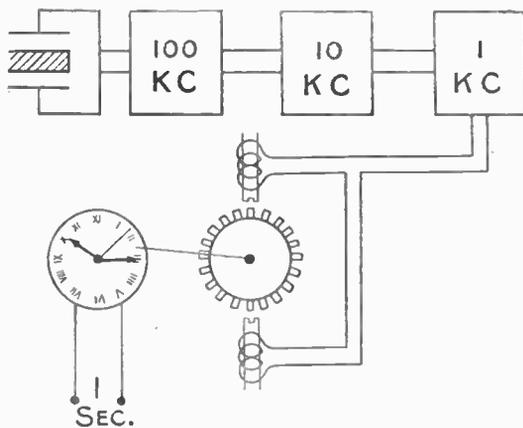


Fig. 25.—Multivibration for frequency division

oscillation is nearly equal to a sub-multiple of the quartz frequency, and a small e.m.f. from the amplifier is injected into the multivibrator, which is thereby locked into step, and oscillates at an exact sub-multiple of the quartz frequency for a considerable range of its capacitances.

nously or does not run at all, so that there is no doubt that the period of rotation of the wheel is an integral multiple of the period of the oscillations of the multivibrator. In practice, unless the period of rotation of the wheel is an integral multiple of the period of the quartz, it falls out of step and stops.



By means of appropriate gearing the wheel is made to drive pinions once every nominal second, nominal minute and nominal hour. The "second" thus obtained is not in general an astronomical second ; it would be so only if the frequency of the quartz were exactly 100 kc/astr. sec. Actually it might differ from this value by one part in ten million, or perhaps even one part in a million. The interval obtained from the quartz clock will therefore be called a "quartz" second. The pinions are made to close contacts once every revolution, and thus provide quartz seconds, minutes and hours. The pinions also turn the hands of a clock dial.

Fig. 26.—Frequency division by multivibrators and synchronous motor

The chronograph shown in Fig. 27 was used by Dye in original comparisons of the rates of a pendulum clock, a valve-maintained tuning fork and a quartz oscillator. The seconds pinion of the synchronous motor also turns the drum of the chronograph, and the hour pinion turns a worm of

1 mm. screw pitch. This worm translates a carrier provided with a pen actuated by an electromagnet ; if the electromagnet is energized for an instant, the pen is depressed and marks the paper cover of the rotating drum. The carrier takes a week to travel from one end of the drum to the other.

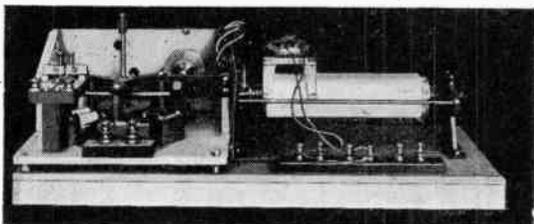


Fig. 27.—Dye's Chronograph used for comparing rates of a pendulum clock, a valve-maintained tuning fork and a quartz oscillator

Suppose it is desired to compare the rates of a pendulum clock and a quartz vibrator : the quartz is made to drive the motor, and a contact closed once every second by the pendulum is included in the electromagnet circuit. But as it stands this device would actuate the pen once every clock second, whereas all that is required is a mark every hour. A contact closed by the hour pinion of the motor for an interval of three or four seconds once every hour is therefore also inserted in the circuit. The three or four marks made every hour are merely superimposed, and appear as a single dot on the record.

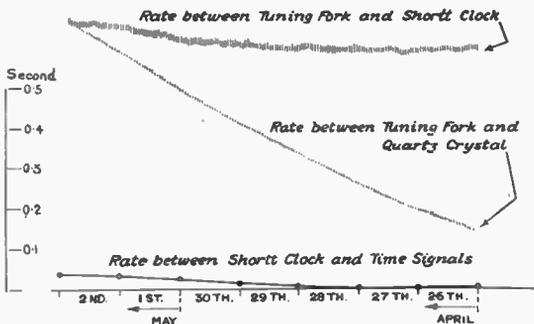


Fig. 28.—Comparison of the rates of pendulum clock, tuning fork and quartz oscillator

Suppose the rate of the vibrator and the rate of the clock were identical ; the dots would all be on a generator of the drum. If the rates were constant but differed slightly, a spiral would be formed on the drum. When the paper is unrolled and laid flat, this spiral becomes a straight line at an angle with the edge of the paper.

Fig. 28 is an interesting record of a comparison of the rates of the pendulum clock, the tuning fork and the quartz oscillator. The tuning fork drove the synchronous motor, and the clock seconds and quartz seconds actuated the pen. The line of dots marked by the clock is on the whole very nearly parallel with the base, indicating that the average rate of the clock is very nearly the same as the average rate of the fork ; but the line is not very smooth. On the other hand, the line of dots formed by the quartz, although indicating by its slope a rate difference of about one part in a million, is very smooth, thus showing that the difference between quartz and fork is subject to none of the irregularities exhibited by the difference between clock and fork. As the two vibrators were driven quite independently, it is legitimate to conclude that they constitute, at least for periods of a few days, better standards than even the best clocks.

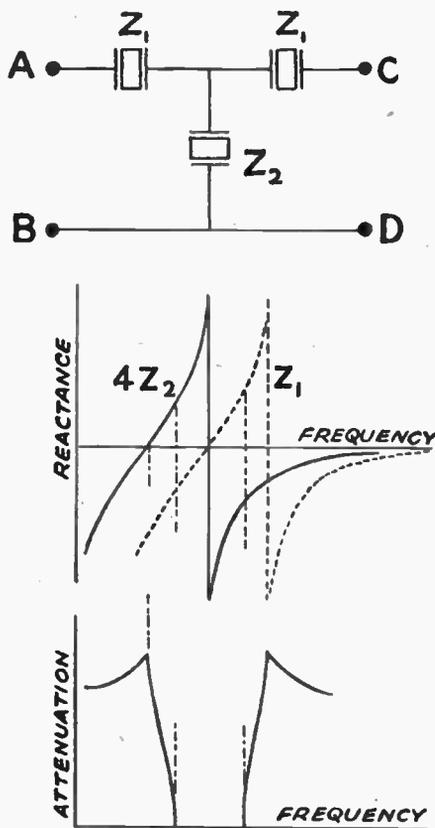


Fig. 29.—Ladder type filter with quartz oscillator elements

These conclusions have, of course, been abundantly verified since for periods of months rather than of days, but an account of these quasi-pioneer investigations on the subject has been given here because they seem to be not without interest, and they provide a good illustration of the principles involved.

which a large number of different conversations are transmitted along a single core of a cable, by modulating a number of different carriers and transmitting only one side band of each modulation. A ring modulator and filter for suppression of the carrier and removal of all frequencies except those of the required side band is shown in Fig. 30.

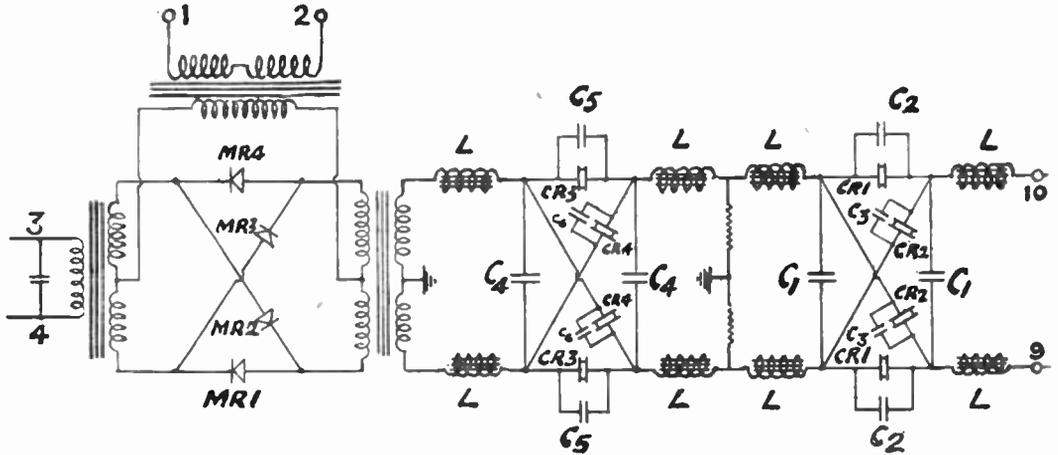


Fig. 30.—Ring modulator and filter for multi-channel telephony

11. Quartz Filters

Filters are an important application of quartz vibrators. Figure 29 (p. 57) shows a filter of a type suitable for heterodyne wave analysers, where a narrow band of frequency is sampled by heterodyning and filtering, and the output measured. The identical crystals 1, 1 resonate at a frequency slightly higher than the crystal 2 : if Z_1 and Z_2 are their reactances, which, of course, vary rapidly with frequency as indicated in the figure, the filter passes all frequencies such that $Z_1/4Z_2$ lies between 0 and -1 . At all other frequencies there is attenuation, in particular at the frequency for which Z_2 is nearly zero, and that for which Z_1 is very large, so that the attenuation varies as indicated in the lower diagram of the figure. This type of filter would also do for fixing the intermediate frequency in a superheterodyne receiver.

In the case of a communication receiver where a "single signal" is required and no band for accommodation of conversation is wanted, a simpler filter is obtained by balancing a single crystal against a capacitor in adjacent arms of a bridge.

But, perhaps the most important application of crystal filters is in multi-channel telephony, in

Filter insertion loss-Frequency characteristic measured between 300Ω terminations.

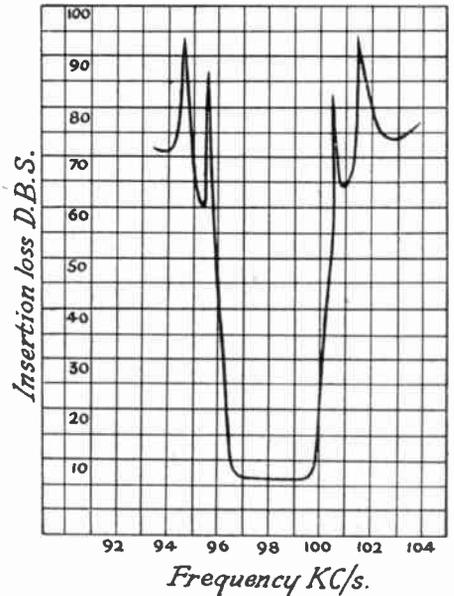


Fig. 31.—Insertion loss of multi-channel telephony filter

A number of such filters for contiguous bands of about 4 kc/s are connected to one end of the cable, and identical devices to the other end, so that although the cable itself carries all frequencies between, say, 100 kc/s and 200 kc/s, yet the output from any filter at the far end contains only

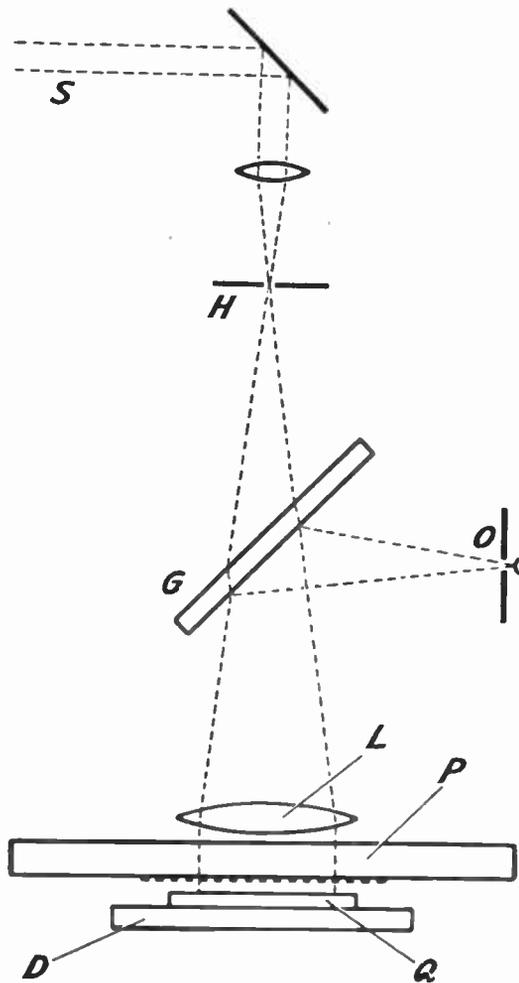


Fig. 32.—Interferometric examination of type of vibration (Dye)

the frequencies introduced via the corresponding filter at the near end. Naturally, the filters must be very permanent and reproducible, and quartz crystals are used. The filter in Fig. 30 (p. 58) is a two-section lattice filter, with small capacitors across the crystals for adjustment of the band width. Figure 31 (p. 58) shows that very high attenuation is secured.

12. Nature of Vibrations of Plates

It is interesting to examine the vibrating surfaces of quartz plates. Fig. 32 shows an interferometer developed for the purpose by the late D. W. Dye of the National Physical Laboratory. The quartz plate Q under investigation rests on a brass plate D which serves as lower electrode. Less than $1/10$ th

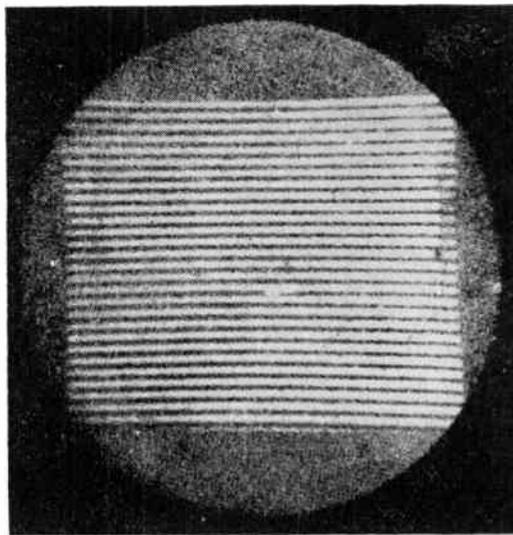


Fig. 33.—Fringes for stationary plate

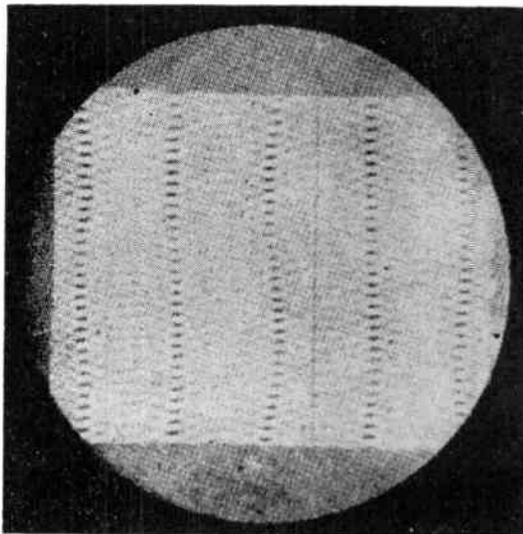


Fig. 34.—Flexural vibration of quartz plate examined with continuous light

inch above is a glass plate P supported on levelling screws ; its lower face is polished flat, and constitutes a reference surface, but the upper face is purposely given a slight inclination to the lower, so that any light reflected from it is deviated from the line of vision. To the lower surface of the glass plate is stuck a thin frame across which silver wires a thousandth of an inch in diameter are stretched—about 25 wires to the inch. This grid forms an upper electrode which only slightly obstructs the light, and the quartz can be excited into resonant vibrations in the ordinary way.

The object of the instrument is to provide interference patterns between light reflected from a fixed surface—here the lower surface of the glass plate or reference surface—and the moving surface of the quartz. For this purpose, mono-chromatic light in parallel rays is thrown on to the surfaces from above. Part is reflected by the reference surface, but most of it reaches the upper surface of the quartz. Again, part of this light is reflected, and the two beams of reflected rays, after deviation by an inclined glass plate G, are viewed or photographed through the pinhole O.

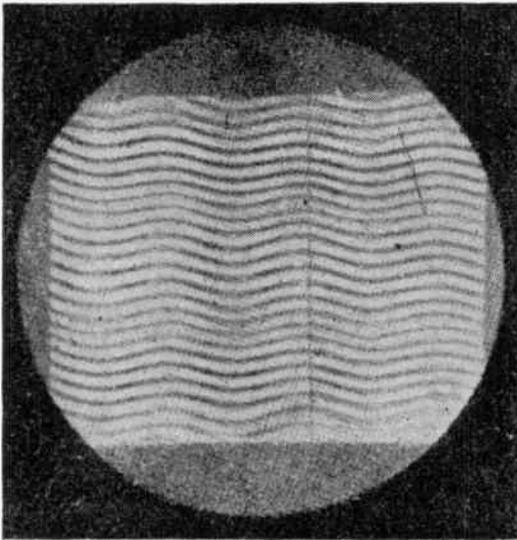


Fig. 35.—Flexural vibration of quartz plate examined with intermittent light

When the quartz is motionless, and its surface flat and truly parallel to the reference surface, the field of view is more or less dark or bright depending on the distance between the two surfaces, for one beam of light is longer than the other by

double that distance, and may interfere with it or reinforce it. When the reference plate P is given a slight inclination, parallel fringes appear as in Fig. 33 (p. 59).

If, now, the quartz vibrates, the interference fringes become blurred wherever there is up-and-

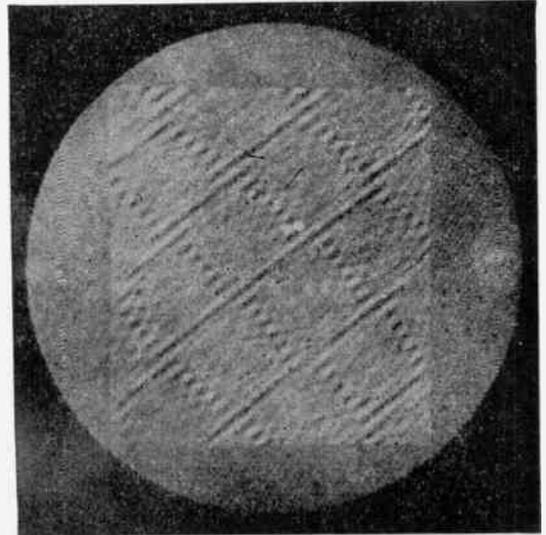


Fig. 36.—Overtone vibration of quartz plate, continuous illumination

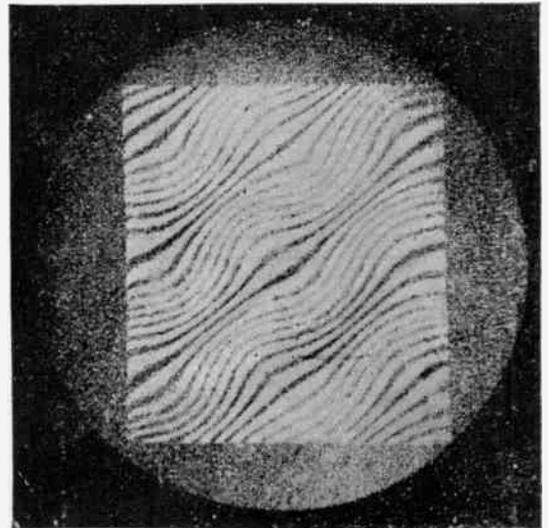


Fig. 37.—Overtone vibration of quartz plate, stroboscopic examination

down motion of the surface, since this motion causes a to-and-fro motion of the fringes. The pattern will, of course, remain unchanged wherever the surface is stationary. The nodes and antinodes can thus be put in evidence, as shown in Fig. 34 (p. 59) which illustrates flexural vibration of the 5th

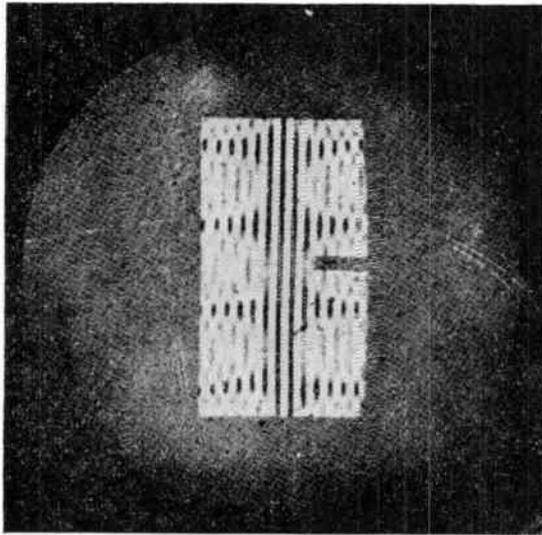


Fig. 38.—Overtone vibration of quartz plate

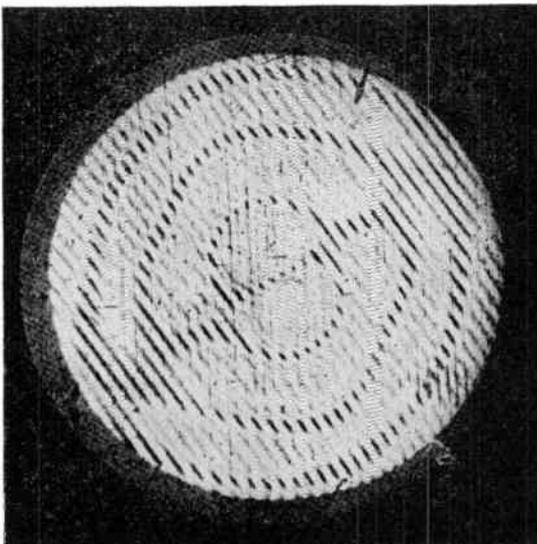


Fig. 39.—Overtone vibration of quartz disc

order. Special electrode arrangements are sometimes required for exciting particular modes of vibration. In the present case, the upper electrode was a single wire, visible in Fig. 34, instead of the usual grid of wires.

Dye also developed a method for the measurement of the amplitude of vibration. In this method, the surfaces, instead of being illuminated with continuous light, are illuminated with light from a helium lamp polarized by direct voltage, and fed from the same valve generator which excites the quartz. Thus, the tube flashes only once every cycle and by adjustment of capacitors the instant of illumination is made to coincide with the instant at which the quartz plate is momentarily at rest at one end of its vibration. In this stroboscopic pattern, the interference fringe becomes displaced, but is not blurred by the vibration. If p be the distance between successive fringes with the plate at rest, and λ be the wavelength of the light used, the amplitude of vibration is equal to the displace-

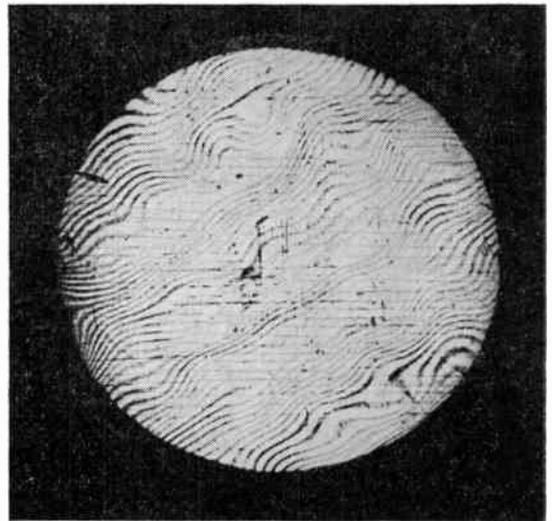


Fig. 40.—Thickness vibration of quartz disc, continuous illumination

ment of a fringe multiplied by $\lambda/2p$. Figure 35 (p. 60) shows the same mode of vibration as before, viewed by this method. Figures 36, 37 (p. 60) and Figs. 38 and 39 show other interesting nodal patterns.

Vibrations corresponding to the thickness, which are the important case in practice, do not as a rule exhibit the beautiful regularity of the low frequency flexural modes. In some plates, a fairly large

proportion of the surface remains stationary, whereas in others almost the entire surface takes part in the vibrations, as in Fig. 40 (p. 61). These are good vibrators. It is precisely for those high-

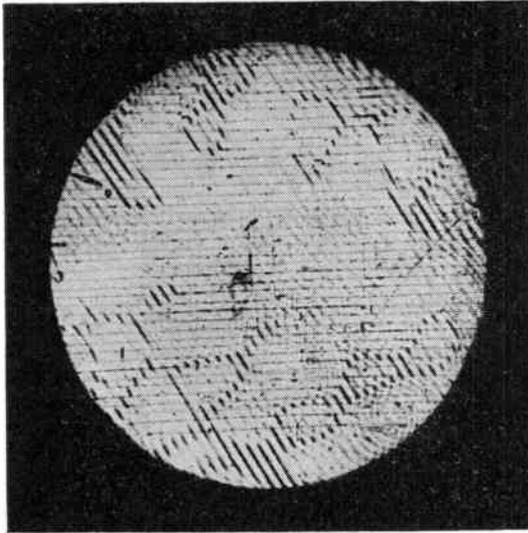


Fig. 41.—Thickness vibration of quartz disc, stroboscopic examination

frequency modes that the stroboscopic pattern, Fig. 41, proves useful, because when the nodal pattern is irregular, it is not easy to tell by examination of the continuous illumination photograph, which parts of the surface are moving up at any one time and which are moving down.

The examination of vibrating surfaces by this method has been mentioned here because, although of more academic interest than the applications described above, it constitutes an advance in the study of vibrators and illustrates useful physical methods of investigation.

13. Acknowledgment

It is desired to thank the Chief of the Royal Naval Scientific Service for permission to give this paper and publish it in this Journal.

References

1. Cady, W. G. "The Piezo-electric Resonator." *Proc. Inst. Radio Engrs.*, N.Y., 1922, 10, 83.
2. Giebe, E., and Scheibe, A. "Piezoelektrische Erregung von Dehnungs, Biegungs und Drillungsschwingungen bei Quarzstäben (Piezo-electric Excitation of Longitudinal, Transverse and Torsional Vibrations in Quartz Rods)." *Zeits. für Phys.*, 1928, 46, 607.

3. Dye, D. W. "The Piezo-electric Quartz Resonator and its Equivalent Electrical Circuit." *Proc. Phys. Soc.*, 1926, 38, 399.
4. Van Dyke, K. S. "The Piezo-electric Resonator and its Equivalent Network." *Proc. Inst. Radio Engrs.*, N.Y., 1928, 16, 742.
5. Pierce, G. W. "Piezo-electric Crystal Resonators and Crystal Oscillators applied to the Precision Calibration of Wavemeters." *Proc. Amer. Acad. Arts and Sci.*, 1923, 59, 81.
6. Marrison, W. A. "A High Precision Standard of Frequency." *Bell Syst. Tech. J.*, 1929, 8, 493.
7. Lack, F. R. "Observations on Modes of Vibration and Temperature Coefficients of Quartz Crystal Plates." *Bell Syst. Tech. J.*, 1929, 8, 515.
8. Bechmann, R. "Temperatur-koeffizienten der Eigenschwingungen piezoelektrischer Quarzplatten und Stäbe (Thermal Coefficients of the Natural Frequencies of Piezo-electric Quartz Plates and Rods)." *Hochfrequenztechn. u. Elektroakust.* 1934, 44, 145.
9. Koga, I. "Thermal Characteristics of Piezo-electric Oscillating Quartz Plates." *Rep. Radio Res., Japan*, 1934, 4, 61.
10. Lack, F. R., Willard, G. W., and Fair, I. E. "Some Improvements in Quartz Crystal Circuit Elements." *Bell Syst. Tech. J.*, 1934, 13, 453.
11. Mason, W. P. "Quartz Crystal Applications." *Bell Syst. Tech. J.*, 1943, 22, 178.
12. Matsumura, S., and Kanzaki, S. "Temperature Coefficient of Frequency of Y-wave in X-cut Quartz Plates." *Rep. Radio Res., Japan*, 1932, 2, 35.
13. Essen, L. "A New Form of Frequency and Time Standard." *Proc. Phys. Soc.*, 1938, 50, 413.
14. Booth, C. F., and Laver, F. J. M. "A Standard of Frequency and its Applications." *J. Inst. Elect. Engrs.*, 1946, 93, III, 223.
15. Essen, L. "The Quartz Clock." *Roy. Astro. Soc. Occas. Notes*, June, 1938, p. 7.
16. Mason, W. P. "Electrical Wave Filters Employing Quartz Crystals as Elements." *Bell Syst. Tech. J.*, 1934, 13, 405.
17. Dye, D. W. "The Modes of Vibrations of Quartz Piezo-electric Plates as Revealed by an Interferometer." *Proc. Roy. Soc.*, 1932, A, 138, 1.
18. Vigoureux, P. "Quartz Oscillators." H.M. Stationery Office, 1939.

NOTICES

Radio Equipment in the Royal Train

British radio manufacturers have been intimately concerned with the provision of the radio equipment for the Royal train convoy now touring South Africa.

The convoy consists of three trains : a Baggage train, a Pilot train and the Royal train.

The Pilot train equipment includes :

(a) V.H.F. transmitters and receivers for inter-train communication and for communication between the Pilot train and signal boxes en route.

(b) A $\frac{1}{2}$ kW transmitter which is used in conjunction with a roof aerial whilst the train is travelling, and when stationary, with a portable mast aerial.

(c) A common receiving aerial, five coaches away from the transmitter, and equipped with a splitter unit incorporating rejector and filter circuits which enable a number of receivers to be used on one aerial with a minimum of mutual interference. Double channel broadcast service is provided by eight-waveband receivers and a microphone is provided for announcements over loudspeakers in each compartment of the train. Additional receivers provide for reception under difficult conditions and as "stand-by" equipment for the communication receiver. The loudspeakers are fed from two 30-watt amplifiers manufactured to Post Office specification.

Radio apparatus in the Royal train is equipped with V.H.F. equipment and a dual channel distribution system similar to those in the Pilot train, and in the King's study is an eight-waveband receiver of approximately two watts output.

Books Received

Television Receiving Equipment by W. T. Cocking, M.I.E.E., Iliffe & Sons, Ltd., price 12s. 6d. (postage 5d.).

This is the second edition of this informative manual, and though there have been no fundamental changes in British television practice since the first edition was published, the present volume covers many of the improvements in technique which have been effected.

The book gives comprehensive details of television receiving apparatus together with

practical details of design. There are chapters on general principles, interference problems, faults and servicing.

In all, there are twenty chapters and four appendices. It contains 354 pages and is well illustrated.

Undoubtedly a book for the student.

Cables for Radio Frequencies by The Telegraph Construction & Maintenance Co., Ltd., Greenwich, London, S.E.10., who invite all who are interested in electronic engineering to write for a copy of this new publication.

Not only does it give details of R.F. cables developed during the war, but also of post-war designs primarily intended for laboratory and domestic use. The book also contains a chapter on the fundamentals of R.F. cables and data for equipment designers is included in another section in which Bending Radii and tolerances for drilling are tabulated.

Television Programming and Production is the title of a new book just published by Chapman & Hall, 37 Essex Street, London, W.C.2, price 15s. The author is Richard Hubbel.

Unlike some books on the same subject, this volume does not attempt to outline the theory of television equipment, it is concerned more with the technique applied to the production of television programmes.

It is a book intended for the producer and the artist rather than for the technician.

Radio Interference Measurement

A draft amendment to B.S. 727/1937 relating to the measurement of radio interference has been prepared by the British Standards Institution.

The draft has been circulated for technical comment in view of the fact that it suggests a radical change in the technique applied to the measuring of radio interference.

The amendment, therefore, should not be regarded or used as a British Standard, and it is now being considered by the Technical Committee of the Institution.

PULSE MODULATION AND DEMODULATION THEORY†

by

M. M. Levy (Member)*

CONTENTS

Notation

1. Summary

2. Introduction

3. Preliminary Study

3.1. Modulation Process

- 3.1.1. Methods of modulation and methods of limiting the modulation width within the boundaries allocated to each channel are very simple.
- 3.1.2. The modulation process introduces a non-linear distortion which is a function of the shape of the modulating signal.
- 3.1.3. In pulse modulation the harmonic distortion increases with the width of modulation and with the frequency of the signal.
- 3.1.4. Some numerical values.

3.2. Demodulation Process

- 3.2.1. The process of demodulation and channel selection is simple in pulse technique.
- 3.2.2. Approximate analysis: to modulate the width of a periodic pulse train is equivalent in first approximation to modulate in amplitude the harmonic components of the train. Demodulation can be obtained by filtering the sideband component produced by modulation of the D.C. component.
- 3.2.3. In first approximation the only distortion produced by the demodulation process is the "cross-distortion" which appears for frequencies greater than half the chopping frequency.
- 3.2.4. A detailed analysis of the demodulation process shows a complex phenomenon.

4. Detailed Study

4.1. Modulation Theory

- 4.1.1. Notations and hypothesis.
- 4.1.2. Modulation formula.
- 4.1.3. Formula in first and second approximation.

- 4.1.4. Expansion of formula (6): the second harmonic produced by the process of modulation is proportional to the product of the modulation width and the frequency of the modulating wave; and the third harmonic to the square of this product.

4.2. Demodulation Theory.

- 4.2.1. Equation of a pulse.
- 4.2.2. Equation of a train of unmodulated pulses.
- 4.2.3. Equation of a train of modulated pulses range.
 - 4.2.3.1. General formula.
 - 4.2.3.2. Study of an important approximation.
 - 4.2.3.3. Solutions of equation (15).
 - (a) First type of solution.
 - (b) Second type of solution.
 - (c) Other types of solution.
 - 4.2.3.4. Summary of solutions.
 - 4.2.3.5. Comparison with phase and frequency modulation.
 - 4.2.3.6. Degeneration of the modulation for the carriers of high order.
 - 4.2.3.7. Practical distortion formula.
 - 4.2.3.8. Amplitude distortion.
 - 4.2.3.9. Amplitude discontinuities at frequencies submultiple of f_c .
 - 4.2.3.10. Harmonic distortion.
 - 4.2.3.11. Cross distortion.

4.3. Combined Modulation and Demodulation.

- 4.3.1. Harmonic distortion.
- 4.3.2. Cross-distortion.
- 4.3.3. General conclusions.

5. First Experimental Verification of the Theory

- 5.1. Numerical Theoretical Data for a 20-Channel System
- 5.2. Some Experimental Measurements
 - 5.2.1. Amplitude distortion.
 - 5.2.2. Harmonic distortion.
 - 5.2.3. Cross distortion.
- 5.3. Conclusions.

† U.D.C. 621.396.619.16 : 621.396.813.
* The General Electric Co. Ltd. of England, and late of Standard Telephones & Cables Ltd.

NOTATION

Symbol	Meaning	Definition in Paragraph
t	Time	
f_c	Chopping frequency	2
T_c	Chopping period	2
ω_c	Chopping angular frequency	
mT_c	Position in time of the m th pulse when unmodulated	3.1.2
Ω	Modulating signal angular frequency	4.1.1.
F	Modulating signal frequency	
ϕ	Phase shift of the modulating signal $\sin(\Omega t + \phi)$	4.1.1
ω	Spectrum angular frequency	3.2.2
$S(\omega)$	Spectrum of any function ..	3.2.2
$P(t)$	Train of pulses	3.2.2
$P_1(t)$	Equation of one pulse	4.2.1
$P_m(t)$	Equation of the m th pulse..	4.2.2
Δt_m	Shift in time for the m th pulse from its unmodulated position $t = mT_c$	4.1.1
τ_m	Width of the m th pulse	4.2.3.1
n	T_c/n is the channel width ..	4.1.1
λ	Fraction of the channel width modulated	4.1.2
$\Delta\tau$	Amplitude of Δt_m for a given modulated signal : $\Delta\tau = \lambda \cdot T_c/2n$	4.1.2
t^1	$t^1 = t - \frac{T_c}{4n}$	4.2.3.1
N	Positive integer	4.2.2
N_1	Positive integer	4.2.3.3
N_2	Positive integer	4.2.3.3
X_o	$X_o =$	4.1.4
X	$X =$	4.2.3.1
X_c	$X_c =$	4.2.3.3
$J_u(x)$	Bessel function of the first kind and u order for the argument X	4.1.4

1. SUMMARY

Study of pulse modulation and demodulation particularly from the point of view of distortions introduced by these processes.

In the modulation, the position in time of the pulses of a periodic train is shifted proportionally to the amplitude of the signal to be transmitted. The pulses are then transmitted to the receiver where they are demodulated. In the demodulation, the position of one edge of each pulse of a periodic train is moved in time in synchronism with the modulated pulse. This gives a train of variable width pulses which has to be detected to yield the modulating signal.

It is shown that the modulation and the demodulation process introduces distortions and the nature and properties of these distortions are studied. *From this study, it is shown that a distortionless pulse communication system can be evolved by following some very simple rules.*

The study is divided into three main parts: qualitative study, theoretical study and conclusions.

I. Qualitative Study

Approximate study of the nature of the distortions produced in the modulation and in the demodulation process.

1.1. Modulation Process.

It is shown that the modulation process produces an amount of harmonic distortion which increases with the width of modulation and the frequency of the signal.

2. Demodulation Process.

It is shown that modulating the width of a pulse train is nearly equivalent to amplitude modulating the D.C. and harmonic components of the pulse train by the same signal.

From these, two important conclusions are derived:—

- (a) To detect the signal from the variable width pulse train it is sufficient to filter the side-band produced by modulation of the D.C. component.
- (b) The highest signal frequency must be smaller than half the chopping frequency (repetition frequency of the pulse train) in order that the lower side-band produced by modulating the chopping frequency does not come inside the signal frequency band. When this is not

the case an unwanted frequency component appears. This phenomenon is called "*cross-distortion*."

II. Theoretical Study

Detailed study of the modulation, the demodulation and the combined modulation and demodulation process.

1. Modulation Process

It is shown that the *main type of distortion introduced by the modulation process is the harmonic type*. It obeys the following laws :

- (a) The amplitude of the second harmonic varies proportionally to the product of the width of modulation, $\Delta\tau$, and the angular signal frequency, Ω :

$$X_0 = \Delta\tau.\Omega.$$

- (b) The amplitude of the third harmonic is proportional to the square of the above product, and is of negligible amplitude.

It is found that in the majority of the practical applications the amplitude of the second harmonic is much greater than the amplitude of the third. For instance, for $\Delta\tau = \pm 5$ microseconds and a signal frequency of 1.5 kc/s, the second harmonic is 5 per cent. and the third about 0.15 per cent. of the amplitude of the signal.

2. Demodulation Process

It is shown that the *main types of distortion introduced by the demodulation and detection process are the harmonic and the cross-distortions*.

(a) Harmonic distortion.

It obeys the same laws as for the modulation process. Compared to the distortion produced by the modulation process the relative phases and amplitudes are as follows: *the second harmonic has the same amplitude and opposite phase in the two processes ;*

the third harmonic has a negligible amplitude as in the modulation process.

These properties are important when studying the overall harmonic distortion of the combined modulation and demodulation process.

(b) Cross-distortion.

The cross-distortion is produced by the modulation of the carrier of frequency f_c in the demodulation process. It is shown that it is similar to the modulation of the D.C.

carrier except that a sort of degeneration appears. This degeneration increases with the carriers of higher order such as $2f_c, 3f_c \dots$. There is, however, an important difference in that the second and third harmonic are in phase with the harmonics produced in the modulation process. This is important when studying the value of each harmonic distortion in the overall system.

The second harmonic (angular frequency $\omega_c - 2\Omega$) has an appreciable amplitude, in practical cases, of the order of some per cent. to 15 per cent. The third harmonic is negligible, usually smaller than 1 per cent.

3. Combined modulation and demodulation processes — overall distortion characteristics.

Summary of the different types and amounts of distortion is as follows :—

(a) Amplitude distortion.

Negligible amplitude distortion for frequencies lower than half the chopping frequency. The amplitude response has discontinuities at frequencies submultiple of the chopping frequency. For these frequencies the amplitude is a function of the phase-shift existing between the signal and the chopping frequency. At $fc/2$, the normal amplitude can vary ± 100 per cent.; the possible variation decreases very quickly with the sub-multiple order. At $fc/3$ the variation is within the limits $+X_0/4$ and $-X_0/4$; at $fc/4$ the limits are $+(X_0/4)^2$ and $-(X_0/4)^2$. This gives in practical cases a variation smaller than ± 5 per cent. at $fc/3$ and ± 0.25 per cent. at $fc/4$. For the other sub-multiples the variation is negligible.

(b) Harmonic distortion.

The second harmonic has the same amplitude and opposite phase in the modulation and in the demodulation process, and they cancel each other to the degree of approximation of the calculations. In fact there is a small residual second harmonic increasing as the square of $X_0 = \Delta\tau\Omega$, but it may only appear because of the approximations made. In any case its value in practical applications is smaller than 0.5 per cent.

(c) Cross-distortion.

For the carrier of frequency f_c the side band of angular frequency ($\omega_c - 2\Omega$) has the

same phase as the second harmonic in the modulation process and it is shown that this results in an increase of the amplitude of this side-band. For the same reason, the amplitude of the side band of angular frequency $(\omega_c - 3\Omega)$ is also increased.

In practical applications, the side band $(\omega_c - 2\Omega)$ has an amplitude which may reach 15 per cent. This produces an appreciable cross-distortion. To eliminate this distortion it is necessary that the signal frequency band be limited to frequencies smaller than $f_c/3$.

The side band $(\omega_c - 3\Omega)$ usually has values less than 0.3 per cent, and introduces a very small cross-distortion.

III. Conclusions

Summarizing the study, it is shown that if the chopping frequency is nearly equal to three times the highest signal frequency to be transmitted, and if the width of modulation is less than some per cent. of the recurrent period, the theoretical distortions are negligible, the amplitude distortion being less than 1 per cent. and the cross-distortion less than 0.1 per cent.

Details are given of some measurements made on the experimental equipment in order to verify the above conclusions.

2. INTRODUCTION

Work on pulse modulation was started by the author in January 1943 and nearly completed by October of the same year. On the experimental side, a multi-channel radio-link has been built up and tested, giving a great deal of experimental information. On the theoretical side, the process of pulse modulation and demodulation has been explored carefully and gave very useful theoretical facts.

A paper on the experimental work will be published later. The present one deals with the theoretical side.

Pulse modulation is very different from other types of modulation in that no attempt is made to supply continuous information as to the amplitude of the signals to be transmitted; the instantaneous amplitude of this wave is signalled to the distant station at regular intervals of time. The number of times per second that information on the instantaneous amplitude is signalled is referred to as the "chopping frequency," since the wave to be signalled is chopped up at this rate. This definition

was suggested by Ullrich in a very interesting memorandum on pulse modulation. If the chopping frequency is high enough compared to the highest frequency to be transmitted, the signals will be reproduced at the receiver with small distortion.

What are the advantages of "chopping" the wave to be signalled? The main advantage with which we are concerned is that the pulse train occupies only short intervals of time so that a multiple of trains of the same chopping frequency can be sent without interfering with one another. For instance, if the chopping period is 100 microseconds, and if this period is divided into 20 equal parts, 20 independent trains can be sent, each one occupying intervals of time of 5 microseconds repetitive each 100 microseconds. Each train will transmit an independent signal, the modulation being done in these short intervals of 5 microseconds.

Modulation can be obtained either by varying the width or by moving the position of each pulse proportionally to the amplitude of the signal. In both cases, the variations must be confined within the boundaries of the intervals of time reserved to the modulated channel. The first type usually is referred to as *width modulation* and the second as *time-phased modulation*.

Ullrich shows in his memorandum that time-phased modulation gives a better signal/noise ratio, when the most important parameters are taken into account. This is the type of modulation which will be considered in this memorandum.

Demodulation is obtained by transforming first, the time-phased modulated pulses in variable width pulses and then smoothing the pulses by means of a low-pass filter.

From the theoretical point of view, it is important to establish the amount of distortion produced by the process of chopping the signal, modulating the pulses, and demodulating and smoothing them. Such is the object of the present study.

3. PRELIMINARY STUDY

The main object of this section is to study the modulation and demodulation processes and to deduce an approximate theory giving the main properties of pulse transmission.

The processes of modulation and demodulation described are those used in our experimental studies. All methods of modulation and demodulation using time-shaped modulation and

converting it in variable width for demodulation purposes differ only in practical details from the methods described in this section. The approximate theory in this section and the detailed theoretical study of the next section can be applied to nearly all experimental methods.

3.1. Modulation Process

3.1.1. *Methods of modulation and methods of limiting the modulation width within the boundaries allocated to each channel are very simple.*

The modulation was produced by means of a circuit shown in Fig. 1 in its simplest form. A

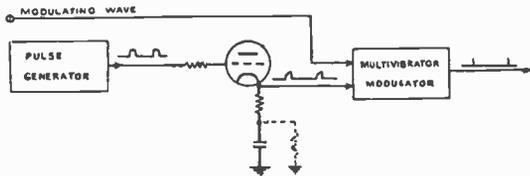


Fig. 1.—*Experimental Modulator.*

multivibrator produces square periodic pulses, having a width equal to the maximum modulation width required and repetitive at a frequency equal to the chopping frequency. These pulses are positive and they are applied on the grid of a valve. Each time a pulse appears the valve is saturated and a constant plate-cathode current flows. This current is sent through a resistor in series with a capacitor. During each pulse the capacitor charges; between two pulses the capacitor discharges. The discharge requires an auxiliary resistor which is not represented in Fig. 1. The pulses produced across the resistor in series with the capacitor have a trapezoidal shape, such as represented in Fig. 2a.

These pulses are applied to a second multivibrator in series with the modulating signals. In this multivibrator one grid is biased negatively far below the cut-off voltage so that normally no current flows in the valve. The multivibrator is then non-sensitive until a pulse appears on the biased grid. Then, if the pulse has a sufficient amplitude, the grid may cross the cut-off point, current will appear in the corresponding valve, and the multivibrator will strike.

For simplicity, we will assume that the multivibrator strikes at the moment when the grid voltage attains, or becomes greater, than a certain

voltage which will be called the "striking voltage." The variations of the grid voltage with respect to the striking voltage are represented in Fig. 2. Fig. 2a shows what happens in the absence of modulating signals and Fig. 2b shows what happens when a modulating signal is present. For simplicity, it has been assumed that the striking line is modulated and the trapezoidal pulses are fixed in position.

Assume the biasing voltage is adjusted so that the multivibrator strikes on the axis of symmetry of each pulse when there is no modulation (point O, Fig. 2a). Then if the modulating wave is applied (Fig. 2B), the striking point will move on the left or on the right of the axis within the boundaries defined by the leading and trailing edges of each pulse. The modulated multivibrator is designed conveniently in order to produce very sharp pulses whose amplitude and width are practically unaffected by the modulation.

The big advantage of obtaining definite boundaries for the modulation is that no cross-talk is possible between adjacent channels if the adjacent space is occupied by other channels. The modulation is linear as long as the striking point remains inside the boundaries.

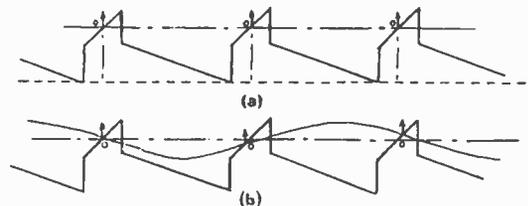


Fig. 2.—*Method of modulation.*

By using a series of trains of modulation pulses, the leading edges of one train coinciding with the trailing edge of the preceding one, a multi-channel system can be built up. In appearance, such a system will require a number of multivibrators equal, at least, to twice the number of channels. But when designing a multi-channel system many combinations are possible and the final number of multivibrators is small. For instance, in a 20-channel system, one master multivibrator and three special types of multivibrator are sufficient. To give details about these improvements would be to go beyond the scope of this memorandum.

3.1.2. *The modulation process introduces a non-linear distortion which is a function of the shape of the modulating signal.*

All experimental methods of time-phase modulation are derived from the same principle; either the striking voltage or the linearly increasing control voltage (such as the trapezoidal pulses) are modulated by the signal. Modulating the striking line or the control line is equivalent, but it is simpler for visual representation to assume that the striking line is modulated.

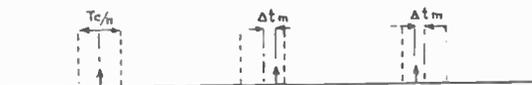


Fig. 3.—Modulated pulse train.

Assume the pulse appears normally at time O , T_c , $2T_c \dots mT_c \dots$ when there is no signal (Fig. 3). When a signal appears, the position of each pulse is moved. Let t_m be the shift in time for the m th pulse. When $t_m = 0$ the pulse appears at time mT_c . Referring to Fig. 4, it can be seen easily that Δt_m is proportional to the value of the signal at time $mT_c + \Delta t_m$. Let $S(t)$ be the signal, then :

$$\Delta t_m = K.S(mT_c + \Delta t_m).$$

This is a complicated equation which shows that Δt_m is not simply proportional to the amplitude of the signal at the mean chopping point $t = mT_c$. In Fig. 4, Δt_m is equal to OC . If Δt_m was proportional to the amplitude of the signal at point O , it would appear equal to OC^1 . The difference

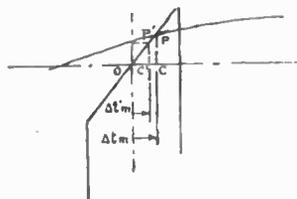


Fig. 4.—Showing how the non-linear distortion is produced in the modulation process.

CC^1 represents the distortion introduced by the process of modulation. It is clear from Fig. 4 that this distortion varies with the shape of the signal in the neighbourhood of point O . More precisely this distortion is a function of the value of the signal and its derivatives at point O .

This characteristic is typical of pulse modulation. In carrier amplitude modulation, the non-linear distortion is produced usually by the curvature of the valve characteristic, and the amount of distortion is a function of the derivatives of the valve characteristic at the working point and is independent of the shape of the signal.

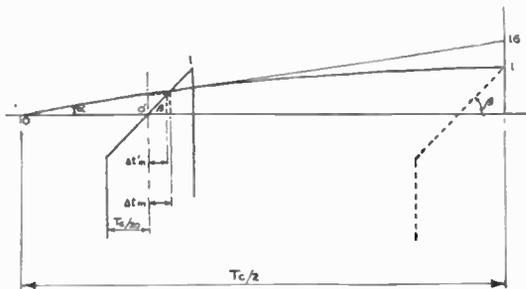


Fig. 5.—Showing how the non-linear distortion varies.

3.1.3. *In pulse modulation the harmonic distortion increases with the width of modulation and with the frequency of the signal.*

For simplicity, we will consider sinusoidal signals only. Figure 4 can then be replaced by Fig. 5.

To vary the maximum width of modulation one has either to vary the amplitude of the modulating signal or vary the slope of the control line. The results will be the same. Assume the slope is varied and assume for simplicity that the portion of signal near point O is replaced by its tangent. Then one gets Fig. 6 which shows by simple inspection that the distortion CC^1 increases with the width of modulation.

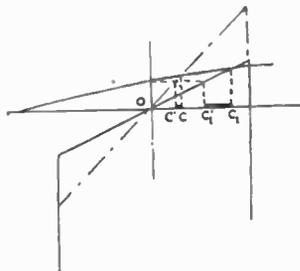


Fig. 6. — Showing enlargement of some details of Fig. 5.

To increase the frequency of the modulating signal is equivalent when drawing the wave and the

control line to decrease the slope of this line and this increases the harmonic distortion.

3.1.4. *Some numerical values.*

To get a numerical idea of the amount of distortion which can be produced in practice, let us make the calculation in a very unfavourable case, that means to say, in taking the highest frequency and the largest modulation width which are used in practical applications. If f_c is the chopping frequency, the highest frequency is usually smaller than $f_c/2$ and the maximum modulation width is usually smaller than $f_c/10$ for a very good signal/noise ratio. Taking $T_c/2$ for the frequency and $T_c/10$ for the trapezoid width, one gets Fig. 5. It can be deduced easily from this figure that

$$\tan \beta = \frac{20}{\pi/2} \cdot \tan \alpha = 13 \tan \alpha.$$

This in turn gives

$$\Delta r_m = \frac{00^1}{12}, \Delta r'_m = \frac{00^1}{13}$$

with the notations of Fig. 5.

Thus :

$$\frac{\Delta r_m}{\Delta r'_m} = \frac{13}{12} = 1.08$$

and CC^1 is equal to about 8 per cent. of Δr_m .

In practice the distortion is much smaller than 8 per cent.

Incidentally, it can be shown that CC^1 is equal to less than 0.7 per cent. of OC. This result will be used in the theoretical study.

3.2. **Demodulation Process**

3.2.1. *The process of demodulation and channel selection is simple in pulse technique.*

Fig. 7a shows a train of pulses produced by a multi-channel system. One synchronizing pulse appears at the beginning of each chopping period. Each period is divided in a certain number of equal parts, each part being the time allocated to one channel. The channel pulse will appear and will move within these limits.

To demodulate a given channel, it is necessary to produce a pulse which defines the boundaries of the channel. This is obtained by using a multivibrator synchronized by the synchronizing pulse. By means of a convenient delay network the position of the new pulse is adjusted in order to start at the beginning of the channel, and the width of

the pulse is adjusted to be equal to the channel width. The process of demodulation is then as follows : the new pulse is applied on the grid of one valve of a second multivibrator. The grid is biased negatively in order to make the multivibrator normally insensitive and conveniently in

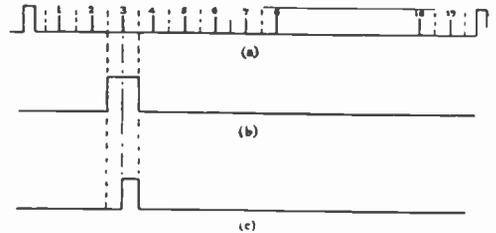


Fig. 7.—Explaining the demodulation process.

order to render the multivibrator sensitive again when the pulse appears. The multivibrator will remain sensitive until the pulse disappears, that is to say, within the boundaries of the selected channel. The square pulse is called "gating pulse" because it makes the multivibrator sensitive for the selected channel only. When the channel pulse appears, the multivibrator strikes. But, contrary to what happens in the modulation process, the multivibrator will produce a pulse having a larger width than the channel if left to itself, but it will stop striking when the gating pulse disappears. This process is shown in Fig. 7.

By this method, time phase modulation is transformed in width modulation. Fig. 8 shows a transmitted signal and the corresponding width modulated pulses. The width of each pulse is proportional to the amplitude of the transmitted signal for the same point.



Fig. 8.—Showing a train of variable width pulses and the corresponding signal.

The question now is : how to demodulate these pulses in order to obtain the transmitted signal ? In order to give a reply it is necessary to make a frequency analysis of the train of width-modulated pulses.

3.2.2. *Approximate analysis: to modulate the width of a periodic pulse train is equivalent in first approximation to modulate in amplitude the harmonic components of the train. Demodulation can be obtained by filtering the side-band component produced by modulation of the D.C. component.*

The spectrum analysis of the train of modulated pulses can be made by means of Fourier Integrals. Let $P(t)$ represent this train of pulses; then the spectrum $S(\omega)$ is given by

$$S(\omega) = \sqrt{A^2_{(\omega)} + B^2_{(\omega)}}$$

where

$$A_{(\omega)} = \int_{-\infty}^{+\infty} P(t) \cdot \sin \omega t \cdot dt.$$

$$B_{(\omega)} = \int_{-\infty}^{+\infty} P(t) \cdot \cos \omega t \cdot dt.$$

If we confine the analysis to the signal frequencies, that is to say to frequencies small compared to the chopping frequency, and if we assume the pulse width small, the value of these integrals will not change too much if each pulse is replaced by another one placed at the same point and having the same surface. Thus a variable width train can be replaced by a constant width-variable amplitude train of pulses. The spectrum analysis of this type is very easy. Assume first that the train is not modulated. Fourier's theorem gives the development:

$$P(t) = A_0 + \sum_m A_m \cdot \cos m \cdot \omega_e \cdot t. \\ + \sum_m B_m \cdot \sin m \cdot \omega_e \cdot t.$$

where $\omega_e = 2\pi f_e$ and f_e is the chopping frequency. If $M(t)$ is the modulating signal, multiplying left and right side of the above relation by $M(t)$ gives

$$P(t) \cdot M(t) = A_0 M(t) + \sum_m A_m \cdot M(t) \cos m \cdot \omega_e \cdot t. \\ + \sum_m B_m M(t) \sin m \cdot \omega_e \cdot t.$$

This relation shows that modulating the pulse train in amplitude is equivalent to modulating the D.C. component and each harmonic component by the same signal.

The D.C. component reproduces exactly the signal; each harmonic component is modulated in amplitude and can be considered as an amplitude modulated carrier. To demodulate the modulated train, the easiest way is to isolate the signal by a

simple low-pass filter. No improvement in signal/noise ratio can be obtained by demodulating each carrier and by adding the signals thus obtained since each carrier is modulated by the same noise.

3.2.3. *In the first approximation the only distortion produced by the demodulation process is the "cross-distortion" which appears for frequencies greater than half the chopping frequency. This distortion is the appearance, inside the signal frequency band, of the lower side band produced by the modulation of the chopping frequency.*

We can now give a reply to the question: what is the distortion introduced by the chopping principle? In the first approximation, there is no distortion as long as the chopping frequency is greater than twice the maximum signal frequency, since then the frequencies of all side bands and carriers are greater than the maximum signal frequency and can be eliminated by a simple low pass filter.

Experiments show this conclusion to be correct. Harmonic distortion is practically negligible, at least, when the modulation width is not too great. When this width of modulation is increased, a special type of distortion appears: when modulating with a pure tone and varying its frequency, a whistle appears in the background showing the presence of unwanted components. It seems as if each carrier is not modulated in amplitude but modulated in frequency or phase so that there are many sidebands, some in the frequency range isolated by the low pass filter. These unwanted components will be said to be produced by "cross-distortion" since they come from modulation of the adjacent carrier.

3.2.4. *A detailed analysis of the demodulation process shows a complex phenomenon.*

The above theory is only approximate and is invalid where small distortions are concerned. From a closer study of the process of demodulation (Fig. 7c) it will be observed that the width modulation is accompanied by time phase modulation (since the centre of each pulse is time modulated on account of the modulation of only one edge of the pulse). It should also be observed that in the process of demodulation, as in modulation, the width of a pulse is not simply proportional to the amplitude of the signal. This complex phenomenon will be studied mathematically in the following sections.

4. DETAILED STUDY

It has been shown in Section 3 that the modulation and demodulation processes each introduce a certain amount of distortion. It is proposed now to determine these amounts and the overall distortion produced by the pulse system.

4.1. Modulation Theory

4.1.1. Notations and hypothesis.

The main notations are represented in Fig. 4 and 5. Let f_c and T_c represent the chopping frequency and the chopping period. If the train of pulses is not modulated, and if the first one appears at time $t = 0$, then the following ones will appear at $t = T_c, 2T_c, (m-1)T_c, mT_c, (m+1) T_c, \dots$. When the train is modulated, the time at which each pulse appears is modulated. The m th pulse will appear at

$$t_m = mT_c + \Delta t_m$$

where t_m is the time shift produced by the modulating wave.

$|\Delta t_m|$ must be small in order to avoid considerable distortion. In general if

$$|\Delta t_m| < \frac{T_c}{20} \dots \dots \dots (1)$$

the distortions are less than about 10 per cent.

$|\Delta t_m|$ must also be smaller than half the channel width in the case of a multi-channel system. If n channels are used, the channel width is equal to T_c/n and the above conditions become

$$|\Delta t_m| \leq \frac{T_c}{2n} \dots \dots \dots (2)$$

This condition holds for (1) if we assume $n > 10$. In practice this condition is always satisfied.

Let the modulating wave be proportional to $\sin(\Omega t + \phi)$ where ϕ indicates the relative phase-shift between the modulating wave and the trapezoidal pulses at time $t = 0$. In Section 3 it has been shown that the frequency of this wave must be smaller than half the chopping frequency in order to avoid a great amount of distortion. For simplicity, we will assume this condition satisfied :

$$\Omega < \frac{\omega_c}{2} \quad (\omega_c = 2\pi f_c) \dots \dots \dots (3)$$

4.1.2. Modulation formula.

Let us now determine the relationship which exists between t_m and $\sin(\Omega t + \phi)$. This can be done with the help of Fig. 4 or Fig. 5. The

modulating wave meets the trapezoidal pulse at point P and the multivibrator producing the modulated pulse strikes at this point. Now OC is proportional to PC, and equal to Δt_m while PC is the amplitude of the modulating wave at time $mT_c + \Delta t_m$. Thus :

$$\Delta t_m \propto \sin \Omega (mT_c + \Delta t_m) + \phi.$$

The maximum amplitude of Δt_m being equal to

$\frac{T_c}{2n}$ we can write

$$\Delta t_m = \frac{T_c}{2n} \cdot \lambda \sin \Omega (mT_c + \Delta t_m) + \phi \dots (4)$$

where

$$0 < \lambda < 1.$$

λ is a quantity proportional to the amplitude of the modulating signal and the last condition ($0 < \lambda < 1$) means that we assume the amplitude of the signal to be smaller than the value corresponding to the maximum width of modulation permissible.

$\frac{T_c}{2n} \cdot \lambda$ represents the maximum amplitude of Δt_m

for a given modulating signal. Put $\frac{T_c}{2n} \cdot \lambda = \Delta \tau$. then

$$\Delta t_m = \Delta \tau \sin \Omega (mT_c + \Delta t_m) + \phi \dots (5)$$

4.1.3. Formula in first and second approximation.

Referring again to Fig. 5, we have seen already, in Section 3.1.4, that CC^1 is equal to 8 per cent. of

OC when $\Delta t_m = \frac{T_c}{20}$ and $\Omega = \frac{\omega_c}{2}$ and decreases

with respect to OC with Δt_m or Ω . Thus, in first approximation, we can replace PC by P^1C^1 and with

$$\Delta t_m = \Delta \tau \sin \Omega . m . T_c + \phi. \dots \dots \dots (6)$$

To get a more accurate formula, we will replace (6) in the right hand side of (5) :

$$\Delta t_m = \Delta \tau \sin \Omega [mT_c + \Delta \tau \sin (\Omega mT_c + \phi)] + \phi \dots \dots \dots (7)$$

What is the maximum error obtained in replacing the correct formula (5) by (7) ?

It has been shown already in Section 3.1.4 that C^1C is always smaller than 0.7 per cent. of OC . Thus $\sin OC$ is equal to $\sin OC^1$ with an error smaller than 0.7 per cent. and formula (7) is correct with the same approximation.

From a figure similar to Fig. 5, except that $T_c/2$

is replaced by $T/4$, one can establish the following relations :

$$\Delta t_m = OC. \frac{T_c \cdot \pi}{T \cdot n} = 2\pi \cdot F \cdot \Delta \tau.$$

$$CC^1 = \Delta t_m \cdot \frac{T_c \cdot \pi}{T \cdot n} = 4\pi \cdot OC \cdot (F \cdot \Delta \tau)^2.$$

These relations show that :

1. The error obtained in using formula (6) decreases proportionally with the decrease of the signal frequency or the modulation width.
2. The error obtained in using formula (7) decreases proportionally with the square of the signal frequency or the modulation width.

- 4.1.4. Expansion of formula (7); the second harmonic produced by the process of modulation is proportional to the product of the modulation width and the frequency of the modulating wave, and the third harmonic to the square of this product.

Let us expand formula (7) :

$$\begin{aligned} \Delta t_m &= \Delta \tau \sin \Omega [m \cdot T_c + \Delta \tau \cdot \sin (\Omega \cdot m \cdot T_c + \varphi)] + \varphi \\ &= \Delta \tau [\sin (\Omega \cdot m \cdot T_c + \varphi) \cos (\Omega \Delta \tau \cdot \sin (\Omega m T_c + \varphi)) \\ &\quad + \cos (\Omega \cdot m \cdot T_c + \varphi) \sin (\Omega \cdot \Delta \tau \sin (\Omega \cdot m \cdot T_c + \varphi))] \end{aligned} \dots \dots \dots (8)$$

Expanding again by means of Bessel relations, we get

$$\Delta t_m = \Delta \tau \left\{ \begin{aligned} &[J_0(x_0) - J_2(x_0)] \sin(\Omega m T_c + \varphi) \\ &+ [J_1(x_0) + J_3(x_0)] \sin 2(\Omega m T_c + \varphi) \\ &+ [J_2(x_0) - J_4(x_0)] \sin 3(\Omega m T_c + \varphi) \\ &+ [J_3(x_0) + J_5(x_0)] \sin 4(\Omega m T_c + \varphi) \\ &+ \dots \dots \dots \end{aligned} \right\} \dots \dots \dots (9)$$

In these equations $J_u(x_0)$ means the Bessel function of the first kind and u order for the argument $x_0 = \Omega \cdot \Delta \tau$.

This equation shows that a time-phase modulation is not linear, the harmonic distortion being a function of the product $X = \Delta \tau \Omega$, that is to say, of the depth of modulation and the frequency of the signal.

It is important to have an idea of the amount of distortion. According to condition (2), $\Delta \tau < \frac{T_c}{20}$;

and according to condition (3), $\Omega < \frac{\omega_c}{2}$; thus

$$X_0 = \Delta \tau \cdot \Omega < \frac{T_c}{20} \cdot \frac{\omega_c}{2} = 0.16.$$

For these small values of x_0 we can replace $J_0(x_0)$ by 1, $J_1(x_0)$ by $x_0/2$ and $J_2(x_0)$ by $0.1x_0$ and $J_3(x_0)$ by zero (curves of Bessel functions are represented in Fig. 9).

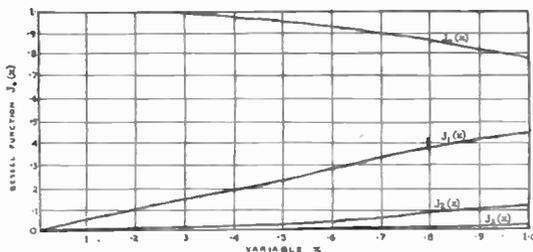


Fig. 9.—Bessel functions of the first kind and order 0, 1, 2 and 3 for the argument x_0 ($x_0 = \Omega \Delta \tau$ and $x_0 < 1$).

Formula (9) becomes

$$\begin{aligned} \Delta t_m &= \Delta \tau \left[\sin(\Omega \cdot m \cdot T_c + \varphi) \right. \\ &\quad + \frac{\Delta \tau \cdot \Omega}{2} \sin 2(\Omega m T_c + \varphi) \\ &\quad \left. + \frac{(\Delta \tau \Omega)^2}{8} \sin 3(\Omega m T_c + \varphi) \right] \dots \dots \dots (9^1) \end{aligned}$$

For $x_0 = 0.16$ this gives 8 per cent. second harmonic and about 0.5 per cent. third harmonic.

Taking $n = 20$, $f_c = 10$ kc/s and $\Omega = 1.5$ kc/s, we find about 1.3 per cent. second harmonic distortion and no third harmonic.

Fig. 10A (p. 74) shows how the harmonic distortion varies with the modulation width for a frequency of 1.5 kc/s. Fig. 10B shows the same type of curve but for a constant modulation width of $\pm 5 \mu$ secs.

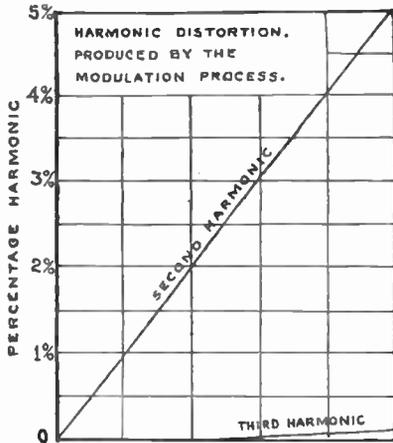
4.2. Demodulation Theory

The demodulation process consists of converting the constant width time-phased modulated pulses to variable width pulses, and detecting the signal from these pulses by means of a convenient low-pass filter.

The process of demodulation introduces three main types of distortion :—linear distortion, harmonic distortion and cross distortion. All of these types of distortion can be evaluated from the spectrum of the train of variable-width pulses. To obtain this spectrum, we will calculate the spectrum of each pulse with the help of Fourier's integrals, and add all the spectrums together.

4.2.1. Equation of one pulse.

Consider a rectangular pulse $P_1(t)$ having an amplitude A_0 for $-\tau < t < +\tau$ and zero for all other values of t (Fig. 11).



A:	0	0.02	0.04	0.06	0.08	0.10	(X_0)
B:	0	2	4	6	8	10	(τ)
C:	0	0.3	0.6	0.9	1.2	1.5	$(\Omega/2\pi)$

- A: VARIABLE $X_0 = \Delta\tau \cdot \Omega$
- B: WIDTH OF MODULATION $\Delta\tau$ IN MICROSECONDS (FOR A FREQUENCY $\Omega/2\pi = 1.5$ K.C.S)
- C: SIGNAL FREQUENCY $\Omega/2\pi$ IN K.C.S FOR A WIDTH OF MODULATION OF $\pm 5 \mu$ SECS.

Fig. 10.—Percentage of second and third harmonic produced in the modulation process for a signal frequency of 1.5 kc/s or a modulation width of ± 5 microseconds.

The spectrum of this pulse can be obtained from Fourier's integrals :

$$P_1(t) = \frac{1}{\pi} \int_0^{\infty} S_1(\omega) \cos(\omega t + \phi(\omega)) d\omega \quad (10)$$

$$A_1(\omega) = \int_{-\infty}^{+\infty} P_1(t) \sin \omega t dt. \dots\dots\dots(11)$$

$$B_1(\omega) = \int_{-\infty}^{+\infty} P_1(t) \cos \omega t dt. \dots\dots\dots(12)$$

$$S_1(\omega) = \sqrt{A_1^2(\omega) + B_1^2(\omega)} \dots\dots\dots(13)$$

For the pulse $P_1(t)$, $A_1(\omega) = 0$ and :

$$S_1(\omega) = B_1(\omega) = A_0 \int_{-\tau}^{+\tau} \cos \omega t dt$$

$$= 2\tau \cdot A_0 \cdot \frac{\sin \omega\tau}{\omega\tau} \dots\dots(14)$$

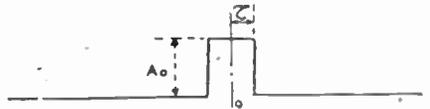


Fig. 11.—Showing the notations for one pulse.

The curve $S(\omega)$ as a function of ω is represented in Fig. 12. The equation of the pulse is, according to (10) and (14) :

$$P_1(t) = \frac{2\tau A_0}{\pi} \int_0^{\infty} \frac{\sin \omega\tau}{\omega\tau} \cdot \cos \omega t. d\omega. (15)$$

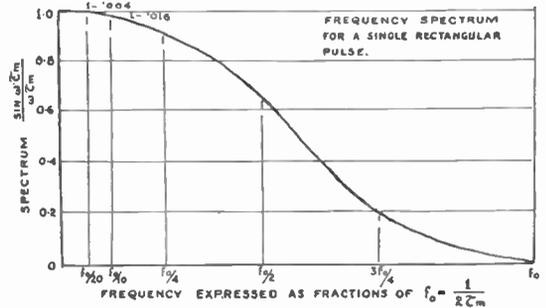


Fig. 12.—Curve $\frac{\sin \omega \cdot \tau m}{\omega \cdot \tau m}$.

4.2.2. Equation of a train of unmodulated pulses.

Before tackling the problem of a modulated pulse train, let us apply first the method to a train of unmodulated pulses.

Consider the periodic train represented in Fig. 13 (p. 75). With the notations of this figure, the equation of the m th pulse is

$$P_m(t) = \frac{2\tau A_0}{\pi} \int_0^{\infty} \frac{\sin \omega\tau}{\omega\tau} \cdot \cos \omega(t - mT_0). d\omega \quad (16)$$

Let $P(t)$ represent the whole train of pulses ; then

$$P(t) = \sum_{m=0}^{\infty} P_m(t) \\ = \frac{2 \cdot \tau \cdot A_0}{\pi} \int_0^{\infty} \frac{\sin \omega \tau}{\omega \tau} d\omega \cdot \sum_{m=0}^{\infty} \cos \omega(t - m \cdot T_c) \dots (17)$$

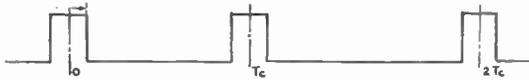


Fig. 13.—Notation for a train of unmodulated pulses.

Because the train is periodic, the sum under the integral sign is zero for all values of ω which are not multiples of ω_c :

$$\text{ie. } \sum_{m=0}^{\infty} \cos \omega(t - m \cdot T_c) = 0 \text{ for } \omega \neq N \cdot \omega_c \dots (18)$$

where N is a positive integer.

For values multiple of ω_c the value of the sum can be determined by direct calculation. However, it is quicker to get it indirectly by evaluating the same sum by means of the Fourier series formula.

Fourier series applied to this train gives for the amplitude of the n th harmonic :

$$\frac{2 \cdot A_0 \cdot \omega_c \cdot \tau}{\pi} \cdot \frac{\sin N \cdot \omega_c \cdot \tau}{N \cdot \omega_c \cdot \tau} \dots (19)$$

The amplitude of the same component can also be obtained by replacing ω by $N\omega_c$ in (17) ; equating the values obtained from (17) and (19), we get

$$\text{Limit } d\omega \sum_{m=0}^{\infty} \cos N\omega_c(t - mT_c) \\ = \omega_c \cos N \cdot \omega_c \cdot t \dots (20)$$

Formula (18) and (20) are very important and will be used in the next section.

4.2.3. Equation of a train of modulated pulses.

4.2.3.1. General Formula.

We will assume that the train of modulated pulses used in the process of demodulation are sinusoidally modulated

$$\Delta t_m = \frac{T_c}{2n} \cdot \lambda \cdot \sin(\Omega m T_c + \phi) \dots (5)$$

This formula is the same as used in Section 4.1.3. It represents the result obtained in first approximation when using a sinusoidal signal. In Section 5

we will start with a sinusoidal signal and use formula (9) in order to take account of the distortion introduced in the process of modulation.

The train of variable width pulses obtained in the process of demodulation is represented in Fig. 14A and details concerning the m th pulse are given in Fig. 14B.

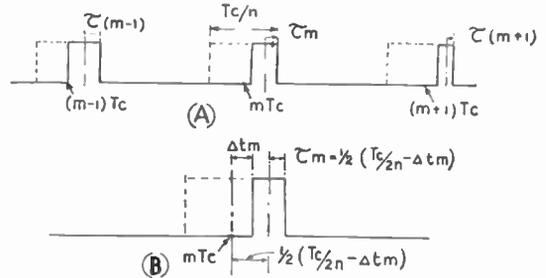


Fig. 14.—Notations for a train of modulated pulses.

Using the notations of Fig. 14, formula (15), giving the equation of one pulse, becomes, for the m th pulse

$$P_m(t) = \frac{2A_0}{\pi} \int_0^{\infty} \tau_m \cdot \frac{\sin \omega \tau_m}{\omega \tau_m} \\ \cos \omega \left[t - mT_c - \frac{T_c}{4n} - \frac{\Delta t_m}{2} \right] d\omega \dots (21)$$

This formula is correct for all values of ω . However, we are interested only in the frequencies contained in the signal transmitted, that is to say, in frequencies which according to condition (3) are smaller than $f_c/2$. Now according to condition (2) f_c is smaller than $1/20 \cdot \Delta t_m = 1/10 \cdot f_0$, f_0 being equal to $1/2 \cdot \Delta t_m$. Thus, the highest frequency in the signal band is lower than $1/20 \cdot f_0$. Referring to Fig. 12, it can be seen that the curve is flat for low frequencies and that for a frequency of $1/20 \cdot f_0$ the drop in amplitude is 0.4 per cent.

The calculation can be done more simply with numerical values. For $\tau_m = 5$ microseconds and a frequency of 3 kc/s, the drop of amplitude is 0.7 per cent.

These values being very small, it seems that we can neglect them and replace $\frac{\sin \omega \tau_m}{\omega \tau_m}$ by 1. This is, of course, correct if we consider only one pulse. But, if we add a great number of variable width pulses, the value of the above expression varies

from one to the other and care should be taken when saying that a sum is zero that the approximation has not altered the result. Making this approximation and writing t^1 for $(t - T_c/4n)$ formula (21) becomes

$$P_m(t) = \frac{2A_0}{\pi} \int_0^{\infty} \tau_m \cos \omega \left(t^1 - mT_c - \frac{\Delta t_m}{2} \right) d\omega. \quad (22)$$

τ_m is given by

$$\tau_m = \frac{1}{2} \left(\frac{T_c}{2n} - \Delta t_m \right) = \frac{T_c}{4n} \left[1 - \lambda \sin(\Omega m T_c + \phi) \right] \quad (23)$$

Replacing in (22) we get

$$P_m(t^1) = \frac{A_0}{n\omega_c} \int_0^{\infty} \left[1 - \lambda \sin(\Omega m T_c + \phi) \right] \cos \omega \left(t^1 - mT_c - \frac{\Delta t_m}{2} \right) d\omega \quad (24)$$

We can now write the equation of the train of pulses :

$$P(t^1) = \sum_{m=0}^{\infty} P_m(t) = \frac{A_0}{n\omega_c} \int_0^{\infty} d\omega \sum_{m=0}^{\infty} \left[1 - \lambda \sin(\Omega m T_c + \phi) \right] \cos \omega \left(t^1 - mT_c - \frac{\Delta t_m}{2} \right) \quad (25)$$

In order to go further, we need to expand the expression under the \sum sign.

It can be shown by means of Bessel relations that this expansion is :

$$\begin{aligned} & \left[1 - \lambda \sin(\Omega m T_c + \phi) \right] \cos \omega \left(t^1 - mT_c - \frac{\Delta t_m}{2} \right) = \\ & \left. \begin{aligned} & I_0 \left(\frac{x}{2} \right) \left[\cos \omega + \frac{1}{2} \lambda [\sin(\omega - \Omega - \phi) - \sin(\omega + \Omega + \phi)] \right] \\ & + I_1 \left(\frac{x}{2} \right) \left[\begin{aligned} & -\lambda \sin \omega + \\ & \cos(\omega - \Omega - \phi) - \cos(\omega + \Omega + \phi) + \frac{1}{2} \lambda [\sin(\omega - 2\Omega - 2\phi) + \sin(\omega + 2\Omega + 2\phi)] \end{aligned} \right] \\ & + I_2 \left(\frac{x}{2} \right) \left[\begin{aligned} & -\frac{1}{2} \lambda [\sin(\omega - \Omega - \phi) - \sin(\omega + \Omega + \phi)] + \cos(\omega - 2\Omega - 2\phi) + \cos(\omega + 2\Omega + 2\phi) \\ & + \frac{1}{2} \lambda [\sin(\omega - 3\Omega - 3\phi) - \sin(\omega + 3\Omega + 3\phi)] \end{aligned} \right] \\ & + I_3 \left(\frac{x}{2} \right) \left[\begin{aligned} & -\frac{1}{2} \lambda [\sin(\omega - 2\Omega - 2\phi) + \sin(\omega + 2\Omega + 2\phi)] + \cos(\omega - 3\Omega - 3\phi) - \cos(\omega + 3\Omega + 3\phi) \\ & + \frac{1}{2} \lambda [\sin(\omega - 4\Omega - 4\phi) + \sin(\omega + 4\Omega + 4\phi)] \end{aligned} \right] \\ & + I_4 \left(\frac{x}{2} \right) \left[\begin{aligned} & -\frac{1}{2} \lambda [\sin(\omega - 3\Omega - 3\phi) - \sin(\omega + 3\Omega - 3\phi)] + \cos(\omega - 4\Omega - 4\phi) + \cos(\omega + 4\Omega + 4\phi) \\ & + \frac{1}{2} \lambda [\sin(\omega - 5\Omega - 5\phi) - \sin(\omega + 5\Omega + 5\phi)] \end{aligned} \right] \\ & + \dots \end{aligned} \right\} \quad (26) \\ & \left. \begin{aligned} & I_0 \left(\frac{x}{2} \right) \left[\sin \omega - \frac{1}{2} \lambda [\cos(\omega - \Omega - \phi) - \cos(\omega + \Omega + \phi)] \right] \\ & + I_1 \left(\frac{x}{2} \right) \left[\begin{aligned} & -\lambda \cos \omega - \sin(\omega - \Omega - \phi) + \sin(\omega + \Omega + \phi) \\ & + \frac{1}{2} \lambda [\cos(\omega - 2\Omega - 2\phi) + \cos(\omega + 2\Omega + 2\phi)] \end{aligned} \right] \\ & + I_2 \left(\frac{x}{2} \right) \left[\begin{aligned} & -\frac{1}{2} \lambda [\cos(\omega - \Omega - \phi) - \cos(\omega + \Omega + \phi)] + \sin(\omega - 2\Omega - 2\phi) + \sin(\omega + 2\Omega + 2\phi) \\ & + \frac{1}{2} \lambda [\cos(\omega - 3\Omega - 3\phi) - \cos(\omega + 3\Omega + 3\phi)] \end{aligned} \right] \\ & + I_3 \left(\frac{x}{2} \right) \left[\begin{aligned} & -\frac{1}{2} \lambda [\cos(\omega - 2\Omega - 2\phi) + \cos(\omega + 2\Omega + 2\phi)] - \sin(\omega - 3\Omega - 3\phi) + \sin(\omega + 3\Omega + 3\phi) \\ & + \frac{1}{2} \lambda [\cos(\omega - 4\Omega - 4\phi) + \cos(\omega + 4\Omega + 4\phi)] \end{aligned} \right] \\ & + I_4 \left(\frac{x}{2} \right) \left[\begin{aligned} & -\frac{1}{2} \lambda [\cos(\omega - 3\Omega - 3\phi) - \cos(\omega + 3\Omega + 3\phi)] + \sin(\omega - 4\Omega - 4\phi) + \sin(\omega + 4\Omega + 4\phi) \\ & + \frac{1}{2} \lambda [\cos(\omega - 5\Omega - 5\phi) - \cos(\omega + 5\Omega + 5\phi)] \end{aligned} \right] \\ & + \dots \end{aligned} \right\} \\ & + \sin \omega t^1 \end{aligned}$$

where $x = \Delta t \omega$ and where mT_e has been omitted temporarily for simplicity.

Referring again to equation (25), it can be seen that we have to make the summation of formula (26) for all positive integer values of m . In making this summation we will find everywhere expressions which can be derived from formula (18) and (20)

by replacing t by $t - \frac{\phi}{\omega}$;

$$\text{Limit } d\omega \sum_{m=0}^{\infty} \cos \omega(t - mT_e) + \phi = 0 \text{ for } \omega \neq N \cdot \omega_e \dots\dots\dots(18)$$

$$\text{Limit } d\omega \sum \cos N\omega_e(t - mT_e) + \phi = \omega_e \cos N\omega_e t + \phi \dots\dots\dots(20) \text{ where } N \text{ is positive integer or zero.}$$

By replacing successively t by 0 and $\frac{\pi}{2\omega}$ in each of these formulæ we get the important relations :

$$d\omega \sum_{m=0}^{\infty} \cos \omega mT_e + \phi = d\omega \sum_{m=0}^{\infty} \sin(\omega mT_e + \phi) = 0 \text{ for } \omega \neq N\omega_e \dots\dots\dots(27)$$

$$d\omega \sum_{m=0}^{\infty} \cos(N\omega_e mT_e + \phi) = \omega_e \cos \phi$$

$$d\omega \sum_{m=0}^{\infty} \sin(N\omega_e mT_e + \phi) = \omega_e \cdot \sin \phi. \dots\dots\dots(28) \text{ and } (29)$$

We are now in a position to find the components of the frequency spectrum of $P(t)$. If in equation (25) we replace the expression under the summation sign by its value given by (26), we will have a large number of expressions identical to (27), (28) and (29). Only expressions identical to (28) and (29) will remain since (27) is zero. Thus the spectrum will be composed of components the frequencies of which are the value of ω contained in the solution.

4.2.3.2. Study of an important approximation.

In formula (25) we have assumed that

$$\frac{\sin \omega \tau_m}{\omega \tau_m} = 1.$$

In equation (26) when making the summations, we will get expressions identical to (27), (28), and (29). It can be shown that the above approximation

does not modify the value of these sums. For simplicity we will not reproduce the demonstration here.

4.2.3.3. Solutions of Equation (25).

Solutions will be found for one of the following values of ω :

$$\omega \pm N_1 \Omega = N_2 \omega_e.$$

where N_1 and N_2 are positive integers or zero.

First type of solution : $\omega = N_1 \Omega$

This type of solution will give the signal component ($\omega = \Omega$) and the harmonic components produced by distortion in the process of demodulation.

1. $\omega = \Omega.$

This gives the signal frequency. If Ω is not a sub-multiple of ω_e , its amplitude is proportional to

$$-\frac{1}{2} \lambda \left[I_0\left(\frac{x_0}{2}\right) - I_2\left(\frac{x_0}{2}\right) \right] \sin(\Omega t^1 + \phi) - I_1\left(\frac{x_0}{2}\right) \cos(\Omega t^1 + \phi)$$

If Ω is a submultiple of ω_e , we have to add to the preceding expression one of the following :

for $\omega_e = 2\Omega$:

$$\frac{1}{2} \lambda \left[I_0\left(\frac{x_0}{2}\right) - I_2\left(\frac{x_0}{2}\right) \right] \sin(\Omega t^1 - \phi) + I_1\left(\frac{x_0}{2}\right) \cos(\Omega t^1 - \phi) + 2 I_3\left(\frac{x_0}{2}\right) \sin 3\phi \sin \omega t^1 - \lambda I_2\left(\frac{x_0}{2}\right) \sin 3\phi \cos \omega t^1.$$

and for $\omega_e = 3\Omega.$

$$\frac{1}{2} \lambda I_1\left(\frac{x_0}{2}\right) \cdot \sin(\Omega t^1 - 2\phi) + I_2\left(\frac{x_0}{2}\right) \cos(\Omega t^1 - 2\phi)$$

In the last expression all terms containing Bessel coefficients of the order greater than the second have been neglected.

2. $\omega = 2\Omega.$

This gives the second harmonic component. If Ω is not a submultiple of ω_e , the amplitude of this component is proportional to

$$\frac{1}{2} \lambda \left[I_1(x_0) - I_3(x_0) \right] \sin 2(\Omega t^1 + \phi) + I_2(x_0) \cos 2(\Omega t^1 + \phi).$$

If Ω is a submultiple of ω_e , we have to add to

the above expression one of the following ones :

$$\frac{1}{2}\lambda [I_{1(x_c)} - I_{3(x_c)}] \sin 2(\Omega t^1 - \phi) + I_{2(x_c)} \cos 2(\Omega t^1 - \phi)$$

if $\omega_c = 4$. This expression is the same as the one above for $\phi = 0$ and equal in amplitude but opposite in sign if $\phi = (2k + 1)\pi/4$; and other expressions for $\omega_c = 5\Omega$ and $\omega_c = 6\Omega$.

$3_0 \omega = 3\Omega$.

This gives the third harmonic component :

$$-\frac{1}{2}\lambda \left[I_2\left(\frac{3x_c}{2}\right) - I_4\left(\frac{3x_c}{2}\right) \right] \sin 3(\Omega t^1 + \phi) - I_3\left(\frac{3x_c}{2}\right) \cos 3(\Omega t^1 + \phi).$$

Corrections have to be added if Ω is a submultiple of ω_c :

$4_0 \omega = N\Omega$, where $N > 3$.

All these harmonic components have a negligible amplitude.

Second type of solution : $\omega = \omega_c \pm N\Omega$.

This gives the modulation of the carrier frequency, f_c .

$1_0 \omega = \omega_c - \Omega$:

$$\frac{1}{2}\lambda \left[I_0\left(\frac{x_c - x_c}{2}\right) - I_2\left(\frac{x_c - x_c}{2}\right) \right] \sin [(\omega_c - \Omega)t^1 - \phi] + I_1\left(\frac{x_c - x_c}{2}\right) \cos [(\omega_c - \Omega)t^1 - \phi]$$

where $x_c = \omega_c \Delta t$.

$2_0 \omega = \omega_c - \Omega$:

$$-\frac{1}{2}\lambda \left[I_0\left(\frac{x_c + x_c}{2}\right) - I_2\left(\frac{x_c + x_c}{2}\right) \right] \sin [(\omega_c + \Omega)t^1 + \phi] - I_1\left(\frac{x_c + x_c}{2}\right) \cos [(\omega_c + \Omega)t^1 + \phi].$$

This gives almost the same amplitude as the lower side band but is reversed in phase.

$3_0 \omega = \omega_c - 2\Omega$:

$$\frac{1}{2}\lambda \left[I_1\left(\frac{x_c - 2x_c}{2}\right) - I_3\left(\frac{x_c - 2x_c}{2}\right) \right] \sin [(\omega_c - 2\Omega)t^1 - 2\phi] + I_2\left(\frac{x_c - 2x_c}{2}\right) \cos [(\omega_c - 2\Omega)t^1 - 2\phi]$$

$4_0 \omega = \omega_c + 2\Omega$:

$$\frac{1}{2}\lambda \left[I_1\left(\frac{x_c + 2x_c}{2}\right) - I_3\left(\frac{x_c + 2x_c}{2}\right) \right] \sin [(\omega_c + 2\Omega)t^1 + 2\phi] + I_2\left(\frac{x_c + 2x_c}{2}\right) \cos [(\omega_c + 2\Omega)t^1 + 2\phi]$$

The phase here is not reversed.

$5_0 \omega = \omega_c - 3\Omega$:

$$\frac{1}{2}\lambda \left[I_2\left(\frac{x_c - 3x_c}{2}\right) - I_4\left(\frac{x_c - 3x_c}{2}\right) \right] \sin [(\omega_c - 3\Omega)t^1 - 3\phi] + I_3\left(\frac{x_c - 3x_c}{2}\right) \cos [(\omega_c - 3\Omega)t^1 - 3\phi]$$

$6_0 \omega = \omega_c + 3\Omega$:

$$-\frac{1}{2}\lambda \left[I_2\left(\frac{x_c + 3x_c}{2}\right) - I_4\left(\frac{x_c + 3x_c}{2}\right) \right] \sin [(\omega_c + 3\Omega)t^1 + 3\phi] - I_3\left(\frac{x_c + 3x_c}{2}\right) \cos [(\omega_c + 3\Omega)t^1 + 3\phi]$$

This gives almost the same amplitude as the lower side-band but is reversed in phase.

Other types of solution : $\omega = N_1\omega_c \pm N_2\Omega$.

These will give similar results as in the case above.

4.2.3.4. *Summary of solutions.*

If Δtm is proportional to $\lambda \sin \Omega m T_c + \phi$, the side band components are given by the following equation :

(a) *D.C. Carrier.*

$$-\frac{1}{2}\lambda \left[I_0\left(\frac{x_c}{2}\right) - I_2\left(\frac{x_c}{2}\right) \right] \sin (\Omega t^1 + \phi) - I_1\left(\frac{x_c}{2}\right) \cos (\Omega t^1 + \phi) + \frac{1}{2}\lambda [I_{1(x_c)} - I_{3(x_c)}] \sin 2(\Omega t^1 + \phi) + I_{2(x_c)} \cos 2(\Omega t^1 + \phi) - \frac{1}{2}\lambda \left[I_2\left(\frac{3x_c}{2}\right) - I_4\left(\frac{3x_c}{2}\right) \right] \sin 3(\Omega t^1 + \phi) - I_3\left(\frac{3x_c}{2}\right) \cos 3(\Omega t^1 + \phi) + \frac{1}{2}\lambda I_{3(2x_c)} \sin 4(\Omega t^1 + \phi) + I_{4(2x_c)} \cos 4(\Omega t^1 + \phi) \dots (30)$$

where all Bessel coefficients of order higher than the fourth have been neglected.

(b) *Carrier of frequency f_c .*

The left-hand side band components have the following amplitudes :

$$\begin{aligned} & \frac{1}{2} \lambda \left[I_0\left(\frac{x_c - x_0}{2}\right) - I_2\left(\frac{x_c - x_0}{2}\right) \right] \cdot \\ & \quad \sin [(\omega_c - \Omega) t^1 - \phi] \\ & \quad + I_1\left(\frac{x_c - x_0}{2}\right) \cos [(\omega_c - \Omega) t^1 - \phi] \\ & + \frac{1}{2} \lambda \left[I_1\left(\frac{x_c - 2x_0}{2}\right) - I_3\left(\frac{x_c - 2x_0}{2}\right) \right] \\ & \quad \sin [(\omega_c - 2\Omega) t^1 - 2\phi] \\ & \quad + I_2\left(\frac{x_c - 2x_0}{2}\right) \cos [(\omega_c - 2\Omega) t^1 - 2\phi] \\ & + \frac{1}{2} \lambda \left[I_2\left(\frac{x_c - 3x_0}{2}\right) - I_4\left(\frac{x_c - 3x_0}{2}\right) \right] \\ & \quad \sin [(\omega_c - 3\Omega) t^1 - 3\phi] \\ & \quad + I_3\left(\frac{x_c - 3x_0}{2}\right) \cos [(\omega_c - 3\Omega) t^1 - 3\phi] \\ & + \frac{1}{2} \lambda \left[I_3\left(\frac{x_c - 4x_0}{2}\right) \sin [(\omega_c - 4\Omega) t^1 - 4\phi] \right. \\ & \quad \left. + I_4\left(\frac{x_c - 4x_0}{2}\right) \cos [(\omega_c - 4\Omega) t^1 - 4\phi] \right]. \end{aligned} \quad (31)$$

The right-hand side band components have the same expression as for the D.C. carrier side band components except for Ω which must be replaced by $\omega_c + \Omega$.

4.2.3.5. *Degeneration of the modulation for the carriers of high order.*

The theory is not very accurate when applied to the carriers of high order. However, it can be seen that, when applying (30) and (31) to a carrier of frequency f_c , the amplitude of the side bands are functions of the Bessel functions for one of the arguments ($x_c \pm Nx_0$). More generally for a carrier of order K , the amplitude of the side bands are functions of the Bessel functions for one of the arguments ($Kx_c \pm Nx_0$). The Bessel functions vary a lot with the value of the argument so that the amplitude distribution for the side bands varies.

4.2.3.7. *Practical Distortion formula.*

In practical applications, for speech transmission, Ω is usually smaller than 3 kc/s and $\Delta\tau$ smaller than 10 microseconds. This gives for $\frac{\Delta\tau\Omega}{2}$ a value smaller than 0.1. Referring to Fig. 9, it can

be seen that $I_3(x)$ has a negligible value for $x < 0.5$. $I_1(x)$ can be replaced by $\frac{x}{2}$ and $I_2(x)$ by $\frac{x^2}{8}$. Replacing by these values in (30) gives : when x_0 is replaced by $\lambda \frac{\pi}{n} \frac{T_c}{T}$ in some expressions :

$$-\frac{1}{2} \lambda \left[\begin{aligned} & \left(1 - \frac{3}{16} x_0^2 \right) \sin (\Omega t^1 + \phi) \\ & + \frac{\pi}{4n} \cdot \frac{T_c}{T} \cos (\Omega t^1 + \phi) \\ & - \frac{x_0}{2} \sin 2(\Omega t^1 + \phi) \\ & - \frac{1}{8} \left(\frac{\pi}{n} \right)^2 \left(\frac{T_c}{T} \right) \cos 2(\Omega t^1 + \phi) \\ & + \frac{9x_0^2}{32} \sin 3(\Omega t^1 + \phi) \end{aligned} \right] \quad (32)$$

For values of x_0 smaller than 0.2 or 0.1 we can make a new approximation and replace (32) by :

$$-\frac{1}{2} \lambda \left[\begin{aligned} & \sin (\Omega t^1 + \phi) - \frac{x_0}{2} \sin 2(\Omega t^1 + \phi) \\ & + \frac{9}{32} x_0^2 \sin 3(\Omega t^1 + \phi) \end{aligned} \right] \quad (33)$$

Equations (32) and (33) give all the distortion properties of the demodulation process.

4.2.3.8. *Amplitude distortion.*

Equation (32) shows that the amplitude distortion is very small, nearly proportional to x_0^2 . If $x_0 < 0.2$, the amplitude distortion is smaller than 1 per cent.

(In practical applications the amplitude distortion produced by the demodulation process is negligible.)

4.2.3.9. *Amplitude discontinuities at frequencies submultiple of ω_c .*

However, for values of Ω submultiple of ω_c , we must take account of the corrections given in Section 4.2.3.3. These corrections are a function of the relative phase between the train of pulses and the modulating signal. There is a sort of discontinuity in the amplitude response at these frequencies.

For $\Omega = f_c/2$ the amplitude can vary from 0 to 200 per cent. of the normal value. For $\Omega = f_c/3$, the variation is equal to $\pm x_0/4$ from the normal value and for $\Omega = f_c/4$ the variation is equal to

$\pm (x_0/4)^2$. For $x_0 = 0.2$, this gives a variation of ± 5 per cent. at $f_c/3$ and less than 0.25 per cent. at $f_c/4$. Discontinuities at other submultiples are negligible.

In practical applications the only important discontinuities in the amplitude response appear at $f_c/2$ and $f_c/3$ and are equal respectively to ± 100 per cent. and ± 5 per cent. of the normal amplitude. These results are represented on Fig. 15.

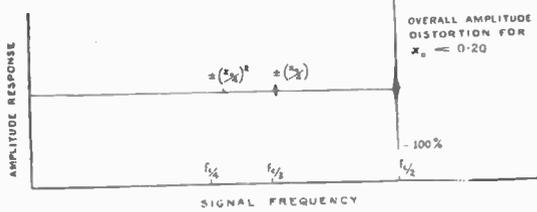


Fig. 15.—Amplitude distortion in the demodulation process.

4.2.3.10 Harmonic distortions.

Equation (33) is very similar to equation (9¹) which gives the amount of harmonic distortion in the modulation process; the second harmonic varies proportionally to x_0^2 and the third to x_0 , that is of negligible amplitude.

Equations (33) and (9¹), although very similar in form, contain a very important difference—the sign of the second harmonic is opposite.

The demodulation process produces nearly the same kind and amount of second harmonic distortion as the modulation process, but with opposite sign.

These results are represented on Fig. 16.

4.2.3.11. Cross distortion.

The cross-distortion is produced by lower side bands of the carrier f_c mixing with the side bands of the D.C. carrier. The lower side bands are given by (31) and the side bands of the D.C. carrier by (30). These side bands can mix with the D.C. side bands and produce cross distortion. The amplitude of these side bands is very small. In first approximation (31) can be written:

$$\frac{1}{2} \lambda \left[\begin{array}{l} \sin [(\omega_c - \Omega) t^1 - \phi] \\ + I_1 \left(\frac{x_c - 2x_0}{2} \right) \sin [(\omega_c - 2\Omega) t^1 - 2\phi] \\ + I_2 \left(\frac{x_c - 3x_0}{2} \right) \sin [(\omega_c - 3\Omega) t^1 - 3\phi] \end{array} \right] \dots (34)$$

It is important to note that the second harmonic side band is of the same sign as for the modulation process, while for the D.C. carrier it is of opposite sign. This property will be used in the combined modulation and demodulation processes.

It can be seen that the distortion is smaller than about 3 per cent. when $n > 20$.

4.3. Combined Modulation and Demodulation

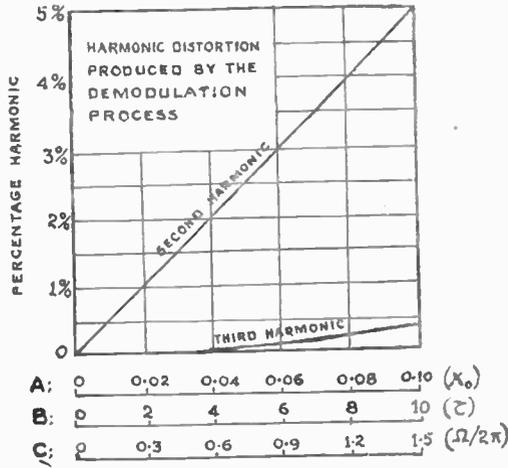
4.3.1. Harmonic distortion.

All the demodulation theory has been developed in assuming

$$\Delta tm = \Delta \tau \sin (\Omega m T_c + \phi) \dots \dots \dots (5)$$

when a signal, proportional to $\sin(\Omega m T_c + \phi)$ is used to produce the pulse modulation, equation (6) must be replaced by (9) or by the more approximate one:

$$\Delta tm = \Delta \tau \left[\begin{array}{l} I_0(x_0) \sin (\Omega m T_c + \phi) \\ + [I_1(x_0) + I_3(x_0)] \sin 2(\Omega m T_c + \phi) \\ + I_2(x_0) \sin 3(\Omega m T_c + \phi) \end{array} \right] (35)$$



- A: VARIABLE $x_0 = \Delta \tau \cdot \Omega$
- B: WIDTH OF MODULATION $\Delta \tau$ IN MICROSECONDS (FOR A FREQUENCY $\Omega/2\pi = 1.5 \text{ K.C.}$)
- C: SIGNAL FREQUENCY $\Omega/2\pi$ IN K.C., FOR A WIDTH OF MODULATION OF $\pm 5 \times 10^{-6}$ SECS.

Fig. 16.—Harmonic distortion in the demodulation process for a modulation width of ± 5 microseconds.

The equation of one impulse is obtained by replacing τm and Δtm in (22) by their new values. Δtm can be replaced by the first component of (34) without making any serious error, as has been already explained. τm must be replaced by $\left(\frac{1 \cdot T_c}{2 \cdot 2n} - \Delta tm \right)$ but here Δtm must be replaced

by (34). The pulse equation (32) being linear with respect to τm , we get the solutions by replacing $\Delta t m$ successively by each component of (35).

For the first component we can assume $I_0(x_0) - I_2(x_0) = 1$. The solution is then given by (30). The second component will give a component of the same frequency, and harmonics. We can neglect the harmonics. Doing the same for the third component of (33) and neglecting $I_3(x)$, we get finally,

$$\begin{aligned}
 &-\frac{1}{2}\lambda \left[I_0\left(\frac{x_0}{2}\right) - I_2\left(\frac{x_0}{2}\right) \right] \sin(\Omega t^1 + \phi) \\
 &\quad - I_1\left(\frac{x_0}{2}\right) \cos(\Omega t^1 + \phi) \\
 &-\lambda I_3(x_0) \cdot \sin 2(\Omega t^1 + \phi) \\
 &\quad + [I_1^2(x_0) + I_2^2(x_0)] \cos 2(\Omega t^1 + \phi) \\
 &-\frac{1}{2}\lambda \left[I_2(x_0) + I_2\left(\frac{3x_0}{2}\right) \right] \sin 3(\Omega t^1 + \phi) \dots\dots\dots(36)
 \end{aligned}$$

which, by replacing the Bessel coefficients by their approximate values, can be written :

$$\begin{aligned}
 &-\frac{1}{2} \left[\sin(\Omega t^1 + \phi) - \frac{3}{4} \frac{x_0^2}{\lambda} \cos 2(\Omega t^1 + \phi) \right. \\
 &\quad \left. + \frac{13}{32} x_0^2 \sin 3(\Omega t^1 + \phi) \right] \dots\dots\dots(37)
 \end{aligned}$$

This equation shows that the combined modulation and demodulation process eliminates practically all harmonic distortion. The residual harmonic terms are not greater than the approximations made.

Recent studies, undertaken a year after the present one, come to a more precise conclusion and show that no harmonic distortion is produced.

4.3.2. Cross-distortion.

If we consider the lower side bands corresponding to the carrier of frequency f_c , their amplitudes are given by (34) when $\Delta t m$ is sinusoidal. If we consider the combined modulation and demodulation process, we have to add expressions like (34) for each component of (9¹). This gives :

$$\frac{1}{2} \lambda \left[\begin{aligned}
 &\sin [(\omega_c - \Omega) t^1 - \phi] \\
 &+ \frac{x_c - x_0}{4} \sin [\omega_c - 2\Omega)t^1 - 2\phi] \\
 &+ \frac{1}{8} \left[\frac{(x_c - 3x_0)^2}{4} + x_0^2 \right] \\
 &\quad \sin [(\omega_c - 3\Omega)t^1 - 3\phi]
 \end{aligned} \right] \dots\dots\dots(38)$$

Figures 17 and 18 represent the second and third harmonic amplitude in percentage of the fundamental when Ω and $\Delta \tau$ varies, as given by equation (38).

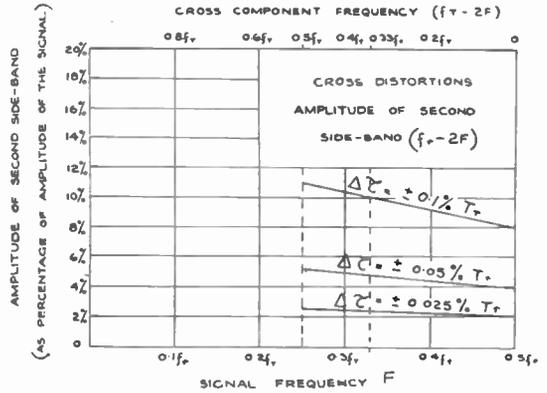


Fig. 17.—Cross-distortion—Amplitude of the side-band of angular frequency equal to $(\omega_c - 2\Omega)$.

Equation (38) shows that the amplitude of the second side band is predominant in practical applications. (X_0 and X_c small). Fig. 17 shows the value of this amplitude in percentage of the amplitude of the modulating wave when Ω and $\Delta \tau$ vary. It can be seen from this figure that the amplitude of this component is always appreciable. The same figure shows that if we confine the signal frequency band to frequencies lower than $f_c/3$, this component has a frequency higher than $f_c/3$ and can be eliminated by a low-pass filter having a cut-off frequency equal to $f_c/3$.

The amplitude of the third side band is represented in Fig. 18. This amplitude is much smaller than the amplitude of the second harmonic. In

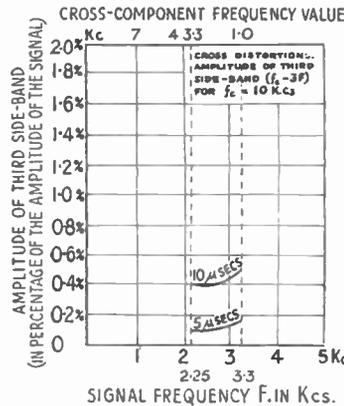


Fig. 18.— Cross-distortion; amplitude of the side-band of angular frequency equal to $(\omega_c - 3\Omega)$.

practical applications this amplitude is smaller than .3 per cent. Although the analysis has been done in making many approximations, the curves of Fig. 19 give at least an order of the magnitude of this component. As this magnitude is very small it can be neglected in practice.

4.3.3. General conclusions.

Comparing Figs. 16, 17 and 18, it can be seen that the most important distortion is due to the cross distortion produced by the side band of angular frequency $(\omega_c - 2\Omega)$. If a small percentage of this distortion is tolerable, then the signal frequency band can be extended to $f_c/2$ with negligible amplitude distortion.

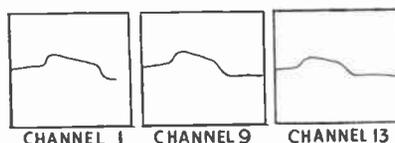


Fig. 19.—
Shape of some
trapezoidal
pulses.

For high quality transmission, it is better to keep the transmission band below $f_c/3$ in order to suppress completely the cross-distortion produced by the $(\omega_c - 2\Omega)$ side band component. In this case the amplitude distortion is always smaller than 1 per cent., the harmonic distortion is negligible and the cross distortion smaller than 0.3 per cent. These values show that a pulse transmission system can give a very high quality and can be used for broadcasting.

5. FIRST EXPERIMENTAL VERIFICATION OF THE THEORY

5.1. Numerical theoretical data for a 20 channel system.

The experimental work has been carried out with a 20-channel pulse transmission system. The chopping frequency was about 8.4 kc/s which corresponds to a chopping period of about 120 microseconds. This gives a channel width of about 6 microseconds. All the channel width was not used so that we can take for $\Delta\tau$ a value of about 2 microseconds. The signal frequency band was 0 to about 3 kc/s which is very near $f_c/3 = 2.8$ kc/s. With these values we get the following theoretical distortions :

Amplitude distortion : negligible (<1 per cent.).

Harmonic distortion : Second and third harmonics smaller than 0.1 per cent.

Cross distortion : Less than 0.1 per cent.

These values are remarkable and are due to the fact that the modulation width of 2 microseconds is very small. But the smaller the modulation width, the worse the signal/noise ratio, since the noise can be represented as a time modulation of the position of the pulse. We measured the signal/noise ratio and we found that for this modulation width the ratio could be made greater than 55 db, which is good enough for practical applications.

5.2. Some experimental measurements

Measurements have been made in order to make a comparison with the theoretical values given above.

The values above are obtained by assuming that the trapezoidal and demodulation pulses are of perfect shape. In the equipment on which the measurements were done, this was not the case.

Figure 19 shows examples of modulating trapezoidal pulses used. They are far from being perfect. Now the measurements gave the following values :

5.2.1. Amplitude distortion.

A typical overall amplitude response is given in Fig. 20. The modulating signal is sent to the

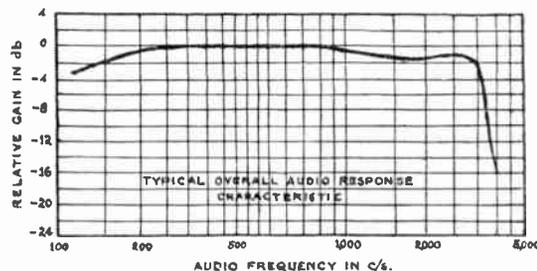


Fig. 20.—Shape of some demodulation rectangular pulses.

transmitter through a transformer and a resistance capacity network, and at the receiver is filtered by means of a low-pass filter and amplified by means of a transformer. Of course, all these components introduce appreciable amplitude distortion and, although no attempt has yet been

made to measure the amplitude distortion produced by each element, these results can be considered as confirmation of the theoretical conclusion, that the amplitude distortion is small.

5.2.2. Harmonic distortion.

Two typical sets of curves for a good and a bad

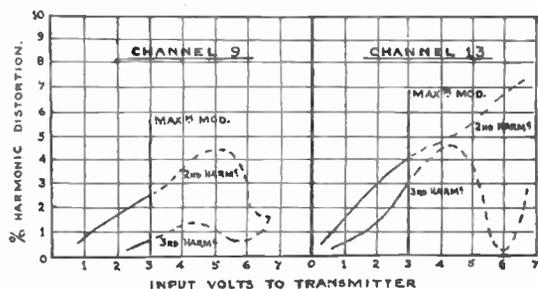


Fig. 21.—Typical overall experimental amplitude response curves.

channel are given in Fig. 21. Limiters were used in the circuit so that only the beginning of these curves can be considered. In the best case, the distortion is smaller than 3 per cent, and it is very probable that a large fraction of this amount is due to the imperfect shape of the trapezoidal and rectangular pulses.

5.2.3. Cross distortion.

A signal having an amplitude such that the modulation width was about 1 to 2 microseconds was applied and the cross-distortion measured. These are the results :

Signal frequency	2 kc/s		2.5kc/s
Theoretical cross-distortion components (Frequency) :	2.3 kc/s	0.3 kc/s	0.8 kc/s
Experimental values found (frequency and percentage amplitude) :	2.4 kc/s	0.3 kc/s	0.9 kc/s
	0.3%	0.43%	0.35%

These values are very approximate and although greater than the theoretical values, they are of the same order of magnitude and seem to be due mainly to the distortion produced by the bad shape of the pulses (this affects, of course, the harmonic distortion (represented by side bands) produced by modulating the carrier f_c).

5.3. Conclusions

The above measurements show that although all kinds of distortions are very small, considerable improvements could be obtained by carefully engineering the system in order especially to improve the shape of the modulating and demodulating pulses. The main importance of the above theoretical study is to show that improvements are possible and to lead to improvements in the circuit design to get a high quality transmission.

I am indebted to Messrs. Standard Telephones & Cables Ltd. for permission to publish this study and to Mr. John Brown for reading and correcting the manuscript early in 1944 after the study was completed.

ELECTIONS AND TRANSFERS TO MEMBERSHIP

The following elections and transfers were recommended by the Membership Committee at their meetings on February 25th and March 25th, 1947.

Elected to Full Member

DORLING, Vice-Admiral Fareham, Hants
James Wilfred Sussex, C.B.

Transferred from Associate Member to Full Member
LEAK, Harold Joseph London, W.5

Elected to Associate Membership

DALE, Reginald Newcastle-on-Tyne

CROSSLAND, Walter Richmond, Surrey

MORGAN, Sydney George Sanderstead, Surrey

SIMONSEN, Willy Oslo, Norway

Transferred from Associate to Associate Member
DOHERTY, Edmund Joseph London, S.E.22

Elected to Associateship

BEERE, Graeme Maunsell, B.Sc. New Zealand
FAULKNER, James Thomas Londonderry, N. Ireland

LOBB, Kenneth London, S.E.6

ROBERTSON, William Barton Newcastle-on-Tyne

ROBINSON, John James Greenford, Middlesex

SMITH, Roy Edward Portsmouth
VIRARAGHAVAN, Triplicane Madras, India
Asuri, B.Sc.

DIXON, Geoffrey Charles Southend-on-Sea

Transferred from Student to Associate

CANNON, Charles William Clacton-on-Sea
Douglas

HAMILTON, William Ian Newcastle-on-Tyne

HIPPLE, Henry Liverpool
MICHIE, Joseph Lawrence Skelmersdale, Lancs

MILLER, George Bertram Roodpoost, South Africa

MILLETT, David Trevor Westcliff-on-Sea
MOTT, Richard John Edgware, Middlesex

SCADENG, Peter London, W.13

WHYTE, Ian Bletchley, Bucks

Elected to Graduate

WILKINSON, Leslie Sheffield

Transferred from Student to Graduate

BROADBERRY, Noel Edward Dublin

Studentship Registrations

ANGELONI, Hernando London, N.W.3

Laurence Glasgow

BOGDANOWICZ, Czeslaw Ilminster

BRIGGS, Robert Frederick London, N.16

BRULEY, John Christchurch,

BURGESS, Philip Hugh Hants

George Shepperton-on-

CARTER, Donald John Thames

CROSS, Edward Charles Hailsham, Sussex

CROSS, Peter Dorrien London

DEEGAN, Peter Anthony London, N.1

de SILVA, Appu Hennedige Ceylon

Derrick

DIPPLE, John Aubrey Walthamstow, E.17

Schofield

FOGG, Harold Joseph Penrith

GILMOUR, George Douglas,

Isle of Man

HARE, Ronald Edwin Mill Hill, N.W.7

HARVEY, James Queensland, Australia

HUMPHREYS, David John Ilford, Essex

KING, Kenneth William London, S.E.12

KIRYLUK, Wlodzimierz London, W.14

MOLD, Donald Douglas Bingley, Yorks

NARASIMHACHAR, Bangalore

Mandayam Kannappan

OLOWU, Adebayo Ayo-Ola Lagos, Nigeria

PELLS, Gordon Hambley Edgware, Middlesex

PORRITT, Brian Longshaw Ashted, Surrey

Day

PRIOR, Horace Walter Harrow, Middlesex

RAMADAN, Mohammed Mansoura, Egypt

SLON, Rickaed Glasgow

STEVENS, Max Mathew Singapore

ZYJEWSKI, Joseph Glasgow

Re-instatement to Associateship

BROWN, Philip Thomas London, S.W.11