

The Bell System Technical Journal

Vol. XXII

October, 1943

No. 3

Effect of Feedback on Impedance

By R. B. BLACKMAN

THE impedance of a network is defined as the complex ratio of the alternating potential difference maintained across its terminals by an external source of electromotive force, to the resulting current flowing into these terminals. If the network contains active elements such as vacuum tubes, the resulting current (or potential difference if the input current is taken as the independent variable) may be due in part to the excitation of the active elements. The definition of impedance does not discriminate between the part of the current (or potential difference) due directly to the external source of electromotive force and the part due to the excitation of the active elements by the external source. Hence the impedance will in general depend upon the degree of activity of the active elements.

These observations were made early in the development of feedback amplifiers by H. S. Black¹ who made two important uses of the effect of feedback on impedance. In the first place it afforded a method of measuring feedback which has some advantages over the method which involves opening the feedback loop, providing proper terminations for it and measuring the transmission around it. In the second place the effect of feedback on impedance was used to control the impedances presented by a feedback amplifier to the external circuits connected to it.

Relations between impedance and feedback were derived by Black and others for a number of specific feedback amplifier configurations. In some cases these relations turned out to be very simple. For the most part, however, these relations were so complicated that they defied reduction to a common form.² The difficulty seems to have been due, in part at least, to the attempt to formulate the relationship, in each case, in terms of the normal feedback of the amplifier. In some cases the difficulty seems to have been due partly also to the valid, but, as it turns out, irrelevant observation that the feedback is affected by the impedance of the measuring circuit as

¹ H. S. Black, "Stabilized Feedback Amplifiers", *B.S.T.J.*, January, 1934.

² Shortly after the general relationship between feedback and impedance was derived, it was independently established by H. W. Bode and J. M. West by examination of a variety of feedback amplifier designs. The generality of the relationship was also independently proved for amplifiers with a single feedback path by J. G. Kreer and by C. H. Elmendorf.

well as by the removal of any impedance elements or circuits which are normally connected to the amplifier.

These difficulties are avoided by the method of derivation adopted in this paper. Illustrative examples are then given of some of the uses to which the general relationship between feedback and impedance may be put.

DERIVATION

The derivation of the general relationship between feedback and impedance will be made here with reference to the diagram shown in Fig. 1.

One of the vacuum tubes in the network, namely that one to which the feedback is to be referred, is shown explicitly at the top of the box in the diagram. The grid lead to this tube is broken at terminals 2, 2'. In practice, the break in the grid lead would leave the grid still coupled to some

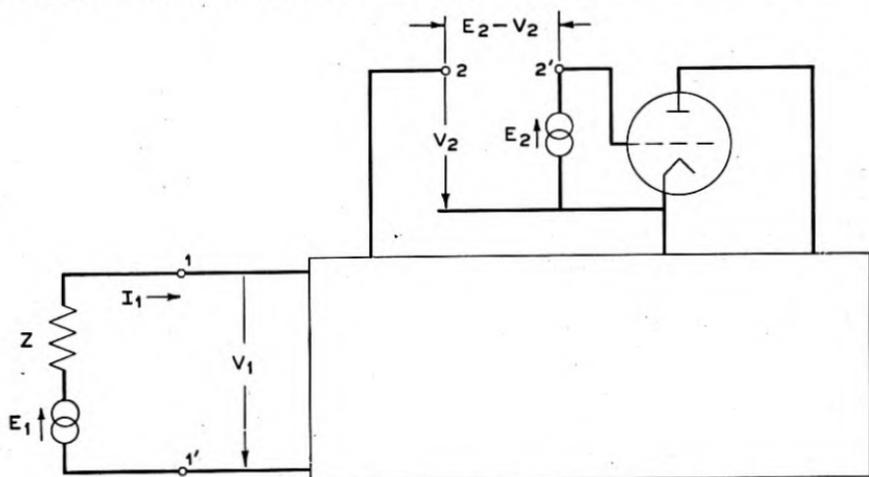


FIG. 1—Relation between feedback and impedance.

degree to the other electrodes of the tube through parasitic interelectrode admittance. For analytical purposes, however, it may be assumed that the parasitic admittances between the grid and the other electrodes of the tubes are connected not directly to the grid within the tube but to some point farther out along the grid lead. Under this assumption the break in the grid lead not only removes the feedback to the tube completely, but also leaves the parasitic admittances connected in the network in such a way that their contribution to the feedback is implicitly taken into account. Furthermore, the impedance looking into the grid of the tube is now infinite so that if a voltage is applied to the grid no current will be drawn from the source of the voltage.

At the left-hand side of the box in the diagram, terminals 1, 1' are brought out. These are the terminals to which the impedance is to be referred. In

the normal condition of the network these terminals may be connected through an external impedance branch. This is the case, for example, when terminals 1, 1' are the input terminals of a feedback amplifier whose input impedance is under investigation. However, this external impedance may also be zero or infinite according as terminals 1, 1' are "mesh-terminals" obtained by breaking open a mesh of the network, or "junction-terminals" obtained by bringing out two junctions of the network.

It is assumed that the network, including all of the vacuum tubes, is a linear system in which, therefore, the Superposition Principle holds. Hence, if an e.m.f. E_1 is applied in series with terminals 1, 1' and a second e.m.f. E_2 is applied between the grid and the cathode of the tube, the potential difference V_1 developed across the input terminals 1, 1' and the potential difference V_2 developed between the terminal 2 and the cathode of the tube will be linearly related to E_1 and E_2 . If the source of E_1 has internal impedance the coefficients in these relations will depend upon this impedance. However, if the input current I_1 is used as an independent variable in place of the e.m.f. E_1 the coefficients will not depend upon the impedance of the source of the current I_1 . It is also convenient to consider the potential difference $E_2 - V_2$ developed across the terminals 2, 2' as one of the dependent variables in place of V_2 . Therefore,

$$\left. \begin{aligned} V_1 &= AI_1 + BE_2 \\ E_2 - V_2 &= CI_1 + DE_2 \end{aligned} \right\} \quad (1)$$

where the coefficients are independent of Z .

From these equations we obtain

$$\begin{aligned} \left(\frac{V_1}{I_1} \right)_{E_2=V_2} &= \frac{AD - BC}{D} \\ \left(\frac{V_1}{I_1} \right)_{E_2=0} &= A \\ \left(\frac{E_2 - V_2}{E_2} \right)_{V_1=0} &= \frac{AD - BC}{A} \\ \left(\frac{E_2 - V_2}{E_2} \right)_{I_1=0} &= D \end{aligned}$$

Hence

$$\frac{\left(\frac{V_1}{I_1} \right)_{E_2=V_2}}{\left(\frac{V_1}{I_1} \right)_{E_2=0}} = \frac{1 - \left(\frac{V_2}{E_2} \right)_{V_1=0}}{1 - \left(\frac{V_2}{E_2} \right)_{I_1=0}} \quad (2)$$

This equation expresses the relationship between feedback and impedance. To make this more apparent the physical significance of each of the factors in this equation will be examined and suitable symbols will be substituted for them.

In equations (1) E_2 and I_1 were regarded as independent variables. However, the ratio $\left(\frac{V_1}{I_1}\right)_{E_2=V_2}$ implies that E_2 is adjusted to be equal to V_2 . This means that E_2 is dependent upon I_1 . The reason for the imposition of this dependence is that with E_2 equal to V_2 the terminals 2, 2' may be connected together and the source of E_2 may be removed without affecting, in particular, the potential difference V_1 across terminals 1, 1' and the current I_1 into these terminals.

Obviously, therefore, the ratio $\left(\frac{V_1}{I_1}\right)_{E_2=V_2}$ is the impedance which will be seen at the terminals 1, 1' when terminals 2, 2' are connected together and the only source of e.m.f. acting on the network is the external circuit connected to the terminals 1, 1'. This ratio will be symbolized by Z_A .

The ratio $\left(\frac{V_1}{I_1}\right)_{E_2=0}$ implies that no voltage is applied between the grid and the cathode of the tube. However, it is immaterial whether or not a voltage is applied to the grid of the tube if the amplification of the tube is nullified. Obviously, therefore, this ratio is the impedance which will be seen at the terminals 1, 1' when terminals 2, 2' are connected together and the amplification of the tube is nullified. This ratio will be symbolized by Z_P .

Finally, the ratios $\left(\frac{V_2}{E_2}\right)_{V_1=0}$ and $\left(\frac{V_2}{E_2}\right)_{I_1=0}$ are readily recognized from the definition of feedback to be the feedback to the vacuum tube with the terminals 1, 1' connected together in the first case, and left open in the second. These ratios will be symbolized by F_{sh} and F_{op} respectively.

Hence, equation (2) may be written in the more significant form

$$\frac{Z_A}{Z_P} = \frac{1 - F_{sh}}{1 - F_{op}} \quad (3)$$

DETERMINATION OF FEEDBACK

One of the uses to which the relationship (3) may be put is in the determination of feedback by impedance measurement. However, since this relationship involves two feedbacks, only one of which may be identified with the feedback to be determined, one of these feedbacks must be known.

In the most common types of feedback amplifiers it is possible to choose

terminals 1, 1' so that either F_{Sh} or F_{Op} is zero. If $F_{Op} = 0$ and $F_{Sh} = F_N$ where F_N is the normal feedback, then

$$F_N = 1 - \frac{Z_A}{Z_P} \quad (4)$$

On the other hand, if $F_{Sh} = 0$ and $F_{Op} = F_N$ then

$$F_N = 1 - \frac{Z_P}{Z_A} \quad (5)$$

Fig. 2 shows a feedback amplifier in which the μ -circuit and the β -network are connected in series at one end and in parallel at the other end. At terminals 1, 1' in this figure the conditions for formula (4) are obviously fulfilled. Hence, if the impedance measurements are made at these ter-

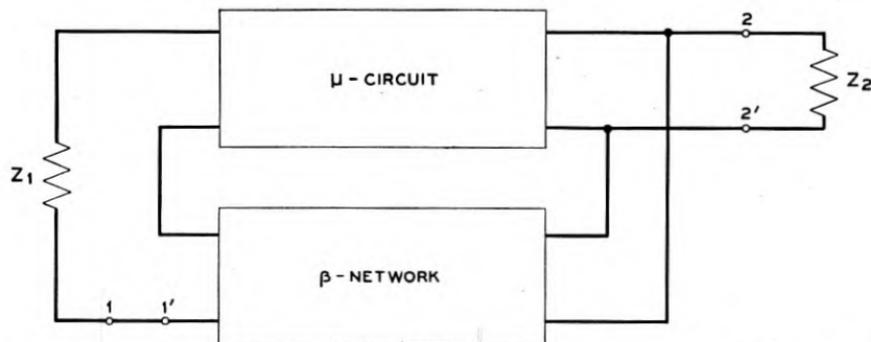


FIG. 2.—Feedback amplifier with series feedback at one end and shunt feedback at the other end.

minals, the feedback is given by formula (4). On the other hand, at terminals 2, 2' in Fig. 2 the conditions for formula (5) are obviously fulfilled. Hence, if the impedance measurements are made at these terminals, the feedback is given by formula (5).

If the grid-plate parasitic admittance of a tube in a feedback amplifier is not negligible it is not possible to open any physical mesh in the amplifier so that $F_{Op} = 0$ for that tube. In such a case, therefore, (4) is not applicable. However, if the impedance measurements are made between the grid and the cathode of that tube the conditions for formula (5) are obviously fulfilled, and the feedback is given by formula (5). Hence, of the two particular forms (4) and (5) of the general relationship (3), only (5) enjoys complete generality in the determination of feedback by impedance measurements.

FEEDBACK DURING IMPEDANCE MEASUREMENTS

While the feedback computed from impedance measurements by formula (4) or (5) is the normal feedback, the feedback during the impedance measurements may be quite different, due to the impedance of the impedance measuring circuit. Referring to Fig. 1 we see that the feedback during measurement is by definition

$$F_z = \left(\frac{V_2}{E_2} \right)_{V_1 = -ZI_1}$$

where Z is the impedance of the impedance measuring circuit. By equations (1) this is easily reduced to

$$F_z = \frac{Z_P F_{sh} + Z F_{Op}}{Z_P + Z} \quad (6)$$

Under the conditions to which formula (4) applies

$$F_z = \frac{F_N}{1 + \frac{Z}{Z_P}} \quad (7)$$

It is clear therefore that even if F_N satisfies Nyquist's Stability Criterion, Z may be of such a character that F_z violates that criterion. In that case it will be impossible to make the impedance measurements.

Contrariwise, if F_N violates Nyquist's Stability Criterion, it is possible to choose Z so that F_z satisfies that criterion and make it possible to measure the impedance. Substituting (4) into (7) we find that a sufficient but not necessary condition in order that $|F_z| < 1$ is that

$$|Z| > |Z_A| + 2|Z_P|$$

Under the conditions to which formula (5) applies

$$F_z = \frac{F_N}{1 + \frac{Z_P}{Z}} \quad (8)$$

Similar observations may be made with respect to (8) as were made with respect to (7). Substituting (5) into (8) we find that a sufficient but not necessary condition in order that $|F_z| < 1$ is that

$$\frac{1}{|Z|} > \frac{1}{|Z_A|} + \frac{2}{|Z_P|}$$

FEEDBACK CONTROL OF IMPEDANCE

The application of the relationship (3) to the feedback control of impedance may be illustrated by a few concrete examples.

Let us assume that we are interested in the impedance faced by the line impedance Z_1 in Fig. 2. If the terminals 1, 1' in Fig. 3 are left open the feedback is obviously zero. Let the feedback when the terminals are shorted together be denoted by F_{sh} . If the impedances of the μ -circuit and the β -network are denoted by Z_μ and Z_β , respectively, then

$$Z_A = Z_P(1 - F_{sh}) \quad (9)$$

where

$$Z_P = Z_\mu + Z_\beta$$

This shows the now well-known fact that series feedback may be used to magnify impedance.

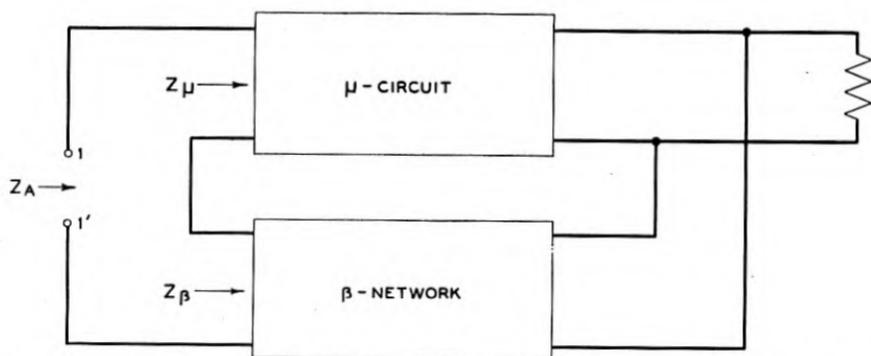


FIG. 3—Impedance faced by the line at the series feedback end of a feedback amplifier.

However, it should be noted that the feedback F_{sh} involved in (9) is not now equal to the normal feedback F_N as it was when the terminals 1, 1' were taken as in Fig. 2. The relation between F_N and F_{sh} may be obtained from (6) by identifying F_N with F_z , and Z_1 with Z . Hence

$$F_N = \frac{F_{sh}}{1 + \frac{Z_1}{Z_P}} \quad (10)$$

From (9) and (10) it follows that even with a very modest amount of normal feedback the magnification of the impedance may be very large. For example, if $Z_P = 1000$ ohms, $Z_1 = 1$ megohm and $F_{sh} = -1000$, then Z_A is better than 1000 times as large as Z_P although F_N is not quite unity in magnitude.

Similarly, the impedance faced by the line impedance Z_2 in Fig. 2, as shown in Fig. 4, is

$$Z_A = \frac{Z_P}{1 - F_{Op}} \quad (11)$$

where

$$Z_P = \frac{Z_\mu Z_\beta}{Z_\mu + Z_\beta}$$

This shows the now well-known fact that shunt feedback may be used to reduce impedance.

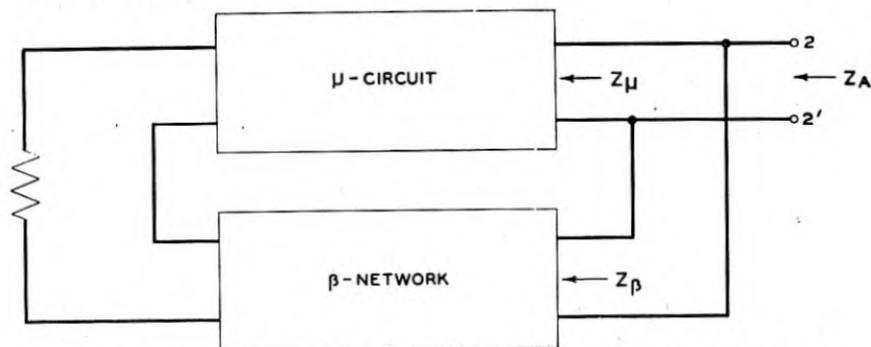


FIG. 4—Impedance faced by the line at the shunt feedback end of a feedback amplifier.

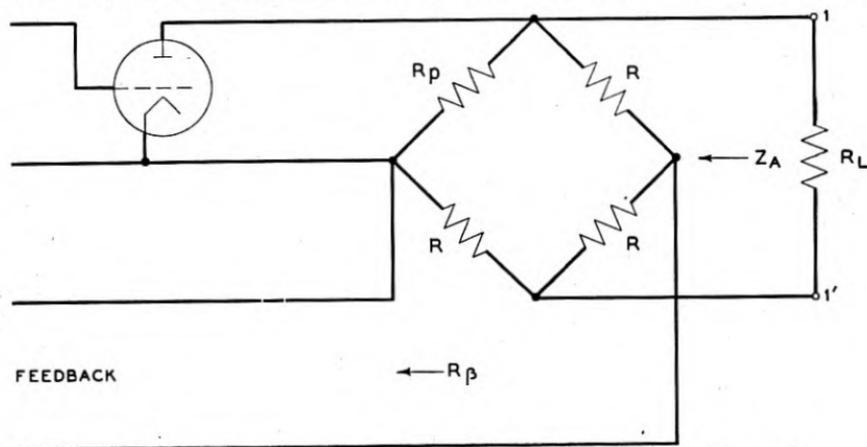


FIG. 5—Impedance faced by the line connected to a bridge feedback amplifier.

The relation between the normal feedback F_N and the feedback F_{Op} involved in (11) is, by (6)

$$F_N = \frac{F_{Op}}{1 + \frac{Z_P}{Z_2}} \quad (12)$$

From (11) and (12) it follows that even with a very modest amount of normal feedback the reduction in impedance may be very large. For example,

if $Z_p = 100,000$ ohms, $Z_2 = 100$ ohms and $F_{Op} = -1000$, then Z_A is less than 100 ohms although F_N is not quite unity in magnitude.

The two examples given above illustrate the use of feedback to magnify or to reduce the impedance of a network. This impedance, however, will be correspondingly sensitive to changes in the characteristics of the vacuum tubes. A third example of the use of the relationship (3) will show that feedback may also be used to make the impedance of a network less sensitive to changes in the characteristics of the vacuum tubes.

In the case of the bridge-type feedback network shown in Fig. 5 we have, with respect to the terminals 1, 1'

$$Z_p = R(1 + Q)$$

$$F_{Sh} = A \frac{1 + \frac{2R + 3R_\beta}{R + R_\beta} Q}{1 + Q}$$

$$F_{Op} = A \left(1 + \frac{2R + 3R_\beta}{R + R_\beta} Q \right) = (1 + Q)F_{Sh}$$

where A is the feedback designed for the condition $R_p = R$, and

$$Q = \frac{(R_p - R)(R + R_\beta)}{(R + R_p)(2R + R_\beta) + 2RR_\beta}$$

Then, by (3)

$$Z_A = R \left(1 + \frac{Q}{1 - F_{Op}} \right)$$

Hence, if the feedback F_{Op} is very large the effect of bridge unbalance on the impedance presented to the line will be very small. If, for example, the design feedback is 40 db the output impedance cannot change more than 1 per cent however severely the bridge might be unbalanced by R_p being larger than R .

The feedback when the line impedance R_L is connected may be obtained by identifying R_L with Z in formula (6). It is

$$F = A \frac{1 + \frac{2R + 3R_\beta}{R + R_\beta} Q}{1 + \frac{R}{R + R_L} Q}$$

whence

$$\frac{\partial \log F}{\partial \log R_L} = \frac{RR_L}{R + R_L} \frac{Q}{R_L + Z_p}$$

The effect of bridge unbalance is to make the feedback sensitive to changes in the line impedance R_L .

Design of Two-Terminal Balancing Networks

By K. G. VanWYNEN

This paper describes a simple graphical method for designing a two-terminal network, which will simulate a given line impedance to such a degree that return losses of the order of 25 db or better will be readily obtained. The method is particularly useful in those problems in which a reasonably accurate balancing network is adequate, but a high degree of precision is not required.

GENERAL

IT IS the purpose of this paper to describe a graphical method which has been found useful in the design of simple two-terminal networks to simulate the impedance of transmission lines or equipment. The discussion which follows is intended to emphasize the simplicity of the method and the rapidity with which it may be employed to arrive at a solution; it will also indicate the analytical background without attempting to develop or establish the rigor of the procedure involved. A solution can frequently be obtained in a fraction of an hour and it is thought that the graphical analysis will appeal to the pragmatist and the engineer who has a job to do, but very little time in which to accomplish his aim, rather than the person interested in the rigor of the solution.

The problem which is considered may be stated as follows: Design a two-terminal network with the minimum number of elements which will give a desired degree of approximation to a given impedance function $Z(\lambda)$, where $Z(\lambda)$ is a fraction whose numerator and denominator are polynomials in frequency in accordance with the customary usage in such problems.

ORIGIN OF PROBLEM

This problem has arisen most generally in providing balancing networks which will give satisfactory return losses against various types of telephone facilities. It is obvious that for a given impedance, $(r + jx)$, at a given frequency there are an infinite number of networks which will satisfy the given impedance. It has also been pointed out that the network which simulates a given impedance function is not unique. Hence there are also a large number of networks which will satisfy a given impedance function.

In designing networks for repeater circuits, it is generally satisfactory

if the return loss is equal to or greater than some specified number of db. This somewhat simplifies our problem and permits a double infinity of solutions. A method has been given by Brune¹ for designing such networks, in which it is pointed out that there is no unique solution to the problem of designing a finite two-terminal network and also states that any network which satisfies the impedance function may be considered a satisfactory solution to the problem. It is thought that the method which is given below will provide a solution which makes maximum use of the number of elements employed. That is, it will provide a given return loss with the minimum number of parts.

The required degree of approximation and the frequency range to be covered determine the number of elements required in this solution. In one simple case which will be discussed below in the first example, the approximation between the impedance of a transmission line and a network designed to simulate it is the approximation between the curvature of the impedance function and the arc of a circle.

GENERATING FUNCTION

The method discussed here differs from that outlined by Brune in that use is made of known generating functions which are added together in series to approximate the total function, similar to the manner in which sine functions may be added to approximate other functions. This series type network can readily be converted to the ladder type by well known network equivalence theorems and the solution will then have the Stieltjes fraction form pointed out by Fry² and Cauer.³

The generating function used here is an impedance consisting of a resistance in parallel with a pure reactance or a special case of this. This function plus a real corresponds to a bilinear transformation, the properties of which have frequently been discussed elsewhere. This particular configuration, for instance, has been pointed out both by Brune and by Guillemin³ at M.I.T. and a discussion of the bilinear transformation has been given by C. W. Carter⁴ of the Bell Telephone Laboratories. The series addition of such generating functions is similar to the form given in Foster's reactance theorem except that there only pure reactances are dealt with. The solution can also be worked out with admittances, but will not be discussed here since the average engineer is more accustomed to dealing with impedances.

In many problems, particularly those involving dissipative transmission

¹ *Jour. Math. & Physics*, Vol X, 1930-1931.

² *Bull. Am. Math. Soc.*, 35, 1929.

³ Guillemin—Vol. II.

⁴ *B.S.T.J.*, July 1925.

lines, the entire impedance function is found in the fourth quadrant of the complex plane. When this is so, the generating function is reduced to a resistance in parallel with a condenser.

GRAPHICAL REPRESENTATION OF FUNCTIONS

As a first step in utilizing the graphical procedure, it will be advisable to acquire some familiarity with the generating function in its general form

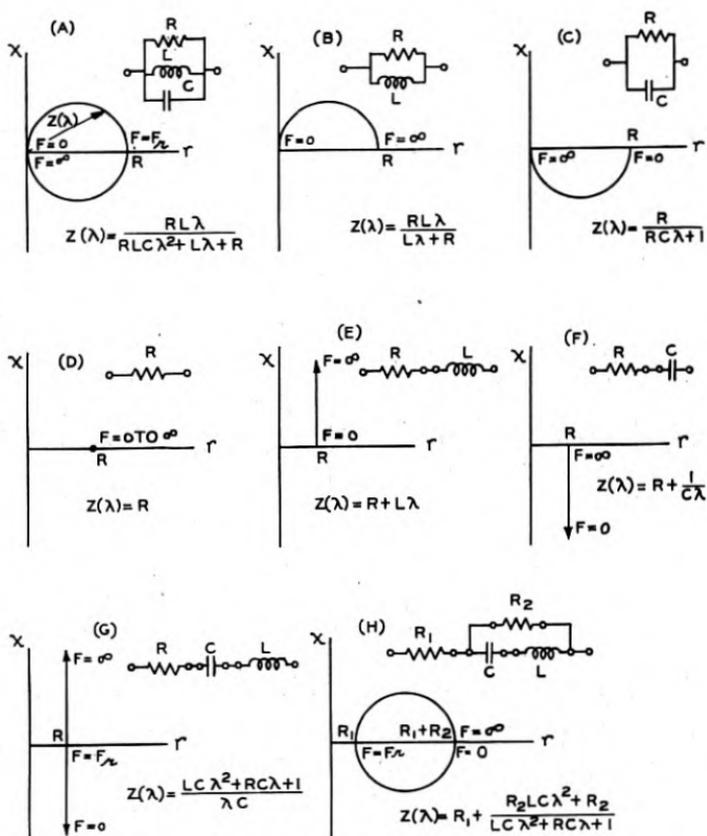


Fig. 1.—The impedance loci, $Z(\lambda)$, for several networks.

and some of its special cases. Plots of various cases are given in Figs. 1(a) through 1(h) together with the network configuration and the impedance function thereof. Obviously the summation of the properly selected generating functions corresponds to the addition of the partial fractions derived by Brune's method. For an accurate solution these partial fractions when combined should approximate the given $Z(\lambda)$.

Figure 1(A) shows the impedance locus of the parallel R, L, C generating

function as frequency varies from 0 to ∞ . At 0 frequency $Z(\lambda) = 0 + j0$ and at ∞ $Z(\lambda) = 0 - j0$. The locus is a circle and crosses the real axis at R and the frequency at which L and C are anti-resonant. The special cases will be readily apparent and without further discussion attention will be shifted to Fig. 1-c which is the generating function applied to obtain solution of the examples listed below, all of which are located in the fourth quadrant.

The impedance of this function (Fig. 1-c) at zero cycles is a real and has the value R , and at infinite frequency its impedance is $0 - j0$. The locus traced by this function in the fourth quadrant of the complex plane as f varies from zero to infinity is a semicircle of radius $R/2$ whose center is at $R/2$ on the axis of reals. Obviously, the impedance for any given frequency depends only on C when R has been fixed.

One of the most useful networks for voice frequency work is that in which two such functions are added together but the second function is the special case in which $C = 0$. We then have a network which consists of a resistance R_1 in series with the parallel combination R_2 and C_2 , and is represented by the semicircle just described but displaced to the right of the origin by the distance R_1 . This form corresponds to a special case of the bilinear transformation previously mentioned.

As stated earlier a given impedance function can be obtained from a large number of networks but when the impedance is to be simulated for a limited frequency range, such as the voice band, the selection of the best network is reduced to sorting through a relatively small range of networks to select that one which is the best compromise for the given conditions. This then is a restatement of the problem: *To find the network having the minimum number of circuit elements which will give the desired approximation to a specified impedance function.*

The other sections of Fig. 1 will be evident upon analysis.

METHOD OF SOLUTION

The first step to be followed in finding the solution to a given problem is to plot in the complex plane the locus traced by the given impedance function as the frequency varies over the range which is to be considered and to mark the frequency at those impedances which are essential to the problem. Having done this, the next step is to draw a semicircle with the center on the real axis such that an arc of the semicircle approximates part or all of the locus of the impedance function. In many cases this semicircle is a sufficiently good approximation but where it is not, it will be necessary to add other functions. The examples given below are illustrative of cases requiring three-, four- and five-element networks.

EXAMPLE 1—104 MIL OPEN WIRE

To demonstrate the method we will now consider the design of a network which simulates a 104-mil copper open-wire line with 12 in. spacing and CS insulators. The impedance function for this particular facility is plotted on Fig. 2. It is perhaps rather obvious that this locus can readily be approximated by a semicircle whose center is on the real axis and whose intercept on the real axis is not at the origin. Such a semicircle has been drawn, but it is recognized that the one shown is not unique, for it would be possible to draw several others which might do equally well. However, they

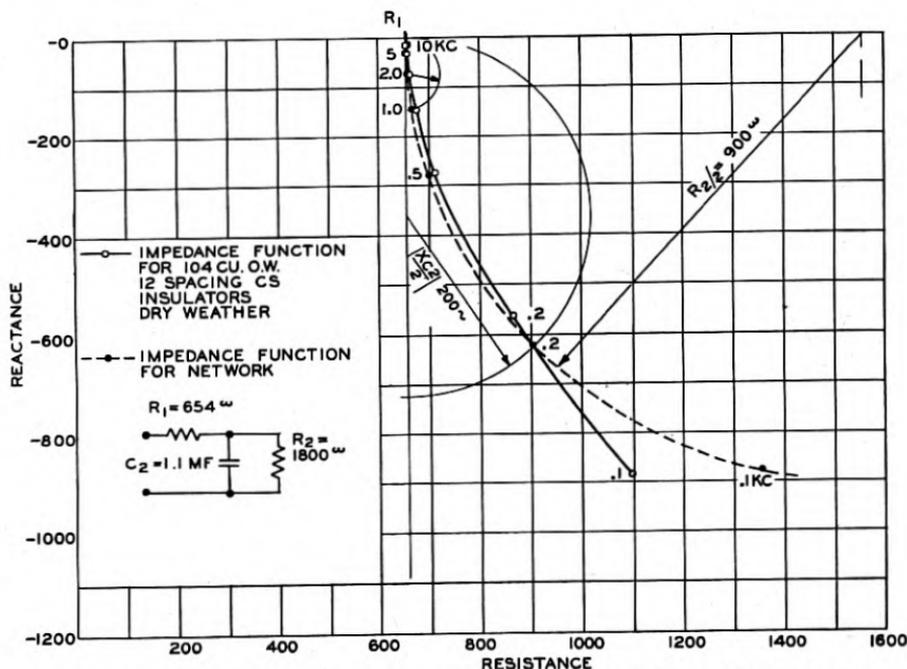


Fig. 2.—Graphical design of two-terminal balancing network for 104-mil. copper open wire.

would in general be fairly close to that shown. Having selected this semicircle, which approximates the impedance function, it is evident that a network consisting of a resistance, R_1 , in series with a parallel R_2C_2 combination will provide a reasonable approximation above 200 cycles. The series resistance R_1 , is, of course, the left-hand intercept of the semicircle and the R axis and the parallel resistance, R_2 , is the diameter of the semicircle. There remains, then, the problem of determining C_2 which obviously governs the distribution of frequencies along the semicircular locus. If C_2 is very small, the 1000 cycle impedance will be near the right-hand end of the locus since

R_2 is controlling and vice versa. The answer as to what value of C should be selected depends on what frequency range we are most interested in approximating closely. Suppose in this case that we say 1000 cycles is the frequency at which we wish to have the best degree of approximation. C_2 will then be determined by drawing a vertical axis passing through R_1 and inscribing a semicircle passing through R_1 and the 1000-cycle impedance of the open-wire line function and having its diameter on the vertical axis. The diameter of this semicircle represents X_C and therefore determines the capacity, C_2 , of the parallel combination.

Carrying out the procedure just described it will be seen by reference to Fig. 2, that $X_{C_2} = 145$ ohms at 1000 cycles and therefore $C_2 = 1.1$ mf. The 3-element network thus determined is a resistance of 654 ohms in series with the parallel combination of 1800 ohms and 1.1 mf. By arbitrary choice the 1000-cycle impedance of the line and network are in good agreement. It is now necessary to determine the network impedance at other frequencies in order to compare them against the open-wire line impedance.

As is well known the parallel impedance at any other frequency is the intersection of the corresponding X_C and R_2 semicircles. At 200 cycles $X_C = 725$ ohms. Drawing a semicircle of diameter 725 ohms on the vertical axis through 654 ohms the network impedance is located at the intersection of this semicircle and the R_2 semicircle, i.e., at $900 - j620$.

Thus the network impedance locus as a function of frequency may be completely determined over the desired frequency range and compared with the given impedance locus of the open wire.

This may be done visually. If corresponding points on the two loci are close together, the simulation will be a good one and vice versa. If it is found that the simulation is too good at one frequency and not good enough at other frequencies, it will be possible to alter the distribution of frequencies along the locus by changing C_2 or the locus may be shifted by changing R_2 or both C_2 and R_2 may be changed. No specific rule can be stated for this but with a little experience considerable dexterity may be acquired in this sort of juggling and a locus found which will give an approximately constant approximation over a reasonably wide frequency range. As may be seen by referring to Fig. 2, it was found that a network consisting of a 654-ohm resistance in series with the parallel combination of 1800 ohms and 1.10 mf. gives a very good simulation of a 104 mil copper open wire line over the voice range. As is obvious from the graphical method, the simulation rapidly deteriorates below 200 cycles due to departure of the network locus from the impedance locus of the open wire line. If it were necessary to improve this low-frequency simulation, it would be necessary to add further generating functions to the design or compromise at the higher frequencies.

Since this network was intended for use as a balancing network, it was

then tested in the laboratory against the open-wire impedances and found to give fairly high return losses as listed in Table I. The corresponding return losses were also computed and tabulated. The impedances are given for both the network and the theoretical line at typical frequencies over the range from 100 cycles to 20,000 cycles.

The impedance function for an open-wire line is given by the equation

$$Z(\lambda) = \left(\frac{R + L\lambda}{G + C\lambda} \right)^{\frac{1}{2}} \quad (1)$$

Expanding this function by the binomial theorem and taking the first approximation and further letting $G = 0$, the impedance function becomes

TABLE I
104 Mil Cu Open Wire, Dry Weather, 12" Spacing, CS Insulators

Freq. Cycles	Impedance				Return Loss of Net- work vs Line-db	
	Network		Line		Measured	Computed
	Rect.	Polar	Rect.	Polar		
100	1360—j878	1620/32.9	1101—j883	1410/38.8	20.9	21.3
200	904—j623	1097/34.7	865—j562	1032/33.0	27.4	29.4
300					34.4	
500	699—j281	754/21.9	712—j273	764/21.0	37.8	39.6
1000	665—j143	681/12.2	674—j144	689/12.0	42.3	43.6
2000	656—j72	660/6.3	662—j74	666/6.4	45.2	47.8
5000	654—j28	654/2.4	658—j32	659/2.8	43.1	48.4
10000	654—j14.5	654/1.0	653—j12	653/1.1	39.5	55.5
20000	654—j7.2	654/0.5	652—j10	652/0.9		51.2

$$Z(\lambda) = \left(\frac{L}{C} \right)^{\frac{1}{2}} + \frac{1}{(LC)^{\frac{1}{2}}(2/R)} \cdot \frac{1}{\lambda} \quad (2)$$

Applying the method of Brune, this equation yields a network consisting of 646.4 ohms in series with a condenser of 1.09 mf. It will also be apparent that eq. (2) has the same form as that of Fig. 1(f), i.e.,

$$Z(\lambda) = R_1 + \frac{1}{C_1} \cdot \frac{1}{\lambda}$$

and by a 1 to 1 comparison of terms it is evident that

$$R_1 = \left(\frac{L}{C} \right)^{\frac{1}{2}} \quad (3-a)$$

and

$$C_1 = (LC)^{\frac{1}{2}} \left(\frac{2}{R} \right) \quad (3-b)$$

Including G , the expression for the first approximation of the impedance function may be written in the form

$$Z(\lambda) = \left(\frac{L}{C}\right)^{\frac{1}{2}} \left[1 + \frac{\left(\frac{R}{2L} - \frac{G}{2C}\right)}{\lambda + \frac{G}{2C}} \right] \quad (4)$$

Following Brune's method or noting the correspondence with the impedance function given in Fig. 1(c), it is apparent that the network is a resistance, R_1 , of 646.4 ohms in series with a parallel R_2C_2 combination. C_2 is 1.09 mf as before but R_2 computes as 312,000 ohms which is so large compared to 1.09 mf. that the additional resistance provides negligible improvement over the previous network for the voice frequency range.

Obviously then, the analytical method requires at least a second order approximation entailing considerable additional analytical work and computation which will not be carried out here. This points out the advantages of the graphical method; namely, it is rapid, requires no special skill, and gives a reasonably accurate answer.

EXAMPLE 2—SIRAL FOUR CABLE—1320 FOOT SPACING—6 MILHENRY LOADING (SP4-1320-6)

In order to indicate the procedure when two complete RC regenerating functions are required, another example is given which covers an impedance simulation of a SP4-1320-6 line. A plot of this impedance function is shown on Fig. 3, and it is at once obvious that two semicircular generating functions should give a reasonably good approximation to the given impedance function.

The method of selecting these functions may be somewhat as follows: Consider first the simulation in the low-frequency range, i.e., 200 cycles to 500 cycles. For this region a semicircle may be selected much as in the first example and the one chosen yields a network consisting of $R_1 = 480$ ohms in series with the R_2C_2 parallel combination in which $R_2 = 1460$ ohms. C_2 was found by choosing an X_{C_2} at 500 cycles close to that of the line and from which C_2 was found to be 1.38 mf.

It is evident that to provide high-frequency simulation a condenser must be placed in parallel with $R_1 = 480$ ohms. Its value is determined by the intersection of the R_1 and X_{C_1} semicircles at 10,000 cycles and C_1 is found to be .0161 mf. The construction lines involved in these determinations are shown as light weight solid lines.

Since there are now two impedance functions to be added in series the locus will depart somewhat from the two semicircles. However, the departure will not be great since the effect of C_1 is small at low frequencies,

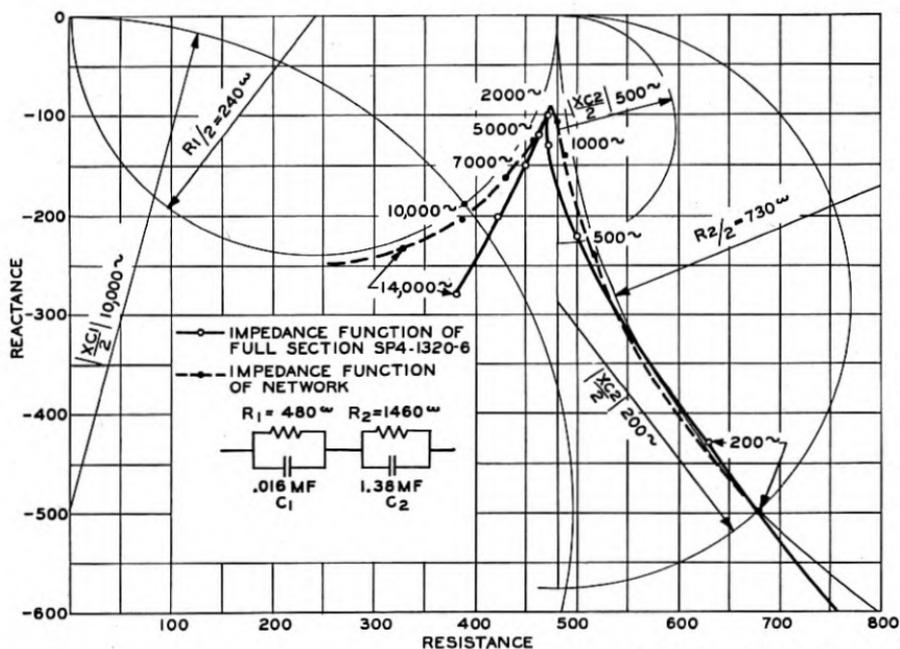


Fig. 3—Graphical design of two-terminal balancing network for spiral four cable.

TABLE II

Spiral Four Cable—1320 Foot Spacing—6 Milhenry Loading—Full Section Termination

Freq. Cycles	Impedance				Return Loss vs Theoretical Line-db		Measured Return Loss vs Artificial* Line-db
	Network		Line		Measured	Computed	
	Rect.	Polar	Rect.	Polar			
100	1046-j713	1266/ $\overline{34.3}$	810-j670	1051/ $\overline{39.6}$	14.8	19.8	21.8
200	701-j528	878/ $\overline{37.0}$	630-j430	763/ $\overline{34.3}$	22.8	22.6	22.9
500	516-j239	569/ $\overline{24.9}$	500-j220	546/ $\overline{23.8}$	33.2	33.0	25.2
1000	488-j139	507/ $\overline{15.9}$	470-j130	488/ $\overline{15.5}$	37.1	33.9	26.2
2000	479-j108	491/ $\overline{12.7}$	470-j100	481/ $\overline{12.0}$	41.2	38.1	26.6
3000							26.2
5000	456-j125	473/ $\overline{15.3}$	460-j120	475/ $\overline{14.6}$	39.9	43.4	26.3
7000	430-j163	460/ $\overline{20.8}$	450-j150	475/ $\overline{18.4}$	36.7	31.8	25.1
10000	389-j201	438/ $\overline{27.3}$	420-j200	465/ $\overline{25.5}$	32.4	29.3	23.5
12000							23.6
15000	328-j231	401/ $\overline{35.1}$	380-j280	472/ $\overline{36.4}$		21.8	

* 120 sections terminated in 450 ohms.

and that of R_2 is small at high frequencies. In this case the two functions may be thought of as virtually independent.

Table II gives the theoretical impedance of this facility and the computed

impedance of the network at frequencies from 100 cycles to 15,000 cycles and, as may be seen by the comparison, a fairly good simulation exists throughout the range. This fact has been verified by making return loss measurements in the laboratory against the theoretical line with the results indicated in the table. Return loss measurements have also been made between the network and an artificial line consisting of 120 sections of this facility terminated in 450 ohms. These results show a fairly constant return loss of about 25 db throughout the frequency range. This seems to indicate that the simulating network is a fairly close approximation to the artificial line so far as frequency is concerned and differs from it by a constant multiplying factor which is of the order of 1.12. It is therefore apparent that whenever it is necessary only to simulate the impedance of this particular facility, this four-element network will provide a fairly adequate simulation. The analytical derivation of this network will be omitted.

EXAMPLE 3A—NON-LOADED EXCHANGE AREA CABLE

Another case will be cited to show the application of the graphical method. This is the simulation of non-loaded cable of which the local plant is largely composed in urban areas. A first approximation of the analytical method does not yield a useful network but the graphical method provides a three-element network of the type discussed above which gives a return of about 20 db in the 300 cycles to 3000 cycles range. The graphical derivation of the three-element network is shown on Fig. 4 which also gives the impedance function for 22 ga. BSA non-loaded cable. This latter function is virtually a straight line in the voice range whereas the network is the arc of a circle. Hence it would be impossible to obtain an appreciably closer approximation throughout the range with a three-element network. However, the addition of elements will improve the match as will be shown in example 3B.

The network just derived can be expressed in terms of the 1000-cycle impedance and applied for any gauge of non-loaded cable as follows:

$$R_1 = .42 K \quad (5-a)$$

$$R_2 = 2.8 K \quad (5-b)$$

$$X_{C_2} = .9 K \quad (5-c)$$

where K is the magnitude of the 1000-cycle impedance and

$$X_{C_2} = \frac{1}{2\pi f C_2} \quad (5-d)$$

Table IIIa gives a comparison of the network and line impedances and the computed return loss for frequencies through the 200 to 3000 cycle range.

EXAMPLE 3B—19-GAUGE QUADDED NON-LOADED TOLL CABLE

Two complete RC functions plus a resistance are required to give a good simulation for non-loaded toll cable when the simulation is carried through the voice and carrier frequency ranges. The impedance function for 19 ga.

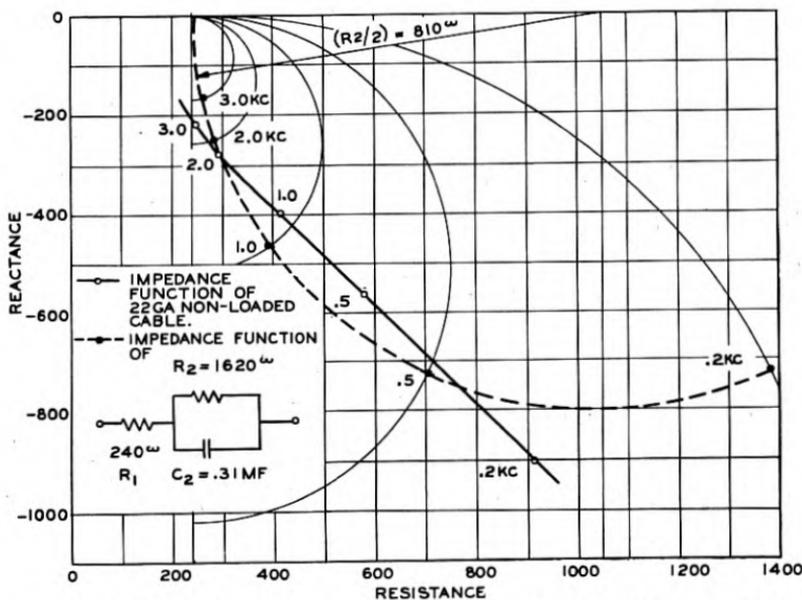


Fig. 4—Graphical design of two-terminal balancing network for 22-ga. non-loaded exchange area cable.

TABLE III-a

22 Gauge Non-Quadded Non-Loaded BSA Exchange Area Cable

Freq. kc	Line Impedance		Network Impedance		Computed Return Loss— db
	Rect.	Polar	Rect.	Polar	
0.2	915—j905	1287/44.7	1380—j725	1555/27.8	15.0
0.5	580—j565	808/44.2	705—j725	1010/45.8	19.1
1.0	415—j400	576/44.0	390—j460	603/49.7	25.2
2.0	295—j280	407/43.5	285—j245	376/40.7	26.6
3.0	250—j220	333/41.3	260—j165	308/32.4	21.2

toll cable is plotted on Fig. 5. The method followed in determining the elements is somewhat as follows: R_1 will be given by the intercept of the function on the R axis and is 130 ohms. Next look at the low-frequency range determined by $R_2 C_2$ and draw a semicircle which approximates the given function in the range of 200–500 cycles. The diameter of this semicircle

determines R_3 as 2100 ohms and R_2 is then automatically determined as the difference between the R -intercept of the R_3 semicircle and R_1 , hence $R_2 = 420 - 130 = 290$ ohms. To determine C_3 , choose the X_{C_3} semicircle at 500 cycles to intersect the R_3 semicircle at a point near the 500-cycle impedance of the cable impedance function, but make some allowance for the added negative reactance of the R_2C_2 generating function. The determination of C_2 can be made in either of two ways. First an X_{C_2} semicircle can be drawn at 5000 cycles which intersects the R_2 semicircle at an impedance near the 5000-cycle impedance of the cable. The impedance at 1000 cycles can then be found graphically for R_2C_2 and R_3C_3 and added together to R_1 . This

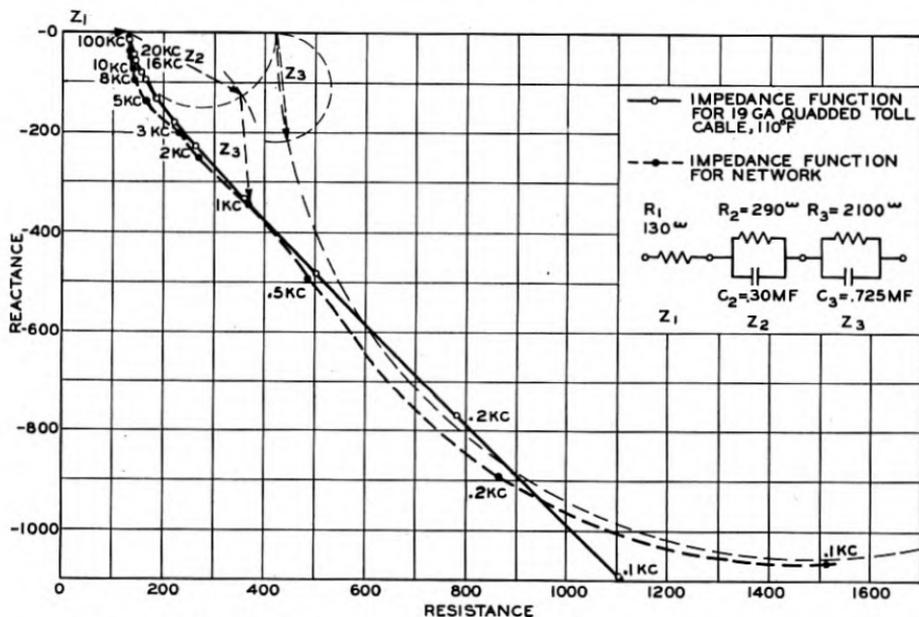


Fig. 5—Graphical design of two-terminal balancing network for 19-ga. quadded non-loaded toll cable.

total impedance at 1000 cycles should provide a good simulation of the 1000-cycle impedance of the cable. A second procedure for finding C_2 would be to follow a somewhat reverse process: Determine the 1000 cycle Z for the R_3C_3 function and subtract it from the 1000 cycle Z of the cable. Choose C_2 such that the intersection of the R_2C_2 semicircles is near the point determined by the subtraction of R_3C_3 from the cable.

To avoid confusion of lines the construction circles have been omitted from this last drawing except to show the addition of the 1000-cycle impedances. As may be seen this network shown in Fig. 5 provides a rather good simulation throughout the frequency range above 200 cycles.

As pointed out earlier, if the first guess is not a sufficiently good approximation a second try can be made based on the evident shortcomings of the first try. In this case if a closer approximation is required up to 20 kc the next step might be to change C_3 to .8 mf which would make the Z_3 contributions above .1 kc somewhat less negative and would therefore raise the network locus. Then changing R_1 to 140 ohms would shift the locus 10 ohms to the right. The resulting locus would be somewhat closer at the upper frequencies but the change would not be necessary unless a rather high degree of balance is required.

TABLE III-b
19 Ga. Quadded Non-Loaded Toll Cable

Freq. kc.	Network Impedance		Line Impedance		Computed Return Loss vs Theoretical Line-db
	Rect.	Polar	Rect.	Polar	
.1	1515 —j1064	1852 /35.9	1103—j1093	1554 /44.7	18.3
.2	867 — j893	1244 /45.8	783— j770	1097 /44.5	24.8
.5	488 — j493	693 /45.3	501— j482	696 /43.9	38.2
1.0	376 — j340	507 /42.1	361— j335	492 /42.8	36.0
2.0	271 — j254	371.6/43.2	265— j229	350 /40.8	29.0
3.0	217 — j203	297.2/43.2	223— j180	287 /38.9	27.8
5.0	166.5— j139	211.3/38.0	187— j131	228 /35.1	26.1
8.0	145.5— j96.8	172.0/32.2	164— j94	189 /29.8	25.8
10.0	140.0— j74.4	158.0/28.0	155— j79.2	174 /27.1	26.5
16.0	134 — j47.2	142.1/19.5	145— j55.0	155 /20.8	27.9
20.0	132 — j37.9	137.9/16.0	141— j45.1	148 /17.7	27.9
100	131 — j7.3	131.0/ 3.2	130— j14.0	131 / 6.2	31.7

In general the success of a trial of the graphical construction may be determined immediately by comparing about three frequencies of the line and network.

Table IIIb gives the computed network impedance and the line impedance. The computed return loss is also given and equals or exceeds 25 db at all frequencies above 200 cycles.

It is apparent that the resistance and condenser elements of the generating functions are in descending order of magnitude with increasing frequency for the non-loaded cable the impedance locus of which is essentially a straight 45° line. As pointed out earlier, the series addition of such generating functions may be converted to a ladder structure⁵, whose sections will have a tapered characteristic rather than repetitive.

⁵ Appendix D of Transmission Circuits by K. S. Johnson.

RETURN LOSS

When designing such networks for balancing purposes, it has been found convenient to plot the function on a sheet such as Fig. 6 which divides the right half of the complex plane into circular regions such that all points on or within the boundary of a given region have a return loss against the network $1 + j0$ equal to or greater than that corresponding to the boundary. These circles are determined by the return loss voltage ratio k and the ratio

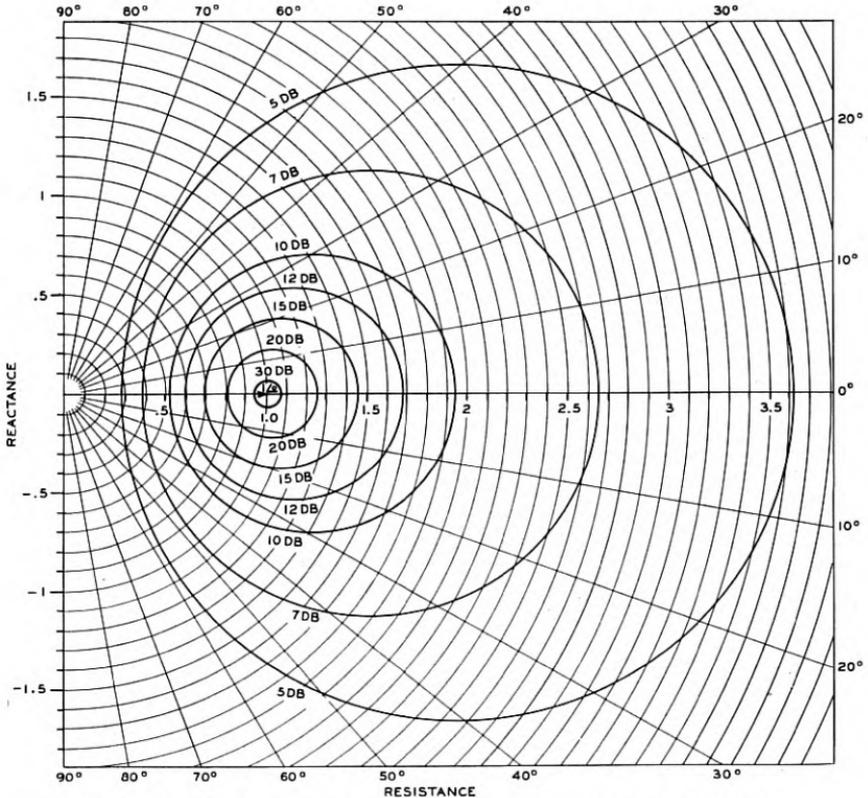


Fig. 6—Curves of constant return loss for the network $1 \angle 0 \equiv 1 + j0$.

of the line and network impedances. They may be computed from the equation

$$(1 + L/N)/(1 - L/N) = k \quad (6)$$

By plotting the line and simulating network loci on such a sheet it is generally possible to observe visually whether or not a given network meets the specified return loss requirement. If visual accuracy is not adequate,

it is always possible to measure off N and L and the angle between them, spot the complex ratio L/N on the complex plane and read immediately the approximate return loss.

CONCLUSION

The examples of the foregoing discussion have been confined to the fourth quadrant. It was shown that by graphical means a number of parallel resistance-condenser functions could be determined which when added together would yield a close approximation to the given function. In the most general case these functions would involve the generating function of R , L and C in parallel, the locus of which is a circle having impedance $+j0$ at zero frequency and $-j0$ at infinite frequency, and crossing the axis of reals at R and the frequency at which L and C are anti-resonant. A case which has been found useful in simulating such things as telephone sets and other inductive elements is the parallel combination of R and L which, of course, is the special case for $C = 0$ and occurs in the first quadrant.

The foregoing has been discussed with the thought that it may be useful where there is limited time and where the required degree of simulation is consistent with a graphical method. At some future time it may be possible to pursue the problem further and devise the analytic counterpart to the somewhat heuristic graphical method.

CHAPTER III

The Use of X-Rays for Determining the Orientation of Quartz Crystals

By W. L. BOND And E. J. ARMSTRONG

THIS paper is one of a series by the Crystal Research Group on the manufacture of quartz oscillator plates. Certain sections of it which are not original, but rather adaptations of text book material to the present problem, are included for purposes of completeness and for the convenience of those readers whose knowledge of the crystallographic literature may be limited.

3.1 PRODUCTION OF X-RAYS FOR QUARTZ CRYSTAL X-RAY WORK

X-rays are produced when electrons strike a metal target at high velocity. The wave-length of X-rays given off from an X-ray tube varies from the longest which can pass through the X-ray tube window to the shortest that can be produced from the given target by the applied peak voltage. By analogy to the visible spectrum this is referred to as "white" radiation. For each different metal, however, there are characteristic radiations of certain wave-lengths whose intensity markedly exceeds those of other wave-lengths (Fig. 3.1). The strongest of these characteristic radiations is known as the $K\alpha_1$, the next strongest (generally half as strong and of slightly longer wave-length) as $K\alpha_2$ and the third strongest (shorter in wave-length than $K\alpha_1$) is $K\beta$. The higher the atomic number of the target, the shorter will be the wave-length of the characteristic radiation. Therefore higher voltages will be required to excite the characteristic radiation from the heavier metals. (The minimum wave-length of X-rays that can be excited by any given voltage is given by the equation $\lambda_{\min.} = \frac{1.234 \times 10^{-4}}{V}$ where V is expressed in volts and $\lambda_{\min.}$ in Ångstrom units).

Higher voltages also raise the intensity of the white radiation and, at any given voltage, the white radiation produced from a heavy metal target is more intense than that produced from a lighter metal target (see Figure 3.1). When "white" radiation is desired, as in Laue photography, heavy metal targets, such as tungsten, are used: when "monochromatic" radiation is desired, as in crystal goniometry, the lighter metal targets, such as copper, are used because, with a lighter metal target (wave-length of characteristic radiation long) the voltage, and therefore the intensity of the

white radiation, cannot be raised very high before exciting the characteristic radiation whereas, with a heavy metal target (wave-length of characteristic radiation short) considerable intensity of white radiation can be produced without exciting the characteristic radiation. This is illustrated in Fig. 3.1 which shows that a potential of 35,000 volts is high enough to excite the K group radiation from molybdenum, but not high enough to excite the shorter wave-length K radiation from tungsten which, further, gives more intense white radiation at this voltage. Even higher voltages, resulting in

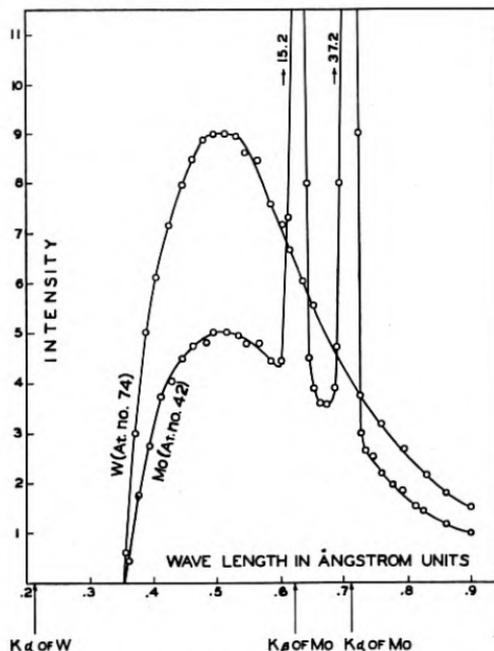


Fig. 3.1—Variation of intensity with wave-length of X-rays from tungsten and molybdenum targets at 35,000 volts

more intense white radiation, could be used with tungsten without exciting the characteristic radiation ($\lambda = .209$).

Figure 3.2 shows the I - λ curve (estimated) for copper, the target metal commonly used in quartz X-ray work, which has a small atomic number and can therefore be used as a source of "monochromatic" X-rays with moderate voltages. (A further advantage of copper for quartz work is pointed out at the end of this section).

The $K\alpha_1$ and $K\alpha_2$ wave-lengths are so close together that for most uses of "monochromatic" radiation no attempt is made to eliminate the $K\alpha_2$ radiation. The $K\beta$ radiation, however, gives a distinct intensity peak of shorter wave-length which must be reduced as much as possible by use of a metal filter having a high absorption coefficient for the $K\beta$ radiation of

the target used. In most cases the best filter-metal for the $K\beta$ radiation is the second element below the target metal in the periodic table. For example, a nickel filter 0.0005 inches thick is used with the copper target which is used for X-ray goniometry of quartz.

The minimum voltage that will excite the $K\alpha_1$ radiation of copper (wavelength $\lambda = 1.5374\text{\AA}$) is 8.86 kv., but a voltage of 30 or 40 kv. is usually used to obtain adequate intensity.

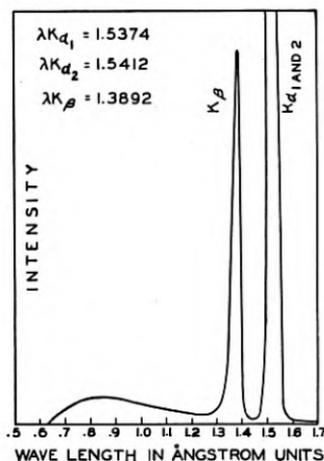


Fig. 3.2—Variation of intensity with wave-length of X-rays from a copper target at 20,000 volts (estimated)

3.2 DETECTION OF X-RAYS

X-rays may be detected by means of heat effects, fluorescence of appropriate screens, photographic effects and by ionization of gases.

Heat measuring devices are not suited to routine intensity measurements. The other three means of X-ray detection are all being used for various types of quartz work but the ionization of gases is used most widely.

The ionization method involves a chamber which consists essentially of a gas-filled metallic cylinder containing an electrode. A potential is maintained between the electrode and the cylinder so that when the gas is ionized by the X-rays the positive and negative ions produced are drawn to the oppositely charged electrodes, thus constituting an electric current, which current is proportional to the intensity of the X-rays entering the chamber. This current is commonly indicated by a current meter preceded by a special vacuum tube amplifier.

The gas used in the chamber is often air but methyl bromide is about 30 times more readily ionized, xenon 155 times. The only advantage of air is that the chamber need not be gas-tight. The entrance to a sealed ionization chamber must be closed by a substance which does not result in undesirable absorption of the X-rays. Since a nickel filter is needed

somewhere in the path of the X-rays to absorb the $K\beta$ radiation from the copper target, it may be used here to seal the entrance of the chamber.

The Geiger-Müller counter is similar to the ionization chamber, but is operated at such high voltage that the gas is always near breakdown, the passing of X-rays supplying the impetus to complete the breakdown. Because of the high voltage the ionization continues to be strong as the intensity of the radiation declines. Such a lag is likely to cause erroneous maximum readings in crystal measurement.

3.3 PHYSIOLOGICAL EFFECTS OF X-RAYS

Even a few minutes direct exposure to X-rays from such a source as the G. E. CA6 tube will result in a burn that will become apparent in a day or so. In most cases such a burn, if not repeated, will heal without ill effects, but because the physiological effect of X-rays is cumulative, repeated exposure to direct radiation could result in a burn that would not heal and might become cancerous.

Pinhole leaks which permit direct radiation to escape are thus exceedingly dangerous. They may be detected by use of a fluorescent screen in a darkened room. A sheet of lead-glass $\frac{1}{4}$ to $\frac{1}{2}$ inch thick should be held between the fluorescent screen and the observer.

X-rays are scattered from all substances which they strike and repeated exposure to this scattered radiation may result in a harmful decrease in the white blood corpuscles, in sterility, and perhaps in serious burns.

To test for scattered radiation dental X-ray films should be arranged as close to the instrument on all sides as any part of the operator's body can get and left for a period of two weeks of normal operation. If there is no position in which the film becomes fogged during the two weeks, the operator is safe. A narrow lead strip across the film will provide an unexposed portion for comparison with the exposed portion. It should be emphasized that a dental film carried in the pocket is an inadequate safety test, since the hands are frequently the most dangerously exposed part of the body.

The absorptive power of shielding materials is proportional to the density of the material. Minimum adequate shielding is provided by 1.5 mm. or about $\frac{1}{16}$ inch of lead or its equivalent for protection against X-rays generated at 70 kv.¹ Equivalent thicknesses of shielding materials are as follows:

Lead.....	$\frac{1}{16}$ inch
Lead rubber.....	$\frac{1}{8}$ — $\frac{1}{4}$ inch
Lead glass.....	$\frac{3}{8}$ — $\frac{1}{2}$ inch
Steel.....	$\frac{1}{2}$ inch
Bricks and concrete.....	6 inches
Woods.....	60 inches

¹ Davey, W. P., "Study of Crystal Structure and its Applications," McGraw-Hill, 1934.

The United States Bureau of Standards has issued a 28-page booklet entitled "X-Ray Protection" which may be obtained from the Superintendent of Documents, Washington, D. C. for 10 cents.

3.4 DIFFRACTION OF X-RAYS BY CRYSTALS

Following the work of the French crystallographer Haüy at the end of the eighteenth century the theory that crystals were made up of small identical units in orderly arrangement was widely held by crystallographers. In the late nineteenth century these units were thought of as intersecting planes of atoms. Then, in 1912, the German physicist von Laue conceived the idea that such a lattice of atoms should act as a three-dimensional diffraction grating for electromagnetic waves of wave-length approximating the interplanar spacing of the lattice. If X-rays were, as was suspected, electromagnetic vibrations of short wave-length, they might be of the right order of magnitude to obtain diffraction from crystals. When the experiment was tried it was found that a beam of X-rays (not monochromatic) passing through a crystal produced an orderly arrangement of spots on a photographic film, the type of photograph now known as a Laue photograph.

The term *reflection* may be used in place of *diffraction* since X-ray diffraction is like light reflection in that the entering and leaving beams make equal angles with the reflecting or diffracting atomic planes.² Since this concept is simpler, X-ray diffraction is commonly referred to as reflection.

Unlike light reflection, X-ray reflection can take place only under the conditions given by the following equation which is known as the Bragg law

$$n\lambda = 2d \sin \theta \quad (3.1)$$

where n = a small whole number,

λ = wave-length of X-rays used (generally stated in Ångstrom units)

d = distance between the atomic planes (generally stated in Ångstrom units).

θ = angle between the X-rays and the atomic planes ("The Bragg Angle").

That is, the angle of incidence must be such that the path-length of two rays reflected from different atomic planes differs by a whole number of wave-lengths so that the emerging rays will be in phase. If the difference in path-length of the two rays is one wave-length the reflection is called the first-order reflection ($n = 1$). At some larger θ angle the path difference will be exactly 2λ and reflection will occur again. This is the second order reflection ($n = 2$). Monochromatic X-rays are used so that only

² Compton, A. H. and Allison, S. K., "X-Rays in Theory and Experiment," D. Van Nostrand, New York, 1935, Pages 340-346,

one value of θ will satisfy the equation for each order reflection. Thus, with a monochromatic X-ray beam one may not obtain reflection from the planes at all angles as with visible light, but only at such specific angles as satisfy the Bragg law. Further, even with optimum conditions for reflection the ratio of the intensity of the reflected X-ray beam to that of the incident beam is of the order of 1:10,000. The Bragg angle θ is highly critical and the reflection of X-rays from atomic planes therefore serves as a precise method of crystal orientation.

Figure 3.3 is a diagrammatic representation of the relation $n\lambda = 2d \sin \theta$. On such a diagram the following laws of X-ray reflection become obvious:

(1) λ must be smaller than $2d$, that is, the wave-length of X-rays used must be less than twice the inter-planar spacing of the atomic planes to be X-rayed.

(2) The number of different orders of reflection n obtainable from atomic planes with interplanar spacing d is fixed by the expression $n\lambda < 2d$. In

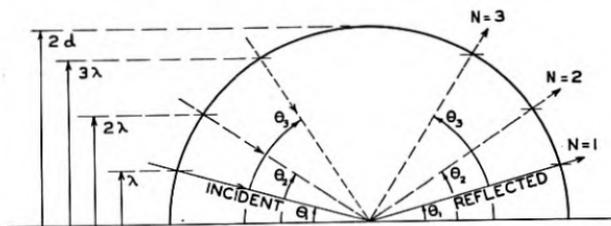


Fig. 3.3—A diagrammatic representation of Bragg's law, $n\lambda = 2d \sin \theta$

other words, since $\sin \theta$ cannot be greater than 1 the value of λ must be less than $\frac{2d}{n}$. The distance between the atomic planes parallel to the hexagonal prism of quartz is 4.2466 \AA and the wave-length of the $K\alpha_1$ radiation from a copper target is 1.5374 \AA . Hence no reflection higher than the 5th order could be obtained from this set of planes using a copper target.

On the other hand, a target metal whose characteristic wave-length is very much shorter than $2d$ is undesirable since it gives so many orders of reflection from each set of atomic planes that the multitude of closely spaced reflections leads to confusion.

(3) Higher orders of reflection occur at larger θ angles.

(4) The relation of θ to λ is not linear but sinusoidal.

The reflected beam can only lie in a plane containing the normal to the atomic plane and the incident beam. Conditions are unchanged by rotating the crystal about the normal to the atomic plane being used.

³ R. B. Sosman, "The Properties of Silica," Chemical Catalogue Co., New York 1927.

3.5 THE NAMING OF ATOMIC PLANES IN CRYSTALS

It is convenient to be able to refer to any atomic plane in a crystal by some symbol that uniquely defines its orientation. The symbols commonly used for this purpose are known as Miller indices (or Bravais-Miller indices for the hexagonal system) and are the reciprocals of the intercepts of the atomic plane on a set of crystallographic axes chosen in accordance with the symmetry of the crystal. In quartz this set of axes is as shown in Fig. 3.4: a vertical axis c and three horizontal axes at 120° a_1 , a_2 , a_3 .

Measurement of the interfacial angles of thousands of quartz crystals has shown that the natural faces have intercepts on the crystallographic

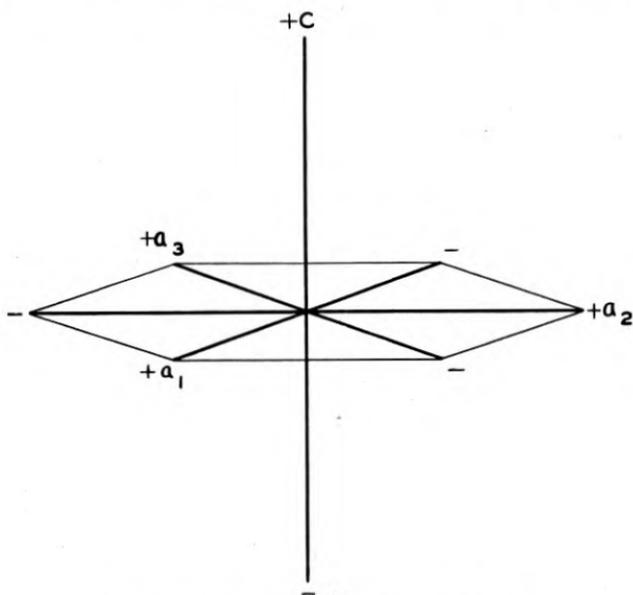


Fig. 3.4—Hexagonal crystallographic axes

axes that are integral multiples of a fixed distance, which is the same in the case of all three a axes and different in the case of the c axis. This fixed distance along the c axis is found to be 1.09997 times the fixed distance along the a axes. Therefore, the "unit length" of each of the a axes is said to be 1; that of the c axis 1.09997 and a face that cuts the c axis at $1.09997a$ from the origin is said to have the c intercept of 1. (This unit axial length is different for different substances but the same for all crystals of the same substance.) For example, the front cap face in figure 3.5 has the axial intercepts $1, \infty, -1, 1$, naming the axes in the order a_1, a_2, a_3, c . The indices for this face are written $(10\bar{1}1)$ (general form $hki\bar{l}$). The front vertical face has the intercepts $1, \infty, -1, \infty$ and the

indices $(10\bar{1}0)$. The first two digits of the indices of any vertical prism plane are the same as those of the adjacent cap faces: the final digit is always zero because they are parallel to the Z (or c) axis. Since the intercepts on a_1 and a_2 uniquely determine the intercept on a_3 , the third digit of the symbol may be omitted. The omission is sometimes indicated by a dot, as $(10\cdot1)$ (general form $hk\cdot\ell$).

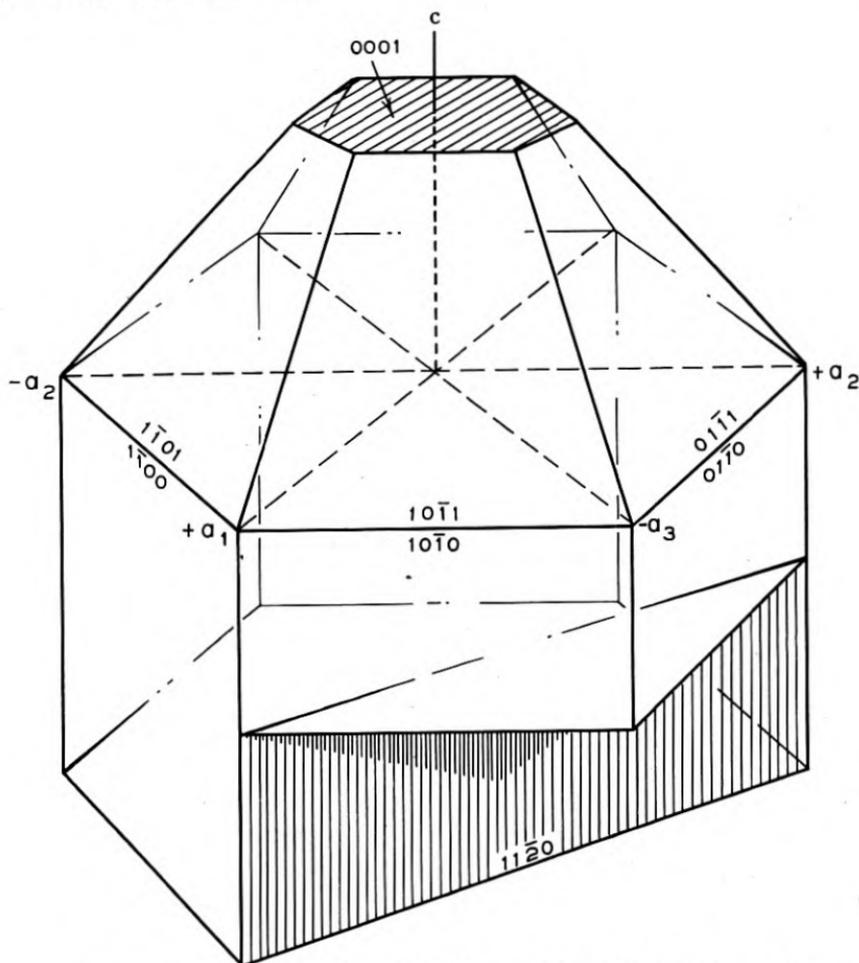


Fig. 3.5—Miller-Bravais indices of certain crystallographic planes in quartz

The use of indices instead of the actual intercepts simplifies the mathematics of crystallography in many ways. For example, the sum of the indices of two planes forms the indices of an intermediate plane that is parallel to the line of intersection of the first two. Thus the sum of $(00\cdot1)$, the atomic plane which is normal to c , and $(10\cdot0)$, prism face, is $(10\cdot1)$, the indices of the cap face directly above the $(10\cdot0)$ prism face.

Further, in the equation $n\lambda = 2d \sin \theta$, the interplanar spacing d of the atomic planes ($hk \cdot \ell$) is

$$d_{hk \cdot \ell} = \frac{a_0}{\sqrt{\frac{4}{3}(h^2 + hk + k^2) + \frac{\ell^2}{(c/a)^2}}} \quad (3.2)$$

Where the axial ratio c/a is 1.09997 in quartz and a_0 , the distance between atoms along the a axis, is 4.903 \AA .⁴

The n order reflection from the $hk \cdot \ell$ plane is sometimes spoken of as the reflection from the $n(hk \cdot \ell)$ plane as in Table I and Fig. 3.8 where, for example, the second order reflection from the $01 \cdot 1$ plane is listed as the reflection from the $02 \cdot 2$ plane. The justification for this notation is that the $02 \cdot 2$ planes would have a d value half that of the $01 \cdot 1$ planes (since the indices are reciprocals of the axial intercepts) and the first order reflection from such planes would have the same θ value as the second order reflections from the $01 \cdot 1$ planes i.e., if

$$2\lambda = 2d \sin \theta$$

then

$$\lambda = 2 \frac{d}{2} \sin \theta$$

Among manufacturers of piezoelectric units a rectangular coordinate system is used in place of the hexagonal one described above, but the indices as derived from the hexagonal axes are retained. One of the a axes is chosen as the X axis. The three a axes are identical (the vertical axis of quartz is an axis of 3-fold symmetry) and therefore any one of the a axes may be chosen as X . The c axis is called the Z axis and a Y axis, normal to X and Z is so chosen as to form a right-handed coordinate system for right-handed quartz or a left-handed coordinate system for left-handed quartz as described in Chapter II.

The indices of the cap faces of right and left-handed quartz crystals (as viewed from above) are given in Fig. 3.6 which should be compared with Fig. 2.4 and 2.6 of Chapter II. (Note that Z in Fig. 3.6 refers to the Z axis whereas z in Fig. 2.6 refers to the z faces or minor cap faces). The cap faces directly beneath those illustrated (i.e., on the other end of a doubly terminated crystal) would have the same indices except that the final digit would be negative since they cut the negative end of the Z (or c) axis. Thus $(0\bar{1} \cdot \bar{1})$ is beneath $(0\bar{1} \cdot 1)$ and parallel to $(01 \cdot 1)$.

Parallel crystal faces lie along the same set of atomic planes. Thus $(01 \cdot 1)$ and $(0\bar{1} \cdot \bar{1})$ represent the same atomic plane. Further, since the

⁴Wyckoff, W. G., "The Structure of Crystals," The Chemical Catalogue Co. (1931).

vertical axis is an axis of 3-fold symmetry, the major cap-face planes $(01 \cdot 1)$, $(\bar{1}0 \cdot 1)$ and $(1\bar{1} \cdot 1)$ have the same properties and this applies also of course to the parallel crystal faces (same atomic planes) which are, respectively, $(0\bar{1} \cdot \bar{1})$, $(10 \cdot \bar{1})$, and $(\bar{1}1 \cdot \bar{1})$. Thus we can choose any one of these symbols to represent this type of plane when speaking of such properties as distance between atomic planes, θ angle, or angle between atomic planes and the Z axis. The symbol that will be used in this paper for this type of plane is $(01 \cdot 1)$.

Similar considerations apply to the set of minor cap faces for which the symbol used here is $(01 \cdot \bar{1})$ which is chosen instead of $(0\bar{1} \cdot 1)$ because it simplifies tabulation, as in Table I and Fig. 3.8.

The indices $(hk \cdot \ell)$ of all faces or atomic planes with the same X-ray properties as those for any given plane may be derived as follows: First replace the omitted digit i in the indices of the given plane, which may be

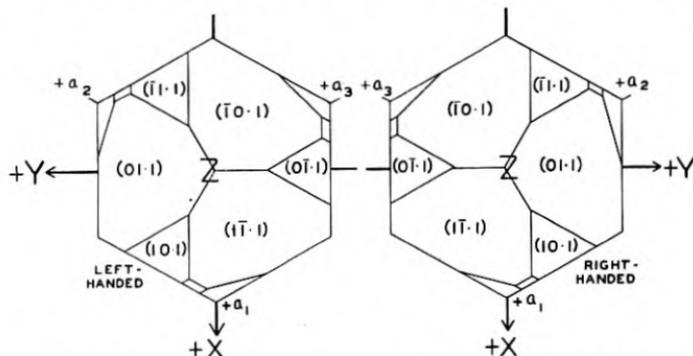


Fig. 3.6—Indices of cap-faces of right- and left-handed quartz crystals

determined from the equation $i = -(h + k)$. Rotary permutation of the first three indices is then permitted, as $abc\bar{\ell}$, $\bar{c}ab\ell$, $b\bar{c}a\ell$.

The other three of the six equivalent faces are found by interchanging any adjacent two of the first three indices and also changing the sign of ℓ , as $abc\bar{\ell}$ becomes $a\bar{c}b\bar{\ell}$ or $ba\bar{c}\bar{\ell}$. For example:

$$\begin{array}{ccc} 1 \ 2 \ \bar{3} \ 4 & \bar{3} \ 1 \ 2 \ 4 & 2 \ \bar{3} \ 1 \ 4 \\ 1 \ \bar{3} \ 2 \ \bar{4} & 2 \ 1 \ \bar{3} \ \bar{4} & \bar{3} \ 2 \ 1 \ \bar{4} \end{array}$$

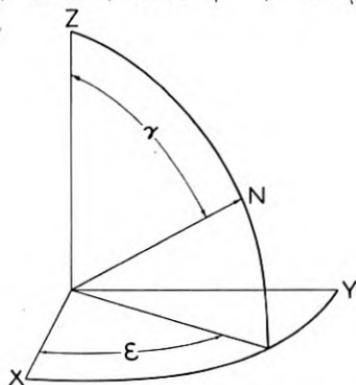
We could consider the $2 \ 1 \ \bar{3} \ \bar{4}$ as derived from cyclic permutation of $1 \ \bar{3} \ 2 \ \bar{4}$, and $\bar{3} \ 2 \ 1 \ \bar{4}$ as similarly derived from $2 \ 1 \ \bar{3} \ \bar{4}$ instead of by interchange of the sets above them.

The six symbols derived from the above by changing the sign of only the final digit refer to faces whose θ X-ray angle is the same as that of the above planes but whose intensity of reflection may be different.

TABLE I
QUARTZ X-RAY REFLECTION ANGLES

The Planes of Quartz from which the $Cu K_{\alpha}$ will reflect
 θ = Bragg angle of X-ray reflection for $Cu K_{\alpha}$ radiation
 γ = Angle that the plane normal N makes with the Z axis
 ϵ = The angle between X and the plane containing the normal and Z , measured in the XY plane
 I = Relative intensity of reflection

$hk \cdot \ell$	θ	γ	ϵ	I for $\ell+$	I for $\ell-$	$hk \cdot \ell$	θ	γ	ϵ
00.3	25°20'	0°	90°	.50	.50	11.0	18°17'	90°	60°
00.6	58°49'	0°	"	.38	.38	11.1	20° 9'	65°33'	60°
01.0	10°26'	90°	"	24.	24.	11.2	25° 5'	47°44'	60°
01.1	13°20'	51°47'	"	100.	75.	11.3	32° 3'	36°15'	60°
01.2	19°45'	32°25'	"	28.	85.	11.4	40°37'	28°49'	60°
01.3	27°41'	22°57'	"	45.	2.7	11.5	51°10'	23°45'	60°
01.4	36°45'	17°37'	"			11.6	65°41'	20° 8'	60°
01.5	47°21'	14°15'	"			12.0	28°38'	90°	70°54'
01.6	60°59'	11°57'	"			12.1	29°59'	73°26'	70°54'
02.0	21°14'	90°	"	28.	28.	12.2	33°53'	59°14'	70°54'
02.1	22°55'	68°31'	"			12.3	39°58'	48°15'	70°54'
02.2	27°27'	51°47'	"	26.	26.	12.4	46°44'	40° 2'	70°54'
02.3	34° 5'	40°16'	"	32.	63.	12.5	58°59'	33°54'	70°54'
02.4	42°30'	32°25'	"			12.6	78°41'	29°15'	70°54'
02.5	53° 6'	26°56'	"			13.0	40°46'	90°	76° 6'
02.6	68°17'	22°57'	"			13.1	41°56'	77°41'	76° 6'
03.0	32°54'	90°	"			13.2	45°35'	66°25'	76° 6'
03.1	34°10'	75°18'	"	40.	1.4	13.3	51°20'	56°46'	76° 6'
03.2	37°51'	62°18'	"	19.	27.	13.4	60° 7'	48°52'	76° 6'
03.3	43°45'	51°47'	"			13.5	75°11'	42°29'	76° 6'
03.4	51°58'	43°37'	"	6.3	3.2	14.0	56° 6'	90°	79° 6'
03.5	63°41'	37°19'	"			14.1	57°22'	80°15'	79° 6'
04.0	46°25'	90°	"			14.2	61°21'	71° 2'	79° 6'
04.1	47°35'	78°52'	"			14.3	69° 1'	62°44'	79° 6'
04.2	51° 8'	68°31'	"			22.0	38°51'	90°	60°
04.3	57°17'	59°26'	"			22.1	40° 3'	77°12'	60°
04.4	67°13'	51°47'	"	12.	0.7	22.2	43°34'	65°34'	60°
05.0	64°54'	90°	"			22.3	47°54'	55°43'	60°
05.1	66°27'	81° 3'	"			22.4	57°59'	47°44'	60°
05.2	71°42'	72°31'	"	.25	23.	22.5	71°45'	41°21'	60°
						23.0	52° 8'	90°	66°35'
						23.1	53°20'	76°40'	66°35'
						23.2	57° 4'	70° 8'	66°35'
						23.3	63°53'	61°33'	66°35'
						23.4	76°52'	54° 9'	66°35'
						24.0	73°24'	90°	70°54'
						24.1	75°40'	81°32'	70°54'
						24.2	88°54'	73°26'	70°54'
						33.0	70°15'	90°	60°
						33.1	72° 8'	81°23'	60°
						33.2	79°32'	73° 9'	60°



The axes (and therefore the indices) given for left-handed quartz are the mirror image of those for right-handed quartz so that similar faces have the same indices in both and the signs of the crystallographic axes are consistent with those of the electrical and physical properties in both. (Crystallographers commonly use the same set of axes for both right and left-handed quartz.) Table I gives the orientation and X-ray properties

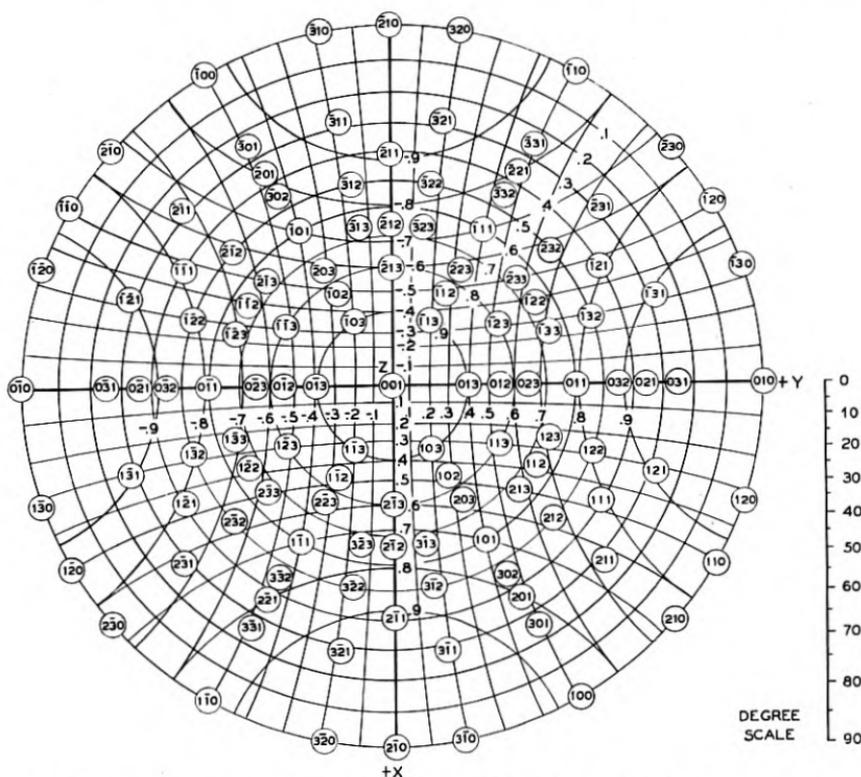


Fig. 3.7—Stereographic projection of normals to the atomic planes in quartz

of the atomic planes in quartz that will reflect X-rays with appreciable intensity.

The θ , γ , and ϵ values have been calculated from the dimensions of the quartz crystal lattice as determined by X-ray investigation. The intensity values were determined experimentally and are only approximate. Figure 3.8 shows the geometrical disposition around the X axis of those planes that are parallel to the X axis (column 1, Table I) together with the X-ray data for these planes and the disposition around the X axis of the common single-rotation cuts. Figure 3.7 is a stereographic projection of the normals of

the more important atomic planes of quartz, identified by their indices. Its usefulness in quartz work will be pointed out in Section 3.7B.

In many cases the reflecting power of an atomic plane differs from that of the symmetrical plane on the other side of the Z axis. (See, for example $01 \cdot 1$ and $01 \cdot \bar{1}$). When this difference is very great as with $04 \cdot 4$ and $04 \cdot \bar{4}$ the planes are useful in determining whether a plate is cut at a positive or negative angle from the Z axis. The intensities of planes that have been found useful for this purpose are circled in Fig. 3.8.

3.6 X-RAY GONIOMETRY

Since the angles θ and the intensities I are different for different planes we can use them to identify these planes, that is, to orient the crystal by measuring angles between recognized atomic planes and plate surfaces.

Figure 3.9 is a diagrammatic representation of an X-ray goniometer where

T is the tube target shown with its intensity pattern,

SS are slits that pass only a narrow beam,

C is the crystal,

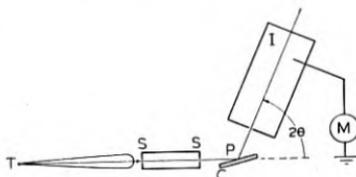


Fig. 3.9—Simplified diagram of X-ray goniometry

I is the ionization chamber

M is the meter that measures the ionization current.

The ionization chamber is placed at an angle 2θ to the incident beam, where θ is the Bragg angle for the atomic plane being used, and is not moved while reflections are being taken from that atomic plane. If C is then rocked about the vertical axis P (normal to the plane of the paper) the ionization chamber registers an electric current when an atomic plane is at the proper angle for reflection.

(a) Atomic plane parallel to plate-face.

Let us examine a simple case, that for which the existing face is parallel to an atomic plane (Fig. 3.10). The crystal is held against the reference points by a coil spring. The crystal holder is free to rotate about the vertical axis P (with respect to the X-rays) and the angle of rotation is read on the graduated scale. If the entering angle (the angle between the entering beam and the plate-face) is one that satisfies the equation $n\lambda = 2d \sin \theta$, we will have a reflected ray which is at a leaving angle of θ . Also the reflected ray is always deviated from the line of the original

ray by the angle 2θ . If, when the crystal gives a reflected ray, it is rotated in its own plane on the reference points the reflection is unchanged.

(b) Atomic plane intersecting plate-face in a line normal to the plane of the instrument.⁵

If in the above case the rotation of the crystal in its own plane does cause a change of ionization current the surface is not parallel to the atomic plane. Figure 3.11 illustrates a case in which the surface is at an angle δ to an atomic plane, this intersection lying normal to the plane of the paper in Fig. 3.11. In order that the angle between the X-ray beam and the atomic plane shall be θ , the angle between the beam and the plate-face must now be $\theta - \delta$ on one side and $\theta + \delta$ on the other. The deviation angle 2θ is the sum of these two. Therefore for atomic planes intersecting

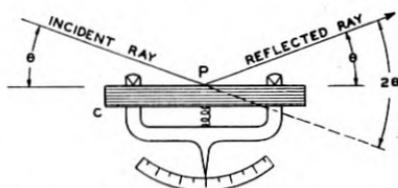


Fig. 3.10—Goniometry case (a): Atomic plane parallel to plate-face

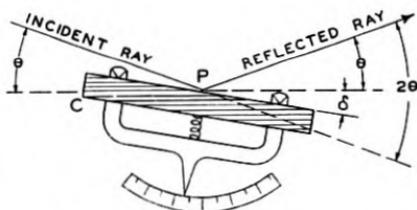


Fig. 3.11—Goniometry case (b): Atomic plane intersecting plate-face in a line normal to "the plane of the instrument"

the plate-face in a line normal to the plane of the instrument the ionization chamber is set at 2θ regardless of the angle between the atomic plane and the surface of the plate.

In Fig. 3.11 the entering angle is shown as $\theta - \delta$, the leaving angle as $\theta + \delta$. If the crystal is now rotated through an angle of 180° in its own plane the entering angle will have to be set at $\theta + \delta$ and the leaving angle will become $\theta - \delta$. The crystal holder will have been moved through an angle 2δ . Thus δ can be found by observing the angle through which the crystal holder must be rotated in order to achieve X-ray reflection from the same atomic plane when the crystal is rotated 180° in its own plane.

⁵ The *plane of the instrument* is a plane normal to the axis of rotation of the instrument and containing the incident ray.

It should be noted that δ should not exceed θ . If δ exceeds θ , it is necessary for the X-ray beam to pass through a great thickness of quartz which would so weaken the beam that the reflected rays could not be detected. Further, even when δ does not exceed θ there is still a variation of the reflected ray intensity with the angle between the entering ray and the plate-face, which may be expressed as follows:⁶

$$I_r = I_R \frac{\sin g_\ell}{\sin g_\ell + \sin g_e}$$

where

I_r = The intensity of the reflected ray

g_e = The "entering angle" = The angle between the entering ray and the plate face

g_ℓ = The "leaving angle" (In Fig. 3.11 $g_e = \theta - \delta$; $g_\ell = \theta + \delta = 2\theta - g_e$)

I_R = The maximum obtainable reflected intensity (i.e., when $g_e = 0$ and $g_\ell = 2\theta$).

or

$$I_r = I_R \frac{\sin (2\theta - g_e)}{\sin (2\theta - g_e) + \sin g_e}$$

The curves in Fig. 3.12 show the variation of I_r/I_R with $g_e/2\theta$ for various values of 2θ . It is seen that for plates as described above and figured in 3.11 a stronger reflection is obtained when $g_e = \theta - \delta$ than when $g_e = \theta + \delta$ (i.e., after 180° rotation in the plane of the plate-face). In other words the smaller g_e angle gives the larger intensity of reflection. It might be added, however, that in practice it may not be possible to obtain the maximum values of reflection intensity due to the fact that, as g_e approaches zero, the width of the reflected beam may exceed the width of the ionization chamber. (See inset, Fig. 3.12.)

Quartz plates containing an X axis are checked as described above (case *b*) for rotation around the X axis. For this check they are oriented as shown in Fig. 3.13, case *b* with the X axis parallel to the axis of the instrument. The stippled face in Fig. 3.13 represents the so-called "reference bevel", the remnant of the surface of the Z section, and is therefore normal to the Z axis.

(c) Atomic plane intersecting plate-face in a line parallel to the plane of the instrument.

The case described above and illustrated in Fig. 3.11 was the case of single rotation plates (AT, BT and others shown in Fig. 3.8) being corrected about the X axis by the use of atomic planes parallel to the X axis. When

⁶ Debye, P. and Menke, H., "Untersuchung der Molekularen Ordnung in Flüssigkeiten mit Röntgenstrahlung," *Ergeb. d. Techn. Röntgenkunde*, B.2, P. 16; Leipzig (1931).

such a plate is being rotated about the X axis as in Fig. 3.11 the normal to the atomic plane lies in the plane of the instrument. We can also use

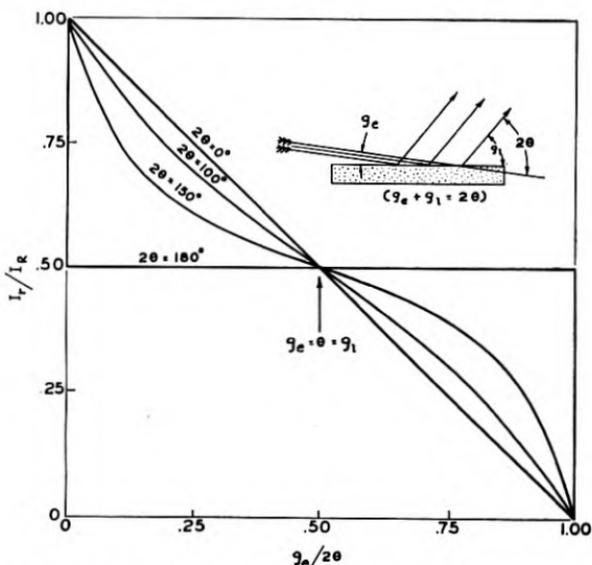


Fig. 3.12—Variation of intensity of reflected ray with entering angle g_e and the Bragg angle, θ

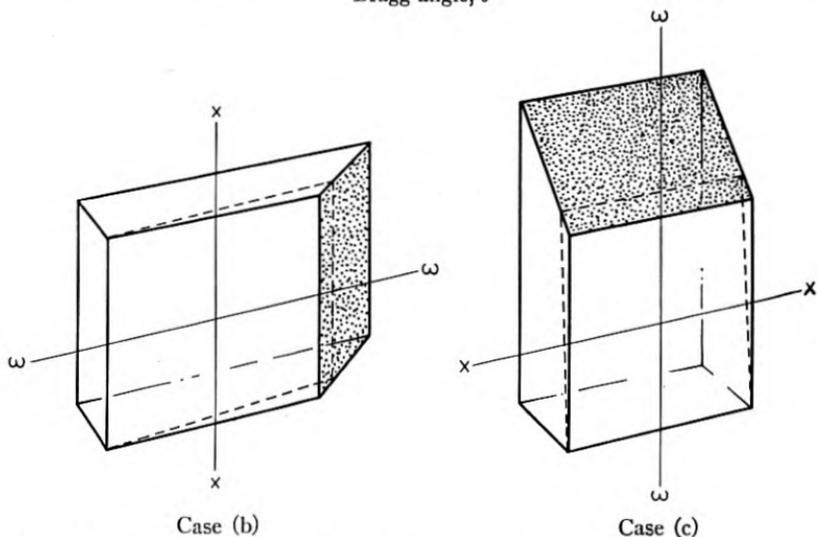


Fig. 3.13—Position of quartz plate relative to vertical axis of instrument in case (b) and case (c). (Stippled face is reference bevel, normal to Z axis.) In the absence of a reference bevel the X direction can be determined with polarized light. (See Chapter II, p. 246.)

the same atomic plane to correct the orientation of this plate about the axis $\omega\omega$ normal to the X axis (Fig. 3.13 and Fig. 3.14). Here, since the

atomic planes are at an angle δ to the axis of the instrument, the reflected ray will not lie in the plane of the instrument. If, however, the vertical angle β , between the reflected ray and the plane of the instrument is small, the reflected ray may still enter the ionization chamber. The maximum angle β which is permissible, is not the same for all instruments, but depends on the vertical length of the slit. In most cases it is about 5° .

The proper angular settings of the instrument are no longer θ for the slab and 2θ for the ionization chamber but are the orthogonal projection of these angles onto the plane of the instrument; g and $g + g'$, respectively.

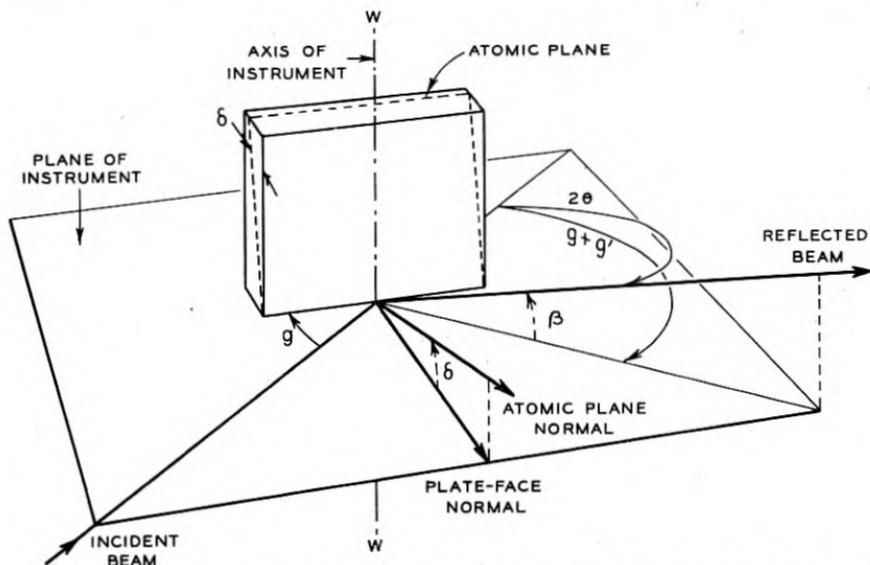


Fig. 3.14—Goniometry case (c): Atomic plane intersecting plate-face in a line parallel to "the plane of the instrument"

Formulae are given below for β , g , and $g + g'$ in terms of the Bragg angle θ and the angle between the atomic planes and the surface of the quartz plate, δ .

$$\sin \beta = 2 \sin \theta \sin \delta \quad (3.3)$$

$$\cos (g + g') = \frac{\cos 2\theta}{\cos \beta} \quad (3.4)$$

$$\sin g = \frac{\sin \theta}{\cos \delta} \quad (3.5)$$

The relations between β , $g + g'$, g , θ , and δ are shown by the curves in Fig. 3.15.

As an example, suppose that a BT plate is to be corrected about the

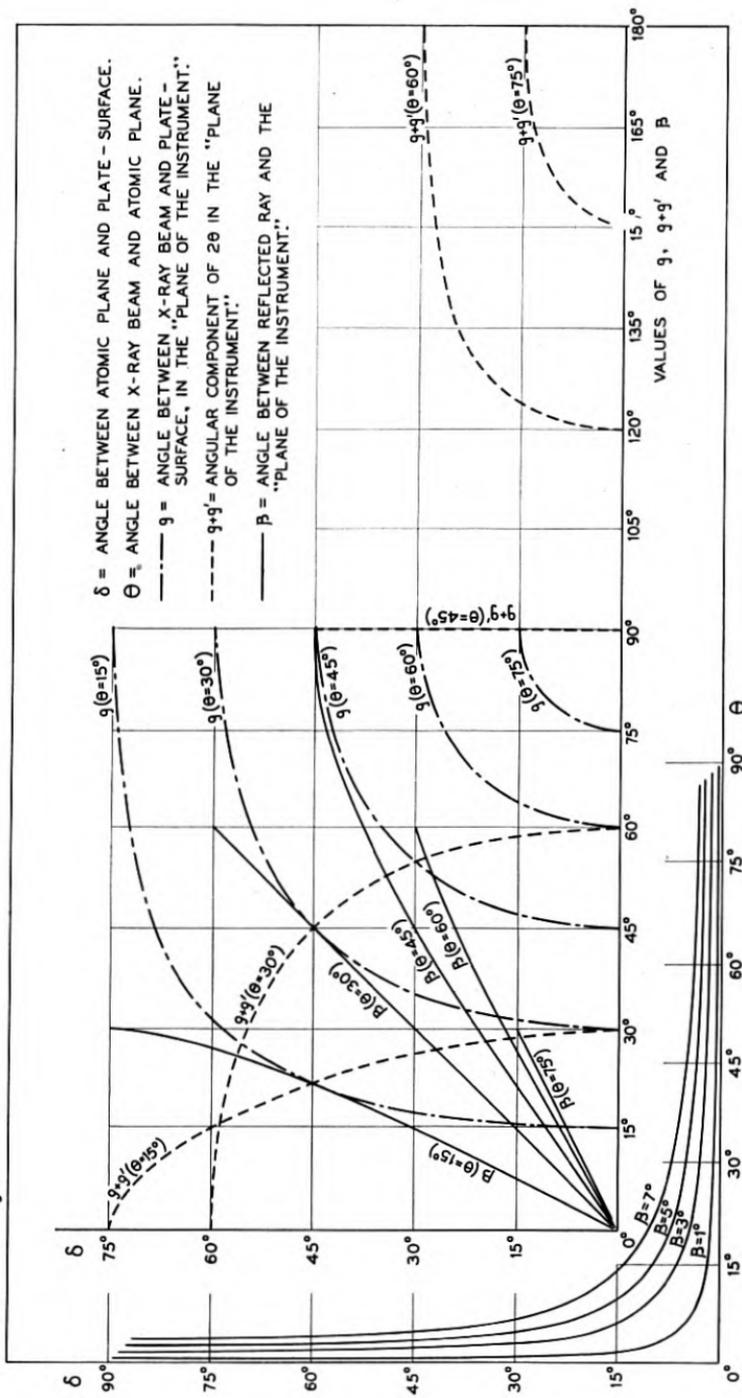


Fig. 3.15—Variation of β , g and $g + g'$ with δ and θ

ww axis using the atomic plane (01·1). (In actual practice the closer plane (02·3) is used.) Here:

$$\delta = 49^\circ - 38^\circ 13' = 10^\circ 47', \quad \theta = 13^\circ 20', \quad 2\theta = 26^\circ 40'$$

$$\text{From which} \quad \beta = 4^\circ 57', \quad g = 13^\circ 35', \quad g + g' = 26^\circ 34'$$

Thus if the plate is correctly cut the reflected ray will enter an ionization chamber whose slit is long enough to receive rays which make an angle of 5° with the plane of the instrument. However, if the plate were in error by more than $6'$ around the XX axis (Fig. 3.13) β would exceed 5° and the center of the reflected beam would not enter the ionization chamber. In actual practice, therefore, the length of the ionization chamber slit should be enough greater than that required for the calculated β to admit reflections from erroneously cut plates.

Since the difference between $g + g'$ and 2θ is only $6'$ and since the width of the ionization chamber slit is usually great enough to accept reflected beams over a range of several times $6'$, no correction of the ionization chamber position may be necessary in this case.

On the other hand, the orientation of the plate with respect to the incident beam is highly critical and since g differs from θ by $15'$ this correction in the orientation of the quartz plate must be made.

Discussion of the general case (d), in which the intersection of the atomic plane and the plate face is neither normal nor parallel to the plane of the instrument, will be found in Section 3.9.

3.7 CHOICE OF AN ATOMIC PLANE FOR CHECKING THE ORIENTATION OF ANY GIVEN FACE

If the plate-face to be checked does not lie parallel to an atomic plane, the nearest usable atomic plane must be found and the orientation of this atomic plane in the plate must be determined. The procedure for these two steps is outlined in this and the following sections.

The problem of the choice of an atomic plane for checking the orientation of one of the faces of a given plate has two parts:

(A) Determination of the orientation of that face with respect to the X , Y and Z axes of the mother crystal and (B) discovery of the atomic plane whose orientation and X-ray properties are most suitable for use with that face.

(A) Determination of the orientation of the face with respect to the X , Y and Z axes of the mother crystal.

The orientation of the plate is commonly given in terms of the A_1 , A_2 and A_3 angular rotations as described and illustrated in Section 2.4 of Chapter II. Briefly, a basal section of crystal is placed initially with its Z axis vertical and its $+X$ axis toward the operator of a horizontal axis saw,

the saw blade being parallel to X and Z . The crystal is then turned through angle A_1 clockwise as seen from above about a vertical axis, then through angle A_2 counterclockwise about the original direction of the X axis. A slab is cut of thickness t and this slab layed down by rotating it 90° clockwise about the original X axis direction. It is then rotated through angle A_3 clockwise about a vertical axis and two cuts are made, separated by a width w (length ℓ being cut last). The rotation through the A_3 angle begins with the linear edge of the reference level of the slab lying parallel to the saw.

The components of the plate edges P_1, P_2, P_3 (length, thickness, and width, respectively) on the $X, Y,$ and Z coordinates after rotation through the angle A_1 are given by the following matrix (See Section 5 of "The Mathematics of the Physical Properties of Crystals" by W. L. Bond, *Bell System Technical Journal*, Volume XXII, No. 1):

$$r' = \begin{matrix} X \text{ component} & Y \text{ component} & Z \text{ component} \\ \left(\begin{array}{ccc} \cos A_1 & \sin A_1 & 0 \\ -\sin A_1 & \cos A_1 & 0 \\ 0 & 0 & 1 \end{array} \right) & \begin{array}{l} \text{of } P_1 \\ \text{of } P_2 \dots \dots \\ \text{of } P_3 \end{array} \end{matrix} \quad (3.6)$$

The components of the plate edges on the X, Y, Z coordinates after rotation through angles A_1 and A_2 are:

$$r'' = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos A_2 & -\sin A_2 \\ 0 & \sin A_2 & \cos A_2 \end{pmatrix} \begin{pmatrix} \cos A_1 & \sin A_1 & 0 \\ -\sin A_1 & \cos A_1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$= \begin{matrix} X \text{ component} & Y \text{ component} & Z \text{ component} \\ \left(\begin{array}{ccc} \cos A_1 & \sin A_1 & 0 \\ -\sin A_1 \cos A_2 & \cos A_1 \cos A_2 & -\sin A_2 \\ -\sin A_1 \sin A_2 & \cos A_1 \sin A_2 & \cos A_2 \end{array} \right) & \begin{array}{l} \text{of } P_1 \\ \text{of } P_2 \dots \\ \text{of } P_3 \end{array} \end{matrix} \quad (3.7)$$

The components of the plate edges on the X, Y, Z coordinates after rotation through angles A_1, A_2 and A_3 are:

$$r''' = \begin{pmatrix} \cos A_3 & 0 & \sin A_3 \\ 0 & 1 & 0 \\ -\sin A_3 & 0 & \cos A_3 \end{pmatrix} \begin{pmatrix} \cos A_1 & \sin A_1 & 0 \\ -\sin A_1 \cos A_2 & \cos A_1 \cos A_2 & -\sin A_2 \\ -\sin A_1 \sin A_2 & \cos A_1 \sin A_2 & \cos A_2 \end{pmatrix}$$

$$= \begin{matrix} X \text{ component} & Y \text{ component} & Z \text{ component} \\ \left(\begin{array}{ccc} \cos A_1 \cos A_3 & \sin A_1 \cos A_3 & \cos A_2 \sin A_3 \\ -\sin A_1 \sin A_2 \sin A_3 + \cos A_1 \sin A_2 \sin A_3 \\ -\sin A_1 \cos A_2 & \cos A_1 \cos A_2 & -\sin A_2 \\ -\cos A_1 \sin A_3 & -\sin A_1 \sin A_3 & \cos A_2 \cos A_3 \\ -\sin A_1 \sin A_2 \cos A_3 + \cos A_1 \sin A_2 \cos A_3 \end{array} \right) & \begin{array}{l} \text{of } P_1 \\ \text{of } P_2 \\ \text{of } P_3 \\ \dots \end{array} \end{matrix} \quad (3.8)$$

Since the edge P_2 (thickness direction) is the normal to the major face of the plate, its X , Y , Z components (second horizontal row) give the orientation of the major face of the plate in terms of X , Y , Z axes of the mother crystal.

For example, suppose the orientation of the major face of an NT plate is desired in terms of the X , Y , Z axes. The shop rotation angles for an NT plate are

$$\begin{aligned} A_1 &= 99^\circ 25' \\ A_2 &= 49^\circ 25' \\ A_3 &= -12^\circ 20' \end{aligned}$$

Substituting these values in the above matrix gives

$$r''' = \begin{pmatrix} 0 & .99027 & -.13917 \\ -.64279 & -.10662 & -.75852 \\ -.76604 & .08946 & .63653 \end{pmatrix} \dots \dots \dots (3.9)$$

the second row of which gives the components (direction cosines) of the unit normal to the plate surface on the X , Y , Z axes:

$$\begin{aligned} X &= .64279 \\ Y &= .10662 \\ Z &= .75852 \end{aligned}$$

(B) Discovery of the plane whose orientation and X-ray properties are most suitable for use with the face to be checked.

The first requisite for the atomic plane to be used is that it shall make as small an angle δ as possible with the face to be checked. This is desirable because (1) δ must be smaller than θ ; (2) if δ is small the same plane can be used for correction around P_1 as around P_3 since the β angle will be small. If δ is very small the difference between g and θ may be within the required limits of error so that the θ value may be used without correction; (3) if δ is large, the difference between g_e and g_ℓ will be large and if the plate is placed in the goniometer in such a way that g_e is the larger angle the reflected ray may be too weak to register on the ionization chamber meter.

By plotting the stereographic projection of the normal to the plate face on Fig. 3.7 the most promising planes may be found quickly. On this chart each concentric circle connects all points with the same Z direction-cosine; each arc that crosses the X axis connects all points with the same X direction-cosine; each arc that crosses the Y axis connects all points with the same Y direction-cosine.

For example, the normal to the major face of the NT cut is plotted as follows:

.64 units are marked off in a positive direction (down) on the X axis (to

a point between $(2\bar{1}\cdot3)$ and $(2\bar{1}\cdot2)$. Then, in the direction of the east-west arcs, .1 unit is marked off to the right of the X axis. The point so located is found to be about half way between the .7 and .8 concentric circles, that is, to have a Z component of about .75, as calculated above. It lies about half way between atomic plane normals $(2\bar{1}\cdot3)$ and $(3\bar{1}\cdot3)$.

The degree scale given in Fig. 3.7 may be used to determine roughly the angle between any two points on the diagram. The scale is non-uniform and for any particular region of the diagram that portion of the scale should be used that falls on that region when the zero mark on the scale is placed at the center of the diagram.

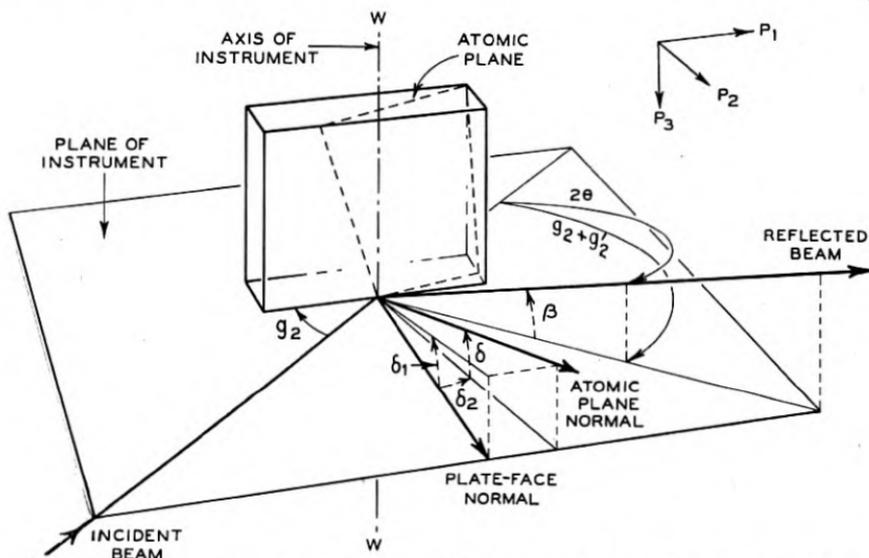


Fig. 3.16—Goniometry case (d): Atomic plane intersecting plate-face in a line which is neither normal nor parallel to "the plane of the instrument" (general case).

In the present example the part of the scale to be used is around the 40° mark and the plate face is found to make angles of less than 10° with both $(2\bar{1}\cdot3)$ and $(3\bar{1}\cdot3)$.

In general the atomic planes with the smaller indices are likely to give the stronger reflections. Therefore the better plane to try first is $(2\bar{1}\cdot3)$. The indices of this plane are not shown in this form in Table I, but by the method described in Section 3.5 it may be seen that a plane with indices $11\cdot3$ would have the same θ value for X-ray reflection. From Table I we see that θ for atomic plane $11\cdot3$ is $32^\circ 03'$.

In order to determine the angles g and $g + g'$ (Fig. 3.16) the orientation of the atomic plane with respect to the plate edges must be found.

3.8 DETERMINATION OF THE ORIENTATION OF AN ATOMIC PLANE WITH RESPECT TO THE PLATE-EDGES, GIVEN ITS MILLER-BRAVAIS INDICES

Since the indices $(hk \cdot \ell)$ are the reciprocals of the intercepts of the atomic plane on the quartz crystallographic axes (See Section 3.5) of which the vertical axial unit is 1.09997 times that of the others, the intercepts on axes which were all divided into the same length unit would be

$$\frac{1}{h}, \quad \frac{1}{k}, \quad \text{and} \quad \frac{1.1}{\ell}$$

(If 1.1 is substituted for 1.09997 the error introduced is never greater than one minute.)

The intercepts on the orthogonal axes may be derived as indicated in Fig. 3.17. They are:

$$\frac{1}{h}, \quad \frac{\sqrt{3}}{h + 2k}, \quad \frac{1.1}{\ell}$$

These intercepts may be taken as the lengths of 3 vectors whose components on the X , Y and Z axes are as follows:

$$i_1(X \text{ axis vector}) = \begin{pmatrix} 1/h \\ 0 \\ 0 \end{pmatrix} \dots \dots \dots (3.10)$$

$$i_2(Y \text{ axis vector}) = \begin{pmatrix} 0 \\ \sqrt{3} \\ h + 2k \\ 0 \end{pmatrix} \dots \dots \dots (3.11)$$

$$i_3(Z \text{ axis vector}) = \begin{pmatrix} 0 \\ 0 \\ 1.1/\ell \end{pmatrix} \dots \dots \dots (3.12)$$

Now the two vectors $(i_3 - i_1)$ and $(i_3 - i_2)$ lie in the plane $(hk \cdot \ell)$ (See Fig. 3.18), and therefore their vector product is the vector normal to the plane $hk \cdot \ell$

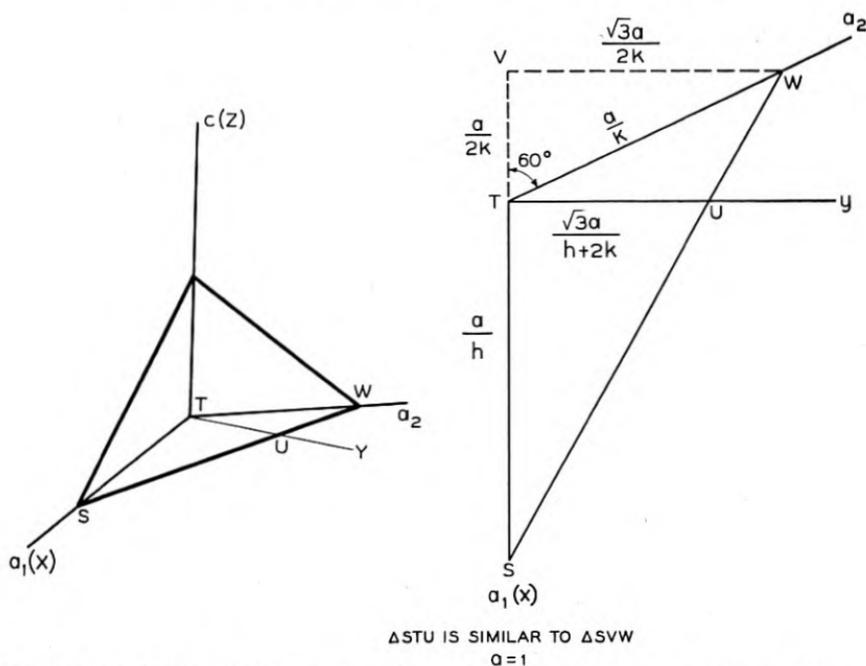
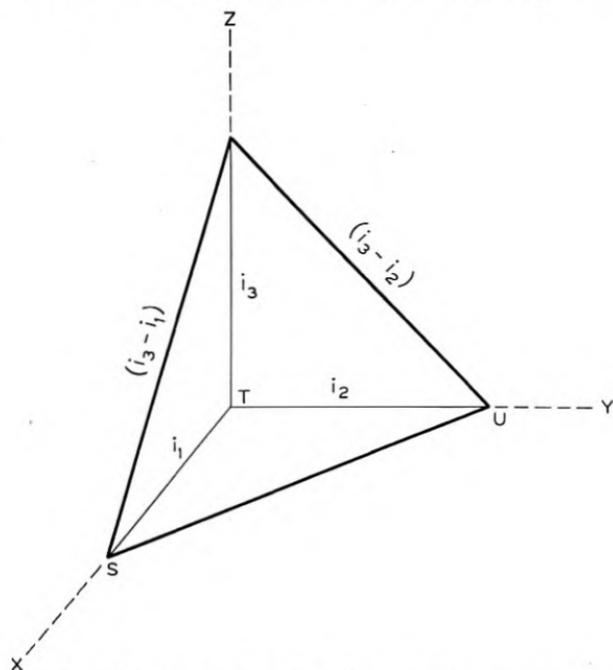


Fig. 3.17—Derivation of intercepts on orthogonal axes from intercepts on hexagonal axes

Fig. 3.18—Orientation of vectors (i_3-i_1) and (i_3-i_2) with respect to orthogonal axes

That is,

\bar{N} (the vector normal to the atomic plane $hk \cdot \ell$)

$$\begin{aligned}
 &= \left[\begin{pmatrix} 0 \\ 0 \\ 1.1/\ell \end{pmatrix} - \begin{pmatrix} 1/h \\ 0 \\ 0 \end{pmatrix} \right] \times \left[\begin{pmatrix} 0 \\ 0 \\ 1.1/\ell \end{pmatrix} - \begin{pmatrix} 0 \\ \sqrt{3} \\ h + 2k \\ 0 \end{pmatrix} \right] \\
 &= \begin{pmatrix} -1/h \\ 0 \\ 1.1/\ell \end{pmatrix} \times \begin{pmatrix} 0 \\ -\sqrt{3} \\ h + 2k \\ 1.1/\ell \end{pmatrix} \\
 &= \begin{pmatrix} 0 & -1.1/\ell & 0 \\ 1.1/\ell & 0 & 1/h \\ 0 & -1/h & 0 \end{pmatrix} \begin{pmatrix} 0 \\ -\sqrt{3} \\ h + 2k \\ 1.1/\ell \end{pmatrix} \\
 &= \begin{bmatrix} 1.1\sqrt{3} \\ \ell(h + 2k) \\ 1.1/h\ell \\ \sqrt{3} \\ h(h + 2k) \end{bmatrix} = \begin{pmatrix} h \\ h + 2k \\ \sqrt{3} \\ \ell/1.1 \end{pmatrix} \dots\dots\dots (3.13)
 \end{aligned}$$

The unit vector $N = \frac{1}{\sqrt{S}} \bar{N} \dots\dots\dots (3.14)$

where S is the sum of the squares of the three terms of \bar{N} . For example, for the (01·1) plane ($h = 0, k = 1, \ell = 1$), we find from equation (3.13) that

$$\bar{N} = \begin{pmatrix} 0 \\ 2/\sqrt{3} \\ 1/1.1 \end{pmatrix} \dots\dots\dots (3.15)$$

whence $S = 4/3 + \frac{1}{1.21} = 2.159 \dots\dots\dots (3.16)$

and the unit vector normal $N = \begin{pmatrix} 0 \\ .7857 \\ .6186 \end{pmatrix} \dots\dots\dots (3.17)$

which means that N is perpendicular to the X axis, makes an angle with Y whose cosine is .7857 ($=38^\circ 13'$) and an angle with Z whose cosine is .6186 ($=51^\circ 47'$).

For the atomic plane ($2\bar{1}\cdot3$) we find from equation (3.13) that

$$\bar{N} = \begin{pmatrix} 2 \\ 0 \\ 3/1.1 \end{pmatrix} \dots \dots \dots (3.18)$$

$$\text{Whence } S = 4 + 9/1.21 = 11.438 \dots \dots \dots (3.19)$$

and, from equation (3.14) the unit vector normal to ($2\bar{1}\cdot3$)

$$N = \begin{pmatrix} .5914 \\ 0 \\ .8064 \end{pmatrix} \dots \dots \dots (3.20)$$

We have now determined the orientation of the atomic plane with respect to the orthogonal axes X , Y and Z . To determine its orientation with respect to the plate edges necessitates the construction of a matrix which expresses the components of the plate edges $P_1P_2P_3$ (length, thickness, and width respectively) in terms of the X , Y , Z axes. When the components of the unit normal to the atomic plane in terms of the orthogonal axes X , Y , Z are acted upon by this matrix they are converted to the components of the unit normal to the atomic plane in terms of the plate edges P_1 , P_2 , P_3 . (For fuller discussion see Section 5 of "The Mathematics of the Physical Properties of Crystals" by Walter L. Bond, *Bell System Technical Journal*, Volume XXII, No. 1, pp. 1-72.) Equation (3.8) gives such a matrix.

To continue with the NT cut as an example, the product of the matrix given in equation (3.9) for the NT cut and the components of the unit vector for the atomic plane ($2\bar{1}\cdot3$) given in equation (3.20) gives us the components of the unit normal to the atomic plane ($2\bar{1}\cdot3$) in terms of the plate edges P_1 , P_2 , P_3 as axes. Thus,

$$\begin{aligned} N_{P_1P_2P_3} &= \begin{pmatrix} 0 & .99027 & -.13917 \\ -.64279 & -.10662 & -.75852 \\ -.76604 & .08946 & .63653 \end{pmatrix} \begin{pmatrix} .5914 \\ 0 \\ .8064 \end{pmatrix} \\ &= \begin{pmatrix} -.11223 \\ -.99180 \\ +.060293 \end{pmatrix} \dots (3.21) \end{aligned}$$

That is, the components of the unit normal to the atomic plane in terms of the plate edges P_1 , P_2 , P_3 are:

$$N_1 = -.11223$$

$$N_2 = -.99180$$

$$N_3 = +.060293$$

When the orientation of the atomic plane with respect to the plate edge has thus been determined the next and final step is the determination of the angles between the incident X-rays and the plate-face.

3.9 DETERMINATION OF ANGLES BETWEEN X-RAYS AND THE FACES OF A FINISHED PLATE

Procedures for determining the angles g and $g + g'$ for certain particular positions of the atomic plane with respect to the plate edges were described in section 3.6 under (a), (b), and (c). For the general case (d) in which the intersection between the atomic plane and the plate-face is neither normal nor parallel to the plane of the instrument, the problem is best solved vectorially, as follows:

(d) Atomic plane intersecting plate-face in a line which is neither normal nor parallel to the plane of the instrument (general case). (Fig. 3.16)

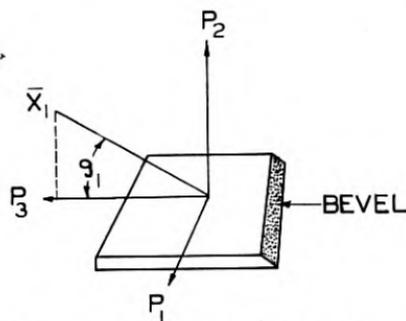


Fig. 3.19—Position 1 for a plate of general orientation (one in which no plate-edge is parallel to a crystallographic axis).

Let N_1, N_2, N_3 be the components of the unit normal \bar{N} to the atomic plane in terms of the plate edges P_1, P_2, P_3 , and X_1, X_2, X_3 the components of the unit vector \bar{X} along the incident beam. Then

$$\sin \theta_{hk \cdot \ell} = X_1 N_1 + X_2 N_2 + X_3 N_3 \quad (3.22)$$

(the inner product of these two vectors, which is thus equal to the cosine of the angle between the incident ray and the normal to the atomic plane or $\cos(90 - \theta)$).

In matrix form this may also be written:

$$\bar{X}_c N_{hk \cdot \ell} = \sin \theta_{hk \cdot \ell} \quad (3.23)$$

Where \bar{X}_c is the matrix X_1, X_2, X_3 ⁷

⁷ See Bond, W. L. "The Mathematics of the Physical Properties of Crystals," *Bell Sys. Tech. Jour.*, Vol. XXII, No. 1.

From Fig. 3.19 we see that, for the beam entering as shown, the components of the unit vector \bar{X}_1 along the beam are

$$\begin{array}{c} 0 \\ \sin g_1 \\ \cos g_1 \end{array}$$

so that equation (3.22) becomes

$$\sin \theta_{hk.l} = N_2 \sin g_1 + N_3 \cos g_1 \quad (3.24)$$

which has the solution

$$g_1 = \theta' - \delta'_1 \quad (3.25)$$

where

$$\tan \delta'_1 = \frac{N_3}{N_2} \quad (3.26)$$

and

$$\sin \theta' = \frac{\sin \theta}{\cos \sin^{-1} N_1} \quad (3.26')$$

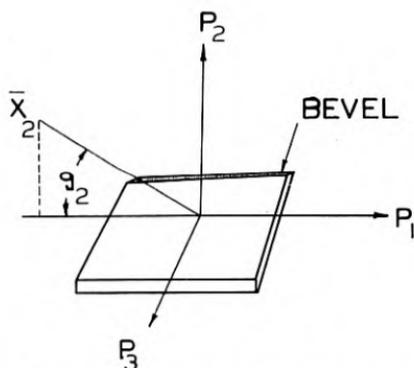


Fig. 3.20—Position 2 for a plate of general orientation

Again, with the plate rotated 90° around its normal (to position 2) so that the entering beam is in the position shown by the unit vector \bar{X}_2 in Fig. 3.20, the components of \bar{X}_2 are

$$\begin{array}{c} -\cos g_2 \\ \sin g_2 \\ 0 \end{array}$$

so that

$$\sin \theta = N_2 \sin g_2 - N_1 \cos g_2 \quad (3.27)$$

or

$$g_2 = \theta'' + \delta'_2 \quad (3.28)$$

where
$$\tan \delta'_2 = \frac{N_1}{N_2} \quad (3.29)$$

and
$$\sin \theta'' = \frac{\sin \theta}{\cos \sin^{-1} N_3} \quad (3.29')$$

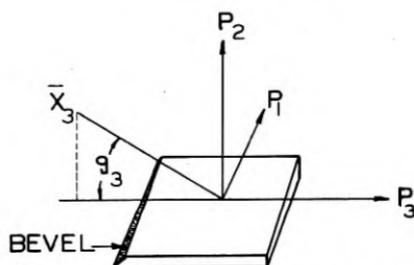


Fig. 3.21—Position 3 for a plate of general orientation

With the plate rotated 90° again (180° from position 1), so that the entering beam is in the position shown in Fig. 3.21, the components of the unit vector \bar{X}_3 along the incident beam are

$$\begin{array}{c} 0 \\ \sin g_3 \\ -\cos g_3 \end{array}$$

so that
$$\sin \theta = N_2 \sin g_3 - N_3 \cos g_3 \quad (3.30)$$

or

$$g_3 = \theta' + \delta'_3 \quad (3.31)$$

where

$$\tan \delta'_3 = \frac{N_3}{N_2} \quad (3.32)$$

and

$$\sin \theta' = \frac{\sin \theta}{\cos \sin^{-1} N_1} \quad (3.26')$$

$$\delta'_3 = \delta'_1 \quad (3.33)$$

Finally, with the X-ray beam entering as shown by the unit vector \bar{X}_4 in Fig. 3.22 (270° from position 1) the components of \bar{X}_4 are

$$\begin{array}{c} \cos g_4 \\ \sin g_4 \\ 0 \end{array}$$

so that
$$\sin \theta = N_2 \sin g_4 + N_1 \cos g_4 \quad (3.34)$$

or

$$g_4 = \theta'' - \delta'_4 \quad (3.35)$$

where
$$\tan \delta'_4 = \frac{N_1}{N_2} \quad (3.36)$$

and
$$\sin \theta'' = \frac{\sin \theta}{\cos \sin^{-1} N_3} \quad (3.29')$$

$$\delta'_4 = \delta'_2 \quad (3.37)$$

For example, suppose an *NT* finished crystal is to be checked by the $(2\bar{1}\cdot3)$ atomic plane as suggested in section 3.7B. The components of the unit normal to the atomic plane $(2\bar{1}\cdot3)$ in terms of the plate edges P_1 , P_2 , P_3 of an *NT* plate were found (at the end of section 3.8) to be

$$N_1 = -.11223$$

$$N_2 = -.99180$$

$$N_3 = .060293$$

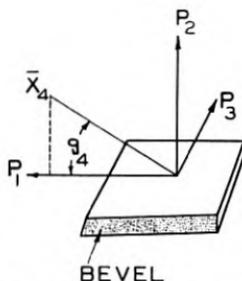


Fig. 3.22—Position 4 for a plate of general orientation

Substituting in equation (3.26):

$$\tan \delta'_1 = \frac{.060293}{-.99180}$$

$$\delta'_1 = -3^\circ 28.7'$$

and in equation (3.26')

$$\sin \theta' = \frac{\sin 32^\circ 3'}{\cos \sin^{-1} -.11223}, \quad \theta' = 32^\circ 16'$$

and in equation (3.29)

$$\delta'_2 = \tan^{-1} \frac{-.11223}{-.99180} = 6^\circ 30'$$

and in equation (3.29')

$$\sin \theta'' = \frac{\sin 32^\circ 3'}{\cos \sin^{-1} .060293}, \quad \theta'' = 32^\circ 7'$$

so that in equation (3.25)

$$g_1 = 32^\circ 16' + 3^\circ 28.7' = 35^\circ 44.7'$$

Similarly

$$g_2 = 32^\circ 7' + 6^\circ 30' = 38^\circ 37'$$

$$g_3 = 32^\circ 16' - 3^\circ 28.7' = 28^\circ 47.3'$$

$$g_4 = 32^\circ 7' - 6^\circ 30' = 25^\circ 37'$$

In the case (d) of an atomic plane which intersects the plate face in a line which is neither parallel nor normal to the plane of the instrument, the angle δ that the atomic normal makes with the plane of the instrument will be different for different positions of the plate.

If N_V is the direction cosine of the normal to the atomic plane with reference to that plate edge (P axis) that is placed in the vertical position,

$$\delta = \cos^{-1}N_V - 90^\circ = -\sin^{-1}N_V$$

This value of δ may be used in determining β according to formula (3.3). For example, when the P_3 axis (width) of an NT plate is placed parallel to the axis of the instrument,

$$\delta = -\sin^{-1}N_3 = -\sin^{-1}.060293 = -3^\circ 27.4'$$

(The negative sign indicates deflection of the normal toward the negative end of the P_3 axis and may be disregarded in determination of β).

Whence $\sin \beta = 2 \sin 32^\circ 03' \sin 3^\circ 27.4'$

$$\beta = 3^\circ 40'$$

This means that the beam reflected from the ($2\bar{1}\cdot 3$) atomic plane when the NT plate is placed with its P_3 (width) axis parallel to the axis of the instrument would be received by an ionization chamber which would accept a beam making an angle of $3^\circ 40'$ with the plane of the instrument.

When the P_1 axis (length) of an NT plate is placed parallel to the axis of the instrument

$$\delta = \cos^{-1}N_1 - 90^\circ = \sin^{-1}-.11223 = -6^\circ 26.7'$$

Since this is a larger β value than most ionization chambers will accept, the ($2\bar{1}\cdot 3$) plane cannot be used in most cases to check an NT plate with its P_1 axis parallel to the axis of the instrument unless the ionization chamber is moved vertically.

(b) and (c) Atomic plane intersecting plate-face in a line which is either normal or parallel to the plane of the instrument.

For plates rotated about X only (as AT , BT , CT and DT) the problem

is so simple (See section 3.6, b) that the above procedure need not be followed.

For such plates the complete transformation is an A_2 rotation around X :

$$r''' = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos A_2 & -\sin A_2 \\ 0 & \sin A_2 & \cos A_2 \end{pmatrix} \quad (3.38)$$

The planes used for checking these plates are of the type $(0k \cdot \ell)$ so that, from equations (3.13) and (3.14)

$$N = \begin{pmatrix} 0 \\ 2k/\sqrt{3} \\ \ell/1.1 \end{pmatrix} \frac{1}{\sqrt{S}} \quad (3.39)$$

whence

$$r'''N = \begin{pmatrix} 0 \\ \frac{2}{\sqrt{3}} k \cos A_2 \frac{\ell}{1.1} \sin A_2 \\ \frac{2}{\sqrt{3}} k \sin A_2 + \frac{\ell}{1.1} \cos A_2 \end{pmatrix} \frac{1}{\sqrt{S}} \quad (3.40)$$

$$\text{but } \frac{2K}{\sqrt{3}\sqrt{S}} = \sin \gamma_{0k \cdot \ell} \quad \text{and} \quad \frac{\ell}{1.1\sqrt{S}} = \cos \gamma_{0k \cdot \ell}$$

where γ = the angle between the normal to the atomic plane and the Z axis (note that for γ angles on the negative side of the Z axis the value given in Table I should be subtracted from 180° , as in the case of the AT cut given below).

Thus (3.40) may be written

$$N = \begin{pmatrix} 0 \\ \sin(\gamma - A_2) \\ \cos(\gamma - A_2) \end{pmatrix} \quad (3.41)$$

and (3.25) becomes

$$g_1 = \theta - \delta'_1 \quad \text{where} \quad \delta'_1 = 90^\circ + A_2 - \gamma \quad (3.42)$$

(Note that where A_2 is positive γ will be on the negative side of the Z axis)

$$\text{whence} \quad g_1 = \theta + \gamma - 90^\circ - A_2 \quad (3.43)$$

$$g_3 = 90^\circ + \theta + A_2 - \gamma \quad (3.44)$$

For position 2, equation (3.40) applied to equation 3.27 gives

$$\sin \theta = \sin(\gamma - A_2) \sin g_2$$

so that we may write

$$\sin g_2 = \sin g_4 = \frac{\sin \theta}{\sin (\gamma - A_2)} \quad (3.45)$$

As an example consider a *BT* plate for which $A_2 = -49^\circ$, corrected around the *WW* axis (Fig. 3.12) from the (02·3) plane. Table I gives us $\theta_{023} = 34^\circ 05'$, $\gamma_{023} = 40^\circ 16'$ whence

$$g_1 = 34^\circ 05' + 40^\circ 16' - 90^\circ + 49^\circ = 33^\circ 21'$$

$$g_3 = 90^\circ + 34^\circ 05' - 49^\circ - 40^\circ 16' = 34^\circ 49'$$

$$g_2 = g_4 = \sin^{-1} \frac{\sin 34^\circ 05'}{\sin 89^\circ 16'} = 34^\circ 05'$$

For an *AT* plate of $A_2 = +35^\circ 15'$ corrected around the *WW* axis from (01·1) we find, under (01·1), that $\theta = 13^\circ 20'$, $\gamma = 180 - 51^\circ 47'$ whence

$$g_1 = 13^\circ 21' + 180^\circ - 51^\circ 47' - 90^\circ - 35^\circ 15' = 16^\circ 18'$$

$$g_3 = 2\theta - g_1 = 10^\circ 22'$$

$$g_2 = g_4 = \sin^{-1} \frac{\sin 13^\circ 20'}{\sin 92^\circ 58'} = 13^\circ 21'$$

A complete determination of the errors in a quartz plate includes the measurement of angle errors about *three* mutually perpendicular axes such as, for example, the plate edges. Therefore, in correcting a plate we must use three different X-ray "shots"; usually two on the major surface with the plate rotated 90° between the two "shots" and a third on a surface normal to the major surface, commonly called an "edge". If the plate is thin, the intensity of reflection from the "edge" is low and this measurement becomes difficult.

The goniometric procedure described above is best adapted to the routine measurement of errors in large numbers of plates which are known to be close to the correct orientation. If the error in a plate is large, however, or if the cut is a rare one or the orientation of the plate completely unknown, analysis by the goniometric method may be extremely difficult.

A simple, direct method of determining the orientation of such plates is that of the Laue camera.

3.10 USE OF LAUE PHOTOGRAPHS IN DETERMINING THE ORIENTATION OF A QUARTZ PLATE

In the original Laue photograph the X-ray beam passed through the crystal and was diffracted so as to give a spot pattern on a photographic

plate beyond the crystal. This necessitated either the use of a very thin crystal or a long exposure.

To avoid the variation due to crystal thickness we adopted the "Back Reflection Laue camera." As shown in Fig. 3.23 the X-ray beam passes through a hole in the photographic film before striking the crystal. It is collimated by two pinholes, one on each side of the film. Spot reflections from many planes fall on the film and in a few minutes exposure leave a record of their points of impingement. Most of these reflections are not due to the peaks of the radiation curve (Fig. 3.2) but each spot is due to a

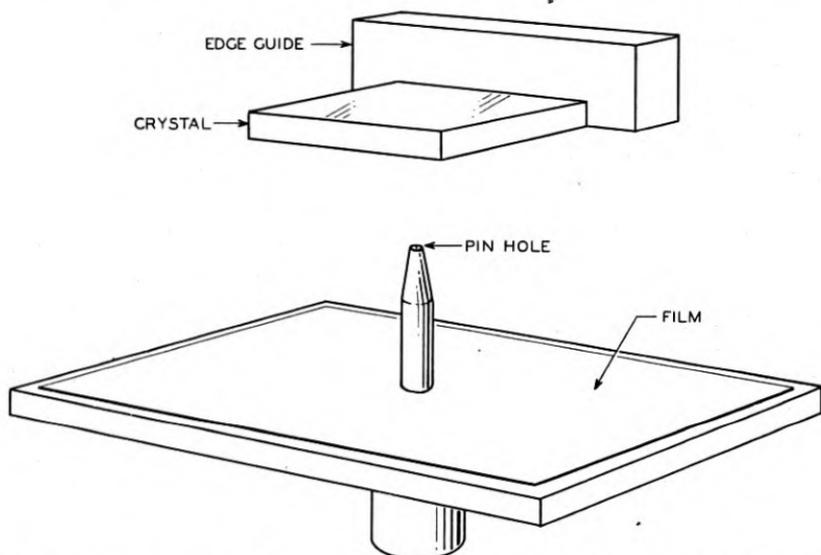


Fig. 3.23—Arrangement of collimator, film and crystal plate for a back reflection Laue photograph

different λ from the continuous background of Fig. 3.2. However, each satisfies the equation $n\lambda = 2d \sin \theta$.

Figure 3.24 is such a record with many of its spots marked. The spot (01·1) is recognized as the point of intersection of the greatest number of rows of spots. The spot (01·2) is the second most obvious intersection point.

On examining such a film we recognize such spot configurations and then mark the indices of the corresponding atomic plane for a few chosen spots. From these we can measure three angle-errors. For example, if the crystal, Fig. 3.25, is rotated about the vector t by amount e_t the spot pattern on the film will be rotated by the same amount e_t' . If the crystal is in error by amount e_w being rotated about w clockwise the spot pattern of the film

will be shifted to the left and the value of e_w' can be measured by means of a specially graduated scale. See Fig. 3.24. (Scale used is of Lucite and

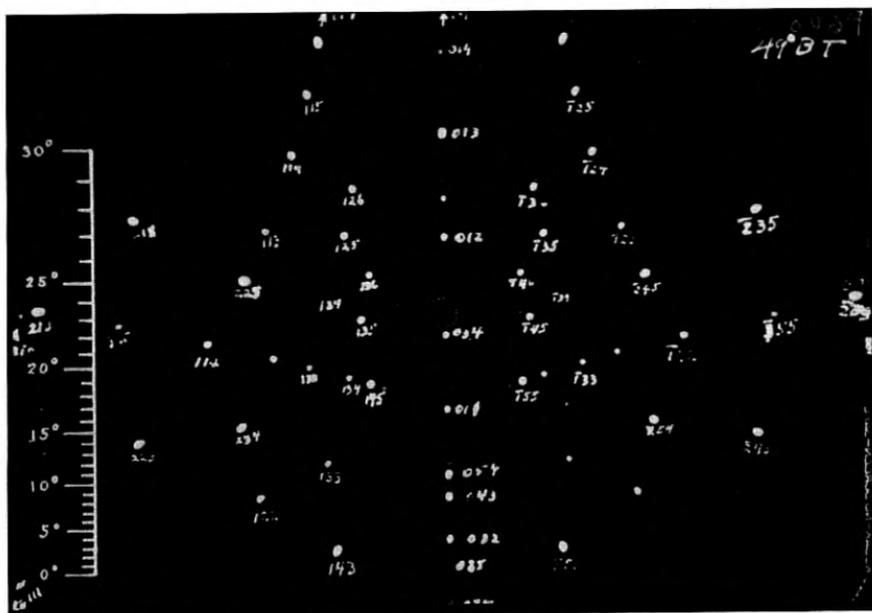


Fig. 3.24—Back reflection Laue photograph of a BT plate

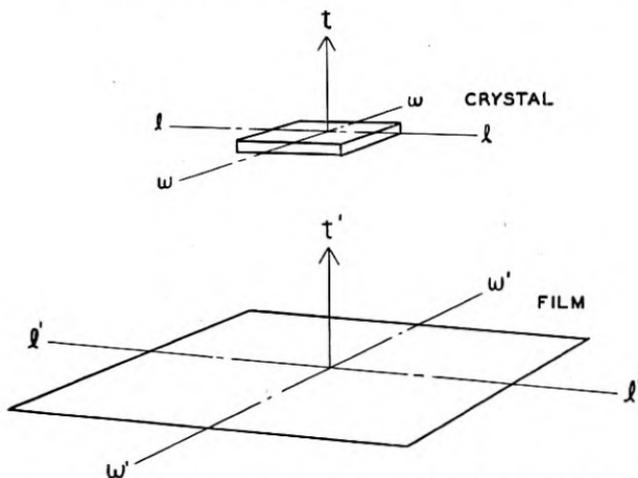


Fig. 3.25—Relation of displacement directions on the film to the $l-l$ and $w-w$ axes

calibrated to 10 minutes.) Similarly, if the crystal is misoriented about ll by amount e_l the spot pattern will shift along $w'w'$ and e_l' can be measured by the same scale used on e_w' . To compensate for film expansion due to

changes in temperature and humidity, measurements are made from two spots about equally removed from the center but on opposite sides.

If all three errors exist simultaneously they may still be measured in this way provided they are small.

Given e_t , e_w and e_l for plates containing the X axis the errors in A_1 , A_2 and A_3 can be computed by the formulae

$$\text{Error in } A_1 = e_w / \cos A_2$$

$$\text{Error in } A_2 = e_l$$

$$\text{Error in } A_3 = e_t - e_w \tan A_2$$

Consider, for example, the negative (Fig. 3.24) for a BT plate. The row of spots $(01 \cdot 4)$, $(01 \cdot 3)$, $(01 \cdot 2)$, $(03 \cdot 4)$, $(01 \cdot 1)$, etc., results from reflections from the group of planes parallel to the X axis (see Fig. 3.8). (Such a group of planes all parallel to a given line is called a zone.) In the BT cut the A_2 angle results from rotation around the X axis. The effect of such a rotation on the photograph (Fig. 3.24) is a shift of the spots of the $(0k \cdot \ell)$ atomic planes along the line on which they lie. Therefore, to measure the A_2 angle from Fig. 3.24, we choose two spots along this line about equidistant from the center of the film such as $(01 \cdot 1)$ and $(01 \cdot 2)$ and measure the angular distance α between the incident X-ray beam and the reflected ray causing each spot by measuring the distance between the center of the film and each spot using the specially calibrated scale.

By "center of the film" is meant the point through which the center of the X-ray beam passed. This point is located as follows. When the film is in place in the camera, two reference points are pricked in it by pins built into the camera. These points may be seen in Fig. 3.24 as small white circles near the center of the right and left edges of the picture. The center of the X-ray beam is at the center of a line between these two points.

When the negative is to be measured it is placed on a glass plate engraved as shown in Fig. 3.26 with the reference points at P, P . The intersection X then marks the center of the film.

If A_2 is the plate angle around the X axis from the Z axis and γ is the angle between the atomic plane normal and the Z axis and α is the angle measured between the center of the film and the spot, then

$$A_2 = 90^\circ - \gamma_{hk \cdot \ell} + \alpha_{hk \cdot \ell} \quad (3.46)$$

For the two spots chosen for measurement of the BT plate of Fig. 3.24, $A_2 = 90^\circ - \gamma_{01 \cdot 2} - \alpha_{01 \cdot 2}$ (Rotation from the BT cut to the $01 \cdot 2$ plane is a negative rotation)

$$A_2 = 90^\circ - \gamma_{01 \cdot 1} + \alpha_{01 \cdot 1}$$

$$A_2 = 90^\circ - \frac{\gamma_{01 \cdot 2} + \gamma_{01 \cdot 1}}{2} + \frac{\alpha_{01 \cdot 1} - \alpha_{01 \cdot 2}}{2}$$

and substituting the values of $\gamma_{01.2}$ and $\gamma_{01.1}$ from Table I,

$$\begin{aligned}
 A_2 &= 90^\circ - \frac{32^\circ 25' + 51^\circ 47'}{2} + \frac{\alpha_{01.1} - \alpha_{01.2}}{2} \\
 &= 47^\circ 54' + \frac{\alpha_{01.1} - \alpha_{01.2}}{2}
 \end{aligned}
 \tag{3.47}$$

For more complicated cuts, such as the *NT* cut, four spots are chosen, two close to the line $l'l'$ and two close to the line $w'w'$. (See Fig. 3.25). The positions these spots should occupy on the film for a correct *NT* cut

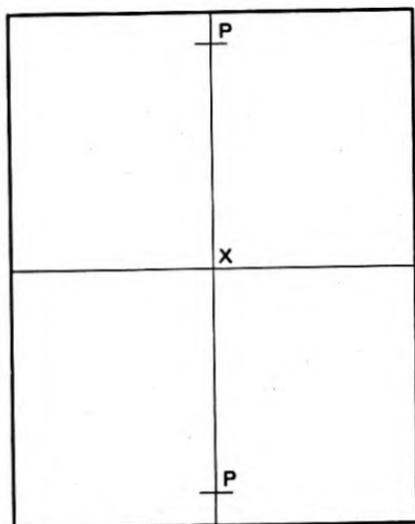


Fig. 3.26—Engraved glass plate for measuring Laue photographs

are computed and compared to the spot-positions on the measured film. From this comparison the following errors are determined:

- e_w error in rotation around the ww axis of the plate from the average displacement along $l'l'$.
- e_l error in rotation around the ll axis of the plate from the average displacement along $w'w'$.
- e_t error in rotation around the t axis by the average angular displacement of the four spots about the center of the film.

The first step in determining the position of a spot for any given atomic plane in a correctly cut plate is the determination of the direction cosines of the normal to that atomic plane with respect to the plate edges. This procedure has been described in Section 3.8.

Figure 3.27 is a Laue photograph of an *NT* plate. Taking this as an example, we may choose the four spots (31.4) , $(3\bar{1}.5)$, $(2\bar{2}.3)$ and $(5\bar{2}.5)$

from which to make the measurements. From equations (3.13) and (3.9) we find that the direction cosines of the normals to these atomic planes in terms of the NT plate edges are

$$(31 \cdot 4) = \begin{pmatrix} .42561 \\ -.90357 \\ .04972 \end{pmatrix}_P, \quad (3\bar{1} \cdot 5) = \begin{pmatrix} -.01112 \\ -.99294 \\ -.11812 \end{pmatrix}_P,$$

$$(2\bar{2} \cdot 3) = \begin{pmatrix} -.42618 \\ -.90421 \\ .02816 \end{pmatrix}_P, \quad \text{and} \quad (5\bar{2} \cdot 5) = \begin{pmatrix} -.00898 \\ -.99140 \\ -.13062 \end{pmatrix}_P$$

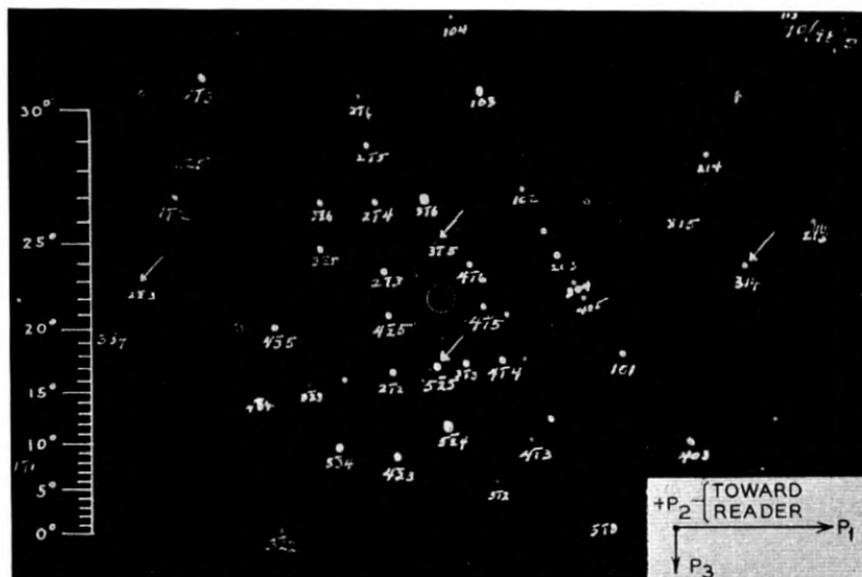


Fig. 3.27—Back reflection Laue photograph of an NT plate

The orientation of the unit normal to the atomic plane $(31 \cdot 4)$ with respect to the plate edges is shown in Fig. 3.28. Since $.90357 = \cos 25^\circ 22'$, the angle between the unit normal N and the P_2 axis is $25^\circ 22'$. Since $\frac{.04972}{.42561} = \tan 6^\circ 40'$ the angular component of the unit normal N in the P_1P_3 plane is $6^\circ 40'$ that is, the $(31 \cdot 4)$ spot should lie in a direction $6^\circ 40'$ from the long axis of the film and should be at a distance from the center corresponding to $25^\circ 22'$ as read from the special scale.

The correct locations of the other three spots are similarly calculated. These values are given below, with the values actually measured from the Laue photograph of the NT plate and the errors determined from a comparison of the calculated and measured values.

(n.b. clockwise and counter-clockwise directions are as seen when looking toward the — end of the axis of rotation)

Distance from the Center

	<i>Calculated</i>	<i>Measured</i>	<i>Difference</i>	<i>Error</i>
(31·4)	25°22'	25°22'	0'	0'
(2 $\bar{2}$ ·3)	25°17'	25°18'	+1'	
(3 $\bar{1}$ ·5)	6°49'	6°18'	-31'	30' counter-clockwise about P_1
(5 $\bar{2}$ ·5)	7°31'	8°00'	+29'	

Azimuth
(counter-clockwise from + P_1)

	<i>Calculated</i>	<i>Measured</i>	<i>Difference</i>	<i>Error</i>
(31·4)	6°40'	5°48'	-52'	7' counter-clockwise about $-P_2$
(2 $\bar{2}$ ·3)	176°12'	177°18'	+66'	or clockwise about $+P_2$ ⁸
(3 $\bar{1}$ ·5)	95°25'	95°36'	+11'	
(5 $\bar{2}$ ·5)	266°03'	267°06'	+63'	

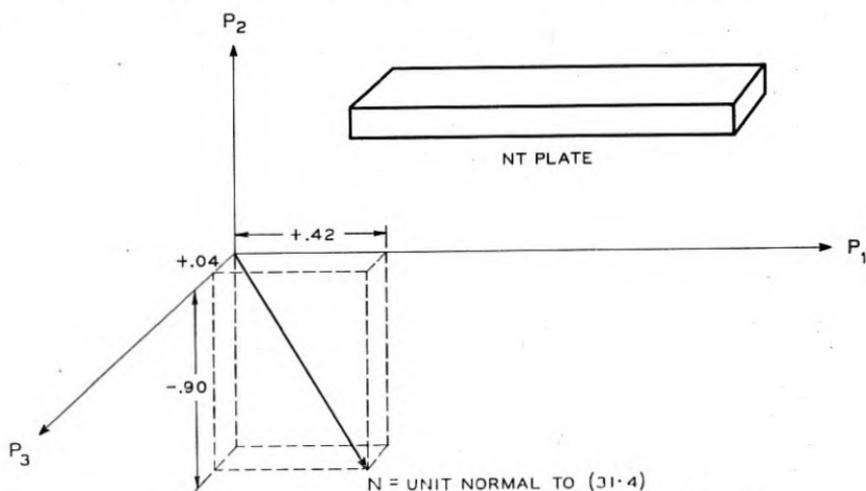


Fig 3.28—Orientation of the (31.4) atomic plane in the NT plate

Now, a small rotation of amount E_1 counter-clockwise about P_1 followed by one of amount E_2 counter-clockwise about P_2 and a similar one E_3 about P_3 is given to the first order approximation by the matrix:

$$e = \begin{pmatrix} 1 & -e_3 & e_2 \\ e_3 & 1 & -e_1 \\ -e_2 & e_1 & 1 \end{pmatrix} \quad (3.48)$$

Where e_1 , e_2 and e_3 are the sines of the rotation angles E_1 , E_2 , E_3 .

⁸ This error is determined by averaging the "difference" values for the two planes.

From (3.48) we see that the order of application of e_1, e_2, e_3 is immaterial.

If the desired transformation is r° (i.e., with zero errors) ($=r'''$, equation 3.9 for the *NT* plate), and the actual one is r , we can consider that r° is made up from the desired rotations $A_1^\circ, A_2^\circ, A_3^\circ$ and the r is made up of the rotations $A_1 = A_1^\circ + \Delta_1, A_2 = A_2^\circ + \Delta_2, A_3 = A_3^\circ + \Delta_3$; or alternatively r is made from the transformation r° followed by the transformation e so that

$$r = er^\circ \quad (3.49)$$

whence

$$\begin{aligned} -\tan A_1 &= \frac{(er^\circ)_{21}}{(er^\circ)_{22}} = \frac{r_{21}^\circ - e_1 r_{31}^\circ + e_3 r_{11}^\circ}{r_{22}^\circ - e_1 r_{32}^\circ + e_3 r_{21}^\circ} \\ -\sin A_2 &= (er^\circ)_{23} = r_{23}^\circ - e_1 r_{33}^\circ + e_3 r_{13}^\circ \\ -\tan A_3 &= \frac{(er^\circ)_{13}}{(er^\circ)_{33}} = \frac{r_{13}^\circ - e_3 r_{23}^\circ + e_2 r_{33}^\circ}{r_{33}^\circ - e_2 r_{13}^\circ + e_1 r_{23}^\circ} \end{aligned} \quad (3.50)$$

where $(er^\circ)_{21}$ is the term in row 2, column 1 of the matrix er° ; r_{23}° is the term in row 2, column 3 of the matrix r° (equation 3.9), etc.

From our unfinished example of the *NT* plate we have $e_1 = .0087, e_2 = -.0018, e_3 = 0$ whence

$$-\tan A_1 = \frac{-.64279 + .0087 \times .766 + 0}{-.10662 - .0087 \times .08946 + 0} = \frac{-.6361}{.1074}$$

$$\text{whence } A_1 = 99^\circ 36'$$

$$-\sin A_2 = -.75852 - .0087 \times .6365 + 0 = -.7640$$

$$\text{whence } A_2 = 49^\circ 49'$$

$$\tan A_3 = \frac{-.13917 - 0 - .0018 \times .6365}{.63653 + .0018 \times .1392 - .0087 \times .7585} = \frac{-.1403}{.6302}$$

$$\text{whence } A_3 = -12^\circ 33'$$

In starting work on a new cut of crystal one may have difficulty in finding the indices of the Laue spots. The easiest method is to photograph a crystal that has been carefully cut at measured angles from known planes (for instance natural faces). For example, from the angles laid off in the shop an *NT* plate such as that described should be sufficiently accurate that when P_2 is located on the atomic plane chart, Fig. 3.7, and several nearby planes of small indices are computed on the P axes, there should be no doubt as to which spots correspond to these locations.

From a few of these spots one can find many others by "zonal" relations. A zone is a family of atomic planes all of which are parallel to one line called the zonal axis. Just as there are indices of a plane there are zonal indices.

Two planes determine a zone. The zonal indices are computed by "cross multiplication" of the indices of the planes. The two planes $(h_1k_1 \cdot \ell_1)$ and $(h_2k_2 \cdot \ell_2)$ determine the zone $(k_1\ell_2 - \ell_1k_2, \ell_1h_2 - h_1\ell_2, h_1k_2 - k_1h_2)$. In practice this is developed by writing the indices of each plane twice; those of one under those of the other, then striking out the two end members of each and taking the difference of the cross products. Whereas plane indices are always enclosed in parentheses (), zonal indices are always enclosed in brackets []. For example (0 0·1) and (0 1·0) are in the zone:

$$\begin{array}{r|cccc|c} 0 & 1 & 0 & 0 & 1 & 0 \\ & \times & \times & \times & & \\ 0 & 0 & 1 & 0 & 0 & 1 \\ \hline & & & & & [1 \ 0 \cdot 0] \end{array}$$

while $(2\bar{1} \cdot 0)$ and $(21 \cdot 2)$ are in the zone $[24 \cdot \bar{4}]$, or reducing, in the $[12 \cdot \bar{2}]$ zone.

This zonal relation is reciprocal in the sense that the plane common to two zones is derived in the same way as were the zonal indices. For example, the plane common to the zones $[10 \cdot 0]$ and $[12 \cdot \bar{2}]$ is

$$\begin{array}{r|cccc|c} 1 & 2 & \bar{2} & 1 & 2 & \bar{2} \\ & \times & \times & \times & & \\ 1 & 0 & 0 & 1 & 0 & 0 \\ \hline & & & & & (0 \ \bar{2} \cdot \bar{2}) \end{array} \quad \text{or,} \quad (0 \ \bar{1} \cdot \bar{1})$$

If a given face with indices xyz lies in a zone with the symbol uvw , the following equation, known as "the zonal equation" will be true:

$$ux + vy + wz = 0$$

In the Laue photograph of the *BT* plate, Fig. 3.24, the row of spots $(12 \cdot 2)$ $(13 \cdot 3)$ $(01 \cdot 1)$ $(\bar{1}5 \cdot 5)$ $(\bar{1}3 \cdot 3)$ etc. is due to reflections from planes which belong in one zone. Other rows such as $(\bar{1}2 \cdot 5)$ $(\bar{1}2 \cdot 4)$ $(\bar{1}2 \cdot 3)$ etc. cross this row at some spot whose indices can be computed if two spots in each row are known. If the row passes through the center of the film all the spots lie on a straight line. If the row is off center it is curved, convex towards the center. Experience soon shows us how much curvature goes with how great a distance from the center for a series of spots to form a zone.

3.11 X-RAY CHECKS OF SLABS IN THE COURSE OF MANUFACTURING QUARTZ PLATE

In the course of manufacturing a quartz plate it is common practice to X-ray check the first sawn slab and correct the orientation of the quartz with respect to the saw before cutting the remaining slabs. Each of these slabs is then cemented to the head of a barrel jig so built that the orientation

of the head can be adjusted through small angles and clamped in the adjusted position. The jig is then placed in an X-ray goniometer set for reflection from the plane to be used and the slab adjusted until maximum reflection is obtained. The jig itself is not moved. If the slab were perfectly cut the maximum reflection would be obtained when the head of the jig was normal to the jig axis. A miscut will give maximum reflection at some other angle and, when clamped at this angle of maximum reflection,

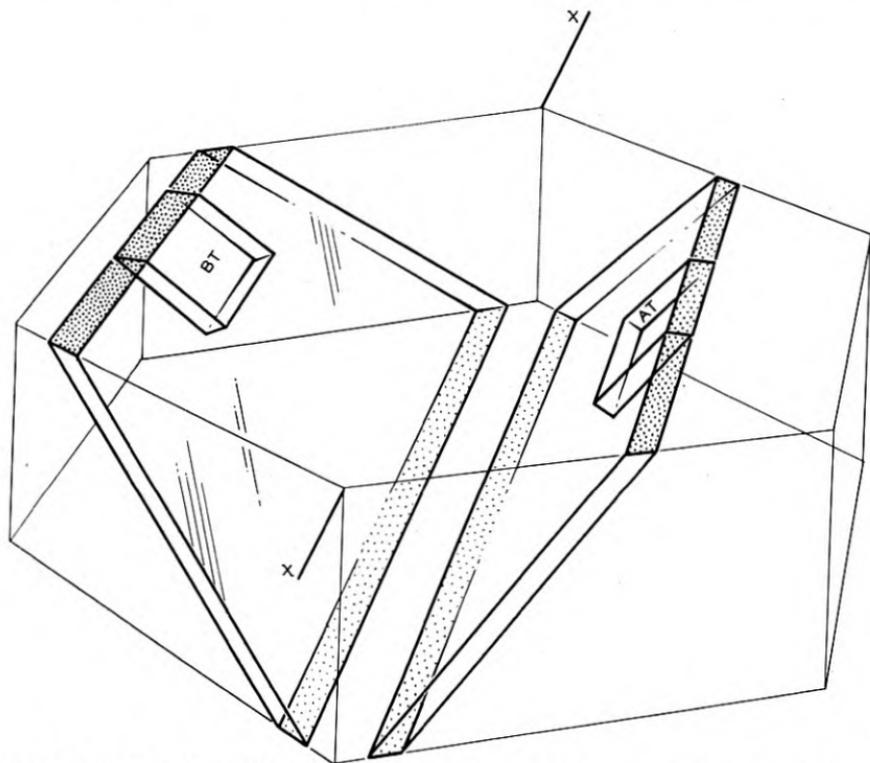


Fig. 3.29—Orientation of AT and BT plates and slabs with respect to Z section (+ X toward reader for right-handed quartz)

will have a surface ground on it normal to the axis of the jig. This surface will have the desired orientation of the plate.

As in the case of the finished plate the slab must be checked for angle-error around two mutually perpendicular axes in the plate surface. The intersection of the slab surface with the Z -cut surface is commonly taken as one axis and the normal to this as the other.

For the AT and BT cuts these two axes are the same ones used in correcting the finished plate and the same settings for jig holder and ionization chamber may be used (See Fig. 3.29). For any cut with an A_3 angle that

is not 0° , such as the *NT* cut, these two axes will not be the same as those used in checking the finished plate and different settings of the jig-holder and ionization chamber will have to be calculated. The procedure is similar to that for the determination of the *g* angles for the finished plate except that the components of the unit vector of the atomic plane used are multiplied by the matrix which expresses the transformation after two rotations (A_1 and A_2) (equation 3.7) instead of the matrix for three rotations.

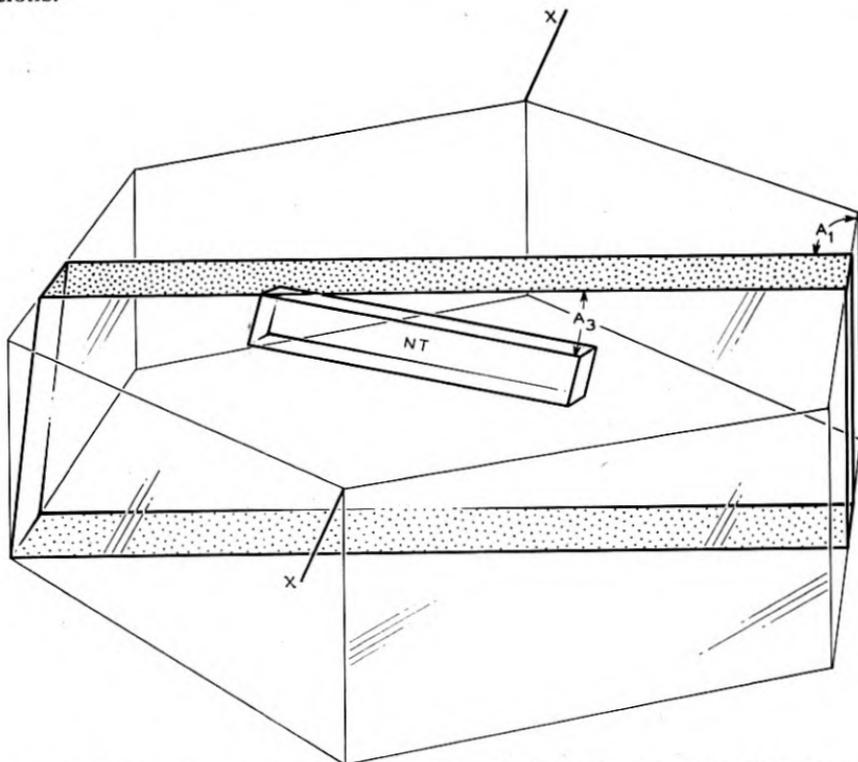


Fig. 3.30—Orientation of *NT* plate and slab with respect to *Z* section (+*X* toward reader for right-handed quartz)

For example, in the case of the *NT* cut, $A_1 = 99^\circ 25'$ and $A_2 = 49^\circ 20'$ so that the $A_1 A_2$ matrix (equation 3.7) becomes

$$r'' = \begin{pmatrix} -.1636 & .9865 & 0 \\ -.6429 & -.1066 & -.7595 \\ -.7483 & -.1241 & .6517 \end{pmatrix}$$

which, multiplied by the components of the unit normal to the atomic plane ($2\bar{1}\cdot 3$) in terms of the *X*, *Y*, *Z* axes $\begin{pmatrix} .5914 \\ 0 \\ .8064 \end{pmatrix}$ gives the components of the unit

normal to the atomic plane ($2\bar{1}\cdot3$) in terms of the reference edges of the slab after rotations A_1 and A_2 (See Fig. 3.30). These are

$$N = \begin{pmatrix} -.0966 \\ -.9919 \\ -.0830 \end{pmatrix}$$

Substituting in equation (3.26)

$$\tan \delta'_1 = \frac{-.0830}{.9919}$$

$$\delta'_1 = -4^\circ 48'$$

and in equation (3.29)

$$\tan \delta'_2 = \frac{-.0966}{-.9919}$$

$$\delta'_2 = 5^\circ 35'$$

and from equations (3.25), (3.28), (3.31), and (3.35):

$$g_1 = 27^\circ 25'$$

$$g_2 = 37^\circ 45'$$

$$g_3 = 37^\circ 1'$$

$$g_4 = 26^\circ 35'$$

which are the g angles for X-raying the NT slab where the reference edge is the intersection between the slab-surface and the Z -section surface.

CHAPTER IV

Raw Quartz, Its Imperfections and Inspection

By G. W. WILLARD

4.1 INTRODUCTION

QUARTZ is one of the commonest of crystalline minerals and occurs in many variations of size, color, purity, and structural perfection. It is used for such varied purposes as jewelry, fusing into heat and chemical resistant dishes, and for optical and piezoelectric units. However, the following discussion will be concerned mainly with such raw quartz as is commercially used in the manufacture of piezoelectric circuit elements. It might be added that the terminology used may be more in keeping with the language of the piezoelectric manufacturers than of the geologist. Further, description of many unusual types of defects, and variations of common types has been omitted. An attempt is here made to describe such defects as are of most interest in the piezoelectric art and in such a manner as to be most widely useful. Following is a description of raw quartz and its defects, the means of observing these defects, their appearance as recorded photographically, and a discussion of their effects on finished plates.

The words *defects* and *imperfections* as used in this article mean a deviation from a perfect specimen of raw quartz; they do not necessarily mean that the material is not entirely satisfactory for the purposes intended.

4.2 SOURCE, SIZE, SHAPE

Quartz crystals of usable quality and size come mainly from the interior of Brazil. From other sources the supply is negligible, or the size too small, or the imperfections too predominant. Even from Brazil only one in a hundred of the mined stones is usable. The size of stones most commonly used run from one-half to five pounds (about one-half cup to one quart size). The shape of the stones varies from well faced material, with all of the original natural faces intact, to stones in which the faces are broken or eroded away. When the faces are entirely broken away by mining operations the stones resemble chunks of broken glass. When the faces are eroded away by having been washed along river beds the stones are called RIVER QUARTZ, and the appearance is that common to river stones. River quartz usually exhibits a network of shallow surface cracks resulting from the continual bumping along a river bed, and hence is more subject to thermal and mechanical shock than uncracked stones.

With defaced quartz (river and broken) the orientation of the crystal structure cannot be determined from the surface shape. Since the stones must be cut at specific orientations relative to this structure, special means must be employed to determine the structure orientation. For this reason many users of quartz prefer faced stones. However, defaced stones are usually more free from defects than faced stones, since optical twinning is commonly concentrated near the natural faces of the original stone and other defects near the base. Thus by making use of special means (inspectoscope, conoscope, oriascope) defaced quartz may be cut to good advantage.

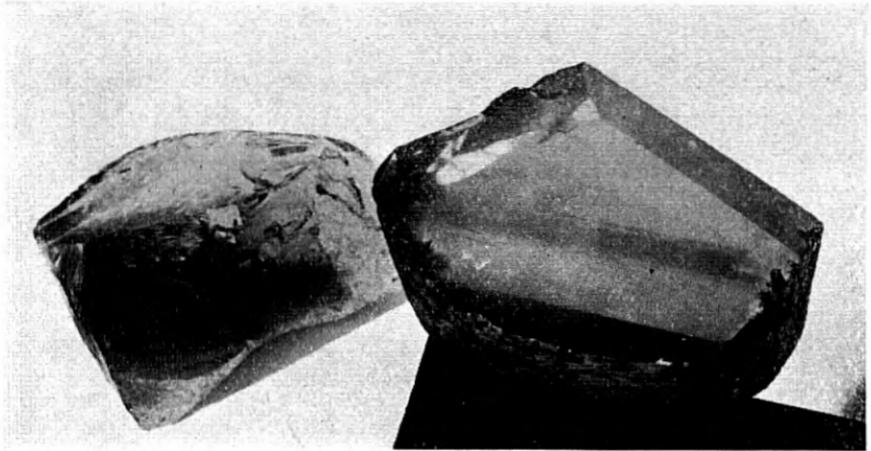


Fig. 4.1—Quartz may show smoky or citrine coloring throughout or only in restricted regions, irregularly as at the left, or in PHANTOM planes as at the right.

4.3 COLOR

Usable quartz is transparent internally (though the exterior surface may be opaquely coated), never translucent (milky). The color of the quartz varies from perfectly clear through slightly smoky to fairly dark. The obviously dark stones are called **SMOKY QUARTZ**. Smokiness may be uniform throughout a piece, or varying from clear to dark, or confined to plane sheets within a single piece, see Figure 4.1. Dark smoky stones are not used because they cannot be inspected for defects and optic axis. With stones that are used this coloration is seldom so dense that it may be detected in the small finished plates with ground surfaces.

Because the smokiness is due to so slight a deviation from the pure quartz its analysis is extremely difficult. The coloration is variously explained, as due to minute traces of impurities (organic or inorganic), as due to the dissociation of a few SiO_2 molecules into free silicon and oxygen, and otherwise.

An important fact about smoky quartz is that it may be cleared of color-

tion merely by heating to 350°C. to 450°C. for a short time¹ (see Fig. 4.2). The clarity of cleared stones, even though originally very dark, rivals that of normally clear quartz, thus leading to the belief that most commercial quartz is colored to a slight degree.² Further, it is claimed that irradiation of either cleared or normally clear stones with radium rays causes them to become

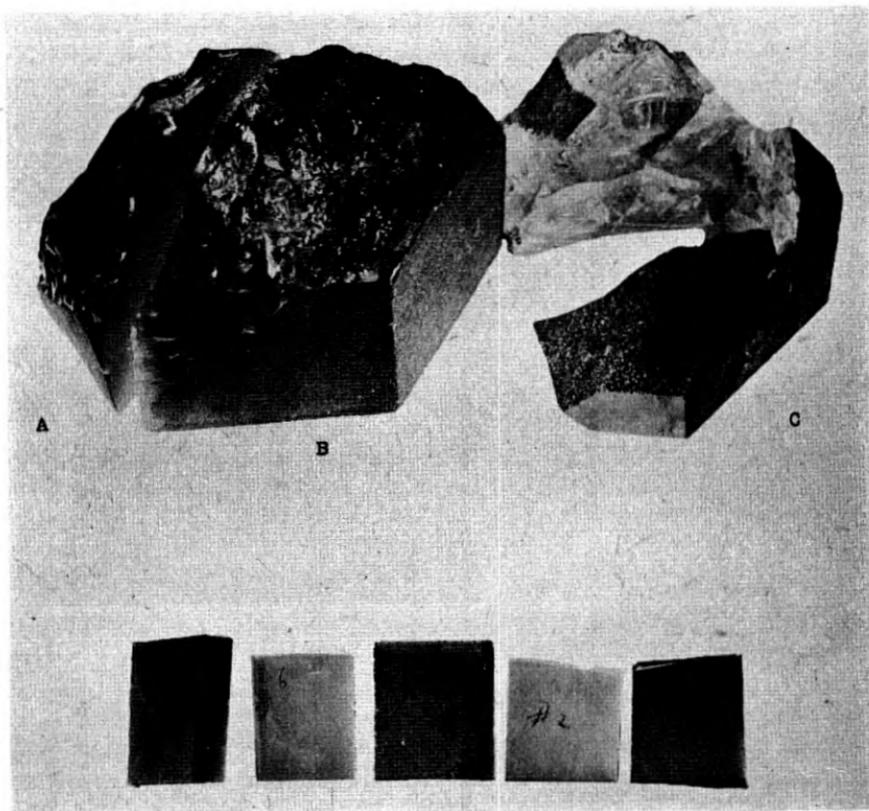


Fig. 4.2—Coloration in quartz may be cleared by heating to temperatures below 500°C. A, B and C are adjacent portions of one dark, smoky stone, only C having been cleared. The five, small blocks were likewise cut from a single stone, the second and fourth having been partially cleared at 350°C and fully cleared at 500°C.

smoky. It is then questionable whether there is a material difference in clear and smoky quartz, or only a difference in condition of the material. By the limited studies that have been made light smokiness has not been found to have any definite effect on finished plates. It might be added that such quartz plates as are normally heated to 400°C. to 500°C. in the process of manufacture, would of course be cleared of any smokiness originally present.

¹ Stones so cleared 8 years ago are still clear.

² Arkansas quartz (not used commercially) is noted for its unusual clarity.

What has been said of smoky quartz applies equally well to CITRINE QUARTZ except that here the coloration is straw colored (yellow to brown). Citrine quartz is even more uncommon, commercially, than smoky quartz.

4.4 TWINNING

Electrical and optical twinning are among the most common defects of crystal quartz. There are few stones without one or both present. Either type is usually difficult or impossible to detect from the exterior form of the crystal. Presence of either type in a finished piezoelectric element interferes with its perfect operation. Twinning is an abnormality of growth, in which an apparently homogeneous crystal is not actually of the same handedness, electrical sense, or orientation throughout. In the case of both electrical and optical twinning (the only common twinning types in quartz), the electric and optic axes in all parts remain parallel each to each.

In a crystal which is only ELECTRICALLY TWINNED, the stone is entirely of one handedness (either right or left), but one portion is of OPPOSITE ELECTRICAL SENSE to another portion. This change in sense of structure is NOT detectable by ordinary optical means. However, at the surface it is detectable by, (1) the piezoelectric effect (determination of electric charge on squeezing); (2) x-ray reflection intensities (using certain sense determining planes, see Chapter III); and (3) most readily and extensively, by etch-pits and etch-pit figure techniques.³ Commonly electrical twins are sufficiently large that they may be separated near a twinning boundary and both parts used.

In a crystal which is only OPTICALLY TWINNED, one portion of the crystal is of OPPOSITE HANDEDNESS and electrical sense to another portion. This change in handedness of structure is detectable by optical means (i.e. by examining between crossed polarizing filters). Optical twinning may also be detected, at the surface, by the etch technique in the same manner as electrical twinning. Usually a stone will be mainly of one handedness with only small, thin, interlayered growths of opposite handedness, thus making it impossible to use both handed portions separately.

Further discussion of both electrical and optical twinning will be found in Chapter V, where means for simultaneously detecting both are described.

4.5 CRACKS

Many quartz stones contain cracks which are not readily seen by a casual surface examination. As mentioned above river quartz commonly exhibits a network of shallow cracks extending inward from the surface, caused by bumping. All types of stones commonly contain one or more cracks, es-

³ In finished plates, of course, the effects of twinning are also determined by measuring the resulting piezoelectric and elastic constants of the plate (i.e., their effect on frequency and activity).

pecially when badly twinned or full of inclusions. Some of these cracks may be due to rough handling, but others are due to growth conditions, or result from temperature changes after growth.

Though large cracks are readily detected by means to be described, it might be noted that when cracks are sufficiently fine (small separation compared to light wave-lengths) they will no longer be seen. Thus every visible crack may be considered to extend beyond its visible range, and some actual cracks will not be visible at all. Because of such cracks and other defects in quartz, special care should be taken in handling uncut and partially cut stones to prevent their subjection to mechanical or thermal shocks. On the other hand, small, flawless, finished plates will stand considerable shock. It is even common practice to solder to finished plates (after metallizing).

4.6 INCLUSIONS (BUBBLES, NEEDLES, PHANTOMS, VEILS, ETC.)

The remaining abnormalities of raw quartz (used for piezoelectric elements) may be classed as inclusions. Among these are inclusions of solid, liquid, and gaseous material. The size of individual inclusions may vary from submicroscopic, to those easily visible with the naked eye. The inclusions may be isolated, or arranged in lines, or planes, or curved surfaces. In many cases the arrangement forms of inclusions (bubbles, needles, phantoms, veils) have been used to describe inclusions, with little regard to the nature or size of the individual inclusions. This is because inclusions which are too small to analyze individually, are still visible when grouped by hundreds in lines or surfaces.

When inclusions are sufficiently small and closely grouped they give a **BLUISH** cast (Tyndall effect) to the group. Thus the bluish cast is recognized as indicating fineness of grouped inclusions. When the individual inclusions are larger, the group appears white. With still larger inclusions one may actually see separate, individual inclusions, looking like minute bubbles. Thus, describing the group as blue, white or bubble textured is of considerable importance, when analyzing or estimating the usability of quartz with grouped inclusions.

BUBBLE INCLUSIONS look like small bubbles (i.e., small spheroidal cavities) in the quartz. When bubbles appear individually, or randomly scattered, they are referred to as just bubbles. When bubbles occur in organized groups the group is referred to as a bubble phantom, bubble veil, etc. Smaller bubbles appear only as light reflection points and their shape is not seen.

In general such bubbles may be filled with gas, liquid, solid, or any combination of phases. They may be of the same nature as rare, large cavities in which one can easily see a liquid moving about. Analysis of the contents of such cavities has indicated the presence of CO_2 , water, salt solutions, and

other substances that might have been present during growth conditions. That these cavities are seldom in the form of negative crystals, i.e. having the plane natural faces characteristic of quartz, is not easily explained.

NEEDLE INCLUSIONS appear as long, thin lines or needles. They may be straight or curved, blue, white or otherwise. Whether they are continuous or composed of rows of individual inclusions is usually not discernable. Needles, visible without concentrated illumination (and fluid immersion), are likely to be inclusions of crystalline material, such as rutile (brown), tourmaline (black). Usually such needles are called rutile needles, because of the commonness of rutile and the difficulty of determining whether they are rutile or some other material. They might better be called **DARK** needles. **BLUE NEEDLES** are fine textured, and may appear singly or in parallel groups, which may be at angles to other parallel groups. In other cases they spread from a bubble point, like a comet. Blue needles may also be feathered (having short feathery rays along the sides), may be hard (very fine and sharply distinguishable), or soft (diffuse). Probably the most important characterization of all blue needles is their blueness, which indicates fine texture. **WHITE NEEDLES** are similarly hard or soft. **CHUVA** is a special type of white needle which would be extremely elusive except for the fact that along its length are small bubbles, giving chuva the appearance of dew drops along a thin fiber. For piezoelectric usage an important distinction between needles is whether they are blue, white, chuva, or dark.

PHANTOMS are an arrangement form of inclusions (or coloration), in plane sheets which are parallel to possible natural crystal faces (usually the prism or pyramid faces). Often several, differently oriented, phantom planes are formed together so as to give the appearance of a crystal within a crystal, thus the name phantom (or ghost). Phantoms may also appear as groups of parallel sheets. Phantoms may be of smoky, blue, white, or bubble texture and should be so noted when describing their effects on piezoelectric elements. That phantoms are closely related to disturbed growth conditions is apparent from their close relationship to crystal faces.

VEILS are an arrangement form of inclusions in curved sheets. They are most commonly of a tenuous bubble texture, but may also be white or bluish. Again this distinction is of importance in estimating their deleterious effects. The cause of inclusions appearing in veil form is not clear.

CLOUDS (a term not widely used) refer to inclusions irregularly distributed in restricted regions of the crystal.

4.7 INSPECTION MEANS

The raw quartz inspectoscope is the name of an instrument used for the inspection of raw quartz. This inspectoscope may, of course, be also used

for the inspection of quartz in various stages of processing, and for other transparent materials than quartz.

By means of a polarized light optical system the stones are examined for optical (but not electrical) twinning. By this same means the direction of the optic axis through the stone is also determined. By means of concentrated high-power illumination the stones are inspected for cracks, color and inclusions. By both means the stones are illuminated and inspected while immersed in an immersion fluid of matching index of refraction.

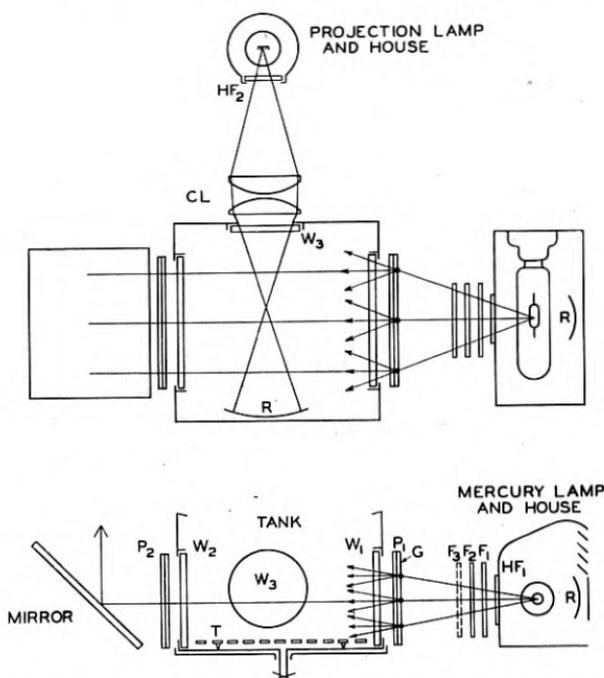


Fig. 4.3—The raw quartz INSPECTOSCOPE, top and front views. The stones are held in the immersion tank and viewed from above. Polarized light from the mercury lamp is used for locating the direction of the optic axis and for detecting optical twinning. Concentrated light from the projection lamp is used for detecting other interior defects. See also Fig. 4.5.

Without such immersion it is difficult or impossible to illuminate the interior properly, or to see into the interior, of oddly shaped or rough surfaced stones. Such immersion eliminates refraction (bending of light rays) at the surface. The interior of stones with ground, fractured, or other surfaces, are as easily examined as a piece of plate glass.

The important design features, maintenance and operation procedures are given below. Figure 4.3 is a diagram of the instrument, and Fig. 4.5 shows an early model utilizing many of the features described below. Three plate glass windows W are cemented into apertures in the open-top, steel

immersion tank, Fig. 4.3. A raised perforated table T rests on the bottom of the tank. For polarized light inspection of twinning and of optic axis direction, the right to left optical system is used. This comprises an AH-4, 100 watt, mercury-vapor lamp (requiring special transformer) isolated in a well ventilated housing with window HF_1 , a set of color filters F_1 , F_2 , (and possibly F_3), a polarizing filter P_1 with ground surface G (for light diffusion), the immersion tank, the polarizing filter P_2 (crossed to P_1), and the mirror (mounted at 45° to the vertical) to reflect light vertically up through a window in the drain pan (not shown in Fig. 4.3), to the eye. For inspection of defects other than twinning, a high powered projection system is used. This comprises a projection lamp (isolated in a housing by window HF_2 , and with forced draft), condenser lenses CL , and the immersion tank. Here one looks down directly into the tank at the stone. Reflectors R may be added to both systems to increase the illumination.

This instrument resulted from a restudy of long-used "inspection tanks" and methods, and includes some features not originated by the author. Since this inspectoscope is believed to be superior to many inspection-tank equipments now in use, the more important design features will be described.

The tank should be large enough to allow easy handling of the stones in the fluid, with allowance for positioning any portion of the stone in the projection beam, and allowance for rise in level of the fluid as the stone is immersed. However, the size should not be made larger than necessary, for it has been found in practice that the fluid very rapidly collects lint and dirt, which (being kept in suspension by agitation) scatters or diffuses the light. This, besides scattering the light from its proper paths, interferes with the polarized light inspection by depolarizing the beam. In large tanks with dirty oil the light becomes almost completely depolarized and no patterns can be seen. If, however, the crystal is large enough to nearly fill the length of the tank (along the polarized light beam), this depolarization is small.⁴ Thus, a tank as small as is consistent with the size of the stones to be examined should be used. The smallest convenient size and shape of tank would be about 8 inches in the polarized light direction, 8 inches in the projection light direction, and $5\frac{1}{2}$ inches high (all elements of Fig. 4.3 are drawn in proportion to these tank dimensions). This permits easy examination of two to three pound stones (pint-sized), and six-inch long stones may be used without great difficulty.

Isolation of lamp heat is an important consideration in both optical systems. In the projection system the high wattage lamp would dangerously heat up the whole instrument if the heat were not properly dissipated. In

⁴In an emergency small stones can be examined in a large tank with polluted fluid by placing the stone at the mirror end of the tank and introducing a polarizing filter directly in the fluid, close to the stone, on the lamp side (Polaroid J-Film is only slowly attacked by many of the immersion fluids).

the polarized light system the heat must be isolated from the color filters and especially from the polarizing filter P_1 , for it is extremely important that polarizing screens should not be overheated. They deteriorate (lose their polarizing property) rapidly above 60°C ., and should not reach a temperature uncomfortable to the touch. Their deterioration by heat or otherwise is not usually discernible except by examining their extinction quality with the aid of a good piece of polarizing material (turned for extinction, they should pass practically no light). Since the polarizing filters and the color filters all absorb some light themselves it is important to ventilate these units, as well as to isolate the mercury lamp heat with a separate housing. The window HF_1 in the mercury lamp housing may be of one-eighth inch pyrex glass, or better the heat filter HF (specified later), and may be a moulded filter (since there is no focusing required here). However, the window for the projection lamp housing should be polished, either pyrex or heat filter as above (since here the light is used in a focused beam).

The polarizing filter P_1 and P_2 may be glass filters (polarizing film cemented between glass plates by the manufacturer) or film filters held between sheets of one-eighth inch plate-glass, with taped edges. The latter arrangement is less expensive, and the film is usually obtainable without delay. Extra filters should be kept on hand. The two filters must be mounted, relative to each other, in a crossed position (for extinction). Since light entering the tank from P_1 must be diffuse it is necessary to introduce a ground glass surface. This is accomplished without adding an extra glass plate by grinding one surface of the polarizing filter, the outer surface if a cemented glass filter, an inner surface if using loose film between glass plates (the inner surface may be used here to provide for protection from dirt). In either case the ground surface must be on the lamp side of the filter or it will depolarize the light.

In the past a carbon arc has been used as the projection light source. Such arcs are not handily turned on and off, nor adjusted, and are now difficult to obtain. An incandescent projection lamp overcomes these difficulties. A 500 to 1000-watt lamp with double-plane filament structure (filament in two planes, and staggered relative to each other, giving a solid square of illumination) is ideal for this purpose. Such lamps operate at high efficiency, are a concentrated source, have a short life, and generally require forced ventilation. A recommended lamp is the Mazda, Clear Projection, 750W—120V, T-12, C-13D Fil., Med. P.F. base. This lamp requires a small blower for ventilation and when operated with a foot switch, only during that part of the inspection that it is needed, gives a satisfactory replacement schedule.

It is important in the projection system to use large, short-focus condenser lenses, and to focus the lamp image near the center of the tank. This allows

the great concentration of light which is necessary for detecting fine-textured defects, and makes it easier to determine just where in the crystal (in depth) the defect lies. Further, with the large angle of illumination available, those defects which require a specific angle of illumination may be found with less hunting. It should be noted that in figuring the object and image distances, the refractive index n of the immersion fluid must be taken into account (the window to image distance with fluid present is about n times that with fluid absent).

The color quality of the light used in the polarized light system has a considerable effect on the ease of observing the light patterns obtained when inspecting for twinning and optic axis. A typical light pattern of a piece of raw quartz viewed along the optic axis is shown in Fig. 4.8. The broad dark and light contours, "thickness-contours," are the ones used in locating the optic axis. The finely "toothed-patterns" at A , B , and C show twinning. The conditions that make the former most pronounced are not necessarily the same as those that make the latter most pronounced. The broad thickness-contours are most pronounced in monochromatic light, but barely visible in white light. This is due to the large variation of rotatory power with color, which, in all but the smallest stones, causes such overlapping of the white-light color contours as to result in practically no appearance of contours at all. This effect does not apply to the twinning regions, since in most cases the thickness of oppositely handed material is too small to develop overlapping. The result with white light is that the stone appears mainly white, except for regions of twinning where the toothed pattern is seen in color. For twinning detection, then, the advantage of white light is largely due to removal of the extraneous thickness contours. This possible advantage for the novice is not obtained without some loss of factors necessary for complete identification.

On the other hand, in determining the direction of the optic axis, the thickness contours are essential, and hence monochromatic (or a restricted spectrum) light is necessary. This illumination is most easily achieved with a mercury arc and color filters. The mercury arc emits a restricted spectrum (mainly $.578\mu$ yellow, $.546\mu$ green, $.436\mu$ far-blue lines and weak red and blue-green bands), and is very efficient. Even without filtering it gives far better thickness-contours for axis determination than does white light. The insensitivity of the eye to blue leaves mainly yellow and green. The yellow may be largely removed without appreciable loss of green by using filter F_2 and the blue and blue-green may be removed with filter F_1 . However, the red can be removed only with considerable loss of green by filter F_3 . The use of F_1 and F_2 alone are recommended as giving sufficient restriction of spectrum and yet high illumination. (All three filters, as used in the conoscope, give a fairly monochromatic green.) The filters need not

be polished, only moulded, since here there is no focusing of light. The filters described above are the Corning glass filters:

F_1 , Code 3484, H.R. Traffic Shade Yellow,

F_2 , Code 5120, Didymium,

F_3 , Code 4303, Dark Shade Blue Green,

HF , Code 3966, Extra Light Shade Aklo (a heat filter).

A corresponding set of filters passing more light, but giving a less monochromatic light are:

F_1 , Code 3486, H.R. Yellow Shade Yellow,

F_2 , Code 5920, H.R. Illusion Pink,

F_3 , Code 4308, Light Shade Blue Green.

It might be added that the pronounced effects of filters are easily observed with any instrument by holding small polished filters over the eye. These same filters give some improvement even with white light, for optic axis detection. Further, if the filter were not so heavy, polished filters might be better applied at the eye than at the light source, since here they would also cut down extraneous illumination from the room. Or the same result might be obtained with large polished filters (expensive) and an eye chute at the viewing end of the system.

Several factors are of importance with regard to the immersion fluid used in the inspectoscope. The fluid should have a refractive index matching that of quartz, and be clear and colorless (to eliminate loss of light). It should be of low viscosity, so that dirt and dust may settle and air bubbles rise, rapidly (to prevent depolarization of the polarized light beam). Low viscosity also aids in the draining of oil from the stones after inspection. Water solubility of the fluid would be an aid to cleaning. Necessarily the fluid must be non-toxic and non-flammable, and preferably odorless, inexpensive and commercial. Various fluids satisfying these requirements to varying degrees have been used. Since there is no majority agreement as to which of the fluids now in commercial use is most satisfactory, no particular fluid can be recommended. (Three are listed in Chapter II, page 258.)

However, a word may be added about the required degree of refractive index match. Mineral oils of index 1.47 to 1.48 are, definitely, very poor immersion fluids for quartz. With them it is difficult to see into the interior of stones without plane polished surfaces. Ground and unpolished surfaces still cause considerable diffusion. For good inspection viewing the fluid should have an index between 1.53 and 1.56 (preferably between 1.54 and 1.55).

The refractoscope is a simple instrument especially designed for the purpose of easily and exactly checking the index match of fluid to quartz. The principle having been already noted (p. 255, Chapter II), it suffices here to describe the use of the instrument. A test tube, Fig. 4.4, filled with the

fluid to be examined, is mounted in an adjustable-height stand, the optical system lowered into the test tube until the flat bottom surface of the lens contacts the fluid, and the stand is placed in an inspectoscope, conoscope, or in front of a lamp (the stand being adjusted to proper height for good illumination). The optical system comprises a lens L (for magnification and elimination of ripples on the liquid surface), a thin Z -cut quartz prism Q ,⁵ and a narrow slit diaphragm S . When the slit is viewed simultaneously through, and at the sides of the prism, a view similar to one of those shown

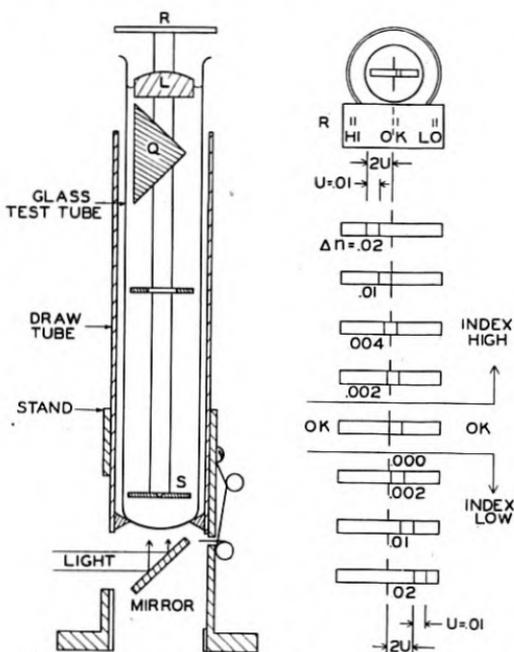


Fig. 4.4—The immersion REFRACTOSCOPE may be used in conjunction with an inspectoscope or conoscope for determining the degree of refractive index match (or mismatch) of the immersion fluid relative to quartz. At the right is shown the manner of reading the instrument.

in the figure may be seen. The two short lines are always the same distance apart and are used as a unit of index mismatch, the unit being $u = .009 \cong .01$. The left-hand short line is the one to be aligned with the long line for perfect match, and the actual mismatch Δn , is measured by the separation of these two lines. The remainder tab is added to remind the operator which indications represent perfect match, too high and too low fluid index. Since the indices of both quartz and fluid vary with temperature and with color of light used, the refractoscope was especially designed for immersion directly

⁵ The larger the prism angle the greater will be the sensitivity of reading, and the smaller will be the readable range of mismatch.

in the instrument using the immersion fluid. When, in such instruments, a polarizing filter is so oriented as to cut off one or both short prism-lines, these lines may be restored to view by rotating the test tube in the holder (upon continuous rotation one line and then the other will disappear). With vertical conoscopes the mirror may be removed for bottom illumination. When monochromatic light is used all three lines are the same color as the illuminant. When using non-monochromatic or white light, the long, slit line is the color of the source, while the short lines develop into two, separate spectra. In this case, that portion of the spectrum is used for alignment which is most predominantly used in the immersion instrument. The sensitivity of the refractoscope, when approaching perfect match, is about $\Delta n = .001$. This sensitivity is of course attained when one adjusts the fluid to match the quartz, by addition of the proper high or low-index component. When the fluid does not match there is a less accurate measure of mismatch, but this measure is still good for determining the degrees of fluid adjustment to be made.

The principle of the refractoscope may be even more simply applied to measuring the fluid to quartz mismatch, by making use of the inspectoscope and a basal section of quartz, using the 120° prism formed between two good, clear, adjacent natural faces. The section is placed base down in the tank at the far side from the mercury lamp, and so positioned that a vertical slit diaphragm, placed on the other side of the tank, may be viewed through the two prism faces. With the polarizing filters removed, the two images of the slit as seen through the prism do not (in general) align with the slit as seen above the prism. The image farthest from the prism vertex is the one that should be aligned with the slit, for perfect match of refractive indices. The necessity of removing the polarizing filters can be obviated by tilting the prism and slit about the line of sight, preferably at 45° from the vertical (or this might have been obviated, if the polarizing screens had been out with their plane of polarization at 45° to the vertical).

Finally, experience indicates that the importance of keeping the immersion fluid clear and clean is not generally realized. As noted above, contamination not only gives bad scattering of the projection beam, but also depolarizes the polarized light. A perforated plate raised from the bottom of the tank is an aid in keeping the settled dirt from being recirculated again. More effective is the provision of simple, easy means for draining, filtering and refilling the tank. One or more thicknesses of chamois makes a good filter, provided the chamois is occasionally washed out with a solvent.

4.8 PHOTOGRAPHIC STUDY OF INTERIOR DEFECTS

The original inspectoscope of Fig. 4.5 was used at the Hawthorne plant of Western Electric Company, in obtaining the accompanying photographs.

For inspecting twinning and direction of optic axis, the stones were viewed horizontally through the polarizing filter *E* and window, with mirror *F* removed (using mercury lamp *A*). Normally the operator looks directly down into the mirror to see the same view. For observing other interior defects, the view is from directly above the tank, through the fluid surface (with projection lamp in housing *H* being used). This is the normal manner of observation. *B* mounts heat and color filters; *C* is a polarizing filter with diffusing surface; *E*, a polarizing filter crossed to polarizer *C*; *G*, a glass window; and *IJ* is a rudimentary lamp house normally fitting over *A* to edge of *B*. The tank *D* has two rectangular windows parallel to *C* and *E*, and a circular window in rear wall for entrance of projection illumination.



Fig. 4.5—The inspectoscope and stones used for the following figures. See also Fig. 4.3

Most of the following photographs are of stones shown, Fig. 4.5, on the drain pan and in the tank. Note that stones 3, 6, 9, 12, 14 are well faced stones; 9, 7, 13 have fractured surfaces; and 15 is typical river quartz (nearly perfect internally). The special manner of orienting the stones in the tank to obtain the desired views will appear from the following descriptions. The views are one-half to full size.

Figure 4.6 is a polarized light view of a wedge shaped basal section (one-fourth inch thick at the left, to three-fourths inch thick at the right). The wedge is viewed along the optic axis with the plane faces approximately perpendicular to the line of sight (i.e., parallel to the polarizer *C* of Fig. 4.5).

The back face is larger than the front face so that outside the borders of the front face the thickness tapers off very rapidly. Since the dark contours here show thickness of the section (as the lines of a contour map show height above sea level) one may easily determine the shape of the wedge. The inner contour *AA* is near the greatest thickness, contour *BB* intermediate thickness, and outer contour *AA* near the thinnest portions at the edges. It might be added that the thickness-contours do not exactly indicate a region of equal thickness unless the eye is distant from the stone, and unless the stone is viewed exactly along its optic axis.

Figure 4.7 is a polarized light view of an inch-thick basal section (i.e., Z-cut) with the parallel ground surfaces perpendicular to the line of sight

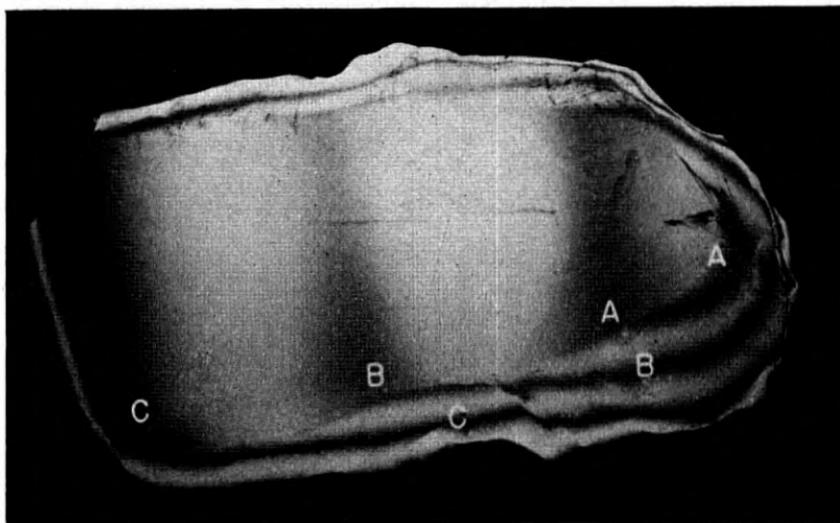


Fig. 4.6—A wedge-shaped basal section of quartz viewed along the optic axis in polarized light. The THICKNESS-CONTOURS locate regions of equal thickness, A-A thickest to C-C thinnest.

(i.e., the stone is viewed along its optic axis). Neglecting for the moment the ring pattern *C*, one observes a wide, diffuse vertical thickness-contour at *B*. With parallel surfaced stones a thickness-contour should cover the whole stone, since the thickness of the stone is uniform. Here the stone is not viewed *exactly* along the optic axis. If the eye be placed close these contours become circular.

Further, this view shows the effect of placing a lens between the tank window and polarizer at *E*, Fig. 4.5. The result is a ring pattern, the real image of which is at the focal distance of the lens on the eye side. The image may also be obtained on a ground glass at this point and its location is independent of the distance between quartz and lens (no rings will appear

if too far separated). This ring pattern is due to conical illumination (or more correctly to conical viewing; here the illumination of the stone is diffuse). This simply illustrates the basic principle of the conoscope. The principle is further illustrated by tilting the section out of its present position, which causes the ring system to move in the direction of tilting. The theory of these effects is given in Chapter II.

Figure 4.8 is a polarized light view of a pyramidal cap of quartz with a fractured back surface (stone 3, Fig. 4.5). Here the stone is viewed along the optic axis (the six natural cap faces making equal angles with the line of sight). The continuous dark bands are thickness-contours, and again

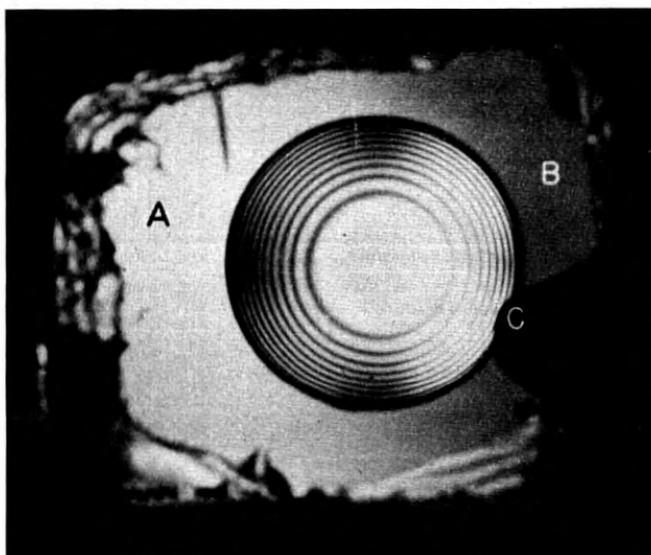


Fig. 4.7—A parallel faced, basal section of quartz viewed along the optic axis in polarized light. Note the vertical thickness-contour at B. The conoscope ring-pattern C may be obtained by introducing a lens.

represent regions of equal thickness in the optic axis direction. Had the fractured surface been flat instead of broken these contours would have been hexagonal and parallel to the hexagonal edges of the cap. The toothed-patterns at A to G are due to optical twinning (thin layers of the quartz whose handedness is opposite to that of the main stone). Although the exact shape and location are not determinable, the approximate location and extent are observed by tilting the crystal while viewing. The contour and pattern changes resulting from angularly moving the crystal (away from the position of viewing directly along the optic axis) are shown by Fig. 4.9, which should be compared with this figure.

Figure 4.9 is a polarized light view of the same stone as shown in Fig. 8,



Fig. 4.8—A broken cap of quartz is viewed along the optic axis in polarized light. Few, broad thickness-contours show good alignment with the optic axis. The fine TOOTHED-PATTERNS are due to optical twinning.

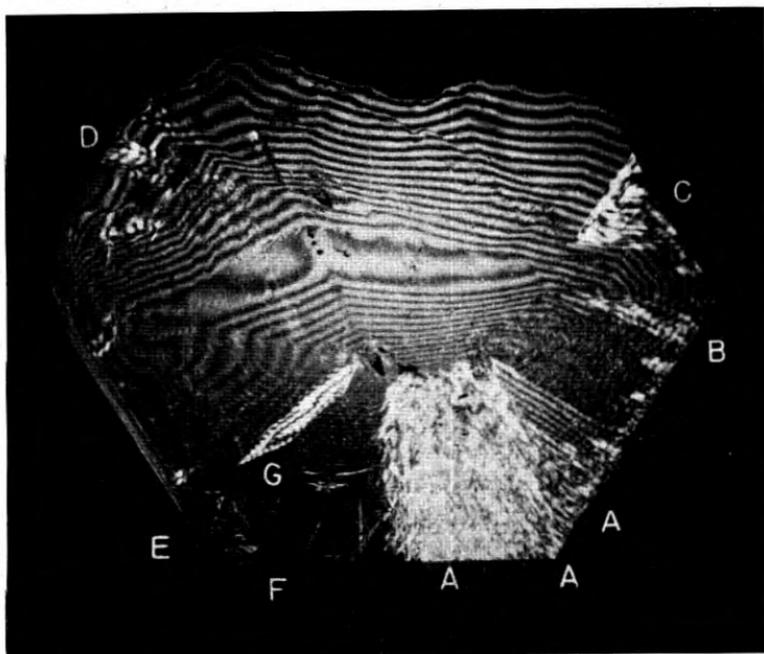


Fig. 4.9—The same cap as in Figure 4.8, but viewed slightly off the optic axis. Note the increased number of thickness-contours, and their fineness, also that some twinning regions are enhanced.

the only difference being that here the crystal is viewed slightly off the optic axis. Comparing the two views it is noted that here there are many more, and narrower, thickness-contours.⁶ As the crystal is further rotated away from optic-axis-viewing, the contours multiply until they are no longer visible. Thus, it is by so positioning the stone as to obtain the fewest and widest thickness contours that the optic axis direction is determined, the axis then being parallel to the line of sight from eye to stone. It might be parenthetically noted that the handedness of raw stones may be determined in the inspectoscope, if the polarizing filter on the viewing side be rotatably mounted, by observing the contour contraction or expansion as the filter is rotated (for stones progressively thicker from outer boundary inward, the handedness rule is opposite to that for the conoscope).

A further effect of tilting the stone away from optic-axis-viewing is to enhance the toothed, twinning-patterns. Certain of these patterns are enhanced by tilting one way, others by tilting differently (note that regions *A*, *C* and *G* are much clearer here than in Fig. 4.8). Also, since the thickness-contours move about and the toothed patterns remain fixed, motion of the stone is an aid to location of twinning (except in the rare cases of large sized twins, where this does not hold). Note that optical twins usually extend inward from the original natural faces. The twin *G* which appears to be internal, actually extends inward from a cap face.

Figures 4.10 and 4.11 show projection illumination views of typical, parallel BUBBLE-PHANTOMS (in stone 7, Fig. 4.5). The light from the left converges into the stone, focuses about centrally, diverges and passes out of the stone at *B*. Due to an internal fracture in the right end the light is also reflected upward at *C*. The light beam is visible in the fluid but not in the stone, because a slightly contaminated fluid scatters far more light than does quartz.

In Fig. 4.10 the stone is held so that the phantom planes *A-A* are viewed edge on (the only way finer textured planes are visible), while in Fig. 4.11 the planes are viewed at a slight angle, to show area of the planes. These planes have a texture of distinguishable bubbles. The planes are long, about an inch wide (with rectangular boundaries at their left end) and are parallel to a possible natural face (no acutal faces present on this broken stone). Such bubble phantoms are probably not permissible in any finished piezoelectric plate.

Figure 4.12 shows bubbles, cracks, veils, and phantoms (in stone 9, Fig. 4.5) and pairs of angularly joined phantom planes, *B-B*, parallel to the natural faces *A-A*; each pair forms two faces of interior phantom crystals. The texture of the phantom planes near *A-A* changes along their length from bubbly at the left to bluish at the right. A dense curved blue veil is seen at *C-C*,

⁶ Actually the thickness contours are not now as closely related to the thickness as before (due to the birefringence effects being added to the rotatory effects; see Chapter II).

while a disperse, curved, bubble veil is shown at *E*. Above and to the right of *D* are two small fractures.

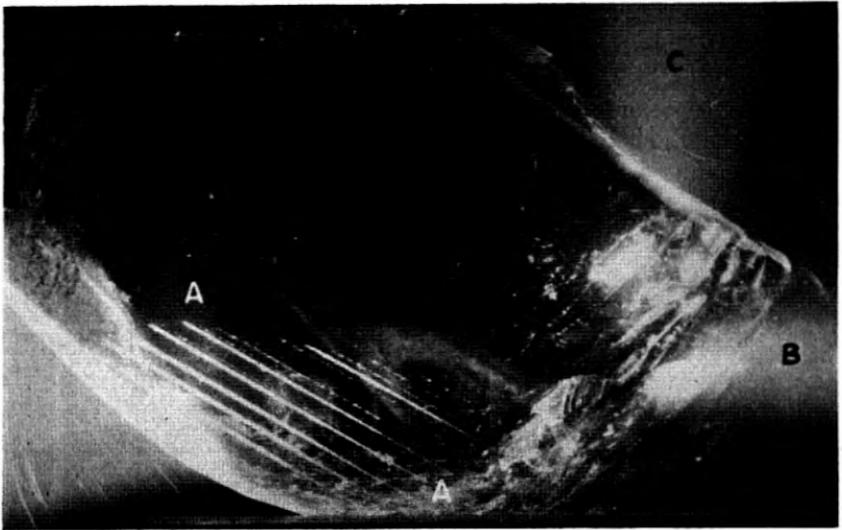


Fig. 4.10—An edge view of BUBBLE-PHANTOMS using the concentrated projection light. See also Fig. 4.11

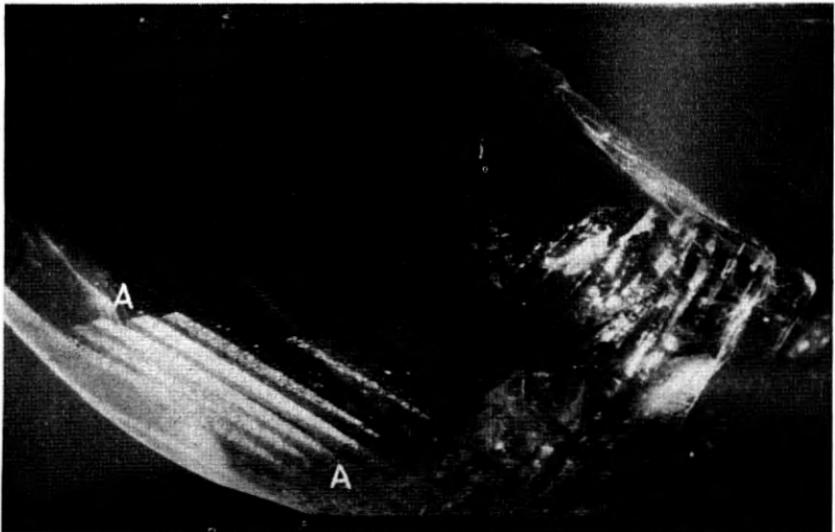


Fig. 4.11—The same stone as in Fig. 4.10, viewed at a slightly different angle. Note the bubble-texture, and the width of the planes.

Figures 4.13 and 4.14 show blue needles in two parts of a single stone which contains needles throughout (stone 10, Fig. 4.5). Only those needles

which are near the concentrated focal point of the projection illumination are visible in each view. The length direction of the needles must be nearly parallel to the direction of illumination to be seen well (thus wide angle of illumination is an aid in finding the needles). Needles elsewhere in the stone may be observed by moving the stone about. The flares at the ends of the stone are due to exterior surface conditions. The needles of Fig. 13 are of the comet type (radiating from a point), while those of Fig. 14 lie in parallel groups. In each case a few of the needles are slightly feathered, and all are soft needles.

Figure 4.15 shows a stone (14, in Fig. 4.5) in which the defects are concentrated in the base, a common occurrence. Were this stone to be proc-

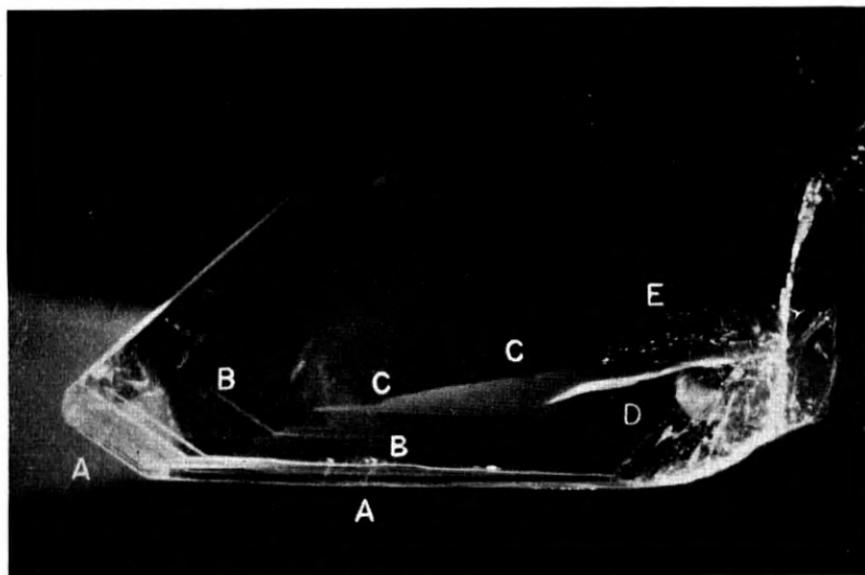


Fig. 4.12—This stone, viewed in the projection light, shows phantoms A-A to B-B, blue-veil C-C, bubble-veil E, and two fractures near D.

essed by *Z*-sectioning, a saw cut near the line C-C would divide the stone into a large, nearly perfect portion, and a small unusable portion which need not be further processed. Otherwise sawn, bad portions will have to be processed, or good portions of a largely bad section would be too small to obtain plates from. This points out the importance of coordinating processing with inspection, even though the stones have been already inspected and judged to be worth processing.

Bubbles and cracks fill the end of the stone at B, scattered bubbles appear in a veil at A, and a few isolated bubbles are at D. Note the clarity of the stone relative to the fluid, as shown by the beam of light entering the stone from the left, not visible internally.

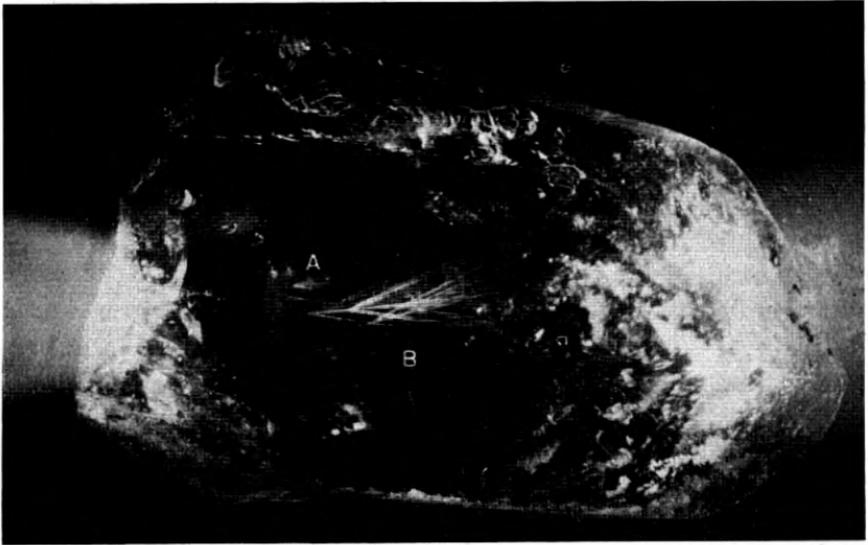


Fig. 4.13—This stone is full of blue needles, though only those located in the focus of the light beam are visible. See also Fig. 4.14.



Fig. 4.14—The same stone as in Fig. 4.13 shows other blue needles in a different arrangement, by holding the stone differently with respect to the light.

4.9 EFFECT OF INTERIOR IMPERFECTIONS ON FINISHED PLATES

The practical effect on the finished piezoelectric plates, of the various types of interior defects, is one of the least understood factors directly related to economic use of the strategic material, raw quartz. This is because of the

wide variation of the types, sizes and concentrations of defects, and the further variation of the types, sizes, and requirements to be met in the finished plates. However, an analysis of the factors involved leads to some important conclusions. The various factors may be correlated as follows.

Fine textured inclusions are less objectionable than coarse textured inclusions (i.e., smokiness, and blue needles and phantoms, than white, or bubble needles and phantoms). Isolated defects are less objectionable than concentrated defects of the same texture size (i.e., isolated bubbles, than bubble phantoms). Cracks should never be permitted in finished plates. Twinning will be discussed in detail in Chapter V.

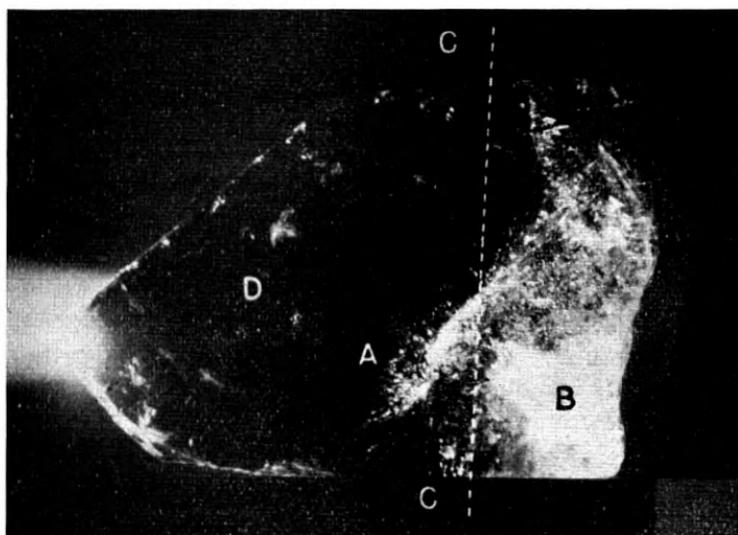


Fig. 4.15. This stone could be economically processed by cutting off the base rear line C-C. A few scattered bubbles appear at D, fractures and bubbles fill the base B, and A is a bubble veil.

Further, a given defect is more likely to be tolerated in: (1) large piezoids (finished piezoelectric elements) than in small ones; (2) in low-frequency-mode plates (*CT* and *DT* types) than in high-frequency-mode plates (*AT* and *BT* types); (3) in plates to be operated at low amplitudes of vibration (filter elements, and oscillator plates with low drive) than those driven to maximum amplitudes (oscillator plates with high drive); and (4) plates having low rather than high-quality requirements (on activity, temperature-coefficient, frequency adjustment).

Thus, blue needles have long been permitted in some types of large, low-frequency-mode filter elements. While breakage has resulted from the use of blue needles in high-frequency oscillators with very high drive, it is not

known that blue needles may not be used in many other types of oscillators. It is likely that smokiness is less objectionable than blue needles (very lightly-smoky material not recognized as such has been widely used).

A further important factor, often disregarded, is related to high-frequency-mode plates and their method of manufacture, and more specifically to the method of finally adjusting the dimensions to give the required activity, frequency, and temperature characteristic. When such plates are made by the PRE-DIMENSIONING technique, which requires very small tolerances on the machined dimensions and orientation, they are finally finished by hand adjustment of only one dimension, the thickness. On the other hand, when the plates are machined to only moderate accuracy of dimension and orientation, they must be finally hand adjusted on all three dimensions to obtain satisfactory characteristics. By this method of adjustment it is possible to correct, not only for misdimensioning and misorientation, but for small defects in the quartz itself. However, with pre-dimensioning and a single dimensional adjustment practically no correction may be made for errors, or quartz defects. Thus, higher quality quartz may be necessary for manufacture by the pre-dimensioning technique than otherwise.

The conclusions that may be drawn from these considerations are: (1) only by a quantitative statistical study can it be determined whether a given type of defect will be permissible in a given type of finished plate, (2) known usability of a given type of defect in a given type of plate does not prove its usability in a different type of plate (the type includes size, mode of vibration, and required electrical operating characteristics), and (3) the method of manufacture is also related to the usability of defective quartz (i.e., pre-dimensioning vs. non-predimensioning).

Since in the past very little defective quartz has been used in the manufacture of piezoelectric elements, especially oscillators, there is little manufacturing experience that may be used as a guide to its introduction now. The quickest means of obtaining this information, and of making use of the reservoir of defective quartz, would seem to result from trial manufacture first of the most likely to succeed types of plates, from quartz with the most likely tolerable types of defects. If and when this utilization is found to be practical the less likely cases may be examined, while at the same time defective material is being used and experience is being gained in grading the raw quartz into usably defective and non-usably defective. This special grading of quartz will be difficult to control exactly. It will be easier to grade into types of imperfections than into quantities of defects per-unit-volume. Further, it will be easier in manufacturing trials to determine whether a given type of defect is permissible if the defects appear in large quantities, than if in very small quantities (where they may actually be absent in some finished plates). For these two reasons it will be preferable

in trial manufacture, to select for processing stones which have the desired type of defect in large (to maximum) quantities per-unit-volume. Due to the many variables involved in both selection of material and in manufacture, a large quantity of stones must be processed for a fair trial. The criterion of usability of the defective material will then be related to number of usable finished plates (satisfying the required electrical and physical specifications on the finished plate) that can be obtained from a given quantity of raw quartz (and thus to the relative costs of producing satisfactory finished plates from defective and from non-defective quartz).

If and when a type of defect has been shown to be harmful by the above method, steps may then be taken to ascertain if some method of selection or measurement can be made on the defective raw quartz to separate the economically usable from the economically unusable material during inspection.

The New Statistical Mechanics

By

KARL K. DARROW

THIS is the second article upon statistical mechanics which I have published this year in this Journal. The first, which appeared in the January (1943) issue, was devoted to the oldest form of the theory, which is variously known as the old, the classical, or the Boltzmann statistics. The word "statistics," I repeat from the former article, is a synonym for statistical mechanics, objectionable but (because of the length of the alternative) hardly to be avoided. The "new statistics," frequently divided into "the Bose-Einstein statistics" and "the Fermi-Dirac statistics," emerged in the middle twenties and ever since it has been gradually pushing its ancestor aside. In this article I propose to expound the new statistics, laying especial emphasis on the theory of monatomic gases, to which the former article was strictly limited.

A definition of statistical mechanics may well be asked for at this point, especially since in the former article I failed to give one. Like many other things either subtle or familiar, statistical mechanics cannot fully be defined till it is fully understood, by which time a definition may seem nugatory. As an attempt at an advance definition, I suggest that *statistical mechanics is the theory which, starting from the assumption that matter (and, in due course, radiation) is an assemblage of particles, undertakes to explain (1) entropy, (2) temperature, (3) specific heats, and (4) the distribution-in-energy of the particles in thermal equilibrium.* The critical reader may justly say that these are four aspects of a single problem, but I think it well to separate them notwithstanding. The word "particle" often has to be construed as standing for an elaborate structure, but in dealing with monatomic gases (and with radiation) we may let it stand for a point endowed with energy and momentum.

How does the classical statistics succeed in handling these four problems? To take them in reverse order: it does very well with the fourth, for material gases (but not for radiation). It does very well with the third, for monatomic gases (but not for polyatomic gases nor for radiation). It produces an adequate theory of temperature for monatomic gases, identifying the temperature with the mean kinetic energy of the atoms multiplied by a certain factor. It has a very strange adventure with entropy, producing a theory which in part is remarkably successful and in part is disconcertingly

fallacious, and has to be altered in awkward and dubious ways to be made completely successful.

To consolidate these statements and introduce the new theory, I review the Boltzmann statistics.

The N atoms of a gas in a container are represented by N numbered balls, identical in every way except the numbering. In the earlier article a game was proposed in which a collection of numbered baskets was provided for these balls, and the balls were tossed into the baskets in a predetermined way: N_1 of them into the first of the baskets, N_2 into the second, and so on until each of the baskets contained its preassigned number of balls with N_M in the M th or last basket. The set of numbers $N_1, N_2, N_3, \dots, N_M$ was called a "distribution," and the question was asked: in how many different ways can this distribution be realized? It demands a previous answer to another question: how can a given distribution be realized in more than one way? It is the numbering of the balls which makes this possible. If for instance we exchange two balls in different baskets, the distribution is not changed, and yet there is a difference between the second situation and the first, for an inventory of all the balls in all the baskets shows that in those two particular baskets the assortment has been changed. We thus have realized the same distribution in two different ways. If we had exchanged two balls in one basket, this would not have been regarded as a change; we should still be realizing the same distribution in the same way, in the sense of the Boltzmann statistics. It was shown in the earlier article that the number W of ways of realizing a distribution—or in more technical language, the number of complexions in the distribution—is given by the formula:

$$W = N! / \prod N_j! \quad (1)$$

I have said that in the Boltzmann statistics, the balls stand for the atoms of a gas. For what then shall the baskets stand? The baskets stand for compartments in space; but "space" may have several different meanings.

Giving "space" the ordinary meaning: imagine the gas contained in a box, and the box divided mentally (not physically!) into M compartments of equal volume. I called these by the name of "cells" in the previous article, but now, for a reason which will shortly appear, I rebaptize them "regions." These are the baskets. W has its smallest value, which is unity, when all of the atoms are in the same region. It has its greatest value, which is $N! / ([N/M]!)^M$, when in each of the regions there is the same number N/M of atoms. But this corresponds, as nearly as the picture is able to correspond, to the uniform spreading throughout the box which by vast experience we recognize as the natural permanent state of the gas "in equilibrium." The uniform distribution is outstanding because it has the

greatest value of W . If now we baptize W with the name of "probability," we may then say that the state which in Nature is the prevailing one is in the Boltzmann statistics the most probable one. In the theory it is described by the equation:

$$N_j = N/M \quad (2)$$

Now think of "momentum-space" in which there is a dot for every atom, and the Cartesian coordinates of the dot are the momentum-components p_x, p_y, p_z of the atom. The coordinates of the dot determine the energy E of the atom, by virtue (for material particles) of the relation:

$$E = (1/2m)(p_x^2 + p_y^2 + p_z^2) \quad (3)$$

This is a fact of the first importance, as will shortly appear. Let us divide the momentum-space into regions of equal volume. Each of the regions will correspond to a small range of energy-values, as a sample of which we may take any particular one among them. Therefore when we distribute the dots—or let me say simply, the atoms—in any manner among them, we have perforce a certain value of the total energy U of the gas, which we may consider as preassigned if we so wish. Now we are to compare this distribution only with such others as show the same value of U . Among these there is one which is outstanding because it has the greatest value of W . This was shown in the earlier article to be the canonical or Maxwell-Boltzmann distribution, described by the formula:

$$N_j = NA \exp(-BE_j) \quad (4)$$

in which N_j stands for the number of atoms in the region numbered j ; E_j for the value of E appropriate to that region, *i.e.* obtained by substituting into (3) the coordinates of some point in that cell; A and B for constants, whereof A depends on B while B depends upon U/N the average energy of the atoms of the gas. This distribution also is attested by experiment as being truly that of a gas in its normal natural abiding state of equilibrium.

Now I mention the concept of a six-dimensional space which comprehends both the ordinary space and the momentum-space, and is divided into six-dimensional regions of equal volume. By this device one is able to speak of (2) and (4) as two aspects of a single distribution in the " μ -space." This is the distribution outstanding among those with which it may legitimately be compared by reason of having the greatest W -value. It is the most probable distribution, in the sense given in the Boltzmann statistics to the word "probable."

This is the first triumph of the Boltzmann statistics, attained by numbering the atoms. Its other triumphs, and its ultimate confusion, come when

it copes with the task of interpreting entropy. But if I continue longer this review of the conclusions of the former article, the reader's tolerance may be exhausted. Let us make haste to find out how the newer statistics sets forth to find out the most probable distribution.

One of the most appealing features of the new statistics is, that it does not impose on atoms of a single kind that peculiar distinction which I described above as "numbering" them. We therefore now remove the numbers from the atoms, restoring thus to atoms of a single kind—it might for example be helium—that quality of absolute indistinguishability which the classical statistics took away from them in order to achieve its aims.

Having de-numbered the atoms, we start anew to play that very game with numbered balls and numbered baskets which we played in the classical statistics with such remarkable but incomplete success. But now that the atoms are de-numbered, they can no longer be the balls (nor, for that matter, the baskets). Something drastically new must now be done, and is.

In the new statistics, the balls stand for the compartments and the baskets for the populations.

I must define the word "population." It means the number of atoms in a compartment, or as I will say from this point onward, in a "cell." The balls which are tossed into the basket numbered 0 stand for the cells containing no atoms; the balls which go into the basket numbered 1 correspond to the cells containing one atom apiece, and so forth indefinitely. C_i shall be the symbol for the number of balls in the i th basket, which is to say, the number of cells containing i atoms apiece. C shall stand for the total number of cells.

Let the cells at first be compartments of equal volume in the ordinary space, obtained by dividing up (mentally) the box containing the gas. For the number of complexions or inventories corresponding to a given distribution, defined by given values of the quantities C_i , we have as before:

$$W = C!/\Pi C_i! \quad (5)$$

and taking the logarithm:

$$\ln W = C \ln C - \Sigma C_i \ln C_i \quad (6)$$

In using this expression I have again, as often in the previous article, assumed the validity of what I there called "the super-Stirling approximation"; but notice that this no longer means that I assume each of the cells to enclose an enormous number of atoms—it means instead that there is an enormous number of cells having each particular population.¹

¹ Clearly this cannot be so for all populations no matter how great! This is a difficulty which also pops up in the old statistics, though there it is not met until the ordinary space is replaced by the momentum-space.

We seek a set of values C_i such that when it is realized, the quantity W shall have a value stationary with respect to all variations δC_i conforming to *two* conditions: first, that the number of cells shall remain the same, which is to say, $\Sigma \delta C_i$ shall vanish; and second, that the number of atoms shall remain the same, which is to say, $\Sigma i \delta C_i$ shall vanish.

Such a set is the following:

$$C_i = C \alpha e^{-i\beta} \quad (7)$$

α and β standing for constants yet to be determined; for taking the first variation of $\ln W$ from (6), we find:

$$\begin{aligned} \delta(\ln W) &= -\Sigma \delta C_i (1 + \ln C_i) \\ &= -\Sigma \delta C_i (1 + \ln C \alpha) + \beta \Sigma i \delta C_i \end{aligned} \quad (8)$$

and the required condition is fulfilled. Assuming without proof that the *stationary* value of W is also a maximum value, and referring to W as the "probability" of the distribution of cells among populations, we have come to the startling conclusion that the most probable distribution is the one given by (7)!

I call this a startling conclusion, because it contravenes our inbred conviction that the natural distribution of a gas in ordinary space is the uniform distribution. Of course, in the last two sentences I have used the word "distribution" in two senses, and this must be rectified at once. What I have just called "the uniform distribution" is the uniform distribution in the *old* sense—the same number of particles in every cell. In the new sense of the word, this is a distribution in which all of the cells have the same population, and therefore in which one basket contains all of the balls. Definitely, this is not, in the new statistics, the most probable distribution! Indeed it is not even a conceivable distribution, for the number of cells is infinite.

To mitigate this clash of theory with experience we can do nothing else than assume our cells to be so tiny that in any region of the gas large enough to be surveyed by observation, there is a mighty number of the cells. Then at worst we can take it from experience that in the normal natural abiding state of the gas the number of atoms in each region will be the same if all the regions are of equal volume, while within each region we can distribute the atoms among the cells as the new statistics tells us to. However, it may yet be possible to come to this conclusion from the theory. In preparation for the effort, I sketch the procedure for evaluating the constants α and β in the distribution (7).

A similar task was set before us in the earlier article: that of evaluating the constants of the Maxwell-Boltzmann law in terms of the total number and the total energy of the atoms. Here for any region we are to evaluate the constants α and β in terms of the number of cells C and the number of

atoms N . The task is greatly eased by the opportunity of using two well-known formulae:

$$1 + x + x^2 + \dots = (1 - x)^{-1} \quad (9)$$

$$1 + 2x + 3x^2 + \dots = (1 - x)^{-2} \quad (10)$$

The start is made from the two self-evident equations:

$$C = \sum C_i = C \sum \alpha e^{-i\beta} \quad (11)$$

$$N = \sum i C_i = C \sum i \alpha e^{-i\beta} \quad (12)$$

By putting x for $e^{-\beta}$ and using (9) and (10), the student can easily win through to the results,

$$\alpha = (1 - e^{-\beta}), \quad C/N = e^{\beta} - 1 \quad (13)$$

and then to the final form of the distribution-law (7):

$$C_i = C \frac{C}{N + C} \left(\frac{N}{N + C} \right)^i \quad (14)$$

and finally, after consulting (5), to the expression for the number of ways W_{\max} in which this the most probable distribution—"most probable" in the eyes of the new statistics—can be realized. Its logarithm is:

$$\ln W_{\max} = C \ln \frac{N + C}{C} + N \ln \frac{N + C}{N} \quad (15)$$

This is the most important formula of the new statistics, as will presently be clear.

Divide now the space containing the gas into "regions" of equal size, each comprising the same number C of cells, which number shall be great. For the benefit of those to whom the memory of the previous article may still be vivid, I say now that insofar as there is any correspondence of the new to the old statistics, these "regions" correspond to the "cells" of the older theory. This is the reason why, in my recent brief synopsis of the old statistics, I used the word "region" to replace the word "cell" used in the prior article. Let the subscript j be the marker for these regions, so that N_j shall stand for the number of atoms in the j th region. Put N_j for N in (15). Now each member of (15) refers explicitly to the j th region, and on the left I should put $(\ln W_{(\max)j})$, but for two purposes—one of which is brevity, while the other will appear in due time—I put $\ln W_j$ instead:

$$\ln W_j = C \ln \frac{N_j + C}{C} + N_j \ln \frac{N_j + C}{N_j} \quad (16)$$

The quantity W_j is an odd sort of "probability" relating only to the contents of the region j . It is, to repeat, the number of ways in which the

most probable distribution of the C cells among the possible populations can be realized, there being N_j atoms in the region. Nothing of the sort appeared in the old statistics.

Now form the product of all of the quantities W_j . This is a "probability" relating to the entirety of all the regions, therefore to the whole of the gas. It is the total number of ways in which the most probable distribution can simultaneously be achieved within each of the regions. It is taken to be the total number of ways in which the most probable distribution of the gas-as-a-whole can be realized; for in the new statistics we have no other way of defining the most probable distribution of the gas-as-a-whole, than this way of subdividing first into cells and then into regions comprising many cells. The symbol for this product shall be W , for although I have already used that symbol in this article, its former meaning is now taken over by W_j , and it is free again. We have:

$$\ln W = \sum \ln W_j \quad (17)$$

the quantities W_j being still those functions of N_j which were shown in (16).

We seek now a set of values of N_j such that when it is realized, $\ln W$, and therefore W also, shall have a value stationary with respect to all variations δN_j conforming to the sole condition that the total number of atoms shall remain the same, which is to say, $\sum \delta N_j$ shall vanish.

It may be recalled that a similar problem arose in the old statistics. I treat it here in a more general and hardly less simple way, by writing the self-evident equation:

$$\delta \ln W = \sum \frac{d(\ln W_j)}{dN_j} \delta N_j \quad (18)$$

For the fulfilment of our wish it is a sufficient condition that all of the derivatives on the right-hand side should have the same value; since than $\delta \ln W$ will be $\sum \delta N_j$ multiplied by a constant, and when one vanishes so will the other. For this it is in turn a sufficient condition that all of the independent variables N_j should have the same value.

Uniform spreading of the atoms among the regions, with equal numbers in all regions of equal size, is therefore the condition in which $\ln W$ has a "stationary" value, which as always is assumed to be a maximum value. With the new definition of probability, the state of uniform spreading becomes the most probable in the new statistics, as with the old definition of probability it was in the old.

We go into the momentum-space to see whether the Maxwell-Boltzmann law results from the new statistics.

The momentum-space is now to be divided into regions of equal size,

each large enough to comprise a great number C of cells and small enough so that the function E of equation (3) may be deemed sensibly constant throughout it; E_j shall stand for the value of E appropriate to the region. (It is convenient to imagine the regions as layers separated from one another by concentric spheres having the origin for their common centre). For each of the regions W_j , the number of ways in which the most probable distribution of the cells among the possible populations can be realized, is given again by (16); and W , the number of ways in which the most probable distribution can simultaneously be achieved within each of the regions, is given again by (17).

We now seek a set of values of N_j such that when it is realized, $\ln W$ shall have a value stationary with respect to all variations δN_j conforming to two conditions: first, that the total number of atoms shall remain the same, which is to say, $\Sigma \delta N_j$ shall vanish; and second, that the total energy of the gas shall remain the same, which is to say, $\Sigma E_j \delta N_j$ shall vanish.

Referring back to (18), we see that for the fulfilment of our wish the following is a sufficient condition:

$$\frac{d(\ln W_j)}{dN_j} = P + QE_j \quad (19)$$

P and Q standing for constants; for when these substitutions are made into every term of the summation on the right of (18), the expression to which $\delta \ln W$ is there equated may be regrouped into one term proportional to $\Sigma \delta N_j$ and one proportional to $\Sigma E_j \delta N_j$, and vanishes when it ought to vanish.

Gone is the comfortable ease with which we disposed of the corresponding problem in the ordinary space! There we did not even have to know what sort of function W_j is of N_j ; whatever it might be, we were able to conclude that N_j must be the same for every region. Here the outcome must depend upon the functional relation between W_j and N_j . There is, however, no ground for apprehension, for though the function in question looks rather involved in equation (16), its derivative is surprisingly simple, and we come with ease to the condition which we seek:

$$\ln(N_j + C) - \ln N_j = P + QE_j \quad (19)$$

which may be rewritten thus:

$$\frac{C}{N_j} = -1 + e^{P+QE_j} \quad (20)$$

This is *not* the Maxwell-Boltzmann law, but approaches that desired law in what I will call the "limit of extreme rarefaction," where the number of cells in the region exceeds manifold the number of atoms. As C/N_j grows

greater and greater, the first term on the right recedes into relative insignificance; and with an ever-increasing degree of approximation, we have:

$$N_j = NAe^{-BE_i} \quad (21)$$

with NA put for Ce^{-P} and B for Q —which is the Maxwell-Boltzmann law and the law confirmed by experiment.

A helpful and troublesome coincidence between two different quantities

When in the earlier article I used the section-heading repeated just above, it referred to the near-equality between the logarithms of two different numbers: one number being that of all the complexions compatible with the *most probable distribution* (of numbered atoms sprinkled among numbered cells in ordinary space) and the other being that of all the complexions compatible with *all conceivable distributions altogether*. The most probable distribution had so great a share of all conceivable complexions, that no grave error was committed in pretending (so long as we were dealing with $\ln W$) that it actually had them all without exception!

A similar coincidence occurs in the new statistics, and will now be set forth.

Consider the j th region by itself. In (15) I have given the expression for $\ln W_{\max}$, the logarithm of the number of ways in which the *most probable distribution* of cells among populations can be realized. This is now to be compared with $\ln W_{\text{tot}}$, the logarithm of the total number of ways in which *all possible distributions* of cells among populations can be realized. Note that I say "all possible" and not "all conceivable" distributions! The only possible ones are those which are compatible with the fixed number N_j of atoms. This limitation prevents us from proceeding by the easy route of the earlier article. Indeed in order to solve the problem "in how many ways can all possible distributions of cells among populations be realized?" it is necessary, or at any rate customary, to restate it in a very different manner, which is the following:

In how many different ways can N_j un-numbered balls be distributed among C numbered baskets? Two ways are considered as different unless $n_i = n'_i$ for every value of i (n_i and n'_i standing for the populations of the i th basket in the two ways).

Notice that again the balls stand for the atoms and the baskets for compartments in space, as they did in the old statistics! We are playing a new game with the old baskets and the old balls, instead of playing the old game with new balls and new baskets as we have just finished doing. It has to be a new game, for the numbers have been removed from the old balls and the old game is therefore unplayable.

This is, to put it mildly, one of the less perspicuous problems of the

"theory of probability." I cannot do better than repeat, with slight changes in wording, the process of solution given by the Mayers in their book.² "A distribution is characterized by the number of balls in each of the numbered baskets, since the balls are indistinguishable. Consider the arrangement in a line of the symbols $z_1, z_2, \dots, z_C, a_1, a_2, \dots, a_N$, as for instance,

$$z_1 a_2 a_4 z_8 a_8 a_9 z_5 z_4 z_9 a_6 \dots$$

Such an arrangement could be used to define an assignment of N numbered balls, the a 's, to C numbered baskets, the z 's, by adopting the convention that the balls to the right of each numbered z belong to the basket of that number. For instance, the above corresponds to balls 2 and 4 in basket 1, balls 8 and 9 in basket 8, no balls in baskets 5 or 4, and ball 6 in basket 9. One must observe the convention that the row starts with a z , and we shall consider only arrangements of the symbols which start with z_1 . However, $(C - 1)!$ such arrangements of the symbols correspond to one arrangement of numbered balls in the same baskets, since permutations of the $(C - 1)$ groups of each z_i with its following a 's correspond to the same arrangement of numbered balls in the baskets. In addition all $N!$ permutations of the a 's correspond to the same distribution of un-numbered and indistinguishable balls in the baskets. In all, each distribution of the indistinguishable balls among the numbered baskets corresponds to $N!(C - 1)!$ arrangements of the symbols, and the $(N + C - 1)$ symbols (after the first) may be arranged in $(N + C - 1)!$ different ways."

Thus we come to the formula³ for W_{tot} , the total number of ways in which *all possible distributions* of the cells of a region among the populations can be realized; it is,

$$W_{\text{tot}} = \frac{(N_j + C - 1)!}{N_j!(C - 1)!} \quad (22)$$

Dropping the "ones" for the amply sufficient reason that they are insignificant by comparison with N_j and C , and taking the logarithm with use of the super-Stirling approximation, we find:

$$\ln W_{\text{tot}} = (N_j + C) \ln (N_j + C) - N_j \ln N_j - C \ln C \quad (23)$$

which with a little regrouping of terms is found to be the very same expression appearing in (15) for $\ln W_{\text{max}}$.

² J. E. and M. G. Mayer, "Statistical Mechanics" (John Wiley & Sons, 1940); p. 438. Reprinted by permission.

³ For the historian of science it is interesting to note that the formula (22) was used by Planck in his earliest derivation of the black-body radiation law. His un-numbered balls were quanta of energy, his baskets were linear oscillators, and his $k \ln W$ was the entropy of the system of C oscillators sharing N_j quanta among themselves. Cf. *Naturwissenschaften*, April 2, 1943.

I review the situation. We began by dividing space (ordinary space, or momentum-space, or μ -space) into what were called cells in the earlier article and are now designated "regions." We wanted to reach, as most-probable-distribution of the atoms among the regions, the uniform spread in ordinary space and the Maxwell-Boltzmann law in momentum-space. To do this by the method of the new statistics, we divided each region into many "cells." The first stage of the argument then consisted in taking a typical region, and ascertaining the most probable distribution of the cells among the populations. We then evaluated W_{\max} , the number of ways in which this distribution could be realized. Inserting $\ln W_{\max}$ into the argument we continued into the second stage, and attained the wanted result. But now it turns out that in the first stage we might have omitted to ascertain the most probable distribution of the cells among the populations in the typical region. Anybody might win through to the same desirable outcome without even suspecting that there is a most probable distribution of cells among populations. All he needs is to evaluate W_{tot} , the number of ways in which all possible distributions of cells among populations within the region can be realized. He may then replace $\ln W_{\max}$ by $\ln W_{\text{tot}}$, and proceed with the second stage as before. Since the two logarithms are practically equal, the outcome is the same.

There are accordingly two routes to the result, which do not merge until the argument is carried partway to the conclusion. Is one of them right and the other wrong? Or to ask a milder question: is either to be preferred to the other?

So far as I can see, neither can be proved wrong, and the question must be asked in the milder form. For myself I stand by the preference exhibited in this article, for the basic reason that along this route each of the stages of the argument consists in finding a most probable distribution: first for the cells among the populations of each region by itself, and then for the atoms among the regions. By the other route the two stages are differently handled, since in the first stage one considers all the distributions (of cells among populations in each region by itself) and then in the second stage the most probable distribution (of atoms among regions). There is also the minor advantage, that the value of W_{\max} is much easier to derive than the value of W_{tot} , or at least so it seems to me.⁴ However, many physicists of eminence have preferred the second route. Anyone may say of course that the question is foolish, since the number of complexions subsumed under the most probable distribution is so large a fraction of the total number of complexions altogether that no danger arises from confusing them. This is what the equations have been saying, and now I have said it again in words.

⁴ It was the other way about in the somewhat similar case which was treated in the earlier article.

Yet any policy which leaves this basic law unsaid, or even fails to emphasize it, is (I think) a bad one for all but the very few to whom it is already obvious.

Another question: may the old statistics be regarded as the limiting form of the new statistics, in the limiting case of "extreme rarefaction" where in every region the number of atoms is very much smaller than the number of cells?

It may seem that this question has already been answered with a *yes*, in view of the fact that in the limiting case the new statistics gives the same distribution-law—in momentum-space as in ordinary space—as the old one does. Nevertheless the answer is *no*. Mathematically this appears in the following way. In the old statistics the Maxwell-Boltzmann law springs from the denominator $\prod N_j!$ in the right-hand member of equation (1), which turns up as $(-\sum N_j \ln N_j)$ in the expression for $\ln W$. Now we look at the equation (23) and see the term $-N_j \ln N_j$ appearing with several other terms in the right-hand member. In the limit of extreme rarefaction it outweighs all the others and survives by itself. We summed it over all the regions and so arrived again at $(-\sum N_j \ln N_j)$, from which again the Maxwell-Boltzmann law emerged. But in this method of the new statistics each term of the summation comes by itself from the corresponding region, whereas in the method of the old statistics the whole summation arrived upon the scene *en bloc* or all in a single piece. The former method does not pass into the latter method in the limiting case. The conclusions agree in the limit, but the methods do not.

I have mentioned this because not infrequently one finds in print the careless statement that the old statistics is the limiting case of the new statistics, or words to that effect. Actually one can find more potent ways of contradicting that statement, as for example by emphasizing that the old statistics numbers the atoms and the new one leaves them un-numbered, and in no way can the one policy be regarded as a limiting case of the other. More convincing yet would it be to show that the new statistics and the old lead to results which definitely differ even in the limiting case of extreme rarefaction. This is what I next undertake to show as an incident of the explanation of entropy which the new statistics affords.

THEORY OF ENTROPY

For a substance of a single kind in a single phase, the basis of thermodynamics is the single equation,

$$dU = TdS - PdV \quad (31)$$

in which there are five variables: pressure P , volume V , absolute temperature T , energy U and entropy S . Two may be varied independently, and

any two of the five may be taken as these two, the remainder becoming the dependent variables.

From (31) we deduce, to begin with,

$$T = (\partial U/\partial S)_V \quad (32)$$

an equation which shows that if ever someone sets up a theory in which entropy is expressed as a function of energy and *vice versa*, it is *per se* a theory of absolute temperature. This, however, will find its due place later. What is of instant importance is a second deduction,

$$(\partial S/\partial T)_V = T^{-1}(\partial U/\partial T)_V \quad (33)$$

for making use of which we take note of the fact (not explicitly stated till now) that when the volume does not change no mechanical work is done upon or by the substance, and therefore all of the change in energy is that brought about by the inflow or outflow of heat. This fact is expressed in another equation,

$$(\partial U/\partial T)_V = H_v \quad (34)$$

H_v standing for the amount of heat that must be fed into the substance to raise its temperature, at constant volume, by one degree—the “heat-capacity at constant volume,” as some would call it. Combining the two,

$$(\partial S/\partial T)_V = H_v/T \quad (35)$$

Envisage now the entropy S as a function of volume and temperature, and view the equation:

$$dS = (\partial S/\partial V)_T dV + (\partial S/\partial T)_V dT \quad (36)$$

An equivalent for the coefficient of dT has been provided, and now it is needful to find one for the coefficient of dV . To do this we use the function $(U - TS)$, to be denoted by A , which by aid of (31) is seen to have the following differential:

$$dA = -PdV - SdT \quad (37)$$

Out of this one draws the following two deductions,

$$(\partial A/\partial V)_T = -P, \quad (\partial A/\partial T)_V = -S \quad (38)$$

Differentiating both sides of each of these equations, the former with respect to T while holding V constant, the latter with respect to V while holding T constant, one gets two expressions for what is one and the same quantity, to wit, the second derivative $\partial^2 A/\partial T\partial V$. Equating these two expressions, and saying goodbye to A which has fulfilled its purpose, one has,

$$(\partial S/\partial V)_T = (\partial P/\partial T)_V \quad (39)$$

which by the way is one of four equations collectively known as Maxwell's relations—a memorial of Maxwell's creative work in early thermodynamics, as the name of the distribution-in-energy law is of his work in early statistics.

Substituting from (35) and (39) into (36), we find:

$$dS = (\partial P/\partial T)_v dV + (H_v/T)dT \quad (40)$$

a usable and a useful expression for the entropy S —usable, that is to say, to anyone who knows the heat-capacity H_v and the derivative $(\partial P/\partial T)_v$, as functions of volume and temperature, for the substance in question.

Now take it on faith that there exists a gas having the following qualities: *first*, its pressure and volume and absolute temperature are linked together by the equation,

$$PV = LT \quad (41)$$

L being a constant; whence follows,

$$(\partial P/\partial T)_v = L/V \quad (42)$$

second, if the attempt is made to express its energy U as function of T and any one of the remaining variables (to wit, P or V or S), then the latter variable drops right out of the picture, leaving U as a function of T alone. I suppose this seems a needlessly longwinded way of saying that U does not depend on P or V , but it is necessary to provide for the fact that U may be expressed as (say) a function of S and V , whereupon it will be found that neither variable drops out of the picture. This is one of the features that make the science of thermodynamics very like a maze.

third, the heat-capacity H_v is independent of all the variables.

With these stipulations, (36) becomes:

$$dS = (L/V)dV + (H_v/T)dT \quad (43)$$

integrating which, we readily find that for the peculiar kind of gas presented just above as an article of faith, the entropy is given by the formula:

$$S = L \ln V + H_v \ln T + C \quad (44)$$

The symbol C stands for one of the most useless things in the world: an arbitrary additive constant of integration. The only purpose normally served by such a constant is, to prevent people from thinking that the equation is right if the constant is left off. Its presence means that the absolute value of S is undeterminable, is beyond the reach of experiment to determine. Nevertheless this constant is one of the principal themes of statistical theory; and we shall see that in defiance of what I have just said, and no part of which I retract, it *does* make sense to assign a particular value to this constant, and remarkably good sense at that.

I propose to begin very soon on the proof that the new statistics, applied to a flock of atoms which are merely mass-points, gives an excellent description of just such a gas. However, there is a detail, or rather an element of the structure, waiting to be inserted correctly—a trivial one in appearance, but in all of thermodynamics and all of statistics there is nothing further removed from the trivial. It is the dependence of the entropy on the quantity of the substance, the dependence of S upon N the number of atoms. To put a question seemingly so simple that it almost answers itself: given two samples of the same kind of gas under identical conditions, one comprising twice as many atoms as the other, what is the ratio of their entropies?

This is a remarkable question, because it seems so absurdly simple and is actually so very complex.

Before the advent of statistical theory, anyone versed in thermodynamics would probably have answered it by replying either that S is proportional to N , or that the question has no meaning. The first reply is suggested by the consequences of the fact that thermodynamics proposes no way of measuring the entropy of a gas (or other substance), but only ways of measuring the entropy-difference between two states. Let V_1, T_1 and V_2, T_2 stand for the values of V and T in two states of one gas. Equation (43) informs us that the entropy-difference is $L \ln (V_2/V_1) + H_v \ln (T_2/T_1)$. The constant C has vanished; the remaining terms are proportional to N because L and H_v are proportional to N . The entropy-difference is therefore proportional to N . It seems reasonable to conclude that S is proportional to N , but so long as there is no specific assertion about C the conclusion is not binding; and the proper reply is actually, that the question has no meaning.

But the statistical theories do make assertions about C , and the question is on the verge of acquiring a meaning; so it might be a good idea to ask in advance what sort of answer we should like to have. It seems natural to expect S to be proportional to N , so that the "double sample" shall have twice the entropy of the "single sample" under identical conditions. But what are "identical conditions?" Here is the catch. No more than two of the three variables P, V, T can be made the same for both the samples. I suppose that almost anyone would choose T for one of these two, so unpalatable would it seem to expect the double sample to have twice the entropy of the single sample if their temperatures differed. But after this is decided, shall we make V the same for both, and accordingly give doubled pressure to the double sample? or shall we make P the same for both, and accordingly give doubled volume to the double sample?

This is no mere quibble, for the choice will determine the dependence of C on N .

The first alternative requires that C be proportional to N . This is

obvious from inspection of (44) when one remembers that L and H_s are both proportional to N . I rewrite that equation accordingly, and put N_0 for the number of atoms in a gramme-molecule, R and C_v and RK_0 for the values of L and H_s and C appropriate to N_0 atoms:

$$S = (N/N_0)R \ln V + (N/N_0) C_v \ln T + (N/N_0)RK_0 \quad (45)$$

S is doubled if N is doubled while V and T stay the same, which is what was intended.

This equation will not suit the second alternative; for if V is doubled along with N , S will be more than doubled. Over and above the doubling, S acquires an extra term $2(N/N_0) R \ln 2$. Now if the constant C just happened to include a term $-2(N/N_0) R \ln 2$, the extra term would be obliterated, and S would just be doubled if N and V were to be doubled while P and T remained the same. Such a term is provided by replacing (45) with the equation:

$$S = (N/N_0)R \ln V + (N/N_0)C_v \ln T - (N/N_0)R \ln N + (N/N_0)RK_0 \quad (46)$$

where the last two terms on the right are to be regarded as forming the constant C . This then is the dependence of C on N which is demanded by the second alternative.

To guide the choice between the two alternatives there is, so far as I know, but the one argument; it is, however, a powerful one, and seems likely to hold the field unchallenged.

We have been thinking of two samples of identical gas at identical temperature. Think of them now as divided by a removable partition. When the partition is taken away, what happens? If the initial pressures are not the same, there is a swirling and a surging, dying away in time into a state in which the pressure is the same throughout the volume now common to the samples, but is not the same as it was before in either separate gas. This is just the sort of trend of events with which one likes to think that an entropy-change, and indeed an entropy-gain, is linked. Notice also that if the partition is replaced, the state of affairs on either side does not become the same as it was before! But now suppose the initial pressures to be the same. The partition can be removed and replaced without entailing any perceptible change in the gas such as one likes to associate with a change in entropy.

The second alternative is in harmony with these facts, the first is not. So to the question "is the entropy of a gas of $2N$ atoms double the entropy of a gas of N atoms?" the acceptable answer is: "yes, if the volume of the double gas is twice that of the single gas, their temperatures being the same." Now, this is also the answer given by the new statistics; for as we shall presently see, it leads to a formula like (46). It is not the answer given by the old statistics, which (as I said in the earlier article) leads to a formula

like (44). This is one of the dominant reasons for preferring the new statistics to the old.

Now I proceed to the theory of entropy and temperature derived from the new statistics.

NEW STATISTICAL THEORY OF ENTROPY

Entropy is identified with the quantity $\ln W$, multiplied by a constant k which as yet is disposable:

$$S = k \ln W \quad (47)$$

It is now to be shown that for the picture of a gas which is a flock of mass-points in the "most probable state" as defined by the new statistics, and in the limit of extreme rarefaction, this expression becomes the same as (46), with further consequences of much value.

As in the previous article, I separate the entropy into S_v the "contribution of volume to entropy" which springs from the sprinkling of the mass-points in ordinary space, and S_m the "contribution of temperature to entropy" which springs from the sprinkling of the mass-points in momentum-space. This is an artificial separation and worse than artificial, for it leads to a fault in a detail which is not trivial. Nevertheless I think that for ease of exposition the procedure is justified, and the detail will be made correct at the end of the argument.

We must now take (16) down to the "limit of extreme rarefaction." I repeat this equation:

$$\ln W_j = (N_j + C) \ln (N_j + C) - C \ln C - N_j \ln N_j \quad (16)$$

The journey toward the limit is menaced by some of the oddest pitfalls, and must be travelled with care. I recall that by Taylor's expansion, $\ln (N_j + C)$ is equal in first approximation to $(N_j/C + \ln C)$ when N_j is small by comparison with C . Making this substitution into (16), one finds that the right-hand member consists of six terms. The two largest of these, $C \ln C$ and $-C \ln C$, destroy one another. The smallest, N_j^2/C , is to be neglected (if we couldn't neglect it, the dependence of entropy upon N would be hopelessly misrepresented). All of the remaining three terms must be kept, for even the smallest—which is N_j —will play a perceptible part in the check of theory with experiment. We have:

$$\ln W_j = N_j \ln C - N_j \ln N_j + N_j \quad (48)$$

The quantity $\ln W$ is the summation of $\ln W_j$ over all the regions. Notice that we are interpreting entropy in such a way, that the entropy of the gas in the container is the sum of the entropies of the portions thereof in the individual regions. This is why we are destined to come to a result

in harmony with the "second alternative" aforesaid, wherein the total entropy of two gases of identical P and T is the sum of their separate entropies! It is otherwise in the old statistics, and that was the source of the troubles of that elder theory. But to proceed:

$$\ln W = \sum \ln W_j = N \ln C - \sum N_j \ln N_j + N \quad (49)$$

In the ordinary space and in the most probable state, N_j is the same for all the regions, as we have found already (page 363) and therefore is equal to N divided by V/V_0 , or to NV_0/V ; here V_0 is the volume of the *region* (not the cell!). The next step is to put this into (49), realizing now that each term becomes the same and the whole summation is V/V_0 times the typical term. One is agreeably surprised to find that V_0 tumbles out of the expression: this is a feature of the new statistics—the regions have but an intermediate and an auxiliary quality, the size assigned to them is gone from the final equations. In its place appears the volume of the *cell*, which is V_0/C , and which I denote by q_c . For S_c we have:

$$S_c = Nk \ln V - Nk \ln N - Nk \ln q_c + N \quad (50)$$

Note the last three terms, for future comparison with the two last of (46); but at this moment note especially the first, and compare it with the first of (46). Entire agreement is attained by assigning to k the value,

$$k = R/N_0 \quad (51)$$

as in the old statistics. The "Boltzmann constant" k is the "gas-constant" R divided by the "Avogadro number" N_0 .

Seeking now the "contribution of temperature to entropy," S_m , we turn to the momentum-space. Here the most probable distribution is given by (21), and is to be inserted into (50):

$$\begin{aligned} \ln W &= N \ln C - \sum N_j \ln N_j + N \\ &= N \ln C - N \ln A + N \sum ABE_j e^{-BE_j} + N \end{aligned} \quad (52)$$

It will be recalled from the earlier article, or failing this can easily be seen, that,

$$N \sum A^{-BE_j} = N, \quad N \sum AE_j e^{-BE_j} = U \quad (53)$$

U standing as heretofore for the total energy of the gas. The expression (52) is simplified of aspect, and multiplying it by k , we find for S_m ,

$$S_m = k \ln W = kN \ln C - kN \ln A + kBU + kN \quad (54)$$

Though I have spoken of this as the contribution of temperature to entropy, the temperature is nowhere to be seen! It is waiting on the doorstep; but before allowing it in, I wish to operate on the quantity $\ln A$.

This quantity, by (53), is given thus:

$$-\ln A = \ln \sum e^{-BE_j} \quad (55)$$

The so-called "partition-function," which is the sum appearing on the right, is made of terms contributed one from each region, the term from the j th region being $\exp(-BE_j)$ —herein E_j stands (it is proper to say) for the average value in the j th region of the function,

$$E = (1/2m) (p_x^2 + p_y^2 + p_z^2) \quad (56)$$

If the regions were of unit volume, this summation would be (approximately) equal to the triple *integral* of $\exp(-BE_j)$ over the whole of momentum-space. But the volume of each region is $C_j q_m$, wherein q_m stands for the volume of the *cell*. This signifies that while the integral contains one term for each unit of volume, the summation comprises $1/C_j q_m$ terms for each unit of volume. The summation is accordingly $1/C_j q_m$ times as great as the integral. Denoting the integral by I , we have in place of (55),

$$-\ln A = \ln I - \ln q_m - \ln C_j \quad (57)$$

and in place of (54),

$$S_m = kN \ln I - kN \ln q_m + kB U - kN \ln N + kN \quad (58)$$

We note with satisfaction that the size of the *region* has disappeared, even as it did while we were operating in ordinary space!

The next step is to consult a table of definite integrals for the value of the integral I (or to work it out one's self, if one's memory of the mathematical technique is vivid). The tables give:

$$I = (2 \pi m/B)^{3/2} \quad (59)$$

Before returning the table of integrals to the library, the student should also look up the value of the definite integral $\int_0^\infty x^2 e^{-ax} dx$; for with its aid he will be able to find a very simple relation between B and U . I have already said that either determines the other, and now for this special case we shall find the relationship. The procedure consists in going back to the second of equations (53), realizing that

$$\sum E_j \exp(-BE_j) = (1/C_j q_m) \iiint E \exp^{-(BE)} d p_x d p_y d p_z \quad (60)$$

and performing the triple integration over the whole of momentum-space, a feat which is not so hard as it looks. Multiplication by A , as indicated in (53), removes the factor $(1/C_j q_m)$, and the simple conclusion is,

$$U = 3N/2B \quad (61)$$

a relation valuable in two ways.

In the first place, (61) enables us to eject either B or U from the expression for S_m . It seems more sensible to do away with B , leaving S_m expressed as a function of the energy of the gas; but I will let the reader do that for his own instruction. In view of the peculiar significance of B this is for our purposes the better one to keep.

Now it is high time indeed to show what is that peculiar significance. Differentiating S with respect to U , we find kB for the derivative. Turning back to (32) we are reminded that this derivative is $1/T$ by the definition of the absolute temperature T . Now the temperature has stepped across the threshold, and S_m assumes the form:

$$S_m = (3/2) Nk \ln T - Nk \ln N + Nk \ln [(2\pi mk)^{3/2} e^{5/2}/q_m] \quad (62)$$

Notice that the term $+ kN$ has been absorbed into the final term, so that $e^{3/2}$ has been replaced by $e^{5/2}$ in the argument of the logarithm: this is a usage with which the student must become familiar. (Some writers also incorporate $-Nk \ln N$ into this final term, which thereby acquires a factor N in the denominator of the argument; the term then ceases to be a constant, which is why I do not follow this policy.)

Comparing (62) with (46) we see that S_m embodies correctly the dependence of entropy on temperature, provided that C_v (the specific heat per gramme-molecule) is equal to $(3/2)kN_0$. Since a value for k —to wit, R/N_0 —has already been forced upon us as a necessary and a sufficient condition for making S depend correctly on the volume, this new requirement is that C_v should be equal to $(3/2)R$. Now this is a fact of experience for the gases called monatomic!

I said that the relation of U and B expressed in (61) is valuable in two ways. The second is only the first seen from a different viewpoint, for which I rewrite (61) in the form:

$$U/N = U = (3/2)kT \quad (63)$$

For the flock of mass-points distributed in momentum-space in the manner indicated as the most probable by the new statistics (as, for that matter, by the old) the average energy is $(3/2)k$ times the absolute temperature. This is the very result obtained from simple kinetic theory for the ideal-gas scale of temperature. The statistical theory therefore identifies the absolute scale of temperature with the ideal-gas scale, which is as it should be. It is therefore an adequate theory of temperature and (as we lately saw) of the specific heat of monatomic gases.

Now I have given an expression for S_v , the "contribution of volume to entropy," which is (50); and an expression for S_m , the "contribution of temperature to entropy," which is (62); and it seems natural to proceed by

adding the two and identifying their sum with the entropy of the gas. But each of the summands contains the so-ardently-wanted term $-Nk \ln N$, and therefore the sum must contain a term $-2Nk \ln N$, which is $-Nk \ln N^2$. This term is not at all of the wanted form, and its mere presence in $(S_m + S_e)$ spoils the chance of identifying that sum with entropy. We have in fact come to a result in contradiction with equation (46) and with the assumption on which that equation was founded, viz. that if two samples of a gas are at the same pressure and temperature their entropies are in proportion to their volumes. The reasoning has not been suited to its aim.

The origin of this final misadventure lies in the circumstance that in the hope of making easier the exposition, I made what now has proved to be an undue separation between the two "contributions" to the entropy. The gas was mentally divided into groups of atoms, each occupying a certain region of limited size and distribution among the cells of that region according to the law of the new statistics. In computing S_e I defined the j th region as a small piece of ordinary space, and then counted all the atoms in that region regardless of the fact that they have very diversified momenta. In computing S_m I defined the j th region as a small piece of momentum-space, and then counted all the atoms in that region regardless of the fact that they are sprinkled all through the total volume of the container. I may properly say that I used a six-dimensional region throughout, but in the first stage it was a region limited in ordinary space and comprising the whole infinity of momentum-space, while in the second stage it was a region limited in momentum-space and comprising the whole volume of the box in ordinary space. I should instead have carried through the operation in a single stage, using a six-dimensional region limited in both ordinary space and momentum-space. It may seem that this procedure must either lead to the same result as the other, or must be much more difficult, or both. Neither is the case.

Instead of writing down a number of new equations which would look precisely like the old ones, I invite the student to go back to page 368 and recommence the argument at the words "We go into the momentum-space. . . ." If he will replace "momentum-space" by " μ -space," he need make no other change as far along as equation (21); the argument is just the same. Now let him turn ahead to page 379, and equation (52): this is valid for the μ -space as it was for the momentum-space, and so are equations (53). The novelty, however, is latent in the first of equations (53), which reappears as (55), and which I now rewrite for one more time:

$$-\ln A = \ln \sum e^{-BR_j} \quad (64)$$

On page 380, the summation was shown to be equal to $(1/C_j q_m)$ times a certain integral denoted by I ; the integral was over the three dimensions of momentum-space; q_m was the size of the elementary cell in momentum-

space. In μ -space, however, the corresponding integral is over the six dimensions, and may be written thus:

$$\iiint dx dy dz \iiint e^{-\beta E} dp_x dp_y dp_z$$

This sixfold integral is nothing but the product of V the volume of the container (resulting from the first three integrations) by the integral heretofore denoted as I (resulting from the last three integrations). It is to be multiplied by $(1/C_j h^3)$, C_j now standing for the number of cells in the six-dimensional region and h^3 for the volume of the six-dimensional cell (I explain the curious symbol later). The product is the reciprocal of A , and therefore:

$$k \ln W = kN \ln V + \ln I + kB U - kN \ln N + kN - kN \ln h^3 \quad (65)$$

The term $(-kN \ln N)$ appears just once, and not twice as it did in the sum S_e and S_m : all is well in this regard. The presentation of I , of B and of U as functions of T follows just the same lines as above.

(Notice, for future reference, that we should have attained to the same result had we ignored the ordinary space, operated in the momentum-space exclusively, and assigned the value h^3/V to the volume of the elementary cell in momentum-space.)

So, identifying $k \ln W$ with S , we come to the consummation of the new statistical theory of entropy, the equation:

$$S = kN \ln V + (3/2)kN \ln T - kN \ln N + kN \ln [(2 \pi m k)^{3/2} e^{5/2} / h^3] \quad (66)$$

The dependence on volume is right; it was qualitatively so to start with, was made exactly so by choice of the value of k as R/N_0 . The dependence on temperature is exactly right, since it is a fact of experience that for monatomic ideal gases the specific heat at constant volume is $(3/2)R$ per gramme-molecule. The dependence on number of atoms is exactly right, that is to say, it makes S proportional to N for given P and T . The additive constant is fixed in value absolutely, or will be when we assign a numerical value to h^3 ; for k is a universal constant, e the base of natural logarithms, and m the mass of an atom of the gas.

For the benefit of such as may still be interested in comparing the old statistics with the new, I recall that the old statistics in its theory of entropy furnished the first and the second terms of (66), and apparently furnished also the fourth term though with $e^{3/2}$ in place of $e^{5/2}$. The third term it omitted, thereby lending itself to the untenable doctrine that entropy should be proportional to N for given V and T (and not for given P and T). Since there was no term $kN \ln N$, I committed no error when in the previous article I deduced S_e and S_m separately and then added them together to get S .

This procedure is right in the old statistics, becomes wrong in the new. I suppose that this is what some expositors mean when they say that in the new statistics there is a correlation between positions and momenta, or words to that effect. I say that the old statistics *apparently* furnished a term equal to the fourth of (66) except for the power to which e is raised. Actually the old statistics gives an additive term $Nk \ln [(2\pi mke)^{3/2}/Q]$ and the new statistics gives an additive term $Nk \ln [(2\pi mk)^{3/2} e^{5/2}/Q]$, but Q in the former case is the volume of the *region* and in the latter case is the volume of the *cell*. Giving the same value h^3 to Q in the two cases is positively *not* doing the same thing. However by doing this notwithstanding, and by "tampering" with the old statistics in a certain way which I described at the end of the previous article, it is possible to produce an expression exactly like (66).

SIZE OF THE ELEMENTARY CELL

We have reached the final step, which consists in assigning a value to the size of the elementary cell in μ -space. For this I have used the symbol h^3 , implying (as everyone has guessed already) that it is taken to be the cube of Planck's constant h so promiscuously found in Nature. What arguments can be advanced to justify this choice?

It may be remarked very simply, that since the volume of the elementary cell has the "dimensions" of the cube of the product of length by momentum, and since these are also the dimensions of h^3 , and since both h^3 and that volume are very fundamental things, what could be more natural than to identify them the one with the other? This was the argument used when formula (66) was first derived from the old statistics with the aid of judicious tampering.

An argument more precise of aspect may be adduced from wave-mechanics. Imagine the box containing the gas to be a cube, its edges—these being of length L so that $L^3 = V$ —being along the coordinate-axes x , y , z . The doctrine of wave-mechanics avers that the momentum-components p_x , p_y , p_z of any atom are perforce integer multiples of $h/2L$; for this is the condition that the waves which are associated with the atom shall form a stationary wave-pattern with nodes at the walls of the cube, and upon this condition wave-mechanics is insistent.⁵ Now let us reenter the momentum-space, and place a dot at every point for which p_x , p_y and p_z are integer multiples of $h/2L$. The dots form a cubic lattice, and it would seem very

⁵ The wave-length of the waves associated with a particle moving parallel to the x -axis is h/p_x , and there must be an integer number of half-wave-lengths between the walls of the cube which are perpendicular to the axis of x and face one another at a distance L . The same may be said, *mutatis mutandis*, of a particle moving parallel to the axis of y or z , with momentum p_y or p_z ; while if an atom is moving obliquely so that two or all three of its momentum-components differ from zero, each of these components is to be treated as if it alone existed.

nice if I could say at once that the elementary cube of this lattice has the volume h^3/L^3 which is h^3/V . However, this cannot be said, for there is an obstinate factor which makes the elementary cube have the volume $h^3/8L^3$. People get around this by remarking that since an atom reverberating to and fro between the walls of the cube changes the sign of one of its momentum-components whenever it strikes against one of the walls, therefore every dot is one of a group of eight dots all of which correspond to the same motion of the atom, and all eight should be counted as though they were one.⁶ Therefore in the region of momentum-space enclosed between any two spheres centred at the origin (such as we used in determining the distribution-in-momentum) we are to count one-eighth of the dots. The number so obtained is the same as the number of cells of volume h^3/V contained in the region. Thus it comes to the same thing whether one says that the atoms are distributed among one-eighth of the dots or among cells of volume h^3/V . Now I recall my remark (page 383) that equation (66) down to the last detail can be derived by playing the game of balls and baskets by the rules of the new statistics *in the momentum-space alone*, provided that to the elementary cell in this space we assign the volume h^3/V . For doing this last, wave-mechanics has now offered a kind of retroactive basis. There seem to be flaws in the basis, but they are of a kind which cannot be mended (if at all) without a thorough study of a very hard subject, to wit, the art of interpreting wave-mechanics in the ordinary language of space and time.⁷ I think it will be better to proceed at once to the test by experiment.

TEST BY EXPERIMENT OF THE NEW STATISTICAL FORMULA FOR ENTROPY

Enough has been said already to cover the first three terms of the formula (66), which correctly give the dependence of entropy S upon volume V , temperature T , and number of atoms N . The present question is: what does experiment say of the fourth term, the additive constant which involves the mass m of the atom and the universal constants k and h ?

Having treated this question at length in the June 1942 issue of this Journal, I will here give only the barest outline. For this purpose I rewrite (66), by the aid of the equation of state of the perfect gas,

$$PV = NkT \quad (67)$$

⁶ If (a, b, c) are the coordinates of one dot, those of the other seven of its group are: $(a, -b, c)$; $(a, b, -c)$; $(a, -b, -c)$; $(-a, b, c)$; $(-a, -b, c)$; $(-a, b, -c)$; $(-a, -b, -c)$.

⁷ In previous pages I said that the proper way of playing the game of balls and baskets is to play it in the six-dimensional space, with N_i representing a definite number of atoms located in a six-dimensional region which is composed of a narrowly-limited region in ordinary space and another narrowly-limited region in momentum-space. Wave-mechanics, in the current interpretation, will not allow this; it claims that, if the N_i atoms are located in a limited region of momentum-space, they are spread all over the box containing the gas.

so as to give entropy S as function of *pressure* and temperature:

$$S = -kN \ln P + (5/2)kN \ln T + kN \ln \left[\frac{(2\pi m)^{3/2} (ke)^{5/2}}{h^3} \right] \quad (68)$$

Notice that here every term is strictly proportional to N , in accordance with the "second alternative" of page 377.

Let P and T be so chosen that the gas is in equilibrium with its solid crystalline phase. To keep this choice in mind, I will replace T by T_s , signifying "temperature of sublimation" at pressure P . Let the N atoms of gas now be cooled to the absolute zero. First they will condense, still at temperature T_s , into the crystalline solid. In so doing they will disgorge the "heat of sublimation," L per gramme-molecule, amounting to NL/N_0 ; and their entropy will decline by NL/N_0T_s , since the process is reversible. Let the cooling continue. As the crystal declines in temperature from any T down to $(T - dT)$, it disgorges heat in the amount of $(N/N_0)C_p dT$ and entropy in the amount of $(N/N_0)(C_p/T)dT$; here C_p stands for the specific heat (per gramme-molecule) of the crystal. The pressure is supposed to remain the same throughout the entire process. When the crystal arrives at absolute zero, its entropy has the value:

$$S_0 = -kN \ln P + (5/2)kN \ln T_s + kN \ln \left[\frac{(2\pi m)^{3/2} (ke)^{5/2}}{h^3} \right] - (N/N_0)(L/T_s) - (N/N_0) \int_0^{T_s} (C_p/T) dT \quad (69)$$

The right-hand member of this equation embodies the new statistical theory of entropy. If on the left I put the value *zero* for S_0 , I express what is known as "Nernst's Heat Theorem" or the "Third Law of Thermodynamics." If experiments say that the right-hand member of (69) is equal to zero, they ratify not indeed the statistical theory by itself or the Third Law by itself, but the assumption that both are true. Now, this is what the experiments *do* say. Better to describe the situation, they say that the first three terms on the right of (69) are equal to the last two terms with sign reversed. All of the noble gases have been tested with suitable accuracy, and eight or nine of the metals with accuracy not so high, yet better than "order-of-magnitude accuracy." For further details I must refer to my article already cited.⁸

⁸ I cannot refrain from mentioning a detail of the statistical theories, which is amusing if one sees it at once and confusing if one sees it belatedly (mine was the latter experience). It pertains to the power to which e is raised in the third term on the right in (69). If in the new statistical theory we leave out the term N in (49), thus stopping with a first approximation instead of going on to the second, we arrive ultimately at $e^{3/2}$ instead of $e^{5/2}$. If in the old statistical theory as modified by Tetrode we use the first-order Stirling approximation instead of the second-order one for $N!$, we arrive ultimately at $e^{3/2}$ instead

THEORY OF RADIATION

Black or total radiation, which is the electromagnetic radiation within a cavity enclosed by walls at a uniform temperature, may be regarded as a monatomic gas of which the atoms are called "photons." It has two peculiarities. First, the relation between energy and momentum is not the same for a photon as for a material atom. If by p I represent the magnitude $\sqrt{p_x^2 + p_y^2 + p_z^2}$ of the momentum, then the energy E is given no longer by the familiar equation (3), but rather by this one:

$$E = pc \quad (70)$$

c standing, of course, for the speed of light. This is no insignificant change, but recedes into secondary importance when compared with the other contrast. Not only the distribution-law for the photons, but the actual total number of photons itself, is fixed by Nature when the temperature of the walls of the cavity is fixed by the observer. To the quantity called N , the number of atoms in a container of volume V , no specific value has ever yet been assigned in these pages; for with a material gas it may be raised or lowered at will, by pumping gas into or out of the box. In this section, however, it will have to have a value, for Nature has given it one.

Can the theory achieve what Nature demands of it? It can, and this is the way.

The momentum-space is divided as heretofore into regions of equal volume, each containing C cells of volume q_m . A distribution is described by giving the number of photons in each region, N_j standing for the number in the j th region. The probability W of a distribution is given as always by the formula (16) and this is it:

$$\ln W = \sum \ln W_j = \sum [(N_j + C) \ln (N_j + C) - C \ln C - N_j \ln N_j] \quad (71)$$

We are not now proceeding to the limit of extreme rarefaction! Radiation presents itself to us under conditions remote from this limit, and must be treated without recourse to the approximation hitherto used in these pages.

When the quantities N_j are altered by the small amounts or "variations" δN_j , W undergoes the slight alteration or variation given thus to first approximation:

$$\delta W = \sum \left(\frac{\partial W}{\partial N_j} \right) \delta N_j = \sum [\ln (N_j + C) - \ln N_j] \delta N_j \quad (72)$$

of $e^{5/2}$ (see the text preceding equation (35) of the prior article on page 134 of the January issue of this Journal). Thus in both cases we arrive at $e^{3/2}$ or $e^{5/2}$, according as we pause at a first approximation or go on to a second; *but* I discern no mathematical or physical similarity whatever in the two situations in which these approximations are made.

In the quest for the "most probable distribution" this quantity is required to vanish for variations which are controlled by a certain condition.

On a previous page (368) where we were dealing with *material* atoms in *ordinary* space, the sole condition was that the total number of atoms should remain the same (and equal to N). This led to the uniform distribution, $N_j = \alpha$; here α stands for a constant, which turns out to be the product of N by the ratio of the volume V_0 of the region to the volume V of the box.

On another previous page (369), where we were dealing with *material* atoms in *momentum* space, the condition imposed was twofold: that the number of atoms N and the total energy U of the atoms should remain the same. This led to the distribution (20), which in the limit of extreme rarefaction became the Maxwell-Boltzmann or canonical distribution $N_j = NA \exp(-BE_j)$; here E_j stands for the energy-value appropriate to the j th region, and A and B for two constants which were shown to be determined by N and U .

In this case where we are dealing with *photons* in *momentum* space, the condition which leads to the right result is simple but surprising. We must admit only such variations as leave the total energy constant, but we must *not* require that the total number of photons should likewise remain the same. Applying this strange condition, we find it taking the form,

$$\ln(N_j + C) - \ln N_j = BE_j \quad (73)$$

with only one constant, which is going to be controlled by the total energy U . Rewriting this:

$$\frac{N_j}{C} = \frac{1}{e^{BE_j} - 1} \quad (74)$$

One sees immediately that N , which is the sum of all the quantities N_j , is no longer at liberty to take whatever value the experimenter pleases!

Hitherto I have assumed that all the regions are of equal volume, but I can free myself from this assumption by pointing out that N_j/C is the average number of photons per cell in the portion of momentum-space where E has the value E_j . Now let us carve up the momentum-space into regions separated by spherical shells all centred at the origin. The region extending from the sphere of radius p to the sphere of radius $p + dp$ will be of volume $4\pi p^2 dp$, and will accordingly contain $4\pi p^2 dp/q_m$ cells, if by q_m I denote the volume of a cell. The appropriate value of E will be pc . The number of photons in the region will accordingly be given thus:

$$dN = \frac{4\pi p^2}{q_m} \frac{1}{e^{Bpc} - 1} dp \quad (75)$$

Unrecognizable as it may seem, this is actually a statement about the spectrum of black radiation! This is because a photon of momentum p and energy pc is associated with light-waves of wave-length given by the "Rule of Correlation":

$$\lambda = \frac{h}{p} \quad (76)$$

If I therefore multiply both members of (75) by pc , I have an expression for the amount of energy associated with the waves ranging in wave-length from h/p to $h/(p + dp)$.

There are instruments able to sort out the waves of different wave-lengths with their associated photons; they are called spectroscopes. There are instruments able to indicate the total energy borne by the photons thus sorted out; they are called by such names as bolometer and thermopile. There are people able to use these instruments; and so (75) can be tested. It is customary to rewrite (75) so that either wave-length or frequency becomes the independent variable, in place of p ; but nothing would be gained for the purpose of this article by doing so. The fact of experience is, that (75) is a correct description of black radiation provided that three modifications be made:

a) For q_m we are to write h^3/V , presuming that this comes to the same as though we had operated in six-dimensional space and put h^3 as the volume of the elementary cell therein (page 383);

b) For B we are to put $1/kT$;

c) We must double the right-hand member of (75), the factor 2 being ascribed to the fact that light is polarizable.

Making these modifications, and putting $V = 1$ so that the forthcoming equation shall refer to the radiant energy contained in *unit* volume, we have

$$dN = \frac{8\pi p^2}{h^3} \frac{1}{e^{pc/kT} - 1} dp \quad (77)$$

for the number of photons in unit volume endowed with momenta between p and $p + dp$, energies between cp and $c(p + dp)$. This is the distribution-formula for black radiation of temperature T , commonly known as "Planck's law."

To have derived this law is the first, the great and the historic achievement of the new statistics. Other ways have indeed been found for deriving it, beginning with Planck's own; but the way of the new statistics is smoothest and quickest. Quite different is this story from that of the theory of material gases! There, the distribution law was correctly given by the old statistics long before it was tested. Here, the distribution-law was found

by experiment years before it was explained, and a great puzzle it was. There, the old statistics and the new (in the limit of extreme rarefaction) led to the same result. Here the old statistics was impotent, and the new had to be invented.

Reverting to the identification of B with $1/kT$: this may be proved in the following way⁹. Refer back to equations (74) and (71), and for ease of operation write x_j for $(e^{B E_j} - 1)^{-1}$. We then have:

$$N_j = C x_j \quad (78)$$

$$S = k \ln W = k \sum \{C(1 + x_j) \ln [C(1 + x_j)] - C x_j \ln [C x_j] - C \ln C\} \quad (79)$$

$$U = \sum N_j E_j = C \sum E_j x_j \quad (80)$$

Differentiate S with respect to B and do the like with U , and divide the former derivative by the latter, so as to get the derivative dS/dU . It will be found that this is equal to kB ; and since by definition of absolute temperature it is also equal to T^{-1} , the identification is made.

THE BOSE-EINSTEIN AND THE FERMI-DIRAC STATISTICS

Hitherto in this article, except for one protective allusion, I have spoken as if the new statistics were one and indivisible. There are, however, two branches of it, known respectively as the Bose-Einstein statistics and the Fermi-Dirac statistics. It is the former of which I have treated throughout this essay. The point at which the latter branches off is to be found on page 365, where I introduced the game of balls and baskets, the balls standing for cells and the baskets for populations. On reaching this point the game is to be played with the supplemental assumption that there are only two baskets, those numbered 0 and 1. That is to say: a cell may either be empty or may contain a single atom, but never more than one.

I leave to the student the task of revising equations (7) to (16) accordingly, but I take it upon myself to point out how easily the problem can be solved by the second method—that of pages 370–71, the method involving the counting of *all* the different ways in which un-numbered atoms can be distributed among numbered cells. In the Bose-Einstein case the fundamental formula is (22), which is not very easy to derive. In the Fermi-Dirac case we proceed by playing anew the game of balls and baskets. There are but the two baskets, one being set out to receive the balls corresponding to the empty cells and the other for the cells containing one atom each—the “filled cells,” we may call them. There being in the j th region N_j atoms and C cells all together, the first basket is destined to contain $(C - N_j)$ balls and the second to contain N_j . The question is then: in how many

⁹I am indebted for this proof, as well as for much other assistance in the preparation of the article, to Dr. L. A. MacColl.

ways can C numbered balls be distributed among two baskets, these to contain $(C - N_j)$ and N_j of the balls respectively, two ways being considered as different unless the inventory of each basket is just the same for both ways? But this is the problem set up and solved in the earlier article, though there the balls stood for atoms and the baskets for regions. The answer is:

$$W_{\text{tot}} = C! / (C - N_j)! N_j! \quad (85)$$

instead of equation (22). Using the first or second order Stirling approximation—it doesn't matter which—one comes to the analogue of (23), which is:

$$\ln W_{\text{tot}} = C \ln C - (C - N_j) \ln (C - N_j) - N_j \ln N_j \quad (86)$$

Different as this looks from (23), the two become alike in the limit of extreme rarefaction, and in this limit equation (48) expresses the result both of the Bose-Einstein and of the Fermi-Dirac statistics. Since equation (48) is the parent of the Maxwell-Boltzmann distribution-law and of the expression (66) for the entropy of a monatomic gas, both of these flow from either type of statistics, and experiment does not decide for either over the other.

When we avoid the limit of extreme rarefaction, the two forms of statistics *do* depart from one another. If photons obeyed the Fermi-Dirac statistics, the distribution-law for black radiation would not be (75). We should be obliged, in the denominator on the right-hand side of that equation, to replace the negative sign of the second term by the positive sign. In so doing we should contradict the data of experiment in an unmistakable way; and for photons accordingly, the Fermi-Dirac statistics is to be rejected.

This form of the new statistics being no better than the other for material gases, and definitely wrong for radiation, where is it to be preferred and why?

To answer the first question, I point to the "electron-gas" which pervades the metals and is accountable for their quality of being excellent conductors. Experiment (as I recounted in these pages fourteen years ago¹⁰) confirms that these intra-metallic electrons form a gas which obeys the Fermi-Dirac statistics. It is not, however, the limit of extreme rarefaction which here we meet but the opposite one, the limit of extreme condensation. These electrons are as densely concentrated as the atoms of the solid itself, a degree of condensation never even approached by any ordinary gases. In this limit the distribution-law attains a form entirely different from both the Maxwell-Boltzmann law and the black-radiation law, and very remarkable. I dare not, however, expose this article to the risk of a doubling in length, which a treatment of this topic would probably entail; and I can avoid it with a fairly clear conscience, for the experimental evidence that electrons

¹⁰ This Journal, 8, 672 (1929); also *Physical Review Supplement* (Reviews of Modern Physics) 1, 90 (1929).

obey the Fermi-Dirac statistics has been enlarged but little since my article of 1929.

As for the second question, I can give only the shadow of an answer. The reason for adopting sometimes the Bose-Einstein and sometimes the Fermi-Dirac statistics springs from wave-mechanics, and that requires an article of its own. I can say, without proof, that the choice depends upon the number of elementary particles in the atom. The gas is supposed to conform to the Bose-Einstein or the Fermi-Dirac statistics, according as that number is even or odd. An electron is an elementary particle all by itself, wherefore the preceding paragraph.

For material gases, the crucial number is obtained by adding up the numbers of the protons and the neutrons in the nucleus, and the number of orbital electrons which surround the nucleus and complete the atom. In nature the atoms for which the crucial number is even vastly outnumber those for which it is odd, and the Bose-Einstein statistics is therefore the prevalent one. The principal isotope of nitrogen and the second isotope of hydrogen do indeed belong to the rarer category, but in the gaseous state their atoms always pair themselves into diatomic molecules, a circumstance which restores these gases to the realm of Bose and Einstein. A detail in the band spectrum of a diatomic molecule is available for telling which form of statistics the individual atom would obey if free; it confirms what I have just been saying—but this is an intricate story.

Electromagnetic Waves

A Textbook* (530 pages, \$7.50 net) by S. A. Schelkunoff, published by D. Van Nostrand, Inc., New York City, 1943.

THIS new addition to a well-known series has been awaited with much interest by all those acquainted with Dr. Schelkunoff's contributions to propagation theory, and it will be found that their expectations have been entirely fulfilled. This monumental piece of work is equally remarkable for the originality and consistency of its approach as for the wealth of information contained in its five hundred densely packed pages.

The author's systematic use of the harmonic oscillation, with complex variables and coefficients, is in line with the marvelous development which has occurred in the communication field during the last fifty years. Alternating current theory, then acoustics, then vibrational mechanics successively dropped the differential equations which physics offered as a basis and systematically restricted themselves to harmonic oscillations. This has resulted in the replacement of the differential operator by $i\omega$, leading to a tremendous simplification of steady-state analysis, which has been reduced to the calculation of amplitude ratios and phase differences. The genuinely difficult problems have not disappeared for all that but are now relegated to Fourier or Laplace transform theory, and it has become apparent that an enormous field of application can be covered by purely algebraic processes.

Not the least advantage of this method has been the unification brought into the three chapters of technical science mentioned above. Electrical impedances gave the model after which acoustical and mechanical impedances were fashioned; and mixed mutual impedances, thereafter, made it possible to write the equations of electro-mechanical or acoustico-mechanical transducers. There was an exciting era of intense development in this field during the twenties; and it was amusing to hear at that time, and even a good deal later, irate die-hards denouncing "impedances" with bitter irony or viewing with alarm the spread of "analogies."

Dr. Schelkunoff has set about to carry this point of view into Electromagnetic Theory, and it may well be that his will be the honor of having brought into the fold of harmonic oscillation theory the last chapter of Physics which still had to be incorporated. (One might think of Optics, but of course half of the book is really Optics.) Having given, in the first pages of his fourth

*This review by P. Le Corbeiller is reprinted from *Quarterly of Applied Mathematics*, Vol. 1, No. 2 by permission of the editors. Mr. Le Corbeiller, until coming to the United States where he has joined the faculty of Harvard University, was an engineer with the Administration Française des Postes, Telegraphes et Telephones.

chapter, a short and quite personal derivation of Maxwell's equations (1-15, p. 69), Dr. Schelkunoff without taking breath adds immediately: "Since we are concerned primarily with fields varying harmonically with time, we replace the instantaneous field intensities and current densities by the corresponding complex variables and write Maxwell's equations as follows:

$$\begin{aligned}\int E_n ds &= -\iint i\omega\mu H_n dS - \iint M_n dS, \\ \int H_n ds &= \iint (g + i\omega\epsilon)E_n dS + \iint J_n dS.\end{aligned}\tag{1-16}$$

Thus the sacrosanct Maxwell equations are swept away with movie-like swiftness, and instead we have the steady-state equations of a medium characterized by a distributed series impedance $i\omega\mu$ and a distributed shunt admittance $g+i\omega\epsilon$ (p. 81).

The analogy with a transmission line whose series inductance is μ , shunt conductance g and shunt capacitance ϵ , all taken per unit length, is inescapable (p. 243). In particular the above primary constants simply beg to be transformed into the familiar secondary constants of transmission line theory; here the intrinsic propagation constant σ and the intrinsic impedance η are defined by

$$\sigma = \sqrt{i\omega\mu(g + i\omega\epsilon)}, \quad \eta = \sqrt{\frac{i\omega\mu}{g + i\omega\epsilon}}\tag{9-1}$$

(p. 81) (σ is in neper/meter, η in ohms; the book is written in MKS—p. 60). For free space we shall have $g = 0$, and the following numerical values of the fundamental constants (p. 82):

$$\begin{aligned}\text{impedance of free space } \eta_0 &\approx 120\pi \text{ ohms,} \\ \text{characteristic velocity } v_0 &\approx 3.10^8 \text{ meters/second.}\end{aligned}\tag{9-4}$$

* * *

Surprising as it may appear to transmission engineers and sound engineers, who daily handle their respective characteristic impedances Z_0 or ρc , there still are very competent physicists who balk at the idea of free space having a characteristic impedance of about 377 ohms. Yet, in the words of Professor Ronold W. P. King:¹ "The existence of such a characteristic resistance for electromagnetic effects is just as mysterious, but not more so, than the existence of the finite velocity v_0 ." Dr. Schelkunoff explains very well how this constant could have been overlooked by the builders of the classical theory: "The physicist concentrates his attention on one particular wave: a wave of force or a wave of velocity or a wave of displacement. His original differential equations may be of the first order and may involve both force and

¹ Mimeographed "Notes on Antennas" for the course of Electronics and Cathode Ray Tubes (Eng. 270), Harvard University.

velocity; but by tradition he eliminates one of these variables, obtains a second order differential equation in the other and calls it the 'wave equation.' Thus he loses sight of the interdependence of force and velocity waves . . ." (p. vii). Still, it is surprising to see that one has started with two constants ϵ_0 and μ_0 , and recognizing the fundamental importance of their product, yet has not enquired about their ratio.

Then, the reader will ask, how can the Theory of Relativity give a leading role to the velocity of light and not mention the impedance of free space. Has Einstein no use for η_0 ? Well, he has, and he has not. First, an essential point in Special Relativity is the merging of the magnetic and the electric fields into one skew-symmetrical tensor. When doing this in the MKS system, homogeneity requires the use of the components of E and of $\eta_0 H$; but the factor η_0 is not apparent, for instance, in the equations on p. 44 of "The Meaning of Relativity" (by A. Einstein, Princeton Univ. Press, 1923) which uses a system of units in which $\eta_0 = 1$. Secondly, if we try to connect the universal constant η_0 with other members of this interesting family, we find that η_0 times a (charge)² has the dimensions of "action," and more precisely that

$$\eta_0 e^2 = \frac{2h}{137}$$

(e = charge of the electron, h = Planck's constant). We see from this that there is more to η_0 than appears in Special Relativity, the first step in the successive Einsteinian extensions of Maxwell's theory.

* * *

We have dealt at length with this question of the "impedance of free space" because it exemplifies the spirit of the whole work. It occurs in the course of a short but apt presentation of the "Fundamental Electromagnetic Equations" (Chapter IV), immediately applied to harmonic oscillations. The book as a whole is devoted not to Electromagnetism in general but, as specified in the title, to *Electromagnetic Waves*.

Three preliminary chapters introduce the more advanced mathematical tools which will be used, but sparingly, in what follows: such topics as contour integration, Bessel and Legendre functions. Chapter V is a short and original presentation of Network Theory.

The central part of the book begins with Chapter VI, "About Waves in General," a sort of preview of the questions which will be treated in detail later, during which we are introduced to radiation from given currents, propagation along wave guides, and to such general tools as electric and magnetic current sheets, the method of images and conformal representation.

In the following four chapters, we meet the most thorough treatment

² See *Quarterly of Applied Mathematics*, 1, 78 (1943).

available of the propagation of waves, guided or bounded, in one, two and three dimensions. It is impossible to do justice here to the richness of the material, which must have cost tremendous labor and which is in great part taken from the author's own publications. We find in Chapter IX, however, classical problems of Fresnel optics, adroitly adapted to contemporary radio needs. Chapter XI is a relatively short treatment of antenna theory, principally of conical antennas, and in the last chapter we return to wave guides and solve various problems involving discontinuities, even to an iris or a transversal wire. This subject is still under development by the author, and the readers of the *Quarterly* have had the benefit of one of its recent extensions.²

The specialist in wave propagation has no need to be told of the value of this book; but the reviewer would like to explain to his fellow non-specialists why it is particularly important that they should not miss it. When the results of much present-day research will suddenly be made available, it will be a hard task to catch up, not only with the new knowledge, but still more with the new modes of attack. The borderland between radio and optics is one of the fields from which great things can confidently be expected. Dr. Schelkunoff's book is a great opportunity for those not at present engaged in research to get familiar with methods which they will want to use tomorrow.

P. LE CORBEILLER.

Abstracts of Technical Articles by Bell System Authors

*Paracon—A New Polyester Rubber.*¹ B. S. BIGGS and C. S. FULLER. Paracons are high molecular weight linear polyesters which are soft enough to be rubbery and are capable of undergoing a vulcanization reaction. They are prepared by the condensation of dibasic acids with glycols or by the self condensation of hydroxy acids, and the name is intended to signify "condensation rubber." Paracon looks and feels like rubber, and in comparison with other rubbers it has some distinct advantages and some definite limitations. Its outstanding properties are oil resistance, high heat and light resistance, lack of odor and fast curing cycle. Tensile strength ranges from 1500 to 3000 pounds per square inch with elongations of 400 to 600 per cent.

Rubber pigments and compounding techniques may be used with it but vulcanization is accomplished in most cases by the action of benzoyl peroxide rather than sulfur.

Aside from its practical aspects the development of paracon is of theoretical interest because of the light it throws on rubber structure and the mechanism of vulcanization processes.

*Unsaturation of Butadiene and Related Polymers as Determined by Iodine Chloride Addition.*² A. R. KEMP and HENRY PETERS. This paper describes procedures which have been developed to determine the unsaturation of various butadiene and related polymers and copolymers, as well as mixed vulcanizates of Buna S and rubber. These methods are based on the use of *p*-dichlorobenzene as a solvent and iodine chloride as the addition agent, following the general technique employed in the standard Kemp-Wijs method for the determination of the unsaturation of natural rubber.

The ratio of butadiene to styrene in copolymers has been calculated from the iodine value and from the carbon-hydrogen ratio; however, the accuracy of these procedures is subject to several variables which are discussed.

Unsaturation data are presented on highly purified emulsion-type polymers of butadiene-isoprene and butadiene-styrene which agree closely with the presence of one double bond for each diolefin molecule present. The reaction rate of Buna S with halogens is shown to agree closely to that of natural rubber hydrocarbon.

*Brittle Temperature of Rubber under Variable Stress.*³ A. R. KEMP, F. S. MALM and G. G. WINSPEAR. This paper supplies the need for a method to

¹ *Chem. and Engg. News*, June 25, 1943.

² *Indus. & Engg. Chemistry, Analytical Edition*, July 1943.

³ *Indus. & Engg. Chem.*, April 1943.

determine the temperature at which rubber and similar materials fracture under variable bending stress. Although the brittle temperature is sharply defined under high-speed bending through a sharp angle, it is lower as the speed of application or the magnitude of the stress is reduced. In some instances decreases of more than 28°C. in brittle temperature resulted from reductions in bending stress such as might be encountered in service.

Vulcanized pure gum natural rubber and plasticized polyvinyl chloride-acetate copolymer showed the largest changes, whereas the compounded and vulcanized natural and synthetic rubbers involved in this study exhibited a reduction in brittle temperature from 5° to 10° C. in going from the highest to the lowest stress employed.

*American Science Mobilizes for Victory.*⁴ ROBERT W. KING. There are no accomplishments of the Bell System in which its men and women take greater pride than those marking the continuous activities in developing and applying the art of communication.

The Bell Telephone Laboratories' accomplishments, reflected for decades in improved instrumentalities and systems for the transmission of electrical signals and speech, have been possible because vast resources of scientific knowledge have been devoted as part of the System's general responsibility to the public, to a broad and fundamental program of exploration, experiment and design.

Today the more than 6,000 members of these Laboratories are engaged on hundreds of development projects requiring research, invention and design, for the Army, the Navy, and the National Defense Research Committee.

That this should be both logical and inevitable will not surprise any one who considers the vital part played by communications in modern warfare. Rapid movement of troops and supplies over far-flung lines of action on land and sea and in the air are possible only when directed through effective communication systems. More and more the electrical transmission of intelligence is becoming the unifying influence pervading all branches of war organizations. It coordinates the movement of naval and aerial fleets; it enables infantry, tank columns and formations of aircraft to operate as a single unit. It shrinks a thousand-mile battle line to the compass of a single sector.

The article by Dr. King points out the place of independent military research, although its actual volume is less than that carried on directly by the Army and Navy. It also draws upon experience in industrial research to show that the sudden solution of war problems by appeal to science is scarcely to be expected.

⁴ *Bell Tel. Mag.*, June 1943.

*Filtered Thermal Noise—Fluctuation of Energy as a Function of Interval Length.*⁵ S. O. RICE. Let a source of thermal noise be connected to the input of a band-pass filter. Consider the energy which would be dissipated during the interval t_1 to $t_1 + T$ if the output current were to flow through a resistance of one ohm. When T is held fixed and t_1 regarded as a random variable, the resulting energies have a distribution whose average and standard deviation depend upon T . Here this dependence is studied. The standard deviation of the difference of the energies of two contiguous intervals, each of length T , is also obtained.

*Ultra-Short Electromagnetic Waves. IV—Guided Propagation.*⁶ S. A. SCHELKUNOFF. Doctor S. A. Schelkunoff presented the material contained in this article as a lecture before the basic science group of the New York Section. He treated the subject in a "non-mathematical" manner. Certainly the electrical engineer will welcome any concept which allows an easier approach to the solution of certain problems involved in wave guides than the more complete equivalent field-theory method. In this fourth article in a series of six on ultra-short electromagnetic waves, Doctor Schelkunoff combines transmission-line theory with optical analogy and derives useful relations for both wave guides and cavity resonators. The three preceding articles appeared in the March, April and May issues of *Electrical Engineering*.

*Variable-Frequency Bridge-Type Frequency-Stabilized Oscillators.*⁷ W. G. SHEPHERD and R. O. WISE. Results are given of a theoretical and experimental investigation into two types of bridge-stabilized oscillators incorporating a thermal device for amplitude control. One circuit employs only resistances and capacitances in the frequency-determining network and consequently is useful for low-frequency operation. The other circuit uses an inductance-capacitance network which is well adapted to the higher-frequency network. Conditions for optimum stability and the variation of the stability with frequency determined experimentally are found to be in general agreement with theoretical results.

*Beyond the Ultra-Short Waves.*⁸ G. C. SOUTHWORTH. This article reviews briefly the work done many years ago by the pioneering physicists with the so-called electric waves as well as the more recent efforts by engineers to put these waves to practical use. It also describes some of the expedients and changes of technic used to overcome difficulties as this work progressed to

⁵ *Jour. Acous. Soc. Amer.*, April 1943.

⁶ *Elec. Engg.*, June 1943.

⁷ *Proc. I. R. E.*, June 1943.

⁸ *Proc. I. R. E.*, July 1943.

higher and higher frequencies. One, of fairly recent origin, is the wave-guide or hollow-pipe technic. The latter not only provides a simple and efficient way of propagating microwave power from one point to another but there have also grown from it some very interesting counterparts of the tuned circuits, the matching transformers, and the filters that have been in common use for some time at the lower frequencies. The possible bearing of this new technic on the future of electrical communications, as, for example, television, is pointed out.

*The Impact of War on Long Distance Service.*⁹ MARK R. SULLIVAN. The article gives in narrative form much of Vice President Sullivan's testimony concerning toll board service before the Federal Communications Commission on December 16, 1942.

Increasing traffic and severe curtailment in additions to plant, make it progressively more difficult to maintain service performance at its usual level. In the last two years Long Lines toll traffic has nearly doubled, this increase being almost equivalent to the total level of business reached over a period of some 65 years. Plant materials available for telephone construction, on the other hand, have been sharply curtailed. Copper, for example, had been used in building telephone plant at the annual rate of more than 90,000 tons; now only about 8,000 tons are used.

The increased volume of long distance calls has been accompanied by increased complexities in the handling of calls; more attempts required per call, a greater proportion of person-to-person calls, and greater length of haul requiring more switching. Over all, the results still average well. However, not all calls fall on or near the average due to the uneven distribution of increased calling which has followed generally the path of war activity and varies tremendously in different localities. Just as growth in calling has not been uniform throughout the System, so has circuit congestion been most pronounced in cities most affected by war activities. A customer whose individual call is delayed well beyond the average may appraise the service by that call, even though he may realize that the average speed of all calls is much faster.

Despite the increased traffic volume, complexities and the shortage of facilities, there have been some truly notable achievements. Accuracy and speed of answer have suffered relatively little. Additions to the operating force of 150,000 were made in the past two years to handle the increased load and to replace losses. Operators, seasoned and new, have faced the challenge to the service and, true to the finest traditions of the service, are giving their best to a difficult job—willingly and cheerfully.

⁹ *Bell Tel. Mag.*, June 1943.

*Drying of Textiles.*¹⁰ A. C. WALKER. In 1937 a textile-drying research project was initiated by the United States Institute for Textile Research and supported by the textile industry, for the purpose of evaluating the effects of temperature and humidity on the physical and chemical properties of important textile fibers. The present paper gives a comprehensive report of the results obtained and points out the basis upon which the industry must proceed in problems relating to textile drying. In all cases, consideration should be given to the theories relating to the form in which the moisture is distributed within the fiber structure and to data of the type discussed by the author.

¹⁰ *Trans. A.S.M.E.*, May 1943.

Contributors to this Issue

ELIZABETH J. ARMSTRONG, M.A. in Geology, Bryn Mawr College, 1934; Ph.D., 1939. Lecturer in Geology, Barnard College, 1938-41; Research Assistant, Columbia University, 1941-42; National Research Council Fellow, 1942-43; Bell Telephone Laboratories, 1943-. Miss Armstrong has been working chiefly on problems related to the orientation of piezoelectric crystal plates.

RALPH B. BLACKMAN, A.B. in Physics and Engineering, California Institute of Technology, 1926. Member of technical staff, Bell Telephone Laboratories, 1927-. Engaged in research and consultation in applied mathematics.

WALTER L. BOND, B.S. in Physics, Washington State College, 1927; M.S. 1928. Member of Technical Staff, Bell Telephone Laboratories 1928-. Studied at Columbia University, New York University, and Stevens Institute. Engaged primarily in studies of the physical properties of crystals.

KARL K. DARROW, B.S., University of Chicago, 1911; University of Paris, 1911-12; University of Berlin, 1912; Ph.D., University of Chicago, 1917. Western Electric, 1917-25; Bell Telephone Laboratories, 1925-. As Research Physicist, Dr. Darrow has been engaged largely in writing on various fields of physics and the allied sciences.

K. G. VAN WYNEN, E.E., Cornell University, 1925; M.E.E., Brooklyn Polytechnic Institute, 1933; American Telephone and Telegraph Company, 1925-1934; Bell Telephone Laboratories, Inc., 1934-. Mr. Van Wynen has worked principally on special problems arising from transmission studies of the local telephone plant.

G. W. WILLARD, B.A., University of Minnesota, 1924; M.A., 1928; Instructor in Physics, University of Kansas, 1927-28; Student and Assistant, University of Chicago, 1928-30. Bell Telephone Laboratories, 1930-. Mr. Willard's work has had to do with special problems in piezoelectric crystals.