The Bell System Technical Journal

October, 1929

A Method of Sampling Inspection

By H. F. DODGE and H. G. ROMIG

This paper outlines some of the general considerations which must be taken into account in setting up any practical sampling inspection plan. An economical method of inspection is developed in detail for the case where the purpose of the inspection is to determine the acceptability of discrete lots of a product submitted by a producer. By employing probability theory, the method places a definite barrier in the path of material of defective quality and gives this protection to the consumer with a minimum of inspection expense.

ONE of the common questions in every day inspection work is, How Much Inspection? The answer must always be arrived at in the light of economy for only the least amount of inspection which will accomplish the purpose can be justified.

We wish here to consider the problem of setting up an economical inspection plan whose immediate purpose is the elimination of individual lots of product which are unsatisfactory in quality. By a lot of unsatisfactory quality is meant one that contains more than a specified proportion of defective pieces. This proportion is usually small and is based on economic considerations. The interest in inspections made for the elimination of such lots is shared by two parties,—the producer and the consumer. The consumer establishes certain requirements for the quality of delivered product. The producer so arranges his manufacturing processes and provides such an inspection routine as will insure the quality demanded. Our problem recognizes that the consumer runs some risk of receiving lots of unsatisfactory quality, if the quality of each lot is judged by the results of inspecting only a sample. The method of attack presumes the adoption of a risk whose magnitude is agreeable to the consumer and the selection of a particular inspection procedure which will involve a minimum of inspection expense on the part of the producer while guaranteeing the protection agreed upon.

Considering various possible concepts of a risk, the consumer would prefer to have one adopted that was worded as follows:

> "Not more than a specified proportion of the delivered lots shall be unsatisfactory in quality."

In other words, he would like to have the assurance that "the risk

of receiving a lot of unsatisfactory quality shall not exceed some definite figure." This risk involves two probabilities:

(*a*) that an unsatisfactory lot will be submitted for inspection.
(*b*) that the inspector will pass as satisfactory, an unsatisfactory lot submitted for inspection.

Without definite information regarding probable variations in the producer's performance and without an absolute assurance that this performance will remain consistently the same, the use of probability (*a*) in stating the risk to the consumer might be misleading. This probability will therefore receive no further consideration in this paper so far as the definition of the risk to the consumer is involved.

For probability (*b*) a reasonably low value for an upper limit can be given without any knowledge of the producer's performance. This upper limit (as defined below) has been taken as the starting point for the inspection method presented in this paper. The concept of risk which has been adopted gives the consumer unconditional assurance that his "chance of getting any unsatisfactory lot *submitted for inspection* shall not exceed some definite magnitude." This magnitude is the Consumer's Risk used.

Our problem may hence be stated as follows:

> Given a product of a specified type of apparatus or material coming from a producer in discrete lots, what inspection plan will involve a minimum of inspection expense, and at the same time insure that under no conditions will more than a specified proportion of the unsatisfactory lots submitted for inspection be passed for delivery to consumers?

This problem is economic in character and its solution involves the use of probability theory for establishing the height of the barrier to be placed in the path of unsatisfactory material.

Our attention will be directed particularly to the inspection of material which is produced more or less continuously on a quantity basis as distinguished from intermittent production in relatively small amounts. Under these conditions the producer is able to secure a continuing record of performance and to set up an inspection program which takes advantage of current quality trends.

General Considerations in Setting Up an Inspection Method

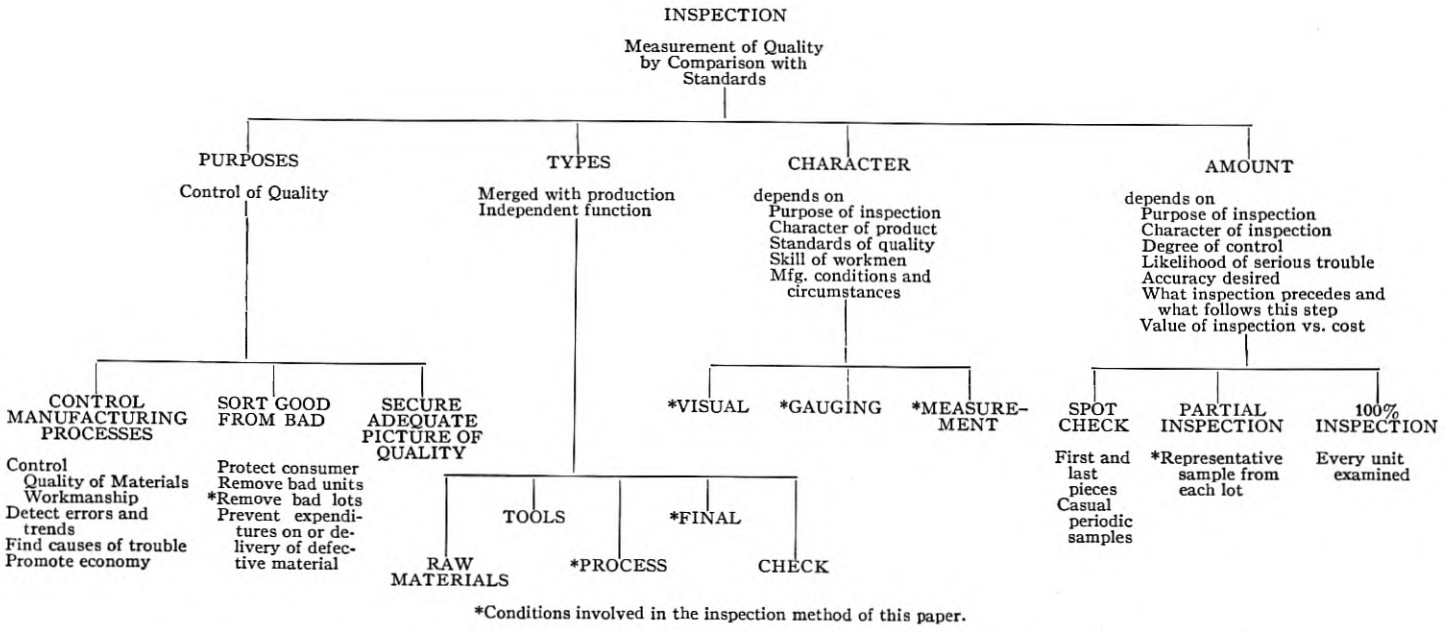
The broad purpose of inspection is to control quality by critical examinations at strategic points in the production process. Raw materials must be inspected. Some of the rough and finished parts must be inspected. In the manufacture of even the simpler kinds of merchandise, inspections dot the chart of progress from raw materials

to the finished product. The distribution of inspection activities throughout any process must be so ordered that the net cost of production will be consistent with the quality demanded by the customer. To determine whether an inspection should be made, or how much should be made at any one of the formative stages, is a major problem involving questions of both quality and economy.

One hundred per cent inspection is often uneconomical at a point in the production process where inspection is clearly warranted, particularly when preceded or followed by other inspections, inasmuch as the cost of more inspection at that point may not be reflected by a corresponding increase in the value of the finished product. In special cases, for example where inspection is destructive, 100 per cent inspection may be totally impracticable. Sampling inspections are often best from the standpoint of both the producer and the consumer when the value of quality and the cost of quality are weighed in the balance.

To arrive at an answer to the question, How Much Inspection, it is first essential to define clearly just what the inspection is intended to accomplish and to weigh all of the important factors both preceding and following the inspection in question which have a direct influence on the quality of the finished product and which as a whole determine how large a part this inspection step must play in controlling quality. Should it serve as an agency for making sure that the product at this stage conforms 100 per cent with the requirements for the features inspected? If so, 100 per cent inspection is required. Or should it serve to make reasonably certain that the quality passing to the next stage is such that no extraordinary effort would have to be expended on defective material? If so, sampling inspection may be employed. Ahead of all else, decisions are needed as to the specific requirements that must be satisfied by the inspection plan itself. This part of the problem is a practical one—one which must be approached in the light of experience, knowledge of conditions and the statistics of past performance. Once the basic requirements of the plan are agreed upon, probability theory can assist in formulating the details which will accomplish the desired results. It is important to hold in mind that statistical methods are aids to engineering judgment and not a substitute for it.

An attempt has been made in Fig. 1 to show schematically some of the outstanding general considerations which must be taken into account in establishing a proper setting for any problem that seeks to determine the economical amount of inspection. Inspections vary widely in purpose, type and character. While their broad purpose is



*Conditions involved in the inspection method of this paper.

Fig. 1—Chart showing some of the factors to be considered in establishing an inspection plan.

to control quality, the immediate objects of individual inspection steps differ. For example, the object of one step may be to secure information which will assist directly in controlling the manufacturing process by detecting errors or trends in performance which would become troublesome if allowed to persist unchecked. In other places the immediate object may be to determine the acceptability of definite quantities of product or to provide a screen for sorting the bad pieces or the bad lots from the good ones. Materials, parts in process and finished units are scrutinized with these objects in view. Depending on circumstances, the character and completeness of inspections vary from visual examination of small samples to careful measurement or testing of each piece.

Conditions Under Which the Present Method Applies

A large amount of industrial inspection work consists in comparing individual pieces with a standard—such as gauging a dimension or measuring an electrical property—to determine whether the pieces do or do not conform with the requirements given in specifications. This is often referred to as inspection by the "method of attributes." Consideration will be directed here to the case where non-destructive sampling inspection of this kind is conducted on discrete lots of product for the purpose of determining their acceptability.

From the standpoint of sampling theory, one of the general requirements is that each lot should be composed of pieces which were produced under the same essential conditions. Practically, this means that an attempt should be made to avoid grouping together batches of material, which, due to manufacturing conditions or methods, are apt to differ in quality. It is presumed, of course, that the sample drawn from any lot will be a random sample so that it may fairly represent the quality of the entire lot.

Summing up the general conditions for which a solution is sought, we assume

(1) The purpose of the inspection is to determine the acceptability of individual lots submitted for inspection, i.e., sorting good lots from bad.
(2) The inspection is made by the "method of attributes" to determine conformance with a particular requirement, i.e., each piece does or does not meet the limits specified.
(3) A lot is homogeneous in quality and the sample from it is a random sample.

The starred items in Fig. 1 indicate the set of conditions involved in our problem.

Protection and Economy Features of the Method

The adoption of sampling inspection at any stage of manufacture carries with it the premise that the product emerging from this point does not have to conform 100 per cent with specification requirements. It is often more economical, all things considered, to allow a small percentage of defective pieces to pass on to subsequent assembly stages or inspections for later rejection than to bear the expense of a 100 per cent inspection. Under these conditions, the status of the inspection can be clarified by establishing a definite tolerance for defects for the lots submitted to the inspector for acceptance. This may be specified as an allowable percentage defective, a figure which may be considered as the border line of distinction between a satisfactory lot and an unsatisfactory one. Thus, if the percentage defective is greater than this "tolerance per cent defective," the lot is unsatisfactory and should be rejected. We say "should be" rejected but this cannot be accomplished with absolute certainty if only a sample is examined. Sampling inspection involves taking chances since the exact quality of a lot is not known when only a part is inspected. According to the laws of chance, a sample will occasionally give favorable indications for bad lots which will result in passing them for delivery to consumers.

The first requirement for the method will therefore be in the form of a definite insurance against passing any unsatisfactory lot that is submitted for inspection.

The second requirement that will be imposed is that the inspection expense be a minimum, subject to the degree of protection afforded by the first requirement.

For the first requirement, there must be specified at the outset a value for the tolerance per cent defective as well as a limit to the chance of accepting any submitted lot of unsatisfactory quality. The latter has, for convenience, been termed the Consumer's Risk and is defined, numerically, as the probability of passing any lot submitted for inspection which contains the tolerance number of defects.

As will be shown further on, the first requirement can be satisfied with a large number of different combinations of sample sizes and acceptance criteria. To satisfy the second requirement, it is necessary then to determine the expected amount of inspection for a variety of inspection plans, determine the cost of examining or testing, add the costs other than those incurred in the simple process of examining samples, and choose among these plans that which involves a minimum of inspection expense.

There are, of course, a number of possible general methods of

inspection procedure, such as single sampling, double sampling, multiple sampling, etc., which allow the examination of only one sample, of two samples, or of more than two samples before a prescribed disposition of the entire lot is made. For each of these general methods, different combinations of sample sizes and acceptance criteria can be found which will satisfy the first requirement. We now prescribe that any lot which fails to pass the sampling requirements shall be completely inspected. Under this condition, one of the above mentioned combinations will give a lesser amount of inspection than the rest. Since a major cost item is that associated with the *amount* of inspection, we will carry through in detail the problem of finding the combination which will result in the *minimum* amount of inspection for one simple general method of inspection.

Single Sampling Method of Inspection

Attention is now directed to what is termed the Single Sampling method of inspection, which involves the following procedure:

(*a*) Inspect a sample.
(*b*) If the acceptance number for the sample is not exceeded, accept the lot.
(*c*) If the acceptance number is exceeded, inspect the remainder of the lot.

The term "Acceptance Number" is introduced to designate the allowable number of defects in the sample.

For this procedure, the first requirement reduces the problem to one which can be solved readily by determining probabilities associated with sampling from a finite lot containing the tolerance number of defects. For any sample size, there is a definite probability of finding no defects, of finding exactly one defect, exactly two defects, etc. If, under the above conditions, the acceptance number were 1, for example, there is one value of sample size such that the probability of finding one or less defects is equal to the value of the Consumer's Risk specified. Since a lot will be accepted if the observed number of defects does not exceed the acceptance number, the probability of finding one or less defects in a sample selected from a lot of tolerance quality is the risk of accepting any lot of tolerance quality submitted to the inspector. It follows that the risk of accepting a lot of worse-than-tolerance quality is less than the Consumer's Risk just defined. If the producer gets into trouble and begins to submit lots of unsatisfactory quality, the consumer has the assurance that his chance of getting them will not exceed this figure. In fact, the worse the quality, the less will be their chance of passing without a detailed

inspection. Thus the amount of inspection is automatically increased as quality degenerates.

For every acceptance number, such as 0, 1, 2, etc., there is an unique size of sample which will satisfy the specified values of tolerance per cent defective and Consumer's Risk. We thus have many pairs of values of sample size and acceptance number from which to choose.

The second requirement dictates which pair shall be chosen. We will select that pair which involves the least amount of inspection for product of *expected* quality. In industry, the quality emerging from any process tends to settle down to some level which may be expected more or less regularly day by day. If this level could be maintained quite constant, if the variations in quality were no larger than the variations that could be attributed to chance, then inspection could often be safely dispensed with. But practically, while such a level may be adhered to most of the time, instances of man failure or machine failure are bound to arise spasmodically and as a result the quality of the output may gradually or suddenly become unsatisfactory. The method of solution takes into consideration this usual or expected quality and requires an estimate of the expected quality under normal conditions. A satisfactory estimate of this can usually be obtained by reviewing data for a past period during which normal conditions existed and by utilizing such other pertinent information as bears on manufacturing performance under present or anticipated conditions. This expected value is defined as the *process average* to be used in the solution. Thus the method under discussion will assure the producer of a minimum of inspection expense so long as he holds to his expected performance. If he gets into trouble and the quality becomes poorer than normally expected, the method automatically increases the inspection by an amount which varies with the degree of quality degeneration. This reacts on the producer directly by increasing his inspection expense and serves as an incentive to the elimination of the causes of trouble. The producer's expected performance is thus made use of in a way that affects the economy of the producer's inspection work but is not used to color or affect the magnitude of the Consumer's Risk.

The amount of inspection, that will be done in the long run for uniform product [1] of process average quality is made up of two parts:

(1) The number of pieces inspected in the samples.
(2) The number of pieces inspected in the remainder of those lots which fail to be accepted when a sample is examined.

[1] By "uniform product" is meant one produced under a constant system of chance causes, giving rise to a quality which is a chance variable. In the present paper, this chance variable is assumed to be the Point Binomial.

But what proportion of the lots will fail to be accepted on the basis of the sampling results? Here is where probability theory comes in again. There will be a definite probability of exceeding the acceptance number in samples drawn from material of process average quality. Since we are interested in the amount of inspection *in the long run*, the sample at this stage of the problem may be regarded as drawn from a very large (mathematically infinite) quantity of homogeneous product whose percentage defective is equal to the process average per cent defective. Thus, for example, with an acceptance number of 1, the average [2] number of pieces inspected *per lot* as a result of *extended* inspections is equal to the number of pieces in the remainder of a lot multiplied by the probability of finding more than one defect in a sample drawn from an infinite quantity of material of this quality. This value plus the number of pieces inspected in the sample gives the average amount of inspection per lot for an acceptance number of 1. Similar results are found for all other acceptance numbers and the desired solution is obtained by choosing that acceptance number for which the average amount of inspection per lot is a minimum.

The plan thus provides the inspector with a definite routine to follow, such that his inspection effort will be a minimum under normal conditions.

Charts for Single Sampling

For any specified value of Consumer's Risk, charts may readily be constructed to give the acceptance number, the sample size, and the average number of pieces inspected per lot for the conditions outlined above. To illustrate the general character of these charts, Figs. 2, 3 and 4 are presented for a Consumer's Risk value of 10 per cent.

In the appendix it is shown that the acceptance number which satisfies the condition of minimum inspection is dependent on two factors, (1), the tolerance number of defects for a lot, and (2), the ratio of the process average to the tolerance for defects. Fig. 2 based on this relationship defines *zones* of acceptance numbers for which the inspection is a minimum.

Fig. 3 gives curves for finding the sample size. The mathematical basis for these curves is likewise given in the appendix. For a given tolerance number of defects and the acceptance number found from Fig. 2, the value of tolerance times sample size is determined. This quantity divided by the tolerance gives the sample size. The curves shown are based on an approximation which is satisfactory for practical use when the tolerance for defects does not exceed 10 per cent and the sample size is not extremely small.

[2] It is to be noted that wherever "average" appears in the paper, "expected" value, in the rigorous probability sense, is meant.

Fig. 4 gives curves which enable one to determine the minimum average number of pieces inspected per lot for uniform product of process average quality. For a given tolerance number of defects and a given ratio of process average to tolerance, a value of tolerance times minimum average number inspected per lot is determined. This when divided by the tolerance gives the desired value of minimum average number inspected per lot.

Illustrative Example

Suppose that lots of 1,000 pieces each are inspected for a characteristic having a specified tolerance of 5 per cent defective and that the process average quality of submitted lots is 1 per cent defective. If it is desired to have a risk of 10 per cent of accepting a 5 per cent defective lot, what single sampling plan should be followed by the inspector to give a minimum amount of inspection, and how much inspection per lot will be required on the average?

Referring to Fig. 2, for "Tolerance Number of Defects" (.05 $\times$ 1,000) = 50 and "Ratio of Process Average to Tolerance" (.01/.05) = .20, we find the acceptance number = 3.

Referring to Fig. 3, for "Tolerance Number of Defects" = 50 and "Acceptance Number" = 3, we find "Tolerance Times Sample Size" = 6.5. Dividing by the tolerance (expressed as a fraction defective) = .05, gives a sample size of 130.

Referring to Fig. 4, for "Tolerance Number of Defects" = 50 and "Ratio of Process Average to Tolerance" = .20, we find by interpolation that the "Tolerance Times Minimum Average Number Inspected" = 8.2. Dividing by the tolerance, .05, gives an average number inspected per lot of 164.

This solution has thus been obtained by the initial specification of first, the tolerance per cent defective for a single lot and a value for the Consumer's Risk, these two factors being combined to give a definite measure of protection against passing faulty material, and second, a minimum amount of inspection for product of process average quality.

These two requirements control the exact details of inspection procedure and must be initially chosen on the basis of practical considerations and circumstances. The specification of these factors lends definiteness to the problem of inspection and provides a rationalized basis of procedure which can be depended on to give the desired degree of protection. Obviously, any value of Consumer's Risk may be chosen according to circumstances. The value which is proper in any case is dependent on the conditions associated with the product

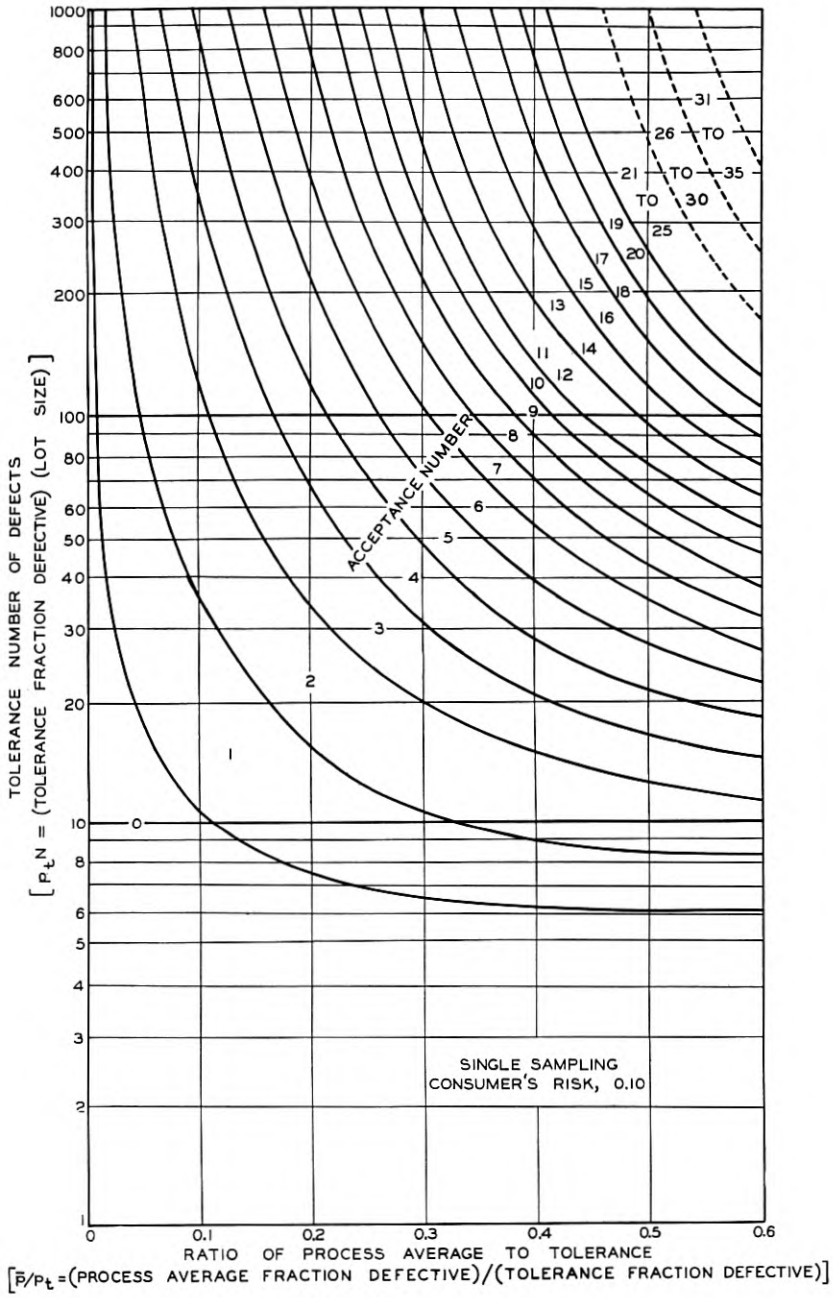


Fig. 2—Chart for finding acceptance number.

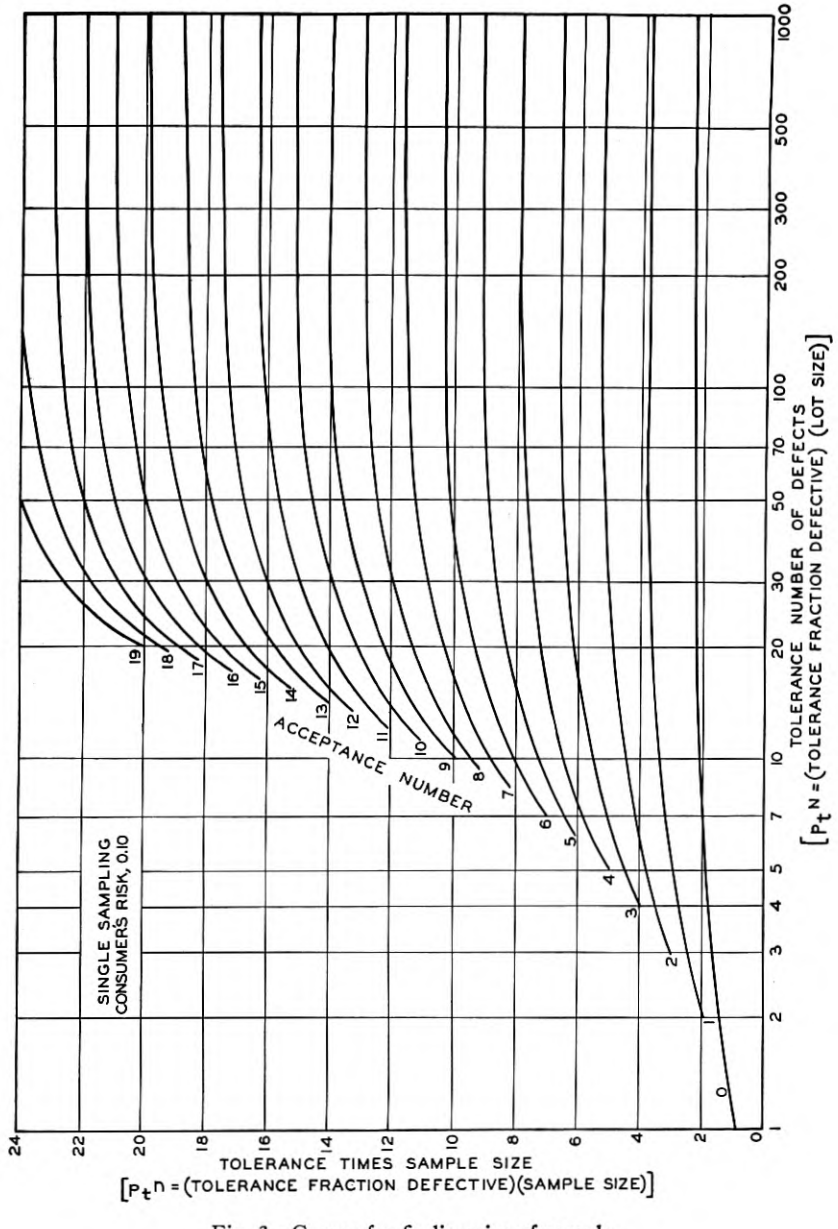


Fig. 3—Curves for finding size of sample.

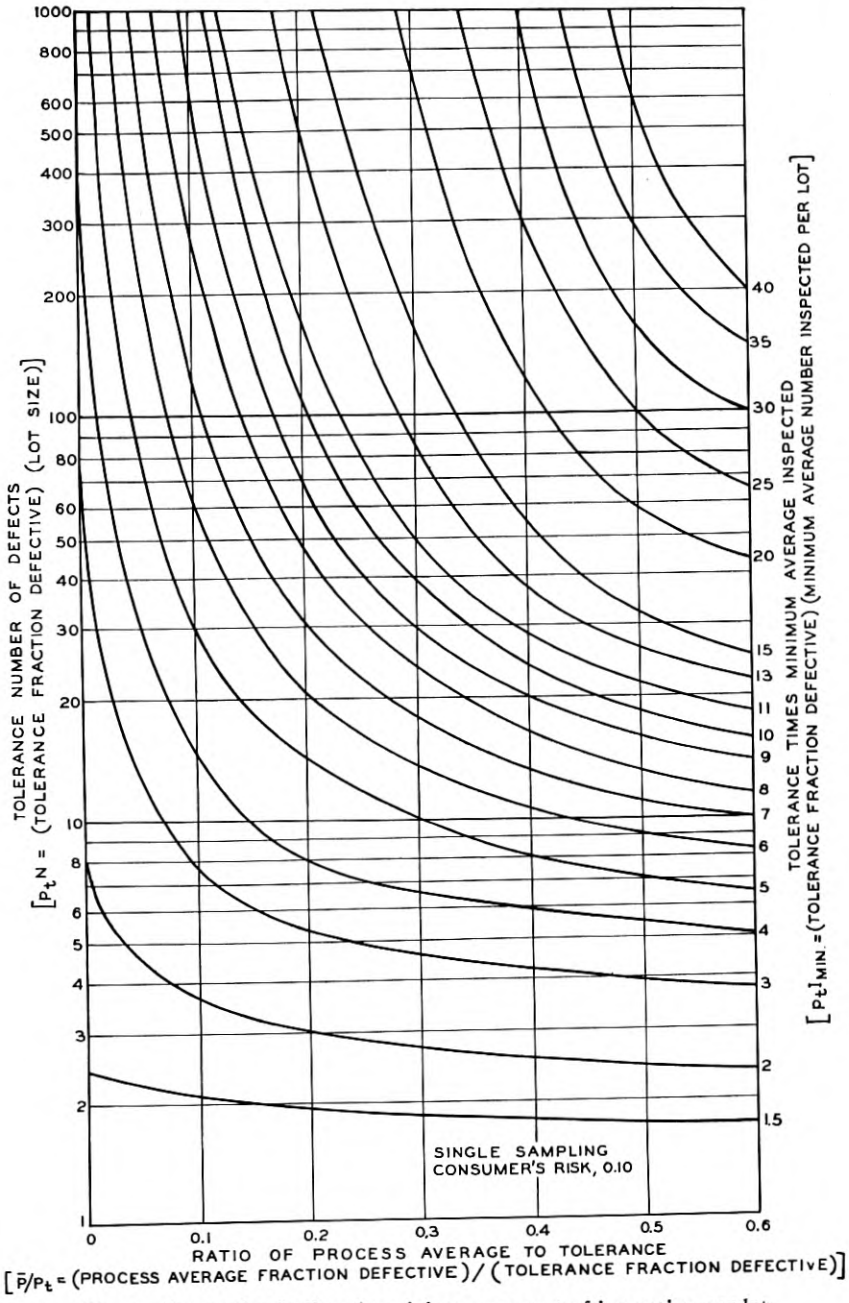


Fig. 4—Curves for finding the minimum amount of inspection per lot.

inspected, such as the degree of control exerted by the producing agencies, or the protective measures which precede or follow the inspection step in question. For any Consumer's Risk value, sets of charts similar to Figs. 2, 3, and 4 may be constructed.

Other Methods of Inspection

The above detailed discussion has been limited to one simple method of inspection, Single Sampling, in order to show how certain concepts and principles may be applied with the aid of sampling theory. The same principles are readily extended to a plan of double sampling or of multiple sampling in which cases a second sample may be examined if the first fails or a third, fourth, etc. examined if the preceding samples fail, before resorting to a detailed inspection of the remainder of a lot. As a matter of fact, for given values of tolerance and risk, the minimum average amount of inspection per lot will be somewhat less for plans which permit the examination of more than one sample before detailing, but when consideration is given to the costs associated with interruption of work, extraction of additional random samples, inconveniences or difficulties in handling the routine called for, etc., it has not been found economical in general to examine more than two samples from any lot.

It may be well to point out that other basically different requirements may be chosen for setting up economical sampling inspection plans. For example a satisfactory method has been devised to meet the following requirements:

(1) A limiting value to the average per cent defective after inspection.
(2) A minimum amount of inspection for product of process average quality.

This method has been found of value in continuous production where the inspection is intended to serve as a partial screen for defective units. It differs from that described above in that it provides a fixed limit to the *average* quality of product after inspection rather than a limit to the quality of each individual lot.

The solution of such problems, which employ probability theory as an aid, always demands a concise statement of the conditions and the specification of numerical requirements which the inspection must satisfy.

MATHEMATICAL APPENDIX

The problem considered is to minimize the average number of pieces inspected per lot in Single Sampling Inspection. The method and equations developed below may be extended to problems involving

2, 3, 4, etc., samples. The method of extension is somewhat complicated although the procedure is identical in nature. For example, in Minimum Double Sampling, first and second acceptance numbers and corresponding first and second sample sizes must be found, and at the same time the total Consumer's Risk must be properly divided between the two samples. We will restrict our attention to the problem of Minimum Single Sampling.

The first seven variables defined below in Table I enter into the two equations needed for the solution of the problem, the first three being fixed by the requirement of the method that a definite protection be provided against accepting faulty material, the fourth being fixed by the requirement that the average amount of inspection shall be a minimum for uniform product of process average quality. Therefore, the three unknown variables are c, n, and I. The five variables N, n, I, $\bar{p}$, and p_t are replaced in the solution by four variables which have been obtained from the original variables by combining p_t with the other four, viz. $M = p_t N$, $a = p_t n$, $z = p_t I$, and $k = \bar{p}/p_t$. Since M, P, and k are specified by the method, the unknown variables are c, a, and z. Two tables showing respectively the notation [3] and the disposition of the variables are presented below.

TABLE I

NOMENCLATURE

N = number of pieces in lot,
P = Consumer's Risk, the probability of accepting a submitted lot of tolerance quality,
p_t = tolerance fraction defective,
$\bar{p}$ = process average (expected) fraction defective,
c = acceptance number, the maximum allowable number of defective pieces in sample,
n = number of pieces in sample,
I = average (expected) number of pieces inspected per lot,
$M = p_t N$ = number of defective pieces in lot of tolerance quality,
$a = p_t n$ = expected number of defective pieces in sample drawn from lot of tolerance quality,
$z = p_t I$ = product of tolerance and average (expected) number of pieces inspected per lot,
$k = \bar{p}/p_t$ = ratio of process average to tolerance,
m = number of defects found in sample,
$C_n^N = \dfrac{N!}{(N-n)!n!}$ = number of combinations of N things taken n at a time.

The solution of the problem requires the consideration of the

[3] The symbols $\bar{p}$ and p_t, as used in this problem, are assumed true parameters of the universe sampled and according to the notation adopted by these Laboratories should be primed, i.e. $\bar{p}'$ and p_t'. For the sake of simplicity here the prime notation has been omitted in the equations. Ref. W. A. Shewhart, "Quality Control," *Bell System Technical Journal*, Vol. VI, p. 723, footnote 3, October, 1927.

following two equations:

$$z = f_1(M, a, c, k), \tag{1}$$

$$P = f_2(M, a, c), \tag{2}$$

where f_1 and f_2 represent symbolic functions which are to be determined later.

We wish to find a pair of values (c, a) which will make z a minimum,

TABLE II

DISPOSITION OF VARIABLES

Initial Variables Involved	Initial Fixed Variables	Initial Unknown Variables	Variables Used in Method	Fixed Variables of Method	Unknown Variables of Method
N	N		$M = p_t N$	M	
P	P		P	P	
p_t	p_t		$k = \bar{p}/p_t$	k	
$\bar{p}$	$\bar{p}$		c		c
c		c	$a = p_t n$		a
n		n	$z = p_t I$		z
I		I			

subject to the condition that this pair (c, a) satisfies equation (2). Hence, due to the discreteness of c, pairs (c, a) satisfying (2) are substituted in (1) until a minimum value of z is found. Thus, for $P = .10$ and for given values of M and k, we read c from Fig. 2, a for this value of c from Fig. 3, and the minimum value of z from Fig. 4.

BASIS OF FIG. 2 GIVING MINIMUM ACCEPTANCE NUMBERS

In determining the function f_1 involved in equation (1), the average number of pieces inspected per lot, I, is treated as the dependent variable. Since a sample is always taken, n pieces will be inspected from every lot submitted. The number of times that the remainder of the lot $(N - n)$ will be inspected on the average is determined from the expression giving the probability that more than c defective pieces will be found in n. The sample is assumed to be drawn from a product of which a fraction, $\bar{p}$, is defective. The probability that c or less defective pieces will be found in n pieces selected at random from a product containing $\bar{p}$ fraction defective pieces is given by the sum of the first $c + 1$ terms of the Point Binomial $[(1 - \bar{p}) + \bar{p}]^n$. Hence the average number of pieces inspected per lot is determined from the relation,

$$I = n + (N - n)\left[1 - \sum_{m=0}^{m=c} C_m^n (1 - \bar{p})^{n-m} \bar{p}^m\right].$$

For the condition, $\bar{p} < .10$, which is usual in practice, it has been found satisfactory to replace the Point Binomial by the Poisson Exponential.[4] By multiplying both sides of the equation by p_t we obtain z in the form,

$$z = M - (M - a) \sum_{m=0}^{m=c} \frac{(ka)^m e^{-ka}}{m!}, \tag{1'}$$

which is the function f_1 desired.

To obtain f_2, we state the probability of finding c or less defects in a sample n taken from a lot N containing $M = p_t N$ defective pieces. This is given by the equation,

$$P = \frac{1}{C_n^N} \sum_{m=0}^{m=c} C_{n-m}^{N-M} C_m^M.$$

But this equation is too difficult to handle in general computations on a large scale. When $p_t < .10$ and n is sufficiently large, a satisfactory approximation to the above equation may be developed from the first $c + 1$ terms of the Point Binomial, that is

$$P = \sum_{m=0}^{m=c} C_m^n \left(1 - \frac{M}{N}\right)^{n-m} \left(\frac{M}{N}\right)^m, \qquad \text{since} \qquad p_t = \frac{M}{N}.$$

An even better approximation is obtained by interchanging[5] n and M in the latter equation giving the expression,

$$P = \sum_{m=0}^{m=c} C_n^N \left(1 - \frac{n}{N}\right)^{M-m} \left(\frac{n}{N}\right)^m.$$

Since $\frac{a}{M} \equiv \frac{n}{N}$, we obtain the final form,

$$P = \sum_{m=0}^{m=c} C_m^M \left(1 - \frac{a}{M}\right)^{M-m} \left(\frac{a}{M}\right)^m, \tag{2'}$$

which is the function f_2 desired.

Now that we have f_1 and f_2 as expressed in equations (1') and (2') we must explain how Fig. 2 was obtained. When $P = .10$, for any pair (M, k), a particular pair (c, a) was found which made z a minimum. The acceptance number c may assume only discrete values since any

[4] G. A. Campbell, "Probability Curves Showing Poisson's Exponential Summation," *Bell System Technical Journal*, Vol. II, pp. 95–113, January, 1923.

[5] Paul P. Coggins, "Some General Results of Elementary Sampling Theory for Engineering Use," *Bell System Technical Journal*, Vol. VII, p. 44, Equation (11), January, 1928.

piece must be considered either as defective or non-defective. Hence minimum values of z ($z_{\text{min.}}$) will be found for many pairs (M, k) for the same value of c. From this it is evident that on an M, k plane there exist zones in which the acceptance numbers are identical. To find the boundary lines of these zones it was noted that for certain pairs (M, k) two pairs of (c, a) exist, giving the same minimum value for z. These values of c were found to differ by 1 in all such cases. Designating in general two such adjacent acceptance numbers as c and $c + 1$ and corresponding values of a which satisfy the Consumer's Risk as a_c and a_{c+1}, we may obtain these boundary curves from the equation,

$$(M - a_c) \sum_{m=0}^{m=c} \frac{(ka_c)^m e^{-ka_c}}{m!} = (M - a_{c+1}) \sum_{m=0}^{m=c+1} \frac{(ka_{c+1})^m e^{-ka_{c+1}}}{m!}.$$

In using the above equation to determine these boundary curves for Fig. 2, the following steps were taken:

(1) Assume values for c and $c + 1$.
(2) Determine a_c and a_{c+1} for a given value of P assuming N to be infinite.
(3) For any given value for k, solve the linear equation in M obtained by substituting the assumed values in the above equation.
(4) Using the value of M thus found, determine the exact values of a_c and a_{c+1} from equation (2′) for $P = .10$ (Fig. 3).
(5) Using the same value of k, again solve the linear equation in M substituting the values of a_c and a_{c+1} obtained from step (4).
(6) If the values a_c and a_{c+1} obtained in step (4) satisfy the value of M thus found, the values of M and k define a point on the boundary curve between two adjacent acceptance numbers. If these values of a do not satisfy the value of M thus determined, steps (4) and (5) may be repeated until the limiting conditions are satisfied.

Basis of Fig. 3 Giving n for Any c

For given values of the Consumer's Risk P and acceptance number c, the sample size n may be obtained from equation (2′) since $a = p_t n$. For the case $P = .10$ values of a are presented in Fig. 3 for selected ranges of M and c.

Basis of Fig. 4 Giving the Minimum Average Number of Pieces Inspected Per Lot

The curves in Fig. 4 represent specific values of $z_{\text{min.}}$ on an M, k plane for $P = .10$. Each curve was obtained by substituting given

values of $z_{\text{min.}}$, k, and (c, a) in equation (1′) and solving for M. If, for this value of M thus found, the selected values of c and a coincide with those read respectively from Figs. 2 and 3, a point was established for the given value of $z_{\text{min.}}$ on the M, k plane. If not, sufficient trials were made until the condition given by Figs. 2 and 3 were met. The curves for $z_{\text{min.}}$ were thus determined. To obtain $I_{\text{min.}}$ it was only necessary to use the relation $I_{\text{min.}} = \frac{z_{\text{min.}}}{p_t}$.

The Frequency Distribution of the Unknown Mean of a Sampled Universe

By E. C. MOLINA and R. I. WILKINSON

In drawing conclusions as to the reliability of the mean of a sample it is important that all relevant information be taken into consideration. The mathematical analysis in this paper is based on the Laplacian Bayes Theorem which implicitly comprehends the results of a sample together with the a priori knowledge available concerning the parameters of the universe.

The discussion is limited to a universe assumed to be normal but whose mean and precision constant are unknown. Several simplifying, yet quite reasonable, assumptions regarding the forms and independence of the a priori frequency distribution of the true mean and standard deviation are incorporated in the analysis so that numerical answers may more easily be deduced.

Conclusions, properly drawn, are usually quite definitely dependent upon the a priori assumptions made, and especially so in the case of small samples. A considerable space is, therefore, devoted to the solution of a problem in which the sample is only five, taking up a wide variety of these a priori assumptions. They give, in consequence, a wide range of numerical results, appearing in the form of probable errors in the mean of the sample. Each set of assumptions is briefly discussed indicating how the sampling technician may be able to make a selection consistent with his a priori knowledge of a particular problem.

EVERY observation or series of observations upon the items composing a "universe" or "population" may be regarded as constituting a sample. We may divide sampling into two broad natural classes, (1) Sampling of Attributes, and (2) Sampling of Variables. The theory of the first class concerns itself with some particular characteristic, such as the color red, which each item of the universe definitely does or does not possess, and endeavors to assign, ultimately, a numerical value to the probability that the number or proportion of the items in the universe having this characteristic lies within any given range. The second division comprehends that wide variety of problems in which each item of the universe displays to a greater or less degree the same particular quality, such as length, weight, or resistance. After having drawn a random sample of items, probability theory is called upon to assert with what likelihood certain important descriptive constants or "parameters" of the universe lie within any given ranges.

In either class the problem is legitimately attacked by means of a posteriori probability theory. This theory makes use of the two important distinct kinds of knowledge which, in varying amounts, are always at hand, namely, (1) a priori or preexisting information regarding the universe and the possible values which the unknown

values of $z_{\text{min.}}$, k, and (c, a) in equation (1′) and solving for M. If, for this value of M thus found, the selected values of c and a coincide with those read respectively from Figs. 2 and 3, a point was established for the given value of $z_{\text{min.}}$ on the M, k plane. If not, sufficient trials were made until the condition given by Figs. 2 and 3 were met. The curves for $z_{\text{min.}}$ were thus determined. To obtain $I_{\text{min.}}$ it was only necessary to use the relation $I_{\text{min.}} = \frac{z_{\text{min.}}}{p_t}$.

parameters may assume, and (2) the actual observed value of the studied characteristic in each item of the sample. The a priori information may be meager, in some instances hardly more than the limits between which the parameters must lie, and again, from past experience a great deal may be known about the universe, such as its general form of frequency distribution, the most likely value for each of its parameters to take, and a general feeling that they will not, except in rare cases, lie outside of certain well defined ranges closely bordering their believed most likely values. When the a priori knowledge is meager, more weight must be attached to the results of the sample, but when considerable a priori information is at hand relatively less reliance should be placed in the sample; and in some rare cases it is conceivable that so much is known before the drawings are made that a particular sample, especially if small, would justifiably be disregarded entirely.

The Sampling of Attributes on the a posteriori basis for both infinite and finite universes has already been set forth in these pages at considerable length.[1] The theory of Sampling of Variables when the samples are large becomes usually a matter of assuming that some of the parameters of the sample are sufficiently close to those of the universe that no sensible error will be made in assuming them to be equal. In this case the a priori knowledge of the universe, unless far more exact than is normally found in practice, would exercise but a slight effect in the conclusions which might be drawn, and is therefore quite often properly neglected.

When, for one reason or another, some conclusions are demanded after having taken a small sized sample, it cannot safely be assumed that the sample itself adequately describes the universe, and what a priori knowledge we have must, of necessity, play an important rôle in the determination of any legitimate statements as to the constitution of the universe.

The purpose of this paper is to study in strict accordance with the theory of probability the conclusions which may be drawn concerning the true parameters of the unknown universe after a "sample of variables" of any size has been examined.

The paper is divided into the following five sections:

I. The general equation is given for the a posteriori probability

[1] "Deviation of Random Samples from Average Conditions and Significance to Traffic Men," by E. C. Molina and R. P. Crowell, January 1924. "Some General Results of Elementary Sampling Theory for Engineering Use," by P. P. Coggins, January 1928. This second paper is based on another by Mr. E. C. Molina presented before the Statistical Section of the International Mathematical Congress, held at Toronto in August 1924.

that the true mean of a sampled normal universe lies within a given range.

II. Certain mild restrictions are placed on the general equation of (I) to facilitate its use in practice.

III. The selection of a priori frequency functions in practice is discussed.

IV. A typical example is selected and solved for various a priori existence probability distributions with a discussion of the ranges of errors.

V. Conclusions.

I. The General A Posteriori Equation

It is common, unless information is known to the contrary, to assume that the universe from which the sample is to be made is composed of an infinite number of items all having a particular characteristic whose numerical value from item to item follows the normal frequency law. In the remainder of this paper we shall limit ourselves to a discussion involving only this type of universe. The problem may now be precisely stated:

A set of n observations has been made on a variable quantity drawn from a universe wherein the normal law of errors

$$\left(\frac{h}{\pi}\right)^{1/2} e^{-h(x-m)^2}, \qquad h = 1/2\sigma^2$$

is satisfied but the values of the mean and the precision constant, or standard deviation, are unknown; *before* the observations were made the probability in favor of the simultaneous existence of the inequalities

$$m < \text{mean} < m + dm \tag{1}$$

$$h < \text{precision constant} < h + dh \tag{2}$$

was some function of m and h, say $W(m, h)dmdh$; what is the probability that *after* the observations were made the unknown mean satisfies the inequality (1)?

Let $x_1, x_2 \cdots x_n$ be the values for x given by the n observations. Set

$$n\bar{x} = \sum_1^n x_i, \qquad ns^2 = \sum_1^n (x_i - \bar{x})^2.$$

Now if m and h were known the probability that a set of n observations, *not yet* made, would give values $x_1, x_2, \cdots x_n$ would be

$$\left(\frac{h}{\pi}\right)^{(1/2)n} e^{-h\Sigma(x_i-m)^2} dx_1 dx_2 \cdots dx_n. \tag{3}$$

Therefore, by the Laplacian generalization of the Bayes formula, the a posteriori probability that

$$m < \text{mean} < m + dm$$

is (cancelling factors which do not involve m or h)

$$P(m)dm = \frac{dm \int_0^\infty W(m, h)h^{(1/2)n}e^{-h\Sigma(x_i-m)^2}dh}{\int_{-\infty}^{\infty} dm \int_0^\infty W(m, h)h^{(1/2)n}e^{-h\Sigma(x_i-m)^2}dh}^2$$

$$= Adm \int_0^\infty W(m, h)h^{(1/2)n}e^{-h\Sigma(x_i-m)^2}dh, \qquad (4)$$

where A is a constant such that

$$\int_{-\infty}^{\infty} P(m)dm = 1.$$

II. Introduction of Restrictions on General Equation

We are now confronted by a difficulty inherent to a posteriori probability problems. What do we know as to the form of the a priori existence probability function $W(m, h)$? If in a specific practical problem the form of $W(m, h)$ is unknown, no conclusions can be drawn from the set of observations *unless some assumptions are made* and then the weight assignable to the conclusions drawn is a delicate question depending on the reasonableness of the assumptions.[3]

The analysis and results given below are based on assumptions which the writers believe will be found justifiable in many problems of practical interest.

A first assumption which suggests itself is that m and h are independent a priori so that we may write

$$W(m, h) = W_1(m)W_2(h). \qquad (5)$$

On this assumption

$$P(m)dm = AW_1(m)dm \int_0^\infty W_2(h)h^{(1/2)n}e^{-h\Sigma(x_i-m)^2}dh. \qquad (6)$$

As a second step toward tentative solutions assume that

$$W_2(h) = Kh^{(1/2)c}e^{-ah}, \qquad (7)$$

[2] See Poincare: "Calcul des Probabilites"; 2d edition; articles 178 and 179.

[3] In this connection see italicized paragraph, page 266, "Probability and Its Engineering Uses," T. C. Fry, 1928.

where K, c and a are constants. This, by means of the change of variable

$$y = h[a + \sum(x_i - m)^2]$$

and throwing the definite integral

$$\int_0^\infty y^{(1/2)(n+c)}e^{-y}dy$$

in with the constant A, reduces (6) to

$$P(m)dm = A'W_1(m)[a + \sum(x_i - m)^2]^{-(1/2)(n+2+c)}dm. \tag{8}$$

We are still confronted with the a priori existence probability function $W_1(m)$.

A plausible form, suggested by the well known "Student" [4] distribution of the ratio $(\bar{x} - m)/s$ for a set of observations *to be* made from a normal universe of *known* mean and standard deviation, is

$$W_1(m) = A_1[1 + B(M - m)^2]^{-(1/2)N}, \tag{9}$$

where M is the value of m which is a priori most probable, N and B are positive constants while the equation

$$\int_{-\infty}^{\infty} W_1(m)dm = 1$$

gives

$$A_1 = \frac{B^{1/2}\Gamma(\frac{1}{2}N)}{\pi^{1/2}\Gamma[\frac{1}{2}(N-1)]}.$$

With this assumed form and noting that

$$\sum(x_i - m)^2 = ns^2 + n(\bar{x} - m)^2$$

equation (8) gives

$$P(m)dm = A''[1 + B(M - m)^2]^{-(1/2)N} \times \left[1 + \left(\frac{ns^2}{a + ns^2}\right)\left(\frac{\bar{x} - m}{s}\right)^2\right]^{-(1/2)(n+2+c)} dm, \tag{10}$$

the integral of $P(m)dm$ between plus and minus infinity determining A''.

Recapitulating: formula (10) gives us the a posteriori frequency distribution for m in terms of the observed data and the arbitrary constants a, c, N, B, M which have entered into the problem in

[4] The writers are aware of the fact that the "Student" frequency function has been put forward in more than one place as the solution for an a posteriori problem. But it should be noted that the various *deductions* of this function which have been given by "Student" and others are entirely a priori.

consequence of the three assumptions made regarding the form of the a priori existence probability function $W(m, h)$; the three assumptions being embodied in equations (5), (7) and (9).

III. Practical Selection of A Priori Frequency Distributions

In equation (10) we have first to assign a numerical value to each of the five constants a, c, N, B, M, before the probability $P(m)$ can be evaluated for any desired range of m. Obviously, in actual practise, the selection of their values is extremely important and too much care cannot be exercised in an attempt to satisfy the engineering judgment that all of the a priori information at hand has been nicely comprehended.

In an endeavor to reduce the number of constants to which we must assign values we shall consider first the a priori function

$$W_2(h) = Kh^{(1/2)c}e^{-ah}.$$

Setting

$$h = c/2a \tag{11}$$

makes $W_2(h)$ a maximum. On the other hand, the value of h which would make the observed set of values of x most probable is given by the equation

$$\frac{1}{h} = \frac{2\sum(x_i - m)^2}{n},$$

or, if m be set equal to $\bar{x}$, we obtain the simpler equation

$$h = 1/2s^2. \tag{12}$$

Upon eliminating h from (11) and (12),[5]

$$a = cs^2. \tag{13}$$

In Fig. 3 are shown four frequency curves of $W_2(h)$. Curve I is plotted for $c = 3$ according to equation (13), and to illustrate the wide possibility of forms, curves II and III have been constructed, keeping $c = 3$, after changing equation (13) to

$$a = \frac{cs^2}{1 - .1s^2} \quad \text{and} \quad a = \frac{cs^2}{1 + .1s^2},$$

respectively. Curve IV again satisfies equation (13) but has c increased from 3 to 6.

[5] It should be carefully noted that there is no necessary relation between the *a priori most probable value of h* and the *value of h which would make the observed event most probable.* The elimination of h between (11) and (12) is justified solely by the practical consideration that a tentative relation between a and c will reduce by one the number of arbitrary constants to which numerical values must be assigned.

For every assumption of a and c in the a priori distribution of h there is, of course, a corresponding a priori distribution, $\phi(\sigma)$, of the standard deviation. Here

$$\phi(\sigma)d\sigma = K'\sigma^{-(c+3)}e^{-(1/2)a/\sigma^2}d\sigma$$

and

$$K' = \frac{a^{(1/2)c+1}}{2^{(1/2)c}\Gamma(\frac{1}{2}c + 1)} .$$

The distributions of σ corresponding to each of the four frequency curves of h in Fig. 3 are shown in Fig. 4 with similar designations.

In many cases, too, it is obvious that very little is known concerning the shape and the parameters of the mean's a priori distribution beyond that it is generally unimodal and quite likely to be fairly symmetrical about its most probable value; a mathematical expression of this has been set up in equation (9). In this circumstance we may not introduce serious restrictions if we make two further assumptions which greatly simplify (9).

The first is that we set $M = \bar{x}$ which says that, a priori, the most probable value of the unknown mean was the same as that which was later calculated as the mean in the sample.[6] It is admitted that the chance of exactly fixing on $M = \bar{x}$ from a priori information is very small, yet if our knowledge is so slight that we must introduce some guesswork here, the selection of the value of $\bar{x}$ at least has the advantage of being a *possible* one which M might assume and, except in rare cases, it will not be greatly distant from the true m in that particular lot. The logical difficulty here also may be minimized by selecting a form of $W_1(m)$ of such flatness that over a considerable range of values in the neighborhood of $\bar{x}$ the existence probability does not take on widely differing magnitudes.[7]

The second assumption can more readily be allowed, and consists in empirically defining

$$B = \frac{n}{a + ns^2} .$$

This removes a degree of freedom from the $W_1(m)$ function but, as far as its form is concerned, except in special cases, the one variable, N, may serve quite well in characterizing the pre-existing information. As is clearly shown in Fig. 1, the increase of N indicates a greater

[6] While it does not matter in this particular problem, the authors wish to carefully distinguish, at least in thought, between an "observed" parameter and a parameter calculated from individual observations.

[7] The setting of $M = \bar{x}$, it should be noted, has no effect if all values of the mean are made a priori equally likely by setting $N = 0$ (that is, $W_1(m) = A_1$).

certainty in the investigator's mind that the true value of m lies closer and closer to the assumed most probable figure, M.

With these two assumptions incorporated in equation (10) we may now write

$$P(m)dm = f(t)dt = A'''(1 + t^2)^{-(1/2)T}dt, \tag{10'}$$

in which

$$t^2 = \left(\frac{\bar{x} - m}{s}\right)^2\left(\frac{ns^2}{a + ns^2}\right) = B(\bar{x} - m)^2, \tag{14}$$

$$T = n + 2 + c + N,$$

$$A''' = \frac{1}{\int_{-\infty}^{\infty}(1 + t^2)^{-(1/2)T}dt} = \frac{\Gamma(\frac{1}{2}T)}{\pi^{1/2}\Gamma[\frac{1}{2}(T - 1)]}.$$

The formula (10') is a "Student" [8] frequency form with the arguments n and $\frac{\bar{x} - m}{s}$ replaced by $n + 2 + c + N$ and $\frac{\bar{x} - m}{s(1 + a/ns^2)^{1/2}}$ respectively.

Fig. 2 shows curves plotted for ranges of t such that

$$A''' \int_{-t}^{+t}(1 + t^2)^{-(1/2)T}dt = .50, .80, .90, \text{ and } .9973,$$[9]

and the errors in the mean corresponding to any of these probabilities, after determining t, may be found by evaluating $\bar{x} - m$ in equation (14).

IV. Solution of a Typical Example

Five samples of retardation coils rated at 47 ohms are taken from a large lot, and careful measurements show them to have resistances of 46.30, 44.40, 47.72, 50.50, and 45.58 ohms respectively. We are asked to determine the probable and 99.73 per cent errors of the average of these resistances, assuming that the samples have been drawn from a normal universe.

The average of these five values is $\bar{x} = 46.90$ ohms and their standard deviation about this average is found to be $s = 2.097$.

From the preceding discussion it is evident that as many answers to this problem may be obtained as there are assumptions made regarding, in general, the a priori distributions of the mean and

[8] Student: "The Probable Error of a Mean," Biometrika, Vol. VI, No. 1, March 1908.

[9] Student: "New Tables for Testing the Significance of Observations," Metron, Vol. V, No. 3, I–XII–1925. Tables I and II, pages 114–118, for values of $n' = 2$ to 21.

precision constant, and in our particular analysis the five constants found in equation (10). In Table I and Fig. 1 we tabulate and portray graphically twenty-one complete solutions of the example based upon as many sets of values given to these constants. A wide

TABLE I

$\bar{x} = 46.90 \quad n = 5 \quad s = 2.097$

No.	Parameters in Existence Probability Distributions							$T = n + 2 + c + N$	A Posteriori Probability Values of Errors in Observed Mean	
	Mean				Precision Constant				Probable or 50% Error	99.73% Error
	M	N	A_1	B	c	a	K			
1	46.9	0	—	.2274	−3	0	—	4	.923	11.16
2	46.9	0	—	.2274	−2	0	—	5	.776	6.94
3	46.9	0	—	.2274	0	0	—	7	.623	4.23
4	46.9	0	—	.1421	3	13.19*	475.5	10	.626	3.61
5	46.9	0	—	.1098	3	23.55†	2,024	10	.712	4.10
6	46.9	0	—	.1605	3	9.163‡	191.2	10	.589	3.39
7	46.9	0	—	.1034	6	26.38*	80,770	13	.625	3.39
8	46.9	1	—	.2274	0	0	—	8	.566	3.59
9	46.9	1	—	.1421	3	13.19*	475.5	11	.587	3.33
10	46.9	2	.1518	.2274	0	0	—	9	.524	3.17
11	46.9	2	.1200	.1421	3	13.19*	475.5	12	.558	3.08
12	46.9	2	.1055	.1098	3	23.55†	2,024	12	.635	3.50
13	46.9	2	.1275	.1605	3	9.163‡	191.2	12	.525	2.90
14	46.9	2	.1023	.1034	6	26.38*	80,770	15	.575	3.01
15	46.9	4	.3036	.2274	0	0	—	11	.464	2.63
16	46.9	4	.2400	.1421	3	13.19*	475.5	14	.511	2.71
17	46.9	4	.2110	.1098	3	23.55†	2,024	14	.581	3.09
18	46.9	4	.2551	.1605	3	9.163‡	191.2	14	.480	2.55
19	46.9	4	.2047	.1034	6	26.38*	80,770	17	.537	2.76
20	41.0	2	.1200	.1421	3	13.19*	475.5	—	.665§	3.87§
21	49.0	2	.1200	.1421	3	13.19*	475.5	—	.635§	3.65§

* $a = cs^2$.

† $a = \dfrac{cs^2}{1 - .1s^2}$.

‡ $a = \dfrac{cs^2}{1 + .1s^2}$.

§ Errors determined by planimeter method.

variety of a priori conditions is assumed giving, in consequence, widely varying probable and 99.73 per cent errors.

The a priori frequency functions, $\phi(\sigma)$, for the standard deviation in the first seven cases are shown in broken lines superposed upon the distributions of precision constants in Fig. 1. The scales of h and σ are not to be confused, the attempt being only to represent the *form* of the $\phi(\sigma)$ frequency curves.

(*a*) If we wish to be very conservative we might select values for the unknown constants which would make all values of m and σ

APRIORI PROBABILITY DISTRIBUTIONS

MEAN

$W_1(m) = A_1\left[1+B(M-m)^2\right]^{-\frac{N}{2}}$

PRECISION CONSTANT

$W_2(h) = Kh^{\frac{C}{2}}e^{-ah}$

ERRORS IN OBSERVED MEAN

PROBABLE ERROR

99.73% ERROR

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 9

0.3 0.2 0.1 (m)

38 40 42 44 46 48 50 52 54 56

m

6 5 4 3 2 1 0 $W_2(h)$

0 0.1 0.2 0.3 0.4 0.5 0.6

h

0 0.5 1.0

0 5.0 10.0

0 0.5 1.0

0 5 10

FIG. 1.

equally likely, that is, $N = 0$, $c = -3$, $a = 0$.[10] Here the precision constant's a priori distribution is decidedly exponential and we might predict the large probable and 99.73 per cent errors in the observed average which actually result.

Case 1 in Table I and Fig. 1 presents the problem in its entirety with the resultant errors tabulated as well as shown graphically.

(*b*) The engineer's knowledge, however, in all probability, is not so limited as in (*a*) above, at least regarding the precision constant

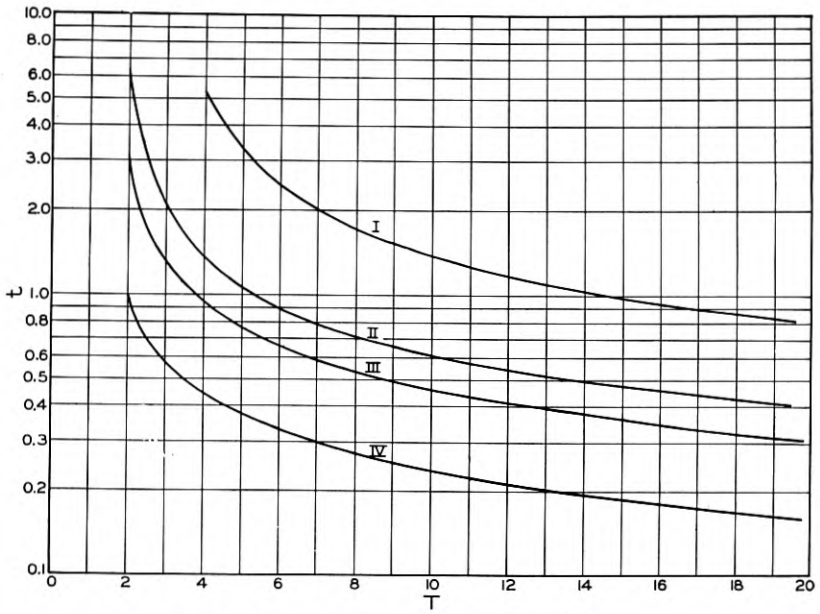


Fig. 2—Errors of averages of samples of size n.
I—99.73 per cent Error.
II—90.00 per cent Error.
III—80.00 per cent Error.
IV—50.00 per cent Error.
Note: Abscissa: $T = n + 2 + c + N$.
Ordinate: t = The Product of the Error of the Average and the Square Root of B.

(or the standard deviation). He knows that extremely small values of the precision constant are less likely than larger ones, and to some extent we picture the transition from (*a*) to this impression in the Cases Nos. 2 and 3 which as before may be found completely portrayed

[10] The formula for $P(m)$ resulting from a substitution of these constants in equation (10′) reproduces the result obtained by Drs. J. Neyman and E. S. Pearson for all values of the a priori function $W'(m, \sigma)$ equally likely: Biometrika, Vol. XX*A*, Parts I and II, July 1928; "On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference," page 196, equation XXXV.

in Table I and Fig. 1. Case No. 2, it is interesting to note, is the familiar "Student" formula; Case No. 3's outstanding characteristic is that all values of h are a priori equally likely. The errors in the mean have here been greatly reduced by merely changing the existence probability distribution of the precision constant.

(*c*) Again, the experienced analyst is quite likely to assume willingly that the distribution of the precision constant (and likewise the standard deviation) is of a unimodal form having its maximum value not greatly distant from the figure determined in the sample. Cases Nos. 4 to 7 inclusive typify this kind of assumption while, at the same time, all values of the true mean are held a priori, equally likely.

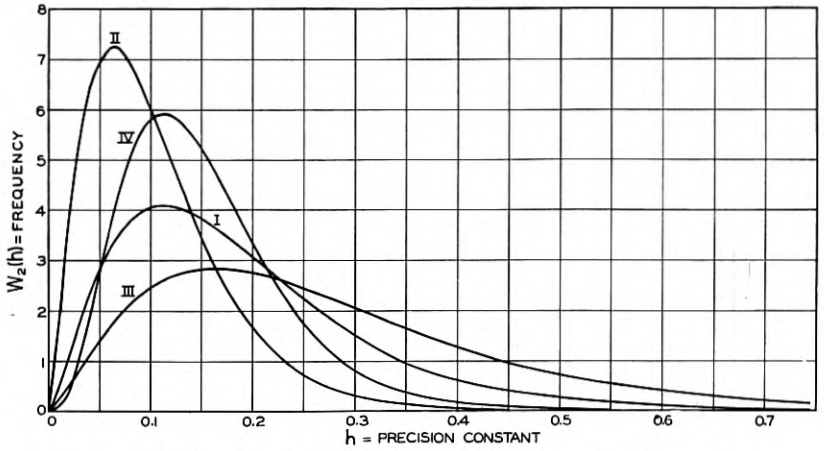


Fig. 3—Typical a priori frequency distributions of the precision constant.

$$W_2(h) = Kh^{(1/2)c}e^{-ah}.$$

$$\text{I}—c = 3,\ a = cs^2 = 13.1922.$$

$$\text{II}—c = 3,\ a = \frac{cs^2}{1 - .1s^2} = 23.5467.$$

$$\text{III}—c = 3,\ a = \frac{cs^2}{1 + .1s^2} = 9.1629.$$

$$\text{IV}—c = 6,\ a = cs^2 = 26.3845.$$

The constants for Cases Nos. 4 and 7 have been so selected as to bring the modal value of h at that found from the sample, that is, that value of h has been made most likely a priori which will make the probability of occurrence of the particular value $(1/2s^2)$ calculated from the observations, a maximum. Case No. 7 is a considerably more peaked distribution than Case No. 4 indicating more faith in the modal figure selected as being close to the true value. Cases Nos. 5 and 6 illustrate how the mode of the $W_2(h)$ function which always lies at $h = c/2a$ may be shifted either down or up and the extent of modification in the resulting errors which may be expected.

The four frequency distributions of h just considered are the same as those shown in more detail in Fig. 3; the corresponding distributions of the standard deviation are found detailed on Fig. 4.

(*d*) If the interpreter of the data is closely familiar with the sampled product and has been observing similar lots for some time he may have a reasonably good idea as to the value of the general average

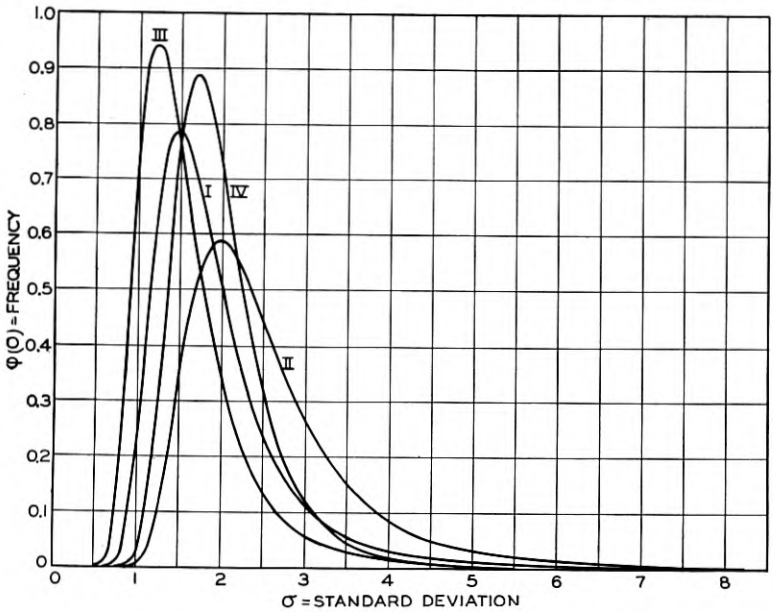


Fig. 4—Typical a priori frequency distributions of the standard deviation.

$$\phi(\sigma) = K'\sigma^{-(c+3)}e^{-(1/2)a/\sigma^2}.$$

$$\text{I—}c = 3,\ a = cs^2 = 13.1922.$$

$$\text{II—}c = 3,\ a = \frac{cs^2}{1 - .1s^2} = 23.5467.$$

$$\text{III—}c = 3,\ a = \frac{cs^2}{1 + .1s^2} = 9.1629.$$

$$\text{IV—}c = 6,\ a = cs^2 = 26.3845.$$

of items produced under these same essential conditions. In Cases Nos. 8 to 19, inclusive, use is made of this knowledge on the assumption that $\bar{x}$, the mean of the sample, turns out to be so nearly equal to M, the most likely a priori value of the true mean m, that we may safely call them identical. Three values of N, regulating the spread of the $W_1(m)$ distribution to conform to the observer's best judgment of the true circumstances have been associated with the same sequence of a priori assumptions regarding the precision constant as were presented in Cases Nos. 3 to 7.

The errors found in Cases Nos. 8 to 19 on the various combinations of a priori frequency curves lie in a fairly narrow band distinctly below those determined from the more conservative assumptions underlying Cases Nos. 1, 2 and 3. This well illustrates the importance of carefully surveying and as far as possible completely utilizing the knowledge available before the sample has been made.

(*e*) Finally, cases are bound to occur in which the engineer can quite definitely say that some value of M other than $\bar{x}$ is a priori most probable; this situation is encountered in Cases Nos. 20 and 21. These are identical with Case No. 11 except that in Case No. 20, M has been reduced about 6 ohms and in Case No. 21 raised about

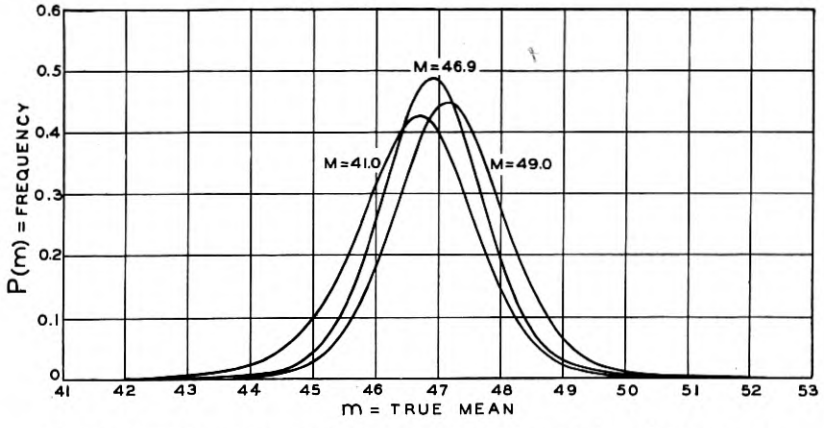


Fig. 5—Typical a posteriori frequency distributions of the unknown mean.

$$P(m) = A''[1 + B(M - m)^2]^{-(1/2)N}\left[1 + \left(\frac{ns^2}{a + ns^2}\right)\left(\frac{\bar{x} - m}{s}\right)^2\right]^{-(1/2)[n+2+c]}.$$

2 ohms. The errors are somewhat increased by these changes in M, as, of course, we should have predicted. Comparisons such as this should help the investigator to decide whether or not his previously selected figure for M is sufficiently close to $\bar{x}$ that they may safely be equated.

In the event that it is decided that M may not be set equal to $\bar{x}$, in any particular problem, as in Cases Nos. 20 and 21, the symmetrical "Student" form of distribution for $P(m)$, (except when $N = 0$) no longer occurs. This is clear from an inspection of Fig. 5 which shows the three cases plotted on the same scale.

It is suggested then, since the integral of $P(m)dm$ here may become difficult to handle, that recourse be had to the use of a planimeter on the distribution plotted from equation (10) on rectangular coordinate paper. In this way may be determined within what range,

equidistant above and below $\bar{x}$, lies the proportion of the total area corresponding to the desired probability.[11]

V. Conclusions

We have presented a general equation for the probability that the true mean of a sampled normal universe lies within a given range, incorporating the kind of knowledge the investigator may be expected to have before the sample was made as well as the information directly presented by the individual observations themselves. It cannot be overemphasized that the problem by its very nature is indefinite since it would be a rare instance indeed to find a mathematical expression which would completely and exactly summarize the a priori knowledge, impressions and beliefs in the mind of any person confronted with its solution. All that can be found is, at best, an approximate probability based upon certain assumptions we are willing to make in order to arrive at a numerical result. And only by utilizing as far as possible all of the available knowledge will the most nearly correct probability values ascertainable be realized.

[11] On certain test cases of "Student" distributions, the error in planimeter readings averaged about one-half of one per cent, and in no case exceeded one per cent.

Speech Power and Its Measurement [1]

By L. J. SIVIAN

The paper is chiefly concerned with the important speech power quantities—frequency spectra, distributions of instantaneous, average, syllabic and peak amplitudes, etc.—as they obtain in actual speech for a large range of voices, talking levels, and subject matters. The analysis is not nearly so complete nor so fine-grained as that which, in principle, can be derived from oscillographic records of individual speech sounds. Its advantage is in the speed with which data can be secured, under widely varying conditions and on a scale which warrants statistical conclusions. Some of the methods in use for measurements of this type are described. A "level analyzer" has been developed, primarily for the measurement of average and peak pressure amplitudes in speech and music, both as to magnitude and as to position in the frequency spectrum. Illustrative results are given for samples of speech, music and noise.

SPEECH sounds are so variable from one to another, from one individual to another, from one conversation to another, as to make it necessary to deal with speech power in a statistical manner. This is particularly true of engineering applications, as distinguished from studies in phonetics and voice dynamics. It is largely from the viewpoint of the former that the subject is here treated.

Speech power occurs in several states, e.g. acoustic, electric, magnetic, mechanical, optical, thermal, etc., but its measurement largely refers to speech in the acoustic or in the electric form. Acoustically, the simplest quantity to define is the instantaneous power transmitted through unit area tangent to the wave front. That power is $L = P \cdot U$, where P is the pressure and U the air particle velocity. With P expressed in bars (dynes per cm.^{2}), and U in cm./sec., L is given in ergs/cm.$^2 \times$ sec. We do not directly measure the product $P \cdot U$. No suitable means for doing that has been developed. In a progressive plane wave, or with good approximation in any progressive wave at sufficient distance from the source, P and U are in phase, and the expression for L simplifies down to $L = P^2/\rho c = U^2 \cdot \rho c$, where ρc is a constant equal to 41.5 mechanical ohms—the sound radiation resistance of air. Hence for the type of waves mentioned the wattmeter type of measurement (pressure $\times$ velocity, corresponding to voltage $\times$ current in the electrical case) may be replaced with the simpler measurement of pressure (voltage) or velocity (current) alone.

What are the acoustic voltmeters and ammeters that are capable

[1] Presented before Acoustical Society of America, May 11, 1929.

of measuring the instantaneous pressures and velocities respectively? First, as to acoustic ammeters: perhaps the two best known forms are the Rayleigh disk and the hot-wire microphone. The disk gives absolute values of U but its response is so slow that it can be used only to measure the effective values in comparatively sustained sound waves lasting, say one second or longer, or for the ballistic integration of shorter pulses. Hence it does not give a measure of the instantaneous velocities for even the slowest audio-frequency vibration. Furthermore, its use when exposed to the speaker in an open room, is rendered difficult by its susceptibility to spurious air currents. The only application of the disk to vocal power measurements which has come to my notice, is one by Prof. Zernov,[2] published in 1908. For sustained loud singing and shouting he found energy densities ranging from 0.3 to 2.0×10^{-4} ergs/cm.^{3}, at 2 meters distance from the singer. Assuming uniform distribution over a hemisphere of 2 m. radius, this gives a total power output of the voice of the order of 50,000 microwatts. At 2 m. distance reflections from the walls of the room materially raise the energy density as compared with that in a progressive wave.

Another acoustic ammeter is the hot-wire microphone. The resistance variations of the wire tend to follow the instantaneous values of the air particle velocity in the sound wave, but the sensitivity is a complicated function of the frequency, rapidly decreasing as the latter increases. Hence it does not give true oscillograms of speech sounds which in general include a frequency range of six or seven octaves.

Still another device in this general class is the glow-discharge microphone. Its response is determined largely by the amplitude of the air particle motion (E. Meyer, E.N.T., v. 6, 17-21, 1929). Hence, for constant sound intensity the response is inversely proportional to the frequency, roughly. No method for its absolute calibration has been proposed other than comparison with a microphone of known performance. The frequency response and the somewhat erratic behavior of the device in its present status, render it rather unsuitable for speech power studies.

Nearly all our information concerning speech power and the waveform of speech sounds has been obtained with acoustic voltmeters, i.e., with devices responding to pressure in the sound wave. To a first approximation, the ear belongs to this class. It includes the resonators which Helmholtz used in his vowel studies. Generally, the vital element in these devices is a diaphragm which vibrates with

[2] *Ann. der Phys.*, *26*, 94, 1908.

the alternating sound pressure and whose motion is converted (indirectly, as a rule) into a visual record. These acoustic voltmeters have gradually evolved from Scott's phonautograph and Koenig's manometric capsules and indicating flames of some 70 years ago to the present day technique. Their extensive use accounts for the fact that sound measurements frequently are expressed in terms of pressure rather than in terms of power. In many cases, when the relation between pressure and velocity is not known or too complex a function of frequency to be manageable,—the description of the sound wave must be confined to pressure values alone.

The first requisite for absolute measurements of speech pressures is a microphone which admits of an absolute electroacoustic calibration over the range of audio frequencies, and which has a substantially uniform sensitivity over the most important part of that range. The development of the condenser microphone and of the thermophone method for calibrating it, supplied this need. This fundamental contribution to the subject is due to E. C. Wente.[3] Using such a calibrated condenser microphone in conjunction with a vacuum tube amplifier and an oscillograph which were uniformly sensitive up to about 6000 p.p.s., I. B. Crandall and C. F. Sacia,[4] and I. B. Crandall [4] obtained oscillograms of the fundamental speech sounds in the English language. Essentially, these oscillograms give a picture of the instantaneous pressures throughout the duration of the sound, impressed on the microphone diaphragm at 2.5 cm. from the speaker's lips. A certain amount of similar work on German speech sounds has been published by Trendelenburg.[5] From the standpoint of phonetics these pressure amplitude oscillograms of individual sounds are a most comprehensive source of information. Their chief use has been in determining the frequency spectra of the fundamental speech sounds. Sacia [6] and Sacia and Beck [7] have used them to compute the mean and the peak powers of those sounds.

From the standpoint of engineering applications, speech power has a twofold interest: (*a*), there is the question of designing microphones, receivers, circuits, room acoustics, of controlling noise levels, etc., (*b*), the control of apparatus (chiefly adjustment of amplification) while it is handling speech from persons who acoustically are not controllable.

[3] *Phys. Rev.*, July 1917, April 1922, June 1922.
[4] *Bell Sys. Tech. Jour.*, April 1924 and Oct. 1925, resp.
[5] Wiss. Veroff. a. d. Siemens-Konzern, 1924 and 1925.
[6] *Bell Sys. Tech. Jour.*, Oct. 1925.
[7] *Bell Sys. Tech. Jour.*, July 1926.

In connection with (*a*) we can use data obtained under "laboratory" conditions, i.e. with the speakers, their voice levels, the talking speed, etc., suitably selected. Perhaps the simplest statistical characteristic

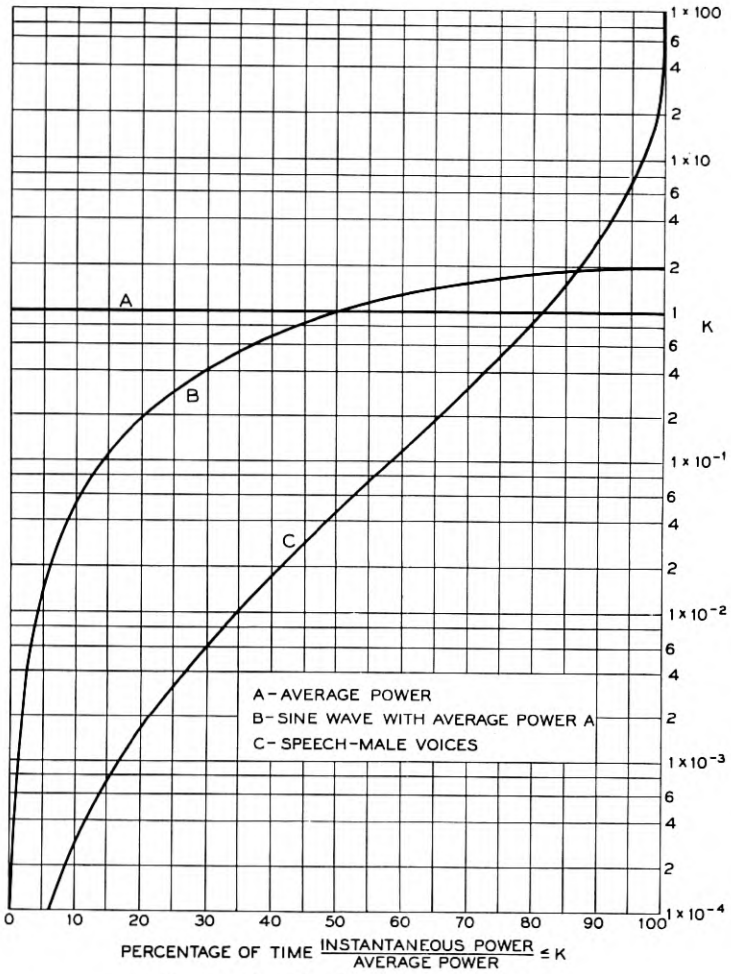


Fig. 1—Time distribution of speech power.

of speech is the average power, i.e. the ratio of the total energy in a large number of speech sounds divided by the total speech time. The quantity immediately measured is the pressure on the transmitter diaphragm. The corresponding power flow through unit area is computed. In doing so it is roughly assumed that the pressure on the diaphragm is twice as great as it would be in free air, which is

assuming total reflection by the transmitter. It is further assumed that the power flow is uniform over a hemisphere whose radius is the distance from the speaker's lips to the diaphragm. For this average power flow with "normally modulated" voices, Crandall and Mackenzie [8] found the value of 12.5 microwatts. More recently Sacia [6] gave 7.4 microwatts, including pauses between speech sounds, which

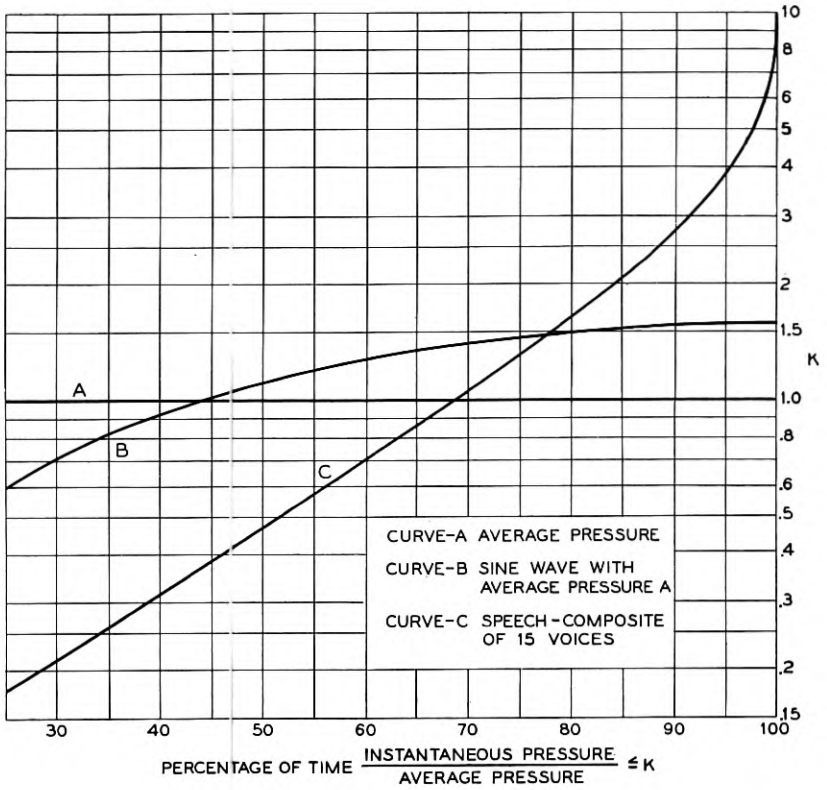


Fig. 2—Time distribution of speech pressures.

occupied about one-third of the total time. Some measurements made since then have led to substantially similar values, and so the value of 10 microwatts may be taken as fairly representative of a normal level. The corresponding average pressure on the diaphragm, at 5 cm. distance from the lips, was found to be about 5 bars. The averaging, of course, is for the absolute values of the pressures. Some notion of the level involved may be gained by noting th it is at

[8] *Phys. Rev.*, March 1922.

decidedly fatiguing to maintain a speech level 20 db higher for more than a few seconds.

It should be noted that the distributions of the instantaneous power and instantaneous pressure values in respect to their average values are quite different for speech from what they are in a sinusoidal wave. Fig. 1 and Fig. 2 show the extent of the difference for powers and pressures respectively. Clearly, speech as a whole is decidedly more "peaked" than a sinusoidal wave. Values of peak factors for

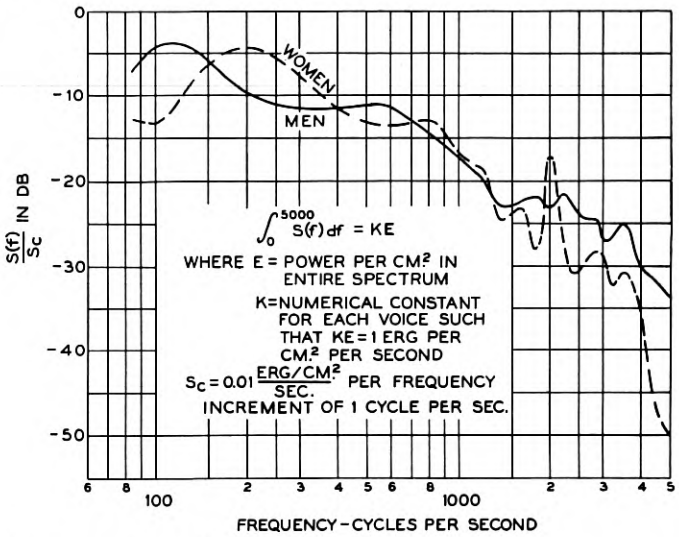


Fig. 3—Energy spectrum of connected speech-composite curves for 4 men and for 2 women.

individual speech sounds, i.e. the ratio of peak amplitude to mean effective amplitude, are given in Sacia's paper.[5]

A paramount characteristic of speech is the distribution of power and pressure in the frequency spectrum. That is so because in general the systems responding to or carrying speech (including the ear) have pronounced frequency characteristics. The same is true of noise sources interfering with speech. Here, as for total powers and pressures, the simplest quantity is the average spectrum for a large number of speech sounds. Crandall and Mackenzie [8] obtained the energy spectrum shown in Fig. 3, based on speech from six voices. They used a condenser microphone whose output was analyzed by a series of resonant circuits covering the range from 75 to 5000 p.p.s. Fig. 4 shows a recent determination in which the average pressure is given as a function of frequency. The apparatus

used to obtain the data will be described in detail elsewhere. Briefly, a series of band-pass filters are used to separate the frequency components. The output of any one filter goes into a rectilinear vacuum

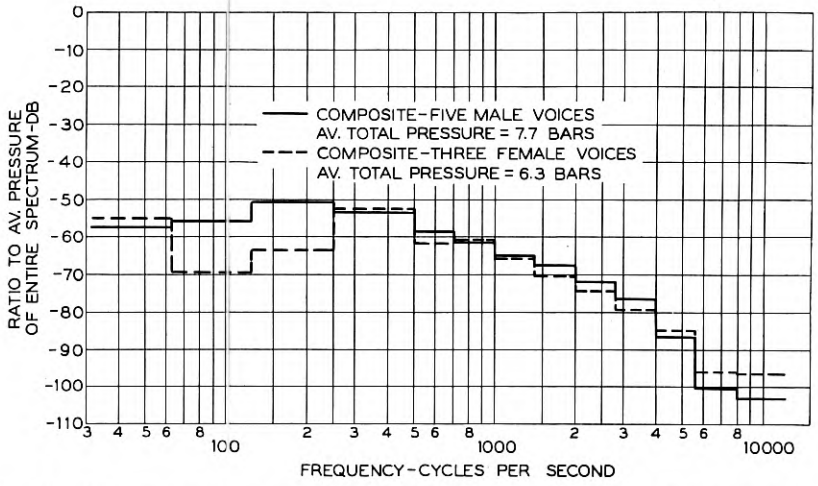


Fig. 4—Average speech pressures per frequency interval of 1 cycle per second—normal conversational voice. Distance 2″.

tube rectifier, and a fluxmeter integrates the rectified current over the duration of the speech. Simultaneously with this measurement the total spectrum also is rectified and integrated. The integration periods used were 15 seconds long. The pressure in any one band is

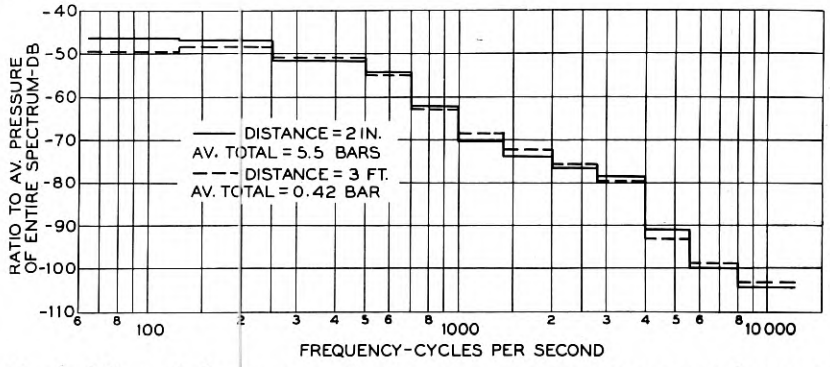


Fig. 5—Effect of distance on voice spectrum—average pressure per frequency interval of 1 cycle per second—normal speech-composite, three male voices.

given by the product: ($10^{0.05\alpha}$ × band width in cycles × total pressure), where α is the ordinate expressed in decibels (db). In using average spectra of this sort it is well to remember that they are

determined not only by the amplitudes in the individual speech sounds but equally by frequency of occurrence. Thus the fact that nearly all vowels for male voices at normal levels have a fundamental frequency between 80 and 150 p.p.s. tends to accentuate that region even though the fundamental amplitudes in individual speech sounds not be outstandingly large.

The data in Fig. 4 are for close talking conditions, 5 cm. from the lips to the diaphragm. Fig. 5 shows the effect of distance on the spectral distribution. The two distances are 5 cm. and 90 cm. respectively, in both cases the transmitter being set into a large fiber-board wall. The shapes of the two spectra are almost identical

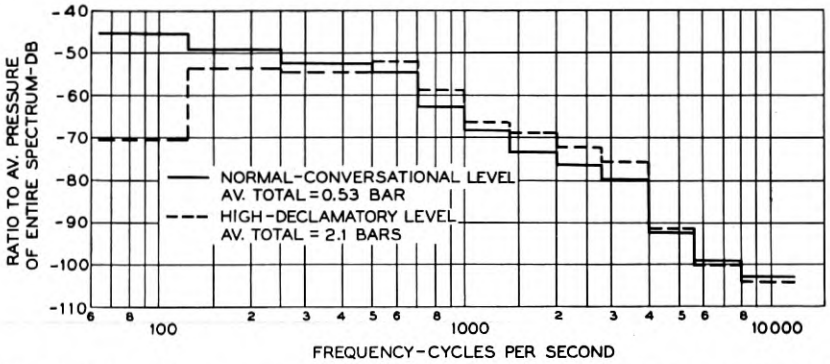


Fig. 6—Effect of level on voice spectrum—average pressures per frequency interval of 1 cycle per second—composite, three male voices, distance 3 ft.

indicating that, on the average, even at 5 cm. the condenser microphone does not greatly affect the voice as a sound generator. The largest difference between the two is in the lowest band, from 62 to 125 p.p.s., perhaps owing to the relatively low radiation efficiency of the voice at those frequencies.

The ratio of the distances is 18 : 1, that of the average pressures 14 : 1. The average pressure is nearly inversely proportional to the distance, part of the difference probably being chargeable to the more nearly total reflection for the distant condition. It is implied, of course, that even for the latter condition the direct sound is large compared with that reaching the microphone by reflections.

So far the spectra discussed were those of normally modulated voices. Fig. 6 shows what happens to the average spectrum when a high, rather declamatory level is adopted. The higher level is relatively poorer in frequencies below 500 p.p.s., relatively richer between 500 and 4000 p.p.s., and above 4000 p.p.s. its spectrum is nearly the

same as for normal levels. The decrease at low frequencies is most pronounced in the band from 62 to 125 p.p.s., i.e. in the region of the voice fundamental frequency at normal levels. This leaves two possibilities. Either at the high speech level the fundamental frequency is the same as for normal, but its amplitude is relatively small; and the spacing of the overtones is the same as at normal levels. Or else—and more probably as indicated by the auditory pitch sense —passage to the high level is attended by an actual upward shifting of the fundamental frequency, and a correspondingly larger spacing

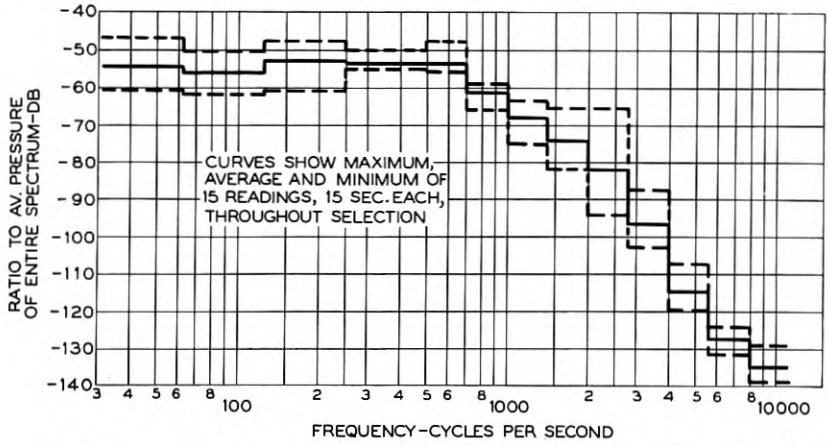


Fig. 7—Average pressures per frequency interval of 1 cycle per second—piano selection—Liszt's "Hungarian Rhapsody No. 2"—average total pressure = 3.5 bars.

of the overtones. In addition the amplitude of this new fundamental would be relatively lower than for the normal level. The average spectrum method is incapable of deciding between the two alternatives. Some oscillograms were made at normal and high speech levels, which clearly indicated that the second alternative is the correct one, or at least the prevalent one. For most vowel sounds, the fundamental frequency was found shifted from about 100 p.p.s. at normal to about 200 p.p.s. at the higher level. This result has an intimate bearing on the loss of naturalness encountered when speech originally picked up at normal voice levels is subsequently reproduced at much higher levels. A corresponding change, though in the opposite direction, probably takes place in going from normal to subnormal levels. Some evidence of this will be seen in the section on peak amplitudes.

The method of average pressure-frequency spectra is equally

applicable to sounds other than speech, provided they are sustained or can be repeated. Two illustrations are given. Fig. 7 is for a piano composition. Characteristic in comparison with speech is the relative smallness of high frequencies. Spectra of radically different types of compositions were found to be rather similar. The instrument, the room acoustics and the position of the microphone largely determine the picture. In this case, the microphone was about 7 ft. from the keyboard, and the reverberation time between 1.2 seconds at 100 p.p.s. and 1.0 second at 4000 p.p.s.

Fig. 8 gives the spectrum of street noise entering a laboratory

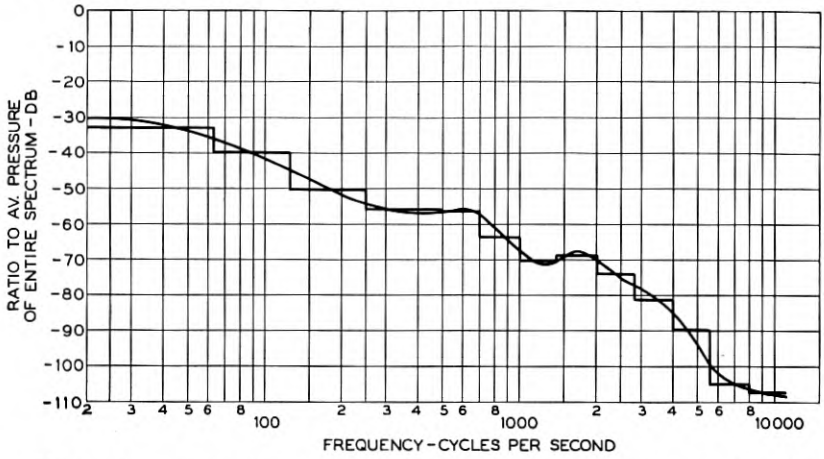


Fig. 8—Average pressures per frequency interval of 1 cycle per second—street noise—10th floor Bell Telephone Laboratories facing east—2.50–3.20 P.M., Oct. 10, 1928—average total pressure = 0.75 bar.

window, ten floors above the street. Street traffic and an elevated railway are the chief contributors, the measurements being taken during traffic peak hours. The relative poverty in high frequencies is even more pronounced than for the piano.

We shall now consider the type of apparatus intended primarily for measurements when the speakers are not acoustically controllable. Instead of averaging over a large number of words the measurement is essentially that of mean power in syllables. Usually it covers the whole spectrum rather than a particular frequency band. It has been widely used for controlling amplification in radio broadcasting, in phonograph and film recording of speech, and for rapid measurement and electrical control of speech levels in telephone conversations. A typical device of this sort is the "volume indicator" [9] shown in

[9] Essential features of this apparatus are shown by E. L. Nelson, U. S. Patent No. 1523827, filed 8/31/22.

Fig. 9, developed about ten years ago. Essentially it is a vacuum tube rectifier with a rapid action d.c. meter in the plate circuit. It is operated on a part of the characteristic such that the rectified plate current is roughly proportional to the square of the speech voltage. The rectifier is preceded by an amplifier of adjustable gain. For the speech level under measurement the gain is adjusted to such a value that the fluctuating meter deflections attain a prescribed

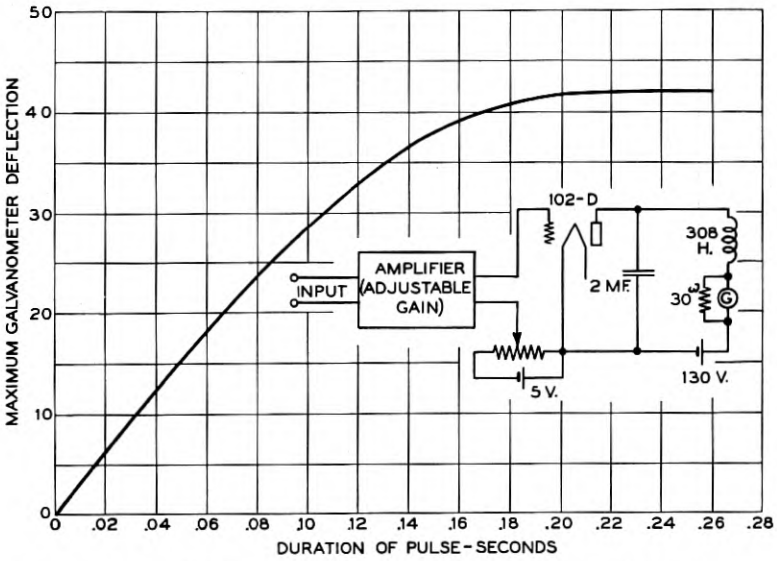


Fig. 9—Volume indicator—deflection as a function of a.c. pulse duration.

maximum value on the average of once in about three seconds. That value of gain, expressed in decibels with respect to a certain normal value of the gain, gives the volume indicator measure of the speech level. The meter, combined with the electric circuit, has a dynamic characteristic as shown in the curve, which gives the maximum deflection as a function of the duration of the a.-c. input. For inputs lasting more than about 0.18 second the maximum deflection remains the same. Since the average syllable duration is of the order of 0.2 second, it follows that the maximum deflection of the "volume indicator" meter is approximately proportioned to the mean power in the syllable. By comparison with oscillograms, or by equivalent methods, the volume indicator readings can be correlated by absolute quantities. Fig. 10 is an example showing the relation between speech level as measured with the volume indicator and the average

instantaneous amplitude of the speech waves on a certain laboratory telephone circuit of the commercial type.

Fig. 11 illustrates the type of data which can be expeditiously secured by means of the volume indicator. Here the levels for a

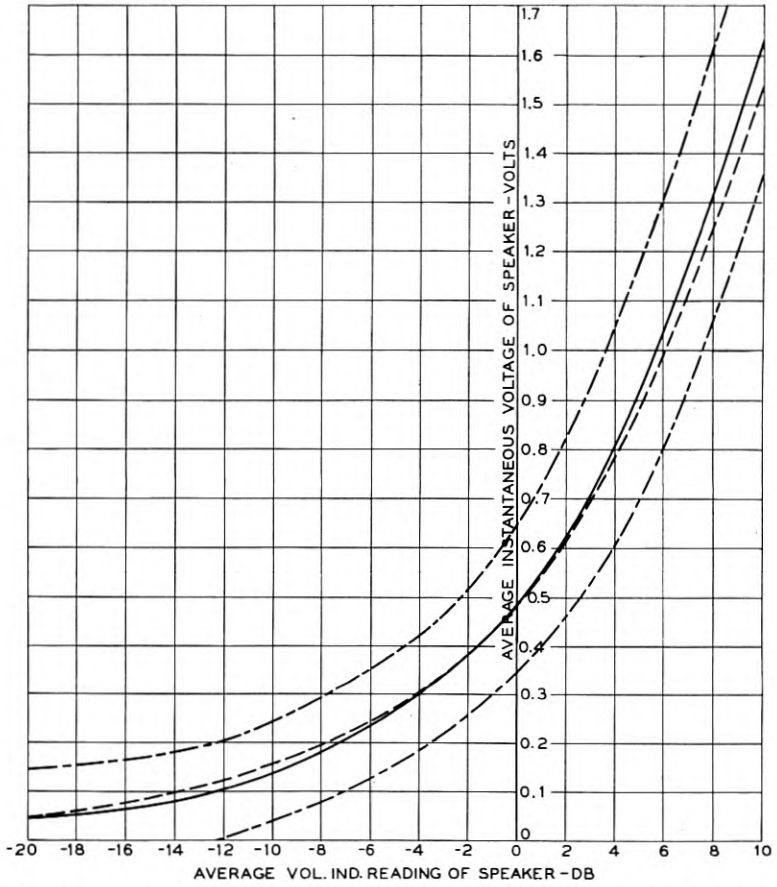


Fig. 10—Instantaneous voltage V.S. Volume indicator reading 85% of observed points included between outer dashed curves—ordinates of middle dashed curve are inversely proportional to gain of volume indicator.

large number of speakers and conversations (over a laboratory telephone circuit of the commercial type) are determined. To cover the same ground by means of oscillographic measurements would have been a decidedly formidable task.

Another device for the rapid measurement of speech levels is the "impulse meter," [10] shown in Fig. 12. This is essentially a peak

[10] D. Thierbach, *Zs. F. Techn. Physik*, No. 11, 1928.

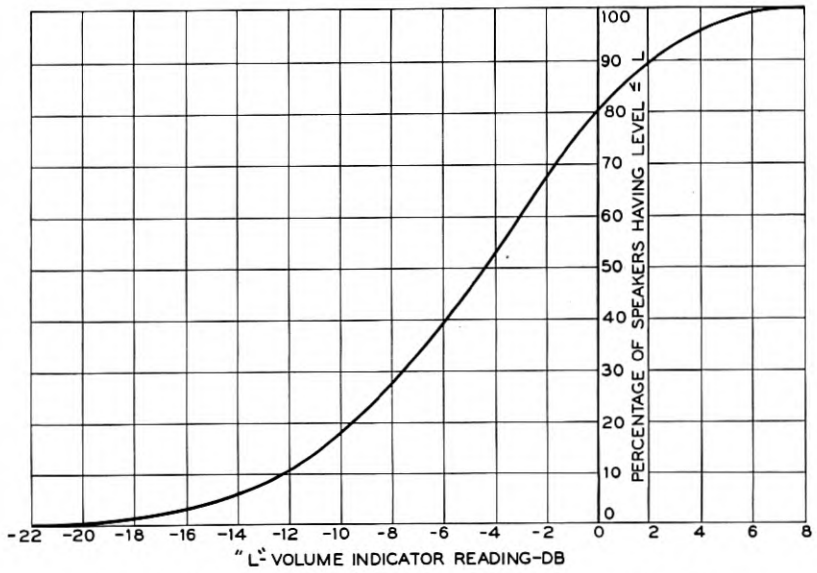


Fig. 11—Distribution of speaker's levels—87 men and 59 women.

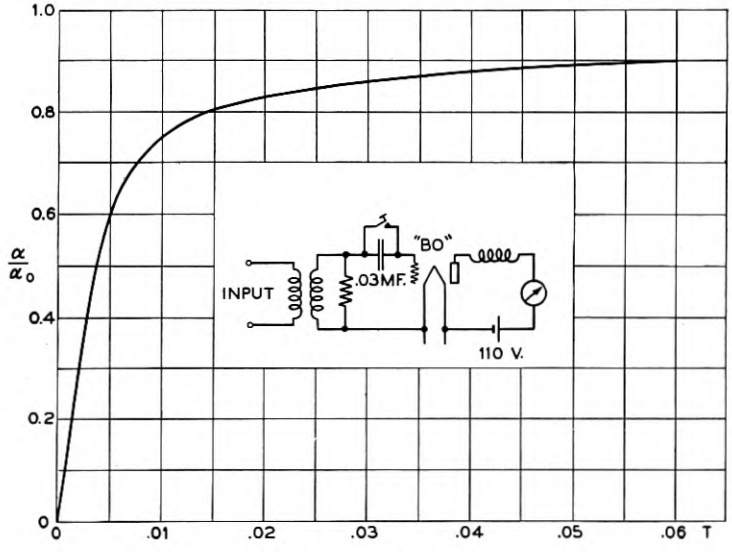


Fig. 12—Impulse meter.

T = Duration of a.c. impulse
α = Corresponding ultimate plate current
α_0 = Ultimate plate current for steadily applied a.c.

voltmeter. The curve shows the time rate at which the potential on the blocking condenser builds up. The time required for the galvanometer to reach its maximum deflection is determined by the dynamic characteristic of the meter and associated plate circuit, as well as by the time constant of the condenser charging circuit.

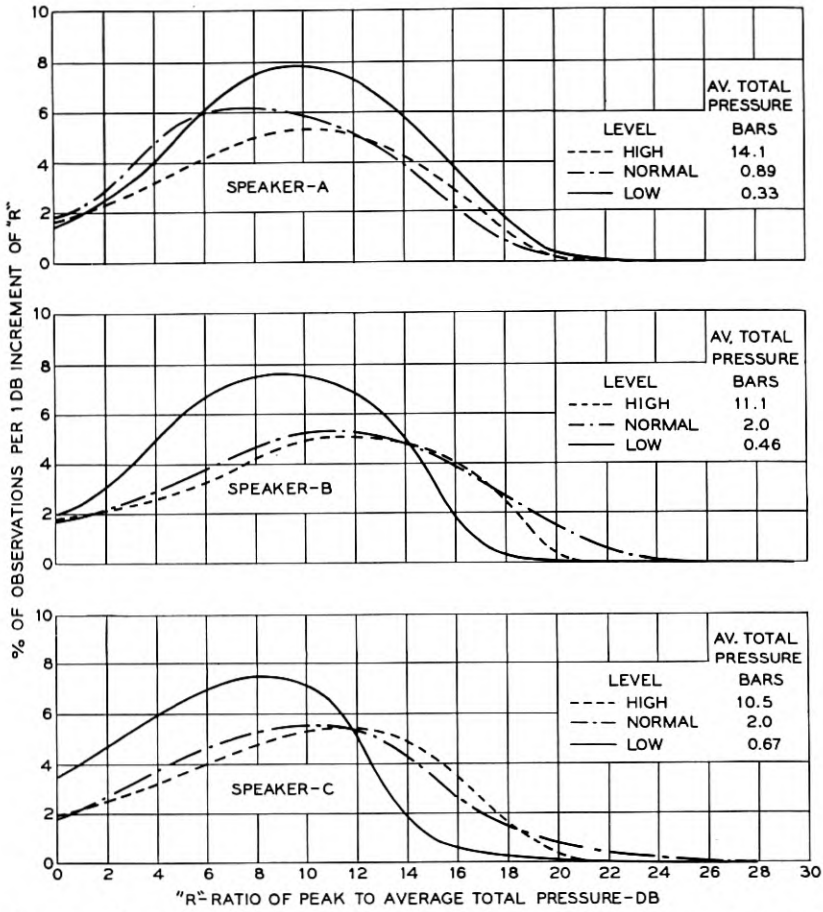


Fig. 13—Distribution of peak pressures for entire speech spectrum—three male voices—each peak is the maximum instantaneous pressure during 1/8 sec.

In using the volume indicator type of instrument in the telephone plant, limits are observed which have been determined by trial to give satisfactory operating results. The fact that it is possible to set such limits indicates a correlation between mean syllabic power and peak pressures. The overload capacity of apparatus is so important that

laboratory investigations of quantities affecting it merit more fundamental study than can be made with instruments of the volume indicator type. The quantities of interest are the instantaneous amplitudes of the wave peaks, the frequency of their occurrence and their distribution in the frequency spectrum. The apparatus used to obtain the data described below will be described in detail elsewhere. It is sufficient to state here that it is capable of directly (and automatically) registering the magnitudes and frequency of occurrence of the in-

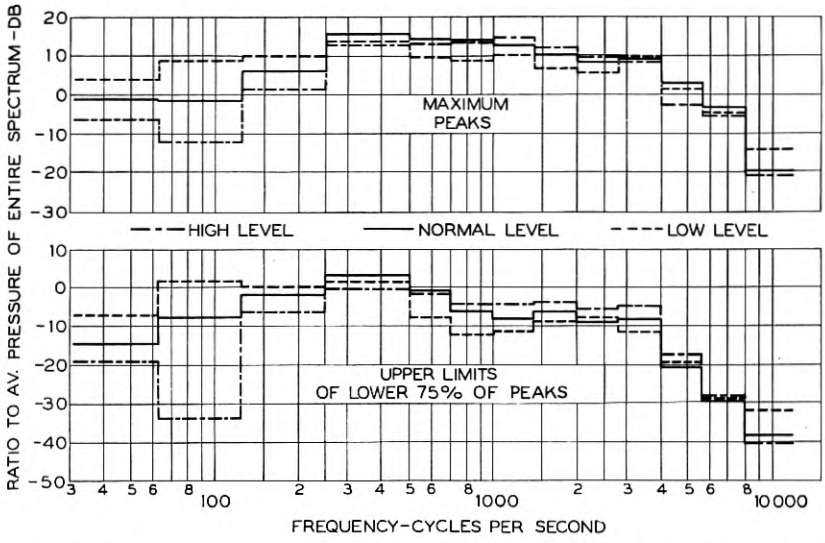


Fig. 14—Peak pressures of speech—composite, three male voices—each peak is the maximum instantaneous pressure during 1/8 sec.—average total pressure:
High (declamatory) level—8.7 bars;
Normal (conversational) level—1.6 bars;
Low (confidential) level—0.5 bar.

stantaneous peaks over a range of 60 db, corresponding to a power ratio of 1,000,000 : 1. The peak amplitudes were measured for the whole spectrum and also for restricted frequency bands selected by filters.

Fig. 13 and Fig. 14 show the results of some measurements with undistorted speech. Each individual observation gives the magnitude of the peak pressure in a 1/8 second interval. This is about as short an interval as one can use and still retain a high degree of probability that the individual measurements will give the maximum peak amplitudes in syllables. Otherwise, from the standpoint of the apparatus, the individual observations could be confined to much shorter time intervals, resulting in many more measurements for a

given length of speech. The ordinates give the ratios of the peak pressures to the average total speech pressure. In Fig. 13 the peaks are those of speech as a whole; in Fig. 14 the ordinates give the peak pressures in the several frequency bands. This is done for three widely different average levels spread over a range of 30 db. From

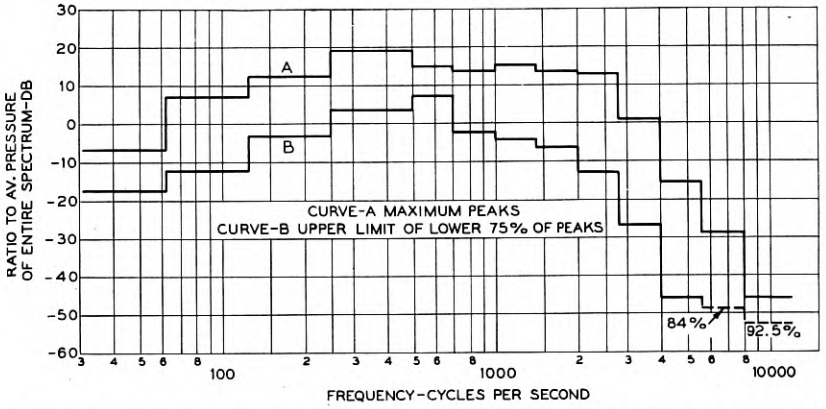


Fig. 15—Peak pressures in piano music—Liszt's "Hungarian Rhapsody No. 2" —each peak is the maximum instantaneous pressure during 1/8 sec.—average total pressure 4.0 bars.

the standpoint of providing apparatus overload capacity it is sometimes important to know not only the maximum values of all peaks (which might be uneconomical to provide for) but the upper limit of a certain percentage of all the peaks. The lower half of Fig. 14 illustrates this method of analyzing the peak data. It is interesting to note how much larger the transmission capacity must be to take care of the highest 25 per cent of the peaks.

Fig. 15 gives a similar picture of the peak amplitude distribution in a piano composition, for which the average amplitudes are given in Fig. 7. The microphone was about 7 ft. laterally from the center of the keyboard, and the measurements were made in a room having an average reverberation of about 1 second.

My thanks are due to Dr. H. K. Dunn and Mr. S. D. White of Bell Telephone Laboratories, who have obtained most of the data discussed in this paper.

Asymptotic Dipole Radiation Formulas

By W. HOWARD WISE

THE analysis of the radiation from dipoles as given by Sommerfeld and by von Hoerschelmann is deficient in one respect: it does not give the true [1] asymptotic expressions for the radiation leaving at a considerable angle from the horizontal. The correct asymptotic formulas have already been easily supplied by an appeal to the Reciprocal Theorem; lately M. J. O. Strutt [2] has got them directly from the boundary conditions and H. Weyl [3] has derived the correct asymptotic formula for a vertical dipole at the surface of the earth by a method quite different from Sommerfeld's. In the present paper it is shown how they can be got by merely improving the rigor of Sommerfeld's analysis.

The present analysis begins with the formulas of von Hoerschelmann for the wave potentials of vertical and horizontal dipoles at a finite distance above the surface of the earth and generally follows Sommerfeld. The derivation of an asymptotic approximation for the wave potential of a vertical dipole is considerably different from Sommerfeld's and results in the simpler and more precise formulas deduced from the reciprocal theorem.

Most of the analysis is somewhat simplified by taking the permeability of the earth to be unity.

The notation used is chiefly that of Bateman.[4]

τ = variable of integration, throughout the paper.

$$k^2 = \epsilon\mu\omega^2 + 4\pi\sigma\mu i\omega$$

$$l = \sqrt{\tau^2 - k_1^2}, \qquad m = \sqrt{\tau^2 - k_2^2}.$$

The subscripts 1 and 2 refer to air and ground respectively.

R_1, R_2, a, ρ, φ, w, x, y and z are adequately defined by Fig. 1.

$$\cos\theta_x = x/R, \qquad \cos\theta_y = y/R, \qquad \cos\theta_z = z/R, \qquad R^2 = x^2 + y^2 + z^2.$$

The wave potential of a horizontal dipole is [5, 4]

[1] See paragraph following equation (8).

[2] M. J. O. Strutt, *Ann. d. Phys.*, Bd. 1, p. 721, 1929.

[3] H. Weyl, *Ann. d. Phys.*, Bd. 60, p. 481, 1919.

[4] "Electrical and Optical Wave Motion," pp. 73–75.

[5] H. v. Hoerschelmann, *Jahrb. d. draht. Teleg.*, Bd. 5, 1912, pp. 14–188.

$$\Pi^h = i \times \left(-\frac{e^{ik_1R_1}}{R_1} + \frac{e^{ik_1R_2}}{R_2} - \int_0^\infty \frac{2}{l+m} J_0(\tau\rho) e^{-wl} \tau d\tau \right) + j \times 0$$

$$+ k \times 2 \cos\varphi \int_0^\infty \frac{k_2^2 - k_1^2}{(l+m)(k_2^2 l + k_1^2 m)} J_1(\tau\rho) e^{-wl} \tau^2 d\tau. \quad (1)$$

$-\frac{e^{ik_1R_1}}{R_1} + \frac{e^{ik_1R_1}}{R_2} = \Pi_x^\infty$ is the wave potential of a dipole placed parallel to a perfectly conducting plane, k_2 infinite. So, writing

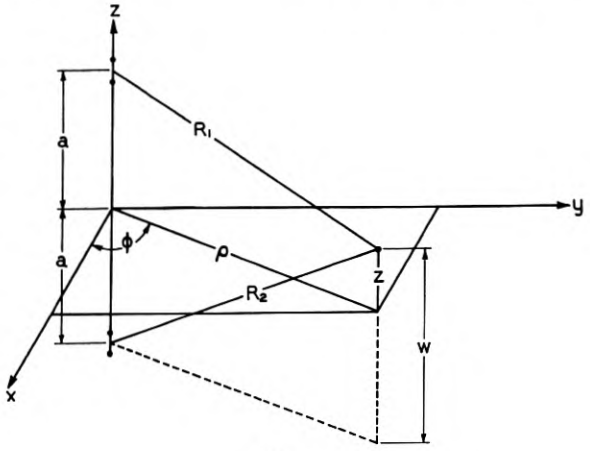


Fig. 1

$\Pi^h = \Pi_x^\infty + \Pi^\Delta$ and correspondingly $E^h = E^\infty + E^\Delta$

$$E_x^\Delta = -i\omega \left[\Pi_z^\Delta + k_1^{-2} \frac{\partial}{\partial z} \left(\frac{\partial}{\partial x} \Pi_x^\Delta + \frac{\partial}{\partial y} \Pi_y^\Delta + \frac{\partial}{\partial z} \Pi_z^\Delta \right) \right]$$

$$= -i\omega 2 \cos\varphi \frac{1}{k_1^2} \int_0^\infty J_1(\tau\rho) \frac{e^{-wl}}{l+m} \left(\frac{(k_2^2 - k_1^2)\tau^2}{k_2^2 l + k_1^2 m} - l \right) \tau^2 d\tau$$

$$= -i\omega 2 \cos\varphi \int_0^\infty J_1(\tau\rho) e^{-wl} \frac{-m\tau^2}{k_2^2 l + k_1^2 m} d\tau,$$

$$= -i\omega \frac{2}{k_1^2} \frac{\partial}{\partial x} \int_0^\infty J_0(\tau\rho) e^{-wl} \left(\frac{-k_2^2 l}{k_2^2 l + k_1^2 m} + 1 \right) \tau d\tau$$

$$= -i\omega \frac{1}{k_1^2} \frac{\partial^2}{\partial x \partial w} \left(V - 2\frac{e^{ik_1R_2}}{R_2} \right), \quad (2)$$

where

$$V = \int_0^\infty \frac{2k_2^2}{k_2^2 l + k_1^2 m} J_0(\tau\rho) e^{-wl} \tau d\tau. \quad (3)$$

When $k_2 = \infty$, $V = 2R_2^{-1} \exp ik_1R_2$ and E_z^{Δ} is zero. When $k_2 = k_1$, $V = 1R_2^{-1} \exp ik_1R_2$ and E_z^{Δ} just cancels the field of the image dipole in E^{∞}.

The wave potential of a vertical dipole is [5, 4]

$$\Pi_z^v = \frac{e^{ik_1R_1}}{R_1} - \frac{e^{ik_1R_2}}{R_2} + V. \tag{4}$$

V is the function Π_0 analysed by Sommerfeld in Riemann-Webers Differentialgleichungen der Physik. Sommerfeld transforms the integration from 0 to ∞ into an integration from $-\infty$ to $+\infty$ by replacing the Bessel function by its equivalent Hankel functions and then wraps the real axis path of integration around the zero of $k_2^2l + k_1^2m$ and the two branch cuts from k_1 and k_2 to $+i\infty$. He thus gets $V = P + Q_1 + Q_2$ where P is the integral around the zero of $k_2^2l + k_1^2m$, Q_1 is the integral around the branch cut from k_1, Q_2 is the integral around the branch cut from k_2 and the function integrated is

$$\frac{k_2^2}{k_2^2l + k_1^2m} H_0^1(\tau\rho)e^{-wl}\tau d\tau. \tag{5}$$

l and m are pure imaginaries on the branch cuts, which are rectangular hyperbolas, from k_1 and k_2 respectively. As one carries τ counterclockwise around the branch cut from k_1 to $+i\infty$, l travels up the right hand side of the imaginary axis from $-i\infty + \epsilon$ to $+i\infty + \epsilon$. A similar statement holds for m.

Multiplying numerator and denominator of equation (5) by $k_2^2l - k_1^2m$ and then taking out a factor $(k_2^4 - k_1^4)^{-1}$ our integrand becomes

$$\frac{k_2^2}{k_2^4 - k_1^4} \frac{k_2^2\sqrt{\tau^2 - k_1^2} - k_1^2\sqrt{\tau^2 - k_2^2}}{(\tau - s)(\tau + s)} H_0^1(\tau\rho)e^{-w\sqrt{\tau^2-k_1^2}}\tau d\tau,$$

where $s = +k_1k_2 \div \sqrt{k_2^2 + k_1^2}$. Integrating around the pole at $\tau = s$, we get

$$\begin{aligned} P &= \frac{k_2^2}{k_2^4 - k_1^4} 2\pi i \frac{k_2^2\sqrt{s^2 - k_1^2} - k_1^2\sqrt{s^2 - k_2^2}}{2s} H_0^1(s\rho)e^{-w\sqrt{s^2-k_1^2}}s \\ &= 0 \quad \text{if} \quad k_2^2\sqrt{s^2 - k_1^2} - k_1^2\sqrt{s^2 - k_2^2} = 0 \\ &= -2\pi \frac{k_2^3k_1s}{k_2^4 - k_1^4} H_0^1(s\rho)e^{-iwsk_1/k_2}, \end{aligned} \tag{6}$$

if

$$k_2^2\sqrt{s^2 - k_1^2} - k_1^2\sqrt{s^2 - k_2^2} = 2k_2^2\sqrt{s^2 - k_1^2} = 2ik_1k_2s.$$

l and m must have their real parts positive and so in taking the square roots of $s^2 - k_1^2 = -k_1^4 \div (k_2^2 + k_1^2)$ and $s^2 - k_2^2 = -k_2^4 \div (k_2^2 + k_1^2)$ one halves the smallest angle with the positive real axis. If they both lie on the same side of the real axis P is zero. In order that they may lie on opposite sides of the real axis it is necessary that

$$\arg k_1^4 < \arg (k_1^2 + k_2^2) < \arg k_2^4.$$

Writing $k_1^2 = \alpha + i\beta$ and $k_2^2 = x + iy$ this means that

$$y > \frac{2\alpha\beta}{\alpha^2 - \beta^2} x + \frac{\alpha^2 + \beta^2}{\alpha^2 - \beta^2} \beta.$$

The goal of the paper being asymptotic formulas for the sky waves of vertical and horizontal dipoles the ground wave, P, will hereafter be ignored. This is possible because at the high frequencies for which dipoles are useful the ground wave is very highly damped.

Sommerfeld gets an asymptotic expression for Q_1 by noting that if we are at a great distance from the source most of the value of the integral comes from that portion of the path of integration very close to k_1. The solution he arrives at is

$$-2(\Omega + \Omega^2 + \Omega^3 + + \cdots)\frac{e^{ik_1R}}{R} \quad \text{where} \quad \Omega = \frac{k_2^2}{k_1^2\sqrt{k_1^2 - k_2^2}} \frac{\partial}{\partial z}.^6 \qquad (7)$$

Neglecting higher powers of $1/R$ than the first, equation (7) sums up into

$$Q_1 \sim \frac{2k_2^2 \cos \theta_z}{k_2^2 \cos \theta_z + k_1\sqrt{k_2^2 - k_1^2}} \frac{e^{ik_1R}}{R}. \qquad (8)$$

But in getting equation (7) Sommerfeld has replaced $\sqrt{\tau^2 - k_2^2}$ by $\sqrt{k_1^2 - k_2^2}$. This is a needless approximation which ruins the symmetry, damages the utility and tends to hide the physical meaning of the final result. To get the true asymptotic formula for Q_1 it is necessary to confine the approximations to the purely operational variety, i.e. make no approximations of substitution before integrating but let the approximation reside wholly in the manner of integrating, as follows

$$Q_1 = \int \frac{k_2^2}{k_2^2 l + k_1^2 m} H_0^1(\tau\rho) e^{-wl} \tau d\tau$$

$$= -\frac{\partial}{\partial w} \int H_0^1(\tau\rho) \frac{e^{-wl}}{l} (C_0 + C_1 l + C_2 l^2 + \cdots) \tau d\tau$$

[6] *Annalen der Physik*, Band 28, 1909, page 705.

where

$$C_0 + C_1 l + C_2 l^2 + \cdots = k_2^2 \div (k_2^2 l + k_1^2 \sqrt{l^2 + k_1^2 - k_2^2}).$$

To the extent that most of the value of the integral comes from that portion of the path of integration very close to $\tau^2 = k_1^2$ the expansion in powers of l is valid. Replacing each l by $-(\partial/\partial w)$ we have then

$$Q_1 \sim -2\frac{\partial}{\partial w}\left(C_0 - C_1\frac{\partial}{\partial w} + C_2\frac{\partial^2}{\partial w^2} - + \cdots\right)\frac{e^{ik_1R_2}}{R_2}. \tag{9}$$

Now

$$-\frac{\partial}{\partial w}\frac{e^{ik_1R_2}}{R_2} = (-ik_1\cos\theta_z)\frac{e^{ik_1R_2}}{R_2} + \frac{\cos\theta_z}{R_2}\frac{e^{ik_1R_2}}{R_2}$$

$$= \left[\gamma + \frac{\cos\theta_z}{R_2}\right]\frac{e^{ik_1R_2}}{R_2} \quad \text{where} \quad \gamma = -ik_1\cos\theta_z,$$

$$\frac{\partial^2}{\partial w^2}\frac{e^{ik_1R_2}}{R_2} = \left(\gamma^2 + 2\gamma^1\frac{\cos\theta_z}{R_2} + 1\cdot 2\gamma^0\frac{ik_1\sin^2\theta_z}{2R_2} + \cdots\right)\frac{e^{ik_1R_2}}{R_2},$$

$$-\frac{\partial^3}{\partial w^3}\frac{e^{ik_1R_2}}{R_2} = \left(\gamma^3 + 3\gamma^2\frac{\cos\theta_z}{R_2} + 2\cdot 3\gamma^1\frac{ik_1\sin^2\theta_z}{2R_2} + \cdots\right)\frac{e^{ik_1R_2}}{R_2},$$

. . . . . .

etc., and so

$$Q_1 \sim 2(\gamma^1 C_0 + \gamma^2 C_1 + \gamma^3 C_2 + + \cdots)\frac{e^{ik_1R_2}}{R_2}$$

$$+ 2\cos\theta_z(1\gamma^0 C_0 + 2\gamma^1 C_1 + 3\gamma^2 C_2 + + \cdots)\frac{e^{ik_2R_2}}{R_2^2}$$

$$+ ik_1\sin^2\theta_z(1\cdot 2\gamma^0 C_1 + 2\cdot 3\gamma^1 C_2 + 3\cdot 4\gamma^2 C_3 + + \cdots)\frac{e^{ik_1R_2}}{R_2^2}$$

$$+ \text{higher order terms}$$

$$= \frac{2k_2^2\gamma}{k_2^2\gamma + k_1^2\sqrt{\gamma^2 + k_1^2 - k_2^2}}\frac{e^{ik_1R_2}}{R_2}$$

$$+ \left(\cos\theta_z\frac{\partial}{\partial\gamma} + \frac{ik_1\sin^2\theta_z}{2}\frac{\partial^2}{\partial\gamma^2}\right)\frac{2k_2^2\gamma}{k_2^2\gamma + k_1^2\sqrt{\gamma^2 + k_1^2 - k_2^2}}\frac{e^{ik_1R_2}}{R_2^2}$$

$$+ \text{higher order terms.} \tag{10}$$

The second order term can be neglected if $k_1R_2 >> 1$, say if $R > 20\lambda$. It will hereafter be supposed that this is the case.

Multiplying numerator and denominator of the first order term by i and canceling a k_1 we have finally

$$Q_1 \sim \frac{2k_2^2 \cos\theta_z}{k_2^2 \cos\theta_z + k_1\sqrt{k_2^2 - k_1^2 \sin^2\theta_z}} \frac{e^{ik_1R_2}}{R_2}. \tag{11}$$

This Q_1 behaves as one would rightfully expect a true asymptotic formula to behave. It is $2(\exp ik_1R_2)/R_2$ at $k_2 = \infty$ and $1(\exp ik_1R_2)/R_2$ at $k_2 = k_1$.

In so far as k_2 is considerably larger than k_1 and the expansion in powers of l is valid Q_2 is negligible in comparison with Q_1.[7] Perhaps the easiest way of seeing this is to note that equation (10) might just as well have been obtained directly from V instead of from Q_1.

Substituting equation (11) in equation (4) we get

$$\Pi_z^v \sim \left(1 + \frac{k_2^2 \cos\theta_z - k_1\sqrt{k_2^2 - k_1^2 \sin^2\theta_z}}{k_2^2 \cos\theta_z + k_1\sqrt{k_2^2 - k_1^2 \sin^2\theta_z}} e^{ik_1(R_2 - R_1)}\right) \frac{e^{ik_1R_1}}{R_1}, \tag{12}$$

whence

$$E_z^v = -i\omega\left(\Pi_z^v + k_1^{-2}\frac{\partial^2}{\partial z^2}\Pi_z^v\right)$$

$$\sim -i\omega \sin^2\theta_z(1 + R_1e^{ik_12a\cos\theta_z})\frac{e^{ik_1R}}{R}, \tag{13}$$

where

$$R_1 = \frac{k_2^2 \cos\theta_z - k_1\sqrt{k_2^2 - k_1^2 \sin^2\theta_z}}{k_2^2 \cos\theta_z + k_1\sqrt{k_2^2 - k_1^2 \sin^2\theta_z}}.$$

Substituting equation (11) in equation (2) and adding E_z^∞ we get

$$E_z^h \sim -i\omega \cos\theta_x \cos\theta_z(1 - R_1e^{ik_12a\cos\theta_z})\frac{e^{ik_1R}}{R}. \tag{14}$$

R_1 is the coefficient of reflection for a plane wave polarized in the plane of incidence.

The horizontal fields of a horizontal dipole are

$$E_y^h = -i\omega\left[\Pi_y^h + k_1^{-2}\frac{\partial}{\partial y}\left(\frac{\partial}{\partial x}\Pi_x^h + \frac{\partial}{\partial y}\Pi_y^h + \frac{\partial}{\partial z}\Pi_z^h\right)\right]$$

$$= -i\omega k_1^{-2}\left[\frac{\partial^2}{\partial y\partial x}\Pi_x^\infty + \frac{\partial}{\partial y}\left(\frac{\partial}{\partial x}\Pi_x^\Delta + \frac{\partial}{\partial z}\Pi_z^\Delta\right)\right] \tag{15}$$

and

$$E_x^h = -i\omega\left[\Pi_x^h + k_1^{-2}\frac{\partial}{\partial x}\left(\frac{\partial}{\partial x}\Pi_x^h + \frac{\partial}{\partial y}\Pi_y^h + \frac{\partial}{\partial z}\Pi_z^h\right)\right]$$

$$= -i\omega\left[\Pi_x^\infty + k_1^{-2}\frac{\partial^2}{\partial x^2}\Pi_x^\infty\right]$$

[7] Riemann-Weber's "Differentialgleichungen der Physik," p. 556.

$$- i\omega \left[\Pi_x^{\Delta} + k_1^{-2} \frac{\partial}{\partial x} \left(\frac{\partial}{\partial x} \Pi_x^{\Delta} + \frac{\partial}{\partial z} \Pi_z^{\Delta} \right) \right]$$

$$\sim - i\omega(\Pi_x^{\infty} + \Pi_x^{\Delta}) + \frac{\cos \theta_x}{\cos \theta_y} E_y^{h}$$

$$= - i\omega(\Pi_x^{\infty} + \Pi_x^{\Delta}) + \cot \varphi E_y^{h}; \tag{16}$$

$$\therefore \quad E_{\varphi}^{h} = E_x^{h} \sin \varphi - E_y^{h} \cos \varphi \sim - i\omega \sin \varphi(\Pi_x^{\infty} + \Pi_x^{\Delta}). \tag{17}$$

Evidently the procedure which yielded the true asymptotic expres-

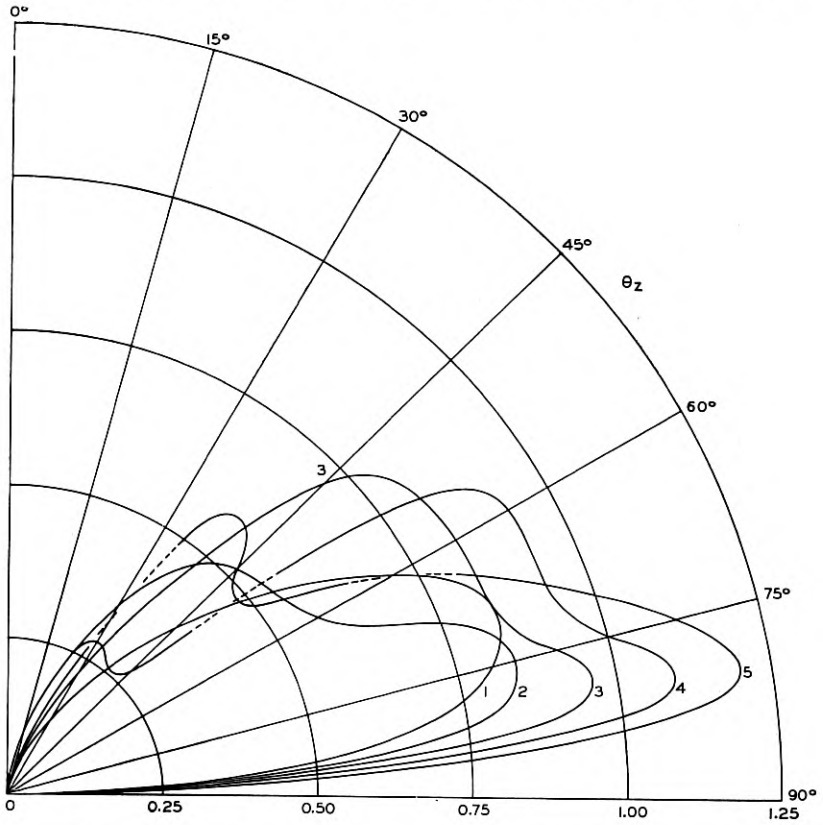


Fig. 2—Vertical dipole polar diagrams computed for $\lambda = 6$ meters, $\epsilon = 9$, $\mu = 1$, and $\sigma = 10^{-13}$

sion for V will do the same for Π_x^{Δ}. The details are not interesting. The result is

$$\Pi_x^{\Delta} \sim - \frac{2k_1 \cos \theta_z}{k_1 \cos \theta_z + \sqrt{k_2^2 - k_1^2 \sin^2 \theta_z}} \frac{e^{ik_1 R_2}}{R_2}. \tag{18}$$

Substituting equation (18) in equation (17) we get

$$E_{\varphi}{}^{h} \sim + i\omega \sin \varphi(1 - R_2 e^{ik_1 2a \cos \theta_z}) \frac{e^{ik_1 R}}{R}, \tag{19}$$

where

$$R_2 = \frac{\sqrt{k_2{}^2 - k_1{}^2 \sin^2 \theta_z} - k_1 \cos \theta_z}{\sqrt{k_2{}^2 - k_1{}^2 \sin^2 \theta_z} + k_1 \cos \theta_z}.$$

R_2 is the coefficient of reflection for a plane wave polarized perpendicular to the plane of incidence.

Formulas (13), (14) and (19) are just what one would get by applying

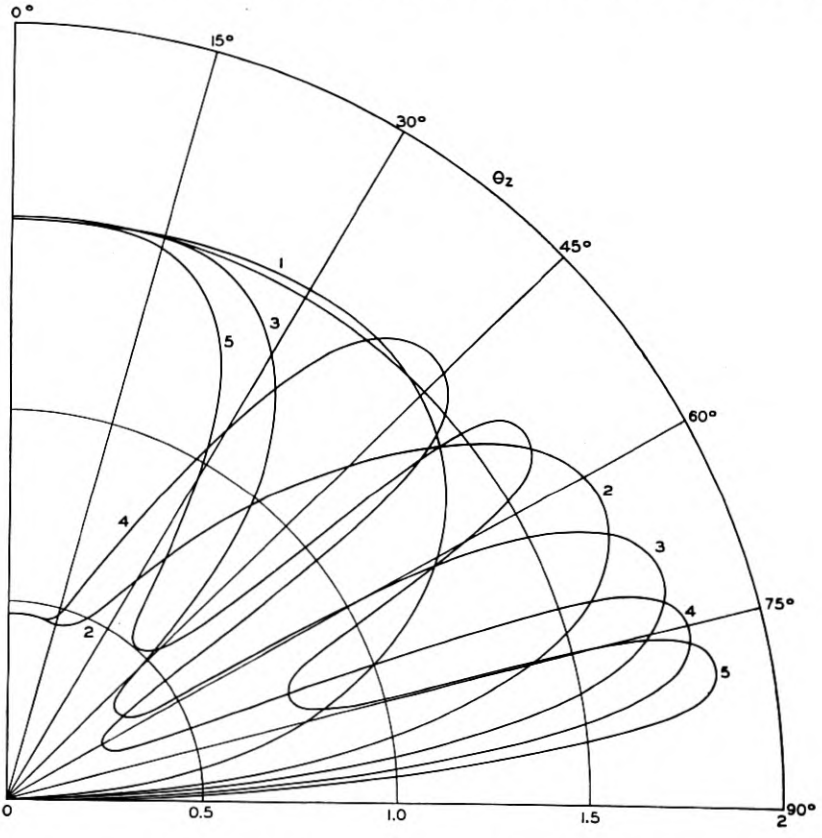


Fig. 3—Polar diagrams of the horizontal fields of horizontal dipoles computed for $\lambda = 6$ meters, $\epsilon = 9$, $\mu = 1$ and $\sigma = 10^{-13}$

the coefficient of reflection from an imperfect reflector to the reflected waves of the corresponding electrostatic formulas.

Returning to equation (15)

$$E_y^h = -i\omega k_1^{-2}\frac{\partial^2}{\partial y \partial x}\left[\Pi_x^\infty - \int_0^\infty \frac{2}{m+l}\left(1-\frac{(k_2^2-k_1^2)l}{k_2^2 l + k_1^2 m}\right) J_0(\tau\rho)e^{-wl}\tau d\tau\right]$$

$$= -i\omega\frac{\partial^2}{\partial y \partial x}(k_1^{-2}\Pi_x^\infty - k_2^{-2}V)$$

$$\sim -i\omega\cos\theta_x\cos\theta_y\left(-1+\left[1-\frac{k_1^2}{k_2^2}(1+R_1)\right]e^{ik_1 2a\cos\theta_z}\right)\frac{e^{ik_1 r}}{R}. \quad (20)$$

Usually one cares only for E_φ^h at $\varphi = \pi/2$ and E_x^h and E_y^h are of no particular interest. Figs. 2, 3 and 4 are polar diagrams in the

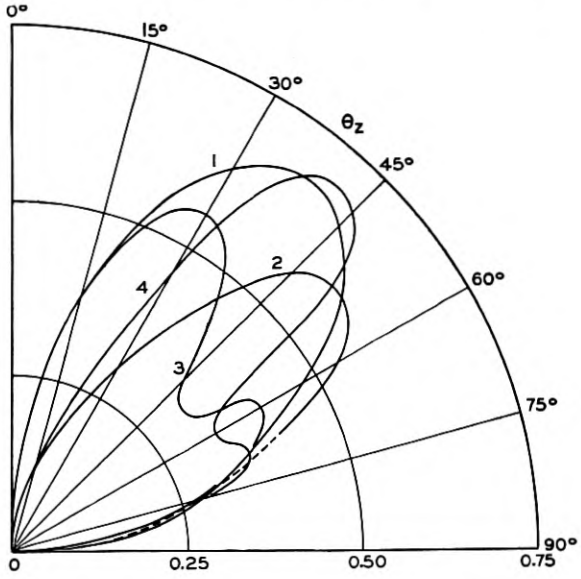


Fig. 4—Polar diagrams of the vertical fields of horizontal dipoles computed for $\lambda = 6$ meters, $\epsilon = 9$, $\mu = 1$ and $\sigma = 10^{-13}$

vertical plane of equations (13), (14) and (19). Assuming the conductivity of the air to be zero and the dielectric constant to be unity

$$k_2^2 = k_1^2(\epsilon + i2c\lambda\sigma), \qquad k_1 = 2\pi/\lambda,$$

where ϵ is the dielectric constant of the ground referred to air as unity and σ is the conductivity of the ground in electromagnetic units. The values of ϵ and σ used in computing the polar diagrams are generally supposed to be somewhere in the neighborhood of their

average values for earth but they vary so much with the locality that the diagrams can scarcely be regarded as giving more than a general idea as to what may be expected of the formulas.

The number attached to each curve is the height of its dipole in quarter wave-lengths.

Formulas (13), (14) and (19) are just what one would get by applying the Reciprocal Theorem to two dipoles, one near the earth and the other far away. The electric field acting on the one near the earth is composed of a direct field and a reflected field which is R_1 or R_2, as the case may be, times the direct field.

When the depth to groundwater, bedrock, an orebody or any marked discontinuity in the electrical properties of the ground is known and is not too great the effect of this discontinuity on the polar diagram ought not to be ignored. The asymptotic formulas for any stratified ground are got by putting the coefficients of reflection for a plane wave reflected from the surface of that ground in place of the corresponding coefficients in equations (13), (14) and (19). For a number of rather obvious reasons it would usually be out of the question to deal with more than one plane of discontinuity; one is bad enough. The coefficients for a single plane of discontinuity at a depth Δ are

$$R_1^1 = \frac{k_2^2 \cos\theta_z - \eta_1 k_1 \sqrt{k_2^2 - k_1^2 \sin^2\theta_z}}{k_2^2 \cos\theta_z + \eta_1 k_1 \sqrt{k_2^2 - k_1^2 \sin^2\theta_z}}$$

and

$$R_2^1 = \frac{\sqrt{k_2^2 - k_1^2 \sin^2\theta_z} - \eta_2 k_1 \cos\theta_z}{\sqrt{k_2^2 - k_1^2 \sin^2\theta_z} - \eta_2 k_1 \cos\theta_z},$$

where

$$\eta_1 = \frac{\mu_2}{\mu_1}\frac{1+\delta_1}{1-\delta_1}, \qquad \eta_2 = \frac{\mu_2}{\mu_1}\frac{1+\delta_2}{1-\delta_2},$$

$$\delta_1 = \frac{k_2^2\mu_3\sqrt{k_3^2 - k_1^2\sin^2\theta_z} - k_3^2\mu_2\sqrt{k_2^2 - k_1^2\sin^2\theta_z}}{k_2^2\mu_3\sqrt{k_3^2 - k_1^2\sin^2\theta_z} - k_3^2\mu_2\sqrt{k_2^2 - k_1^2\sin^2\theta_z}} e^{i2\Delta\sqrt{k_2^2 - k_1^2\sin^2\theta_z}},$$

and

$$\delta_2 = \frac{\mu_3\sqrt{k_2^2 - k_1^2\sin^2\theta_z} - \mu_2\sqrt{k_3^2 - k_1^2\sin^2\theta_z}}{\mu_3\sqrt{k_2^2 - k_1^2\sin^2\theta_z} + \mu_2\sqrt{k_3^2 - k_1^2\sin^2\theta_z}} e^{i2\Delta\sqrt{k_2^2 - k_1^2\sin^2\theta_z}}.$$

If Δ is not large and k_3 is considerably different from k_2 then η_1 and η_2 will differ considerably from unity.

Statistical Theories of Matter, Radiation and Electricity [1]

By KARL K. DARROW

The atomic or "kinetic" theory of gases, with its interpretations of such qualities as temperature, pressure, viscosity and conductivity, has ranked for more than half a century as a very important part of theoretical physics. A corresponding theory for radiation and for negative electricity is much to be desired, since it is known that in many ways each of these entities behaves as though it were atomic. There are, however, differences among the three, and only within the last five years have these been formulated suitably. This article is devoted to the resulting statistical theories.

THE major subjects of this article are two extensions of what formerly was called *atomic theory*—that is to say, the attempt to explain as many as possible of the properties of pieces of matter large enough to be visible and tangible and ponderable, by visualizing these as swarms of tiny particles each endowed with only a very few and simple qualities. Among the properties of gases, for example, are pressure and viscosity and entropy and temperature. Conceivably one might invest the ultimate atoms with all four. The atomic theory of gases as it stands today, however, is the outcome of a very different procedure. It is the achievement of an effort to interpret these four properties and several more as features of a hypothetical assemblage of very many corpuscles all alike, and not possessing them nor any others except position and velocity and mass (and moment of inertia, sometimes) and the liability to make elastic impacts with each other. On the whole the effort has been remarkably successful. Therefore viscosity and temperature and entropy are not attributed to single atoms, but pictures and expressions for them are derived as qualities of the assemblage. The theory which leads to these results is called *statistical;* it is based on certain assumptions which, in the form in which they were originally made, constitute the *classical statistics*. The successes of the classical statistics are a part of the evidence that matter is corpuscular. Once they were nearly the whole of the evidence, for they antedated the striking demonstrations of individual atoms which now spring to the mind whenever one is asked to state the reasons for accepting the atomic theory.

Radiation resembles a gas in some respects. Entropy and temperature and pressure, for example, are properties displayed by radiation when enclosed in a space surrounded by a wall of even temperature, just as they are by a gas in a like situation. It seems quite natural that one should try to interpret them in the same way as for a gas they are interpreted by the atomic theory: imagining the radiation

[1] *Physical Review Supplement*, Vol. 1, July 1929, pp. 90–155.

as an assemblage of innumerable particles, a swarm of photons or corpuscles of light. Nowadays at least the idea seems quite natural; but of course, in the years when no one as yet had broken away from the tradition that light is altogether wavelike, it would doubtless have been thought a very wild one. Even after Einstein had ventured such a breach with the past, nearly a score of years elapsed before there was developed out of the theory of quanta an adequate conception of the "radiation-gas." The historical sequence in the growth of the atomic theory of matter was here inverted: there was abundant evidence for the corpuscular theory of light, in phenomena such as the Compton effect and the photoelectric effect showing the work of individual photons, before the statistical theory of these corpuscles was perfected. We now see that the trouble was, that even when one accepts the notion of corpuscles of light without reserve, and even when one knows the proper values of energy and momentum to be assigned to these corpuscles, it still is not correct to apply to them the same statistics as gives such good results when applied to the atoms of matter. Bose discovered how to remodel the statistics, in order to construct a competent atomic theory of the radiation in thermal equilibrium in an enclosure.

The other of the new extensions of atomic theory is partly a revival —the resurrection of the theory, first proposed some thirty years ago, that part at least of the negative electricity within a metal acts like a swarm of freely-flying corpuscles which collide now and again not with each other but with the atoms. It was of course the classical statistics which was always used in developing this theory. Moribund because of several incurable discordances with fact, the theory was resuscitated by Pauli and by Sommerfeld with a revision of the statistics. It was not quite the same revision as enabled Bose to set up an atomic theory of radiation, but a very similar one, invented first by Fermi and later independently by Dirac. One cannot say that the so-renovated "electron-gas theory" is a perfect explanation of all the multifarious phenomena of the flow of electricity and heat inside of metals and outward through the boundaries of metals. Its initial successes, however, are so auspicious as to suggest that the hope of further progress lies not in renouncing it (as seemed to be almost inevitable before the alterations) but in amending it in its details.

Is the atomic theory of material gases to remain untouched by these novel ideas? Apparently all three forms of statistics, the classical and the two recent types, lead to very nearly the same conclusions when applied to material gases. Only at remarkably low temperatures and remarkably high densities do their predictions diverge; and under

these conditions the experimental data are not easy to interpret for that purpose. Suppose, however, that eventually the data are proved to decide for one of the new forms of statistics against the old: what then? Probably we shall merely remove one of the theoretical foundation-stones of the kinetic theory of gases and insert another to take its place, meanwhile leaving practically intact the great superstructure of formulæ and equations whereby the kinetic theory makes contact with experience. Happily this is an easier process in theoretical physics than in architecture.

Custom has lately changed the meaning of the term *atomic theory*, making it almost synonymous with *theory of the structure of the atom;* but the province which this latter has taken for its own is one to which its forerunner disclaimed all right of entry. It was never supposed that all of the properties of a gas can be interpreted as statistical features of a swarm of corpuscles. The earlier atomic theory conceded some of them to the individual atoms, thus in effect renouncing the ambition to explain them; and among these were the spectra. Where the statistical theory left off, the builders of atom models took up the work. Bohr, for example, designed a model for the individual hydrogen atom, competent—at least to a great extent—to explain the Balmer series and the rest of the line-spectrum of "atomic hydrogen." This model he constructed, following Rutherford, out of a pair of corpuscles. What he and his successors thus developed was in a way an *atomic theory of the atom*—a degree deeper, or further, or higher perhaps, then the atomic theory of matter which had provided him with the notion and the scale of the atom to begin with.

What then distinguishes this new "atomic theory of the atom" from its ancestor? Well, the major differences in method and in aim are traceable to the fact, that in the later theory the number of elementary particles which constitute the system is quite manageably small, while in the earlier, it is inconceivably tremendous.

Bohr constructed his model for the hydrogen atom with only a pair of corpuscles, and those for all the other atoms out of not more than a few dozen each. Now with a model consisting only of two particles, one can specify positions and velocities for these with the utmost of precision, and go merrily ahead predicting and describing orbits with as much exactness as one cares to lavish. Even with dozens of electrons and a nucleus one can attain at least a specious accuracy of detail; remember the portraits of the electron-orbits of massive atoms which six or seven years ago were so profuse. Perhaps it is not wise to make such definite assertions; but it is feasible. Not so, however, with the subjects of the older theory.

The model proposed for a cubic centimetre of gas under the ordinary conditions of temperature and pressure consists of something like 10^{20} particles. Merely the mention of so extravagant a figure is sufficient to persuade that it is vain to dream of making any progress by postulating a definite position and a definite velocity for each of these. The life of the human race would not be long enough to write down even the postulates, to say nothing of the inferences.

This seems a fearful handicap; but it is not so at all. Adopting the statistical method, one does not even begin upon the hopeless task of fixing place and motion for every particle. We content ourselves with writing down a function, which states how many among the multitude of particles we assume to be situated in each small (but not too small) element of volume; and how many we assume to have momenta which lie in each small (but not too small) range of momentum. These are specifications much more modest and vague; but they are ample. For the things which we wish to interpret—entropy and temperature, viscosity and conduction and diffusion—the atomic picture need not be made one whit more definite.

In saying this I am understating the case. If the atomic picture could be made more definite, say by stating the locations and the velocities of all the atoms with absolute precision, the meanings which we shall presently attach to entropy and temperature would be dissolved. Our theory of these entities depends upon the vagueness of the picture. Seemingly they appeal to us as physical realities because our senses and our instruments are too obtuse to perceive the atoms. Our minds must feign a somewhat similar obtuseness, pretending not to fix the particles of the imagined swarm too sharply; therefore it does not matter that they are so numerous that the pretence becomes sincere. Exact knowledge of the individual atoms is unattainable; but it is useless, is not desirable even. One remembers Æsop's tale of the fox and the inaccessible grapes; in this case it is probable that the grapes really *are* sour.

For that matter, perhaps they do not even exist. One of the most striking of the very recent ideas in theoretical physics is the thought, that even for atom-models with but a few particles, even in thinking of an isolated particle, it may be altogether pointless to assign exact positions and velocities. In dealing with a swarm of particles by the statistical way, we do in effect fix the position of each corpuscle, but with a certain latitude; we fix the velocity of each, but again with a certain latitude. Perhaps this latitude, this indefiniteness, is something inherent in nature. Insisting as I am upon the contrast between the theory of the structure of the atom and the statistical

theory of matter and radiation, I may in effect be insisting on the contrast between a faulty way of visualizing some phenomena, and a correct way of visualizing all.

A function of the sort which I just mentioned, a so-called *distribution-function*, is the goal of every statistical theory. I have said that it states how many among the multitude of particles we *assume* to be located in each small element of space and to have momentum comprised in each small range of values of momentum. So it does; but the purpose of a statistical theory is, to derive it from assumptions still more fundamental, in preference to assuming it outright. Of course one might say instead, that the reason for deriving a distribution-function is to put the fundamental assumptions to their test. Whichever viewpoint one prefers, it is the distribution-function which is tested by experiment: indirectly, in that it supplies numerical values for such things as conductivity, viscosity, specific heat; and directly, for there are now immediate ways of observing it in certain cases.

A distribution-function commonly appears in an equation of this form:

$$dN = f(x, y, z, p_x, p_y, p_z) \cdot dxdydzdp_xdp_ydp_z. \qquad (1)$$

Such an equation will as a rule refer to some particular assemblage of particles, say N altogether, occupying some definite region of space: a gas in a tube, radiation in a cavity, electrons in a wire. It is to be read as follows: "dN, equal to $f \cdot dxdydzdp_xdp_ydp_z$, stands for the number of particles having coordinates in dx at x, in dy at y, in dz at z, and components of momentum in dp_x at p_x, in dp_y at p_y, in dp_z at p_z." The phrasing "in dx at x" is a succinct alternative for "between x and $x + dx$."

The function f is the distribution-function in the variables in question—here the *coordinates* of the particles referred to some Cartesian frame in the ordinary or "coordinate" space, and the components of momentum resolved along the axes of that frame, the *momenta*. Heretofore it has been customary to use the components of velocity rather than the momenta, but these are much to be preferred: partly because it is they which figure in the canonical equations, but chiefly because we shall find when we pass over to the study of assemblages of photons that the momenta play the same role in these as they do in assemblages of atoms, while the speeds of all photons are the same. There is a well-known formula for translating a distribution-function from one set of variables to another set dependent

on the first, which we shall use in special cases.[2] It is also well known that to obtain the distribution-function in *some* of the independent variables from the distribution-function of *all* of them, it is necessary to integrate the latter over the entire range of all the *other* variables: in such a case as is symbolized by equation (1), the distribution in p_z would be obtained by integrating f with respect to the first five variables over the entire range of each.

The product $dxdydz$ is an element of volume in ordinary or *coordinate-space;* the product $dp_x dp_y dp_z$ is an element of volume in *momentum-space*, in which each particle is represented by a point having for its coordinates in a Cartesian frame the values of its momenta; the product $dxdydzdp_x dp_y dp_z$ is an element of volume in *phase-space.* The function f describes the distribution of the assemblage in this phase-space of six dimensions. In some cases—for instance, that of electrons in a metal not at an even temperature, and that of oscillators —we shall have to think continually of this six-dimensional space. In others—whenever we deal with photons, and whenever we consider atoms or electrons in a region where neither temperature nor potential varies from place to place—we shall be able to assume that the distribution in the coordinate-space is uniform (that f is independent of x, y, z) and to dismiss it from mind, and to derive the distribution in the three-dimensional momentum-space quite separately as if there were no other. Even in these simplest cases it would no doubt be more consistent to operate always in the phase-space. Unhappily the human mind is so constructed, that no matter how much it may ratiocinate about space of six dimensions or six trillion, it always visualizes in space of three.

In an equation such as (1), the differential element or the product of such elements which terminates the right-hand member must be neither too large nor too small. If it is so large that f varies considerably from one point in it to another, then its multiplier, which is by definition the *mean value* of f in the said element, must be computed by the methods of integral calculus. If on the other hand it is so small that it contains only a few of the corpuscles, then the product of f into its size may be many times as great or many times as small as the number which it does contain. This is easily perceived by proceeding to the absurd limit of dividing the space into say ten times as many elements as there are corpuscles, so that in at least

[2] Let $u_1, u_2, \cdots$ represent the variables of the first set, $v_1, v_2, \cdots$ those of the second; let $f(u_1, u_2, \cdots)$ and $F(v_1, v_2, \cdots)$ stand for the distribution-functions in the two sets; then

$$F(v_1, v_2, \cdots) = f(v_1, v_2, \cdots, \frac{\partial(u_1, u_2, \cdots)}{\partial(v_1, v_2, \cdots)}.$$

nine-tenths of the elements the number of participles is zero while f is greater than zero, and in the others the number is generally much greater than f times the size of the element. To subdivide the space so finely would be to make the atomic picture too definite, and ruin it for the purposes for which we now require it.

It is not too early in this paper for me to say emphatically that the differential elements which figure in equations such as (1) *must not be identified* with the elementary compartments of the phase-space, which we shall presently encounter, and which are so important in the new statistics and in the old alike. It takes a great many of these latter to make up an element large enough to be employed in an equation like (1). Otherwise expressed: the subdivision of the phase-space into the elementary cells or compartments of the forthcoming theory is much too fine to be used in connection with the distribution-function. Much confusion may arise from failing to realize this.[3]

In speaking of the distribution-function, I have been tacitly assuming that there is such a thing as a stable, self-sustaining, changeless distribution of the atoms of a gas, the photons in a cavity, the electrons in a wire. This assumption must now be examined. It is scarcely self-evident; one might guess at first that the more numerous the particles, the more abruptly would the distribution vary from one moment to the next, and that an assemblage of 10^{20} particles would be in such unceasing turmoil that it would be senseless to imagine one single distribution for it.

Experience however shows the reverse. The gas in a tube remains uniformly dense and stationary, it does not surge forever to and fro nor huddle in a corner nor become spontaneously hot at one end and cold at the other. In the radiation in a cavity with heated walls the intensity comprised within any portion of the spectral range remains unchanged so long as the temperature of the walls is constant. The distribution-in-velocity of the electrons streaming from a heated filament does not appear to change. Moreover, when by artifice the gas in a tube is forced to assume uneven density, non-uniform temperature, or any sort of flow or turbulence, it settles down very quickly into a stagnant uniformity as soon as it is left to itself.

Now we know that while a gas is passing from an unstable state—a state of non-uniform temperature, for instance—to its stable and permanent condition, a property which we call its *entropy* and denote

[3] I am thinking particularly of the fact that in most expositions of the classical statistics one is adjured that there must be many particles in each compartment, and then in taking up the Fermi statistics one is told that there must be not more than one in each compartment; yet the two lead to formulæ which in the limiting case are the same.

by S is increasing; in certain simple cases we can evaluate this rate of change of entropy. We know that when a gas is in its stable condition, its entropy is at a maximum; we know how to compute the entropy (except perhaps for an additive constant) of a given quantity of a gas in this condition, as a function of its temperature and others of its measurable properties. And when we have evaluated both the entropy S and the energy E of a gas under any specific conditions, we know that its absolute temperature is determined by the following equation,

$$dS/dE = 1/T, \tag{2}$$

which is the definition of absolute temperature.

If we had obtained by some independent way an adequate atomic picture of entropy, so that whenever a distribution-function was suggested we could compute the value of S: then necessarily the stable distribution would be the one for which S has the greatest value compatible with the given number of particles and the given amount of energy. We do not have an independent way. But if instead we adopt some *tentative* atomic picture of entropy, some function S of which we can compute the value for any given distribution: then the test of our picture will be, whether the distribution for which this tentative S has its greatest value is verified by experiment to be the stable one. It will be found that this distribution "of maximum S" involves the derivative dS/dE, and therefore the absolute temperature; so the temperature enters into the postulated distribution-function in the course of its derivation, not by separate assumption or by an afterthought.

This method is the very notable one invented by Boltzmann, and continued by Planck. One choice of the function S which is to be identified with entropy leads to the classical or Maxwell-Boltzmann distribution-law; another leads either to the Bose or to the Fermi distribution, the difference between these two entering in at another point.

Each of these suggested functions is logarithmic; it is proportional to the logarithm of a function which is called *probability*. In theoretical physics it is a fairly general rule, that when a theorist introduces the word *probability* he is abandoning all hope of explaining by cause-and-effect the phenomena of which he is discoursing. This is the disadvantage of Boltzmann's method. The "distribution of maximum S" is baptized "the most probable distribution"; there is even a numerical estimate of its "probability," and in general it turns out to have so much greater a probability than all the others put together

that one accepts without demur the conclusions that in practice it will be stable. But there is no proof that the "most probable" distribution is always or even usually followed by another exactly like it, nor that an "improbable" distribution is always or even usually followed by another of greater probability; there is no study of the way in which one distribution is transformed into another, there are no assumptions about the collisions or encounters which presumably offer to the particles their means of interchanging speed and energy, and to the assemblage its means of approaching the stable distribution. There are other statistical methods in which account is taken of these things, and we shall have a glimpse of one of them in the last section of this paper; but the notion of causality is absent from the method which will be followed in deriving the distribution-laws of Maxwell and Boltzmann, of Bose, and of Fermi and Dirac.

These three distribution-laws will be applied to freely-flying particles in regions which are either field-free, or else pervaded by a field (electrostatic or gravitational) derivable from a potential. It may surprise the reader to hear so little about oscillators, considering that the statistics which Planck applied to these objects was the first of all the modifications of the classical statistics, was the source of the entire quantum-theory, and therefore the most important advance of the physics of the last quarter-century. The history of this period is very curious; but I cannot mention more than a couple of the salient points.

The Planckian oscillators served two purposes: they enabled Planck to derive the law of distribution of radiant energy at uniform temperature in a cavity, by supposing the radiation to be entirely wavelike and to be in equilibrium with myriads of oscillators in the walls of the cavity; and they enabled various savants to develop, step by step, a progressively improving theory of the specific heat of solids. The Bose statistics made them quite superfluous for the first purpose: by applying this statistics to the radiation supposed to consist of corpuscles, we can derive the same law of distribution without invoking the oscillators at all. As for the second: as early as 1912 (which seems remarkable, now) Debye had replaced the concept of a solid as a latticework of vibrating atoms by the concept of a solid as a system of stationary waves agitating a continuum. I do not mean to imply, of course, that the existence of the atoms was denied; I mean no more than to say, that in these statistical reasonings the individual vibrating atom was replaced by an individual pattern of stationary waves. Today we are becoming familiar with the idea

that in certain reasonings, a freely-flying electron or quantum or even an atom in a region bounded by walls may be replaced by a pattern of stationary waves filling the whole of this region. Thus it seems that the free particle, the oscillator, the stationary wave-pattern, are in close affinity with one another; they may simply represent different ways of looking at the same thing. Though on almost every page of this article I shall write in the language of the strictest corpuscular theory, it is probable that every one of the results could be translated into the language of oscillators or the language of waves.

There are still assumptions to be made about the individual particles. They are to have position, momentum and energy. Momentum may be separated into mass and velocity; often it is better left as an elementary concept. It will turn out that one of the essential differences between photons on the one hand, electrons and atoms on the other—that is to say, between the particles out of which we shall try to build a picture of radiation, and those of which we shall build models of gases and of electricity in metals—lies in the relation between momentum and energy.

Experience with matter in bulk leads to the well-known equations connecting kinetic energy K and momentum p with mass m and speed v:

$$K = \tfrac{1}{2}mv^2, \qquad p = mv \tag{3}$$

and these are supposed to hold for the ultimate particles of matter and of electricity.

During the years in which the corpuscular theory of light was struggling into existence—for, it will be remembered, light was still considered to be entirely wavelike even after Planck had founded the quantum-theory by his statistics of oscillators—Einstein proposed at two different times (1905 and 1917) the following formulæ for the energy and the momentum of photons in terms of their wavelength:

$$E = hc/\lambda, \qquad p = h/\lambda. \tag{4}$$

Historically it is interesting that he proposed the latter formula because of certain statistical studies which he had made of the equilibrium between photons and atoms. The verification of the latter by the Compton effect, of the former by the photoelectric effect and many other phenomena, is too familiar to require comment.

Now from equations (3) we deduce, for particles of matter and of electricity:

$$E = \frac{1}{2m}p^2 = \frac{1}{2m}(p_x^2 + p_y^2 + p_z^2) \tag{5}$$

and from equations (4) we deduce, for particles of light:

$$E = pc. \tag{6}$$

The difference between these two relations is responsible for some of the contrasts between radiation-gas on the one hand, electron-gas and material gases on the other; but by no means for the major part. The major difference lies in the statistical theory, as we shall now find out.

The Classical Statistics

We are going to represent three kinds of objects—ordinary or material gases, radiation in enclosures, negative electricity in metals—as assemblages of particles possessing location and momentum. We may visualize such an assemblage first as a swarm of points in ordinary space, with a coordinate-frame along the axes of which the coordinates x, y, z of the particles are measured; then as a swarm of points in momentum-space with a frame along the axes of which the momenta p_x, p_y, p_z are measured.

I will first illustrate the method of classical statistics by using it to ascertain the most likely distribution of particles in ordinary space, a case where seemingly the result may be foreseen. For it seems a truth of intuition that inside a box of ordinary space, with nothing (*e.g.* no variations of potential) to distinguish one region from another, the particles must tend to distribute themselves uniformly. This is a conclusion to which the statistical method *must* lead. The uniform distribution *must* be the most probable. How then should we define the "probability" of a distribution so that it shall be greatest for the uniform one?

But in the first place, what *is* a uniform distribution? We must divide the space—mentally, of course—into conpartments of equal volume. The distribution will then be called uniform, if the numbers of particles in the various compartments are about the same. But this clearly requires that these subdivisions be of a certain size. Their linear dimensions cannot for example be smaller than the average distance between particles, as then a "uniform distribution" would be impossible. To partition the space too finely would be like studying a painting with a microscope. The quality which we wish to define evades too sharp a scrutiny. The compartments should contain large numbers of particles, both for the stated reason and for the convenience of a certain mathematical approximation which is made.

Denote then by N the total number of particles, by m the number

of compartments into which the volume V is divided, by N_i the number of particles in the ith compartment. A distribution is described by stating all the numbers N_1, $N_2 \cdots$, N_i, $\cdots N_m$.

The basis of the classical statistics is the fact that if the particles have identities—if each of them is labelled by a distinctive letter, for instance—there are different ways of arranging them in the same distribution. One starts with any arrangement compatible with the prescribed "populations" N_1, N_2, $\cdots N_m$, and obtains all the other arrangements by interchanging particles *ad libitum* among the compartments, respecting only the condition that each of these shall always have as many as it had at first. The total number of distinct arrangements, the number of *permutations of the combination* N_1, N_2, $\cdots N_m$, is by a well-known theorem:[4]

$$W = \frac{N!}{N_1!N_2!\cdots N_m!}. \tag{7}$$

This number has its minimum value of unity for a distribution in which all the particles are crowded into one compartment, which would be the most non-uniform conceivable; and its maximum value for the uniform distribution, as I now proceed to show.[5]

Let us use the logarithm of W instead of W itself. If W has a maximum for any distribution so also will its logarithm, which is easier to handle, and will presently be chosen as the representation of entropy. We have:

$$\log W = \log N! - \sum \log N_i!. \tag{8}$$

Now we introduce Stirling's approximation for the factorial of a large number—by far the greatest and the most frequently invoked

[4] Imagine yourself stationed beside a set of m baskets and an urn filled with N lettered but otherwise indistinguishable balls, which are to be lifted out at random and dropped into the baskets under the following rules of the game: the first N_1 which come to your hand are to be dropped into basket 1, the next N_2 to come to your hand are to go into basket 2, and so on to the end. Having acted accordingly, you note down the assortments of balls in the various baskets, and repeat the process *ad infinitum*. Now there are $N!$ different orders in which the balls may come out of the urn. When the inspection of the baskets after two drawings reveals different results, the orders must certainly have been different. But two different orders need not reveal two different results to the inspection. Take any order, to start with; then there are $(Q - 1) = (N_1!N_2! \cdots N_m! - 1)$ others which yield the same result. For there are $N_1!$ orders in which the earliest N_1 balls emerge might come out, without any of them losing its place among the first N_1; there are $N_2!$ orders in which the next N_2 might come, without any losing its place in the second basket; and so forth. Each of the $N!$ orders then is but one among Q altogether which lead to the same result; so that there are only $N!/Q$ different results.

[5] What will actually be shown is that for the uniform distribution the function W is stationary; that it is maximum (not minimum) seems fairly obvious from the physics of the case, and can be proved.

of the mathematical aids in statistical theory. It is:

$$x! = (2\pi x)^{1/2}(x/e)^x, \qquad (9)$$
$$\log x! = x \log x - x + \tfrac{1}{2} \log (2\pi x).$$

The first two terms of this latter expression form an approximation singularly good even when x is no greater than ten or thereabouts. Using it we have:

$$\log W = \text{const.} - \sum N_i \log N_i. \qquad (10)$$

Denote by W^0 the value of W for some particular distribution N_1^0, N_2^0, $\cdots N_m^0$ and by $W = W^0 + \delta W$ its value for some other only slightly different distribution $N_1^0 + \delta N_1$, $N_2^0 + \delta N_2$, $\cdots N_m^0 + \delta N_m$. The difference between the values of $\log W$ for these two distributions is to first order of approximation:

$$\delta \log W = \delta W/W = -\sum_i (1 + \log N_i^0)\delta N_i. \qquad (11)$$

If W^0 is a maximum for the distribution $N_1^0 \cdots N_m^0$, then the difference between $\log W^0$ and the value of $\log W$ for any other slightly different or "slightly varied" distribution must vanish to first approximation. The quantity on the right of (11), the "first variation" of $\log W$, must be zero for any permitted set of values of $\delta N_1, \cdots \delta N_m$; meaning by "permitted" any set of integer values adding up to zero, for we consider an assemblage of an invariable number of particles.

Now one sees immediately that the right-hand side of (11) does vanish, if all the populations N_i^0 have the same value, say α; for then

$$\log W = -\sum_i (1 + \log \alpha)\delta N_i = \text{const.} \sum_i \delta N_i \qquad (12)$$

and the permitted variations are precisely those, for which the summation $\sum N_i$ is zero.

We do therefore reach the result which was desired. Failing it, this mode of "counting the ways in which a distribution may be realized" would have been unprofitable. As it is, the quantities W and $\log W$ are greatest for the uniform distribution which seems intuitively the most probable and is the rule for gases, and least for the utterly non-uniform one which seems the least probable. Tentatively the former is adopted as measure of the "probability" of a distribution.

I point out in passing that while the foregoing result is mathematically valid for any value of the constant α, the total number of particles prescribed for the assemblage determines the value of α which is physically permissible: viz. N/m.

We proceed to apply this method to the swarm of points in momentum-space representing the assemblage.

Like the coordinate-space, the momentum-space is to be divided into equal compartments large enough to contain each a multitude of particles. We are to define a distribution by specifying how many particles are in each compartment, and calculate as before the number W which is to measure the "probability" of the distribution. The values for W, for $\log W$ and for the variation of $\log W$ are obtained just as before. There is however an important novelty. Since the energy of a particle depends on its position in momentum-space, different distributions usually entail different values for the total energy of the assemblage. If we compute the variation of $\log W$ due to a slight change in distribution, we shall usually be computing a variation in $\log W$ correlated with a certain variation of the total energy U of the assemblage.

We now take the very great step of *identifying the quantity log W with entropy.*

More precisely, we assume that the entropy S is proportional to the logarithm of W:

$$S = k \log W, \tag{13}$$

introducing a constant factor k, and relying on subsequent experiments to teach us its numerical value.

Now when a gas being initially in thermal equilibrium at temperature T receives an infinitesimal amount of energy dE, and regains thermal equilibrium with its augmented energy, its entropy ascends by the amount of dS given by the equation:

$$dS/dE = 1/T. \tag{14}$$

If then the foregoing model of the gas and the foregoing picture of entropy are justified, the variation of $\log W$ in passing from the most probable distribution consonant with a total energy E to the most probable distribution consonant with a total energy $E + dE$ (the total number of particles remaining the same) must be equal to $(1/kT)dE$.

If we start from the most probable distribution for energy E and make *any* slight change in it involving an energy-change dE, the new distribution will presumably differ but little from the most probable distribution for $E + dE$. We therefore say: the most probable distribution for energy E is the one of which the first variation is dE/kT. This expression vanishes, if we are comparing distributions for which E is the same; which is as it should be.

It is now easily shown that such a distribution is the following

$$N_i = \alpha \exp(-\epsilon_i/kT), \tag{15}$$

in which α stands for any constant and ϵ_i for the average energy of particles in the ith compartment, which is related to the average momenta of these particles by the equation

$$\epsilon = \frac{1}{2m}(p_x^2 + p_y^2 + p_z^2)' \tag{16}$$

for we have only to write down the expression for δS as furnished by equation (11), and introduce into it the value of $\log N_i$ as supplied by equation (15):

$$\begin{aligned}\delta S &= -k\sum(1 + \log N_i)\delta N_i \\ &= -k(1 + \log \alpha)\sum\delta N_i + \sum\epsilon_i\delta N_i/T = \delta E/T,\end{aligned} \tag{17}$$

the result which was desired.

The value to be chosen for the constant α will be determined as before by the total number of particles. Denote this number by N, and conceive the compartments as tiny cubes of volume H, so that there are $1/H$ of them per unit volume of the momentum-space. The density ρ of the particles in momentum-space, which is no other than the *distribution-function in the momenta*, is given anywhere by the value of N_i/H computed for the value of energy there prevailing:

$$\begin{aligned}\rho = N_i/H &= \frac{\alpha}{H}\exp(-\epsilon/kT) \\ &= \frac{\alpha}{H}e^{-p_x^2/2mkT}e^{-p_y^2/2mkT}e^{-p_z^2/2mkT}\end{aligned} \tag{18}$$

and it is the integral of this expression over the whole of momentum-space which is equal to N:

$$N = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\rho dp_x dp_y dp_z. \tag{19}$$

The integration is easily affected; the triple integral is the product of three identical single integrals, and we have:

$$N = \frac{\alpha}{H}\left[(2mkT)^{1/2}\int_0^{\infty}e^{-w^2}dw\right]^3, \tag{20}$$

w being a symbol for each of the three momenta in turn; so that

$$\alpha = \frac{NH}{(2\pi mkT)^{3/2}}. \tag{21}$$

The expression for the number of particles in any compartment thus becomes:

$$N_i = \frac{NH}{(2\pi mkT)^{3/2}} \exp\,(-\epsilon_i/kT), \tag{22}$$

involving the four constants m, N, k and H. The first three are determinable by experiment, the third is the universal constant known in Boltzmann's honor by his name, though he himself never evaluated it. The fourth, the volume H assigned to the compartments, drops out of the distribution-functions—out of the function ρ, out of the distribution-in-energy soon to be deduced, out of the fundamental distribution-function f in the coordinates and momenta defined by equation (1), and which I now set down in place of ρ:

$$f = \frac{N}{V(2\pi mkT)^{3/2}} \exp\left(-\frac{p_x^2 + p_y^2 + p_z^2}{2mkT}\right), \tag{23}$$

V standing for the volume in ordinary space of the enclosure which contains the assemblage. This evasion of H is very deceptive; for it suggests not merely that the exact volume of the compartments is of no importance, but that the compartments themselves were invented only as a momentary stepping-stone to the distribution-functions, and should be allowed to shrink to zero like the infinitesimals of the calculus. This however is precisely what is not allowed. It is of the essence of the argument that there are compartments of finite size. As will presently transpire, I suspect that the division of momentum-space into compartments should be regarded as a quantum postulate, even in this case of the derivation of the Maxwell-Boltzmann law which seems to be at the opposite extreme from all the notions of quantum-theory.

The next step is the derivation of the *distribution-in-energy*. In preface I point out that the distribution which we are considering is, in respect to the directions of motion of the particles in ordinary space, *isotropic*. Mathematically, this occurs because p_x, p_y and p_z enter symmetrically into all the distribution functions; physically it occurs because we have made no assumption leading to a preference of any direction over any other. Later on we may establish a preferred direction by introducing a field of force, and then the impending steps may have to be reconsidered. Until then the distribution which we shall study will be described completely by saying that they are isotropic and giving the distribution-function-in-energy. This may be obtained from the distribution-function-in-the-momenta

by transforming to a polar coordinate frame in the momentum-space.[6]

I follow practically the same route.

Divide up the momentum-space into spherical "shells" by means of a sequence of spheres all centered at the origin. Each sphere corresponds to a value of ϵ, each shell to a range $d\epsilon$ of values of ϵ. Take one of the latter at random; call it shell s, denote by ϵ_s and by ϵ_{s+1} or $\epsilon_s + d\epsilon$ the energy-values at its boundary spheres, by r_s and $r_s + dr$ the radii of these, by dV the volume of the shell. Then:

$$r_s = (2m\epsilon_s)^{1/2}, \qquad dr = \left(\frac{m}{2\epsilon_s}\right)^{1/2} d\epsilon,$$

$$dV = 4\pi r_s^2 dr = \frac{2}{(\pi)^{1/2}} (2\pi m)^{3/2} (\epsilon_s)^{1/2} d\epsilon. \tag{24}$$

Suppose to begin with that each shell is large enough to contain very many compartments. The number Q_s of compartments in shell s will then be:

$$Q_s = dV/H = \frac{2}{H(\pi)^{1/2}} (2\pi m)^{3/2} \epsilon_s^{1/2} d\epsilon \tag{25}$$

and the average number of particles per compartment in shell s, call it N_s, will be:

$$N_s = \alpha \exp(-\epsilon_s/kT) \tag{26}$$

and the total number M_s of particles in the shell will be:

$$M_s = Q_s N_s = \frac{2N\pi}{(\pi kT)^{3/2}} \epsilon_s^{1/2} e^{-\epsilon_s/kT} d\epsilon = F(\epsilon_s) d\epsilon. \tag{27}$$

This is the number of particles having energy-values in $d\epsilon$ at ϵ_s. Hence the distribution-function-in-energy F is the factor multiplying $d\epsilon$ (it would be well to discard the subscript s in writing it). I have copied the value of α from (21), but it could have been derived by integrating F from $\epsilon = 0$ to $\epsilon = \infty$ and equating the integral to N.

The separation of M_s or $F(\epsilon)d\epsilon$ into two factors—Q_s the number of compartments in the shell s, N_s the average number of particles per compartment—is highly advantageous in searching for the distinctions

[6] Denote by p the quantity $(p_x^2 + p_y^2 + p_z^2)^{1/2}$ which is the magnitude of the momentum; and by θ and ϕ the angles which with p constitute a spherical coordinate system. We have

$$\rho dp_x dp_y dp_z = \rho p^2 \sin\theta d\theta d\phi dp = \frac{\alpha}{H} e^{-p^2/2mkT} p^2 \sin\theta d\theta d\phi dp$$

and the distribution-in-momentum is obtained by integrating over all values of θ and ϕ, the distribution-in-energy from it by means of the relation (5).

between the various proposed statistical laws. We shall see that in passing from one to another sometimes one of the factors is changed, sometimes the other, sometimes both.

In particular, we may pass from the Maxwell-Boltzmann law to a distribution like that which Planck derived for oscillators, simply by changing the factor Q_s. We have been dividing the momentum-space into compartments of equal volume, so that the number comprised in a shell s between spheres ϵ_s and $\epsilon_s + d\epsilon_s$ is proportional to $\epsilon_s^{1/2} d\epsilon_s$. Let us instead divide it into compartments of which the volumes increase steadily from the origin outward, at such a rate that the number in a shell s is proportional to $d\epsilon_s$ without the factor $\epsilon_s^{1/2}$.

This is, of course, not the way in which Planck's postulate is habitually stated, though it is substantially the way in which Planck stated it himself. Usually it is said, that Planck restricted the energy of the particles of the assemblage to a set of "permitted values" spaced at equal intervals: say the values a, $a + b$, $a + 2b$, $a + 3b$, $\cdots$ where a and b stand for constants. Each of these permitted values corresponds to a sphere in the momentum-space. In the shell s there are approximately $d\epsilon_s/b$ of these "permitted spheres"; the approximation being closer, the larger ϵ_s and $d\epsilon_s$ are in comparison to b. Now whether we conceive that the $d\epsilon_s/b$ sets of particles in the shell s are located on the surfaces of as many permitted spheres, or alternatively that they are scattered through as many compartments, is for the statistical results of no importance. There may be other reasons for preferring one picture to the other; but the predictions of the statistical theory are the same, whichever is adopted. I will therefore alternate between the two pictures, retaining for the moment that of a subdivision of the momentum-space into compartments; but now it will be expedient to think of these as thin spherical films, centered at the origin and increasing in volume from the innermost outward at the specified rate.

If the shell s is large enough to contain many of these compartments or permitted spheres, we may use the first approximation for the number which it contains:

$$Q_s = d\epsilon/b, \tag{28}$$

and putting the expression (26) for the number of particles per compartment, we get:

$$M_s = Q_s N_s = \frac{\alpha}{b} \exp\,(-\epsilon_s/kT) d\epsilon_s = F(\epsilon_s) d\epsilon_s \tag{29}$$

for the number of particles having energy-values between ϵ_s and

$\epsilon_s + d\epsilon_s$. The value of the constant is fixed as heretofore by the condition that the integral of F over the entire range of energy from 0 to ∞ shall be equal to N:

$$\int_0^\infty F(\epsilon)d\epsilon = \frac{\alpha kT}{b}\int_0^\infty e^{-w}dw = N, \tag{30}$$

so that we arrive at the following distribution-in-energy function:

$$F(\epsilon) = \frac{N}{kT}\exp\,(-\epsilon/kT). \tag{31}$$

This function certainly does not display any feature which suggests the achievements of Planck! It looks as smooth and continuous as the Maxwell-Boltzmann function itself, and the constant b, the step or interval between the successive permitted energy-values or the boundaries of successive compartments, is nowhere to be seen. The constant b however has slipped out for the same reason as the constant H from the function (23), and the apparent continuity is due in both cases to the same cause. In preparing and effecting the integration (30) in order to obtain a value for the constant α, we assumed that the various permitted energy-values within the range $d\epsilon_s$ are all sufficiently nearly equal to be identified with the single value ϵ_s. That is to say, we smoothed over the discontinuities which had previously been brought in by the assumption of separate compartments. No wonder that there is not a sign of them in the function (31), even as there is not a sign of them in the Maxwell-Boltzmann law!

We might however avoid this smoothing-over, if we could attain the value of α by an actual summation over the various compartments instead of by integration. Now with Planck's postulate this is mathematically feasible and indeed easy. For the number of particles in the ith compartment being

$$N_i = \alpha\exp\,(-\epsilon_i/kT) \tag{32}$$

the total number of particles is computed thus:

$$\begin{aligned} N = \sum N_i &= \alpha e^{-a/kT}\sum_{i=0}^{\infty} e^{-ib/kT} \\ &= \frac{\alpha e^{-a/kT}}{1 - e^{-b/kT}} \end{aligned} \tag{33}$$

by virtue of the very convenient consequence of the binomial theorem that $(1 + x + x^2 + \cdots) = (1 - x)^{-1}$; so that for α we obtain the

exact value:

$$\alpha = Ne^{(a-b)/kT}(e^{b/kT} - 1) \tag{34}$$

and for the populations of the various compartments, the formula:

$$N_i = Ne^{-b/kT}e^{-ib/kT}(e^{b/kT} - 1). \tag{35}$$

Here the discontinuity implied in the classical picture of a momentum-space divided into compartments is admitted and accepted, as it never was in the process of deriving the Maxwell-Boltzmann law. Planck did not put discontinuity into the classical statistics; it was there already; he refrained from disregarding it. Instead of confining his studies to the circumstances in which it can safely be ignored, he extended them to ranges where it had to be taken account of, and he took account of it.

As I intimated, the distribution (35) was proposed by Planck not for freely-moving particles, but for oscillators. The "Planckian oscillator" may be visualized as a particle which executes simple-harmonic vibrations back and forth in a straight line across a position of equilibrium, to which it is attracted by a force proportional to its displacement. It is like a free particle, in that its state at any moment is described by giving the values of its position q and momentum p, q being measured from its point of equilibrium; but it is unlike a free particle in that its energy depends not on p alone but on both p and q, being a function of the form $(Ap^2 + Bq^2)$. Therefore we must envisage not the momentum-space alone but the phase-space of the variable p and q. In principle it would have been better, had we envisaged the phase-space all along; but since for an assemblage of free particles that space has six dimensions, it was impractical to visualize more than the momentum-space, and since the energy depended only on the momenta that compromise was not detrimental except for one feature which I can later introduce. Here the compromise would be ruinous, but it is unnecessary since the phase-space has only two dimensions.

Visualize then this two-dimensional phase-space as a plane with p and q axes at right angles to each other. Suppose all the oscillators to have the same mass and the same natural frequency, which is to say, the same values of the constants A and B in the above-mentioned formula for their energy; but let them differ in amplitude. The point representing any oscillator in the phase-space runs round and round in an elliptical orbit centered at the origin. Different amplitudes correspond to different ellipses. The energy of an oscillator depends on its amplitude; therefore different energy-values correspond

to different ellipses, and reversely. If we divide the phase-space into compartments by a succession of ellipses centered at the origin, each of these compartments corresponds to a specific range of energy-values. If the dividing ellipses are so spaced that these compartments are of *equal* area (equal volume of the phase-space), they correspond to *equal* ranges of energy-values—an important difference between this case and the one which was previously treated.

If the dividing ellipses are spaced to form equal compartments, they themselves correspond to energy-values forming a linear sequence: call these a, $a + b$, $a + 2b$, $\cdots a + ib \cdots$ as before. Whether we call these the "permitted" energy-values and allow the oscillators only the choice among them, or whether we sprinkle the oscillators uniformly through the compartments, makes only a secondary difference. In this case, in fact, we can easily see exactly what difference it makes. If the oscillators are sprinkled uniformly in each compartment, then by applying the classical statistics we get just the same distribution (35) as when we assume them restricted to the energy-values $(a + ib)$. But when we undertake to evaluate the average energy of all the oscillators, then in the one case we must put down the mean energy of those in the ith compartment as the arithmetic mean of the values $a + ib$ and $a + (i + 1)b$, while in the other case we must put down the energy of those at the ith permitted ellipse as $a + ib$. Hence to change over from the picture of permitted energy-values to the picture of compartments is the same thing as to replace the original sequence of permitted energy-values by another sequence of values located midway between them. I mention this chiefly in order to emphasize that the subdivision of phase-space into compartments is *ipso facto* quantum-theory.

As every reader knows, Planck postulated that the quantity b—the interval between the permitted energy-values, or the energy-range within a compartment, whichever picture is chosen—is the product of a universal constant (h) and the frequency of the oscillators (ν). The area of the equal compartments is then equal to the universal constant [7] whatever the frequency of the oscillators. From this latter statement the general principle is derived: To state it one must first adopt a symbol (say n) and a name (say *number of degrees of*

[7] The point in the phase-space representing an oscillator of mass m, frequency ν, and amplitude C describes an ellipse having semi-axes C and $2\pi m\nu C$ and area $2\pi^2 m\nu C^2$; its energy is $U = 2\pi^2 m\nu^2 C^2$; hence the relation between energy U and area F is

$$U = \nu F$$

and the area between two ellipses is equal to h if the energy-difference between them is equal to $h\nu$.

freedom) for the number of distinct coordinates q required to describe the individual member of whatever assemblage one may be considering; this is also the number of distinct *momenta* p, there being one p for each q. Then the principle generalized out of Planck's postulate for oscillators is this: *For an assemblage of individuals with n degrees of freedom the phase-space is to be divided into compartments of volume* h^n.

We will now see what the classical statistics, supplemented by this principle, proposes for an assemblage of particles for which the relation between energy and momentum is $\epsilon = cp$ as it is for corpuscles of light, instead of $\epsilon = p^2/2m$ as it is for corpuscles of matter.

Different energy-values correspond as before to different spheres all centred at the origin of the momentum-space, but the numerical relations are changed. Instead of equations (24), we have:

$$\begin{aligned} \epsilon_s &= cr_s, \qquad d\epsilon_s = cdr_s, \\ dV &= 4\pi r_s^2 dr_s = (4\pi/c^3)\epsilon_s^2 d\epsilon_s, \end{aligned} \tag{36}$$

dV standing for the volume of the shell s covering the energy-range between ϵ_s and $\epsilon_s + d\epsilon_s$. Divide the momentum-space into compartments of equal volume H. We derive the "smoothed-over" distribution-function for the case in which N_i varies so little from one compartment to the next that even when the shell s is thick enough to comprise very many compartments the values of N_i for all of them may be equated to a mean value N_s. Under these conditions we may write for the number of compartments in the shell s,

$$Q_s = dV/H = (4\pi/c^3H)\epsilon_s^2 d\epsilon_s. \tag{37}$$

Putting down the classical value (15) for the number of particles in any of these compartments, remembering that N_i is identified with N_s, we obtain for M_s the number of particles in the shell s:

$$M_s = Q_s N_s = \alpha(4\pi/c^3H)\epsilon_s^2 e^{-\epsilon_s/kT} d\epsilon_s \equiv F(\epsilon_s)d\epsilon_s \tag{38}$$

and evaluate α by the same procedure as before. The result is:

$$F(\epsilon) = \frac{N}{3k^3T^3}\frac{\epsilon^2}{e^{\epsilon/kT}}. \tag{39}$$

This is the smoothed-over distribution-in-energy predicted for the radiation-gas by the classical statistics, it being assumed that the momentum-space is to be divided into compartments of equal volume. Experiment however supplies a quite different distribution-in-energy, to wit:

$$F(\epsilon) = \left(\frac{8\pi V}{c^3h^3}\right)\frac{\epsilon^2}{e^{\epsilon/kT} - 1}. \tag{40}$$

It looks as if (39) might be the limiting form of (40)—as if the actual distribution-law might be obtained by avoiding the approximations whereby we came to the formula (39), as Planck's law of distribution for oscillators was obtained by refraining from approximation. Such however is not the case. True, the second factor in (39) is evidently the limiting form, for very high temperatures, of the second factor in (40). But the first factor in (40) contains nothing but the volume of the gas and some universal constants, while the first factor in (39) contains the temperature and an apparently disposable constant standing for the number of particles in the assemblage. The former is not the limit of the latter. It will be noted also that although I said that the volume of the elements of phase-space was to be set equal to h^3, this assumption in no wise enters into the function (39). Bose in fact found it necessary to upset the basis of the classical statistics, in order to arrive at (40) instead of (39).

The Bose Statistics

The momentum-space of the photons is to be divided as heretofore into equal compartments, and various distributions of the particles among these are to be compared, in order that we may elect one of them as "the most probable" and make a picture of the entropy of the assemblage. But the manner of defining a distribution, the manner of "counting the ways" in which it may be realized and computing its "probability," is to be changed, and changed in a most thoroughgoing and fundamental way.

Start with any distribution of the particles, defined as heretofore: defined that is, by saying that there are N_0 of the particles in the compartment 0, N_1 in the compartment 1, and in general N_i in the compartment i.

Count the number of compartments containing no particle; call it Z_0. Count the number of compartments containing one particle apiece; call it Z_1. In general, let Z_i stand for the number of compartments containing i particles apiece. Put down the values of all the numbers Z_i.

Now change the terminology. Elect some neutral word, "arrangement" say, to denote what we have heretofore denoted as a "distribution," and use the latter word in the following new sense: a *distribution* shall henceforth be described by stating the values of the numbers $Z_0, Z_1, Z_2, \cdots, Z_i, \cdots$ and the total energy of the assemblage.[8]

[8] It is of course confusing thus to change the meanings of words, but in the long run less confusing (I think) than to use some other word than *distribution* for the concept always called by that name in the new statistics.

This means that each distribution in the new sense comprises a number of distinct distributions in the old sense. This we shall regard as the number of different ways in which the distribution in the new sense may be realized. Going over entirely to the new terminology: we shall now identify the probability W^* of a distribution with the number of arrangements which are included in it. Previously we identified the probability W of an arrangement with the number of permutations included in it, according to equation (7). This we now must forget; we must proceed as if the probability of each arrangement were the same.

The number W^* is now to be evaluated. In doing this we must remember that we have to count, not the total number of arrangements yielding the prescribed set of values of the quantities Z_i, but the portion of these which give the prescribed value to the total energy of the assemblage.

As before, we superpose upon the partitioning of the momentum-space into small compartments of equal volume H, another partitioning into spherical shells each of which is sufficiently large to contain many of the compartments, yet sufficiently small so that the same value of ϵ may be assigned to all the compartments within it. The final result is thus to be a "smoothed-over" formula. It is rather singular that whereas Planck introduced the quantum into physics by avoiding the smoothing-over which had been customary in the classical statistics the quantum-formula for radiation is now derived by a method in which it is accepted.

Consider then any shell at random, say the "shell s." Denote by Q_s or by Z the total number of compartments in it; by Z_{is} the number of these compartments which contain i particles apiece; by M_s the number of particles in the shell. According to the scheme now being tried out, the number of ways of attaining the particular distribution characterized by the numbers Z_{is} is given by the formula:

$$W_s^* = \frac{Q_s!}{Z_{0s}!\,Z_{1s}!\,Z_{2s}!\cdots}, \qquad Q_s = \sum_i Z_{is}. \tag{41}$$

As the energy-values for all the compartments in the shell are (by hypothesis) approximately the same, these various ways of attaining the distribution Z_{is} all corresponding to approximately the same total energy, as well as the same total number of compartments and the same total number of particles.

Suppose this process repeated for every one of the shells s. The total number of ways of attaining the actual distribution, compatible

with the conditions of constancy of total energy, total number of particles and total number of compartments, is then the product of all the quantities W_s^*: call it W^*:

$$W^* = \prod_s W_s^*. \tag{42}$$

As before, and with the same end in view, we form the expression for log W^*, and employ Stirling's formula (assuming thus in effect that none of the quantities Z_{is} is smaller than ten or so):

$$\log W^* = \sum_s \log W_s^* = \sum_s [Q_s \log Q_s - \sum_i Z_{is} \log Z_{is}]. \tag{43}$$

We now take the very great step of *identifying* not log W of equation (10), but log W^*, *multiplied by a constant* k, *with the entropy of the assemblage.*

$$S = k \log W^*. \tag{44}$$

Then, when the numbers Z_{is} are changed by small amounts δZ_{is}, the ensuing change δE in the total energy E of the assemblage must be linked to the ensuing change in (k log W^*) by the equation:

$$\delta S = \delta(k \log W^*) = \delta E/T. \tag{45}$$

The first variation of (k log W^*) is given thus:

$$\delta(k \log W^*) = -k \sum_s \sum_i (1 + \log Z_{is}) \delta Z_{is}. \tag{45a}$$

Let us try the distribution:

$$Z_{is} = \alpha_s e^{-i\epsilon_s/kT}. \tag{46}$$

Substituting this expression into (45a), we get:

$$\begin{aligned} \delta(k \log W^*) &= -k \sum_s \sum_i (1 + \log \alpha_s - i\epsilon_s/kT) \delta Z_{is} \\ &= -k \sum_s (1 + \log \alpha_s) \sum_i \delta Z_{is} + \frac{1}{T} \sum_s \epsilon_s \sum_i i \delta Z_{is} \\ &= \delta E/T, \end{aligned} \tag{47}$$

for the summation $\sum_i \delta Z_{is}$ vanishes because the number of compartments in each shell is invariable, while the quantity $\epsilon_s \sum i Z_{is}$ is equal to the total energy of the particles in the shell s.

The new statistics, in proposing a new conception of entropy as embodied in equation (44), therefore leads to a new distribution for thermal equilibrium. This distribution is expressed by equation (46) in the new-fashioned way, by stating the number of compartments

in each shell which contain each of the permissible quotas of particles. We must translate it into a distribution-function-in-energy such as we used to express the results of the old statistics.

Before undertaking this translation, we compute the values of the constants α_s by summing the numbers Z_{is} over all values of i for each shell separately, and equating the sum to the total number Q_s of compartments in the shell. We obtain:

$$Q_s = \sum_i Z_{is} = \alpha_s \sum_i e^{-i\epsilon_s/kT} = \alpha_s(1 - e^{-\epsilon_s/kT})^{-1}. \tag{48}$$

On substituting these values of α_s into (46) we get something which begins to look familiar.

Next for the number M_s of the particles in the shell s, we compute:

$$M_s = \sum_i iZ_{is} = \alpha_s e^{-\epsilon_s/kT}(1 - e^{-\epsilon_s/kT})^{-2}$$

$$= Q_s \frac{1}{e^{\epsilon_s/kT} - 1}, \tag{49}$$

which begins to look very familiar indeed.

Now for Q_s, the number of compartments in the shell s, we put the value already stated in equation (37), derived from the assumption that the compartments are all of the same volume H:

$$M_s = \frac{4\pi}{c^3 H} \frac{\epsilon_s^2}{e^{\epsilon_s/kT} - 1} d\epsilon_s \equiv F(\epsilon_s)d\epsilon_s. \tag{50}$$

Here the function $F(\epsilon_s)$ is the "smoothed-over" distribution-in-energy in which the new statistics culminates. Unlike those which we earlier derived from the old statistics, it involves the volume of the elementary compartments directly. Whereas from the old statistics we obtained formulæ involving the quantum only by avoiding the approximation, here we obtain a quantum-formula even when we admit the approximation—a contrast on which, I think, it is worth while to insist.

Let us then, in preparing for the final assumption, accept the principle generalized from Planck's assumption about oscillators: let the elementary cell of phase-space be given the volume h^3. This is not yet an assumption about the compartment of momentum-space. Supplement it, then, by supposing that the compartment of phase-space h^3 is the product of the compartment H of momentum-space and the entire volume V occupied by the radiation-gas. Then:

$$H = h^3/V. \tag{51}$$

This assumption—let me remark in passing—takes a very elegant form if we replace the compartments by the permitted energy-values, and then the corpuscular picture by the wave picture; for then we have a series of permitted wave-lengths, which are precisely those which can form stationary waves in a cube of volume V.

So the new statistics leads to the distribution-law:

$$F(\epsilon)d\epsilon = \frac{4\pi V}{c^3h^3}\frac{\epsilon^2}{e^{\epsilon/kT}-1}d\epsilon. \tag{52}$$

Dividing out the factor V, we get the number of particles per unit volume having energy-values between ϵ and $\epsilon + d\epsilon$, in an assemblage having the most probable distribution at the temperature T. Multiplying this by ϵ, we get the total energy per unit volume in the possession of such particles. Identifying these particles with photons, we observe that they have wave-lengths between ch/ϵ and $ch/(\epsilon + d\epsilon)$, frequencies between ϵ/h and $(\epsilon + d\epsilon)/h$. Transforming then from the variable ϵ to the new variables λ and ν, we obtain distribution-functions which give the *density of radiant energy* as functions of wave-length and frequency. It turns out that these agree absolutely with the observed distributions, except that they lack a factor 2. This factor is at once imported, and is ascribed to the fact that light is polarizable. So we arrive at the black body radiation-formula:

$$\rho(\nu)d\nu = \frac{8\pi h}{c^3}\frac{\nu^3 d\nu}{e^{h\nu/kT}-1} \tag{53}$$

and the new statistics is justified by its success.

It will be observed that the new statistics leads to a precise value for the number of photons per unit volume, at any prescribed temperature; whereas the old statistics led to nothing of the sort, but to a formula which contained the number of atoms per unit volume as a disposable constant. This corresponds to a profound physical difference between radiation-gas and material gases. When I state the temperature and the volume of a box containing helium, I am not giving data enough to fix the quantity of helium inside the box; on the contrary, the quantity and the density of the helium in the box can be varied *ad libitum* while the temperature and the volume are held constant. But when I state the temperature and the volume of an enclosure containing radiation, I am giving data sufficient to fix the amount of radiant energy and the number of quanta in the enclosure absolutely. This is a fact of experience, and the new statistics is evidently in accord with it. But if one were tempted to

try out the new statistics upon a material gas, would there be any way of avoiding the inadmissible conclusion that the number of atoms in such a gas is also absolutely fixed by temperature and volume?

There is such a way. One might replace the distribution proposed in equation (46) by a more general one involving a disposable constant B, as follows:

$$Z_{is} = \alpha_s e^{-iB - i\epsilon_s/kT}. \tag{54}$$

On substituting this into the expression for $(k \log W)$ we get instead of (47) the equation:

$$\delta(k \log W^*) = -k\sum_s (1 + \log \alpha_s)\sum_i \delta Z_{is} + kB\sum_s\sum_i i\delta Z_{is} + \frac{1}{T}\sum_s \epsilon_s \sum_i i\delta Z_{is}. \tag{55}$$

The right-hand member must as before reduce to $\delta E/T$ if the distribution (54) is acceptable; and this it will do, provided that not only $\sum_i \delta Z_{is}$ but also $\sum_s \sum_i i\delta Z_{is}$ is zero. Now the second of these quantities is zero for all variations in which the total number of particles remains the same. The distribution (54) enjoys a greater entropy than any other which is compatible with the same total energy and the same total number of particles. The distribution (46) was still more exalted; it enjoyed a greater entropy than any other compatible with the same total energy, even including those for which the total number of particles was somewhat different. But the distribution (54) is sufficiently distinguished to be qualified as the most probable distribution for a material gas. It seems rather singular that the distribution (46) is required for radiation-gas. Here is evidently one of the deep differences between matter and radiation.

Following the same routine as before, we arrive at the following expression for the number of particles in the shell s:

$$M_s = Q_s \frac{1}{e^{B + \epsilon_s/kT} - 1}, \tag{56}$$

and in dealing with radiation-gas we have put $B = 0$ and have taken the value of Q_s from equation (37). If in dealing with a material gas we take the value of Q_s from equation (25) instead and put $H = h^3/V$ we obtain:

$$M_s = \frac{2V}{h^3(\pi)^{1/2}}(2\pi m)^{3/2} \frac{(\epsilon_s)^{1/2} d\epsilon_s}{e^{B + \epsilon_s/kT} - 1} \equiv F(\epsilon_s) d\epsilon_s, \tag{57}$$

and now it is obvious that we must evaluate B in terms of the total

number of particles N by the already so familiar way of integrating $F(\epsilon_s)$ over the entire energy-range from 0 to ∞ and setting the integral equal to N.

Einstein proposed this as an alternative to the Maxwell-Boltzmann law derived from the classical statistics. It is not easy to decide which of the two is supported by experiment, as with increasing temperature the formula (57) becomes more and more nearly like the classical one, and it turns out that throughout the convenient ranges of temperature and pressure the two are indistinguishable. It would be very valuable to determine between the two, as then we should know which of the two ways of defining a distribution and estimating the probability thereof, which of the two pictures of entropy, is the proper one for a material gas. The reader may have remarked that if one were to apply Bose's method to the problem of determining the most probable distribution of particles in ordinary space, one would reach a result at variance with that of the classical statistics, and therefore at variance with intuition. One must deal altogether with the six dimensional phase-space, to be perfectly consistent. This is to be regretted.

The Fermi Statistics

The statistics invented by Fermi, and later independently by Dirac, involves the same fundamental assumptions as that of Bose—the same manner of counting the ways in which a distribution may be realized, of defining its probability, of picturing its entropy. But there is an additional assumption, of the nature of a limitation: it is postulated, that a compartment may contain not more than some specific maximum number of particles. In particular for a gas to which no external field is applied, it is postulated that each compartment must either be empty, or else contain one particle only.

The "exclusion-principle of Pauli" gave the hint from which the Fermi-Dirac theory sprang. This principle may be paraphrased as follows. In Bohr's "atomic theory of the atom" the electrons belonging to an atom are forbidden to revolve in any except certain specific orbits, set apart from the rest as the "permitted" orbits, and labelled by specific "quantum-numbers." In later versions of the theory the "permitted orbits" are less conspicuous, the "permitted quantum-numbers" more so; but the picture is acceptable at all events as a beginning. Upon this prohibition, then, Pauli superposed another; not more than one electron is allowed in each orbit or to each set of quantum-numbers. Perhaps it would be better to say "not more than some definite small number of electrons . . ."

instead of "not more than one." The affinity of this to one of Fermi's assumptions will soon be manifest. It would take much too long to give an idea of the successes of the Pauli principle; they are however so great as to increase the inherent plausibility of Fermi's idea very much—or perhaps I should say, so great as to render the idea plausible, which otherwise it might not seem.

The reasoning follows exactly the same course as when we were deriving the distribution-law (56), except that all the summations over the variable i are now summations of two terms only, the term for $i = 0$ and the term for $i = 1$. For each of the shells there are only two numbers Z_{is} required to describe the distribution: viz. Z_{0s} the number of empty compartments and Z_{1s} the number of compartments containing one particle apiece. We try the distribution (54):

$$Z_{0s} = \alpha_s; \qquad Z_{1s} = \alpha_s e^{B-\epsilon_s/kT}, \tag{58}$$

and easily find that it is the distribution of maximum probability, by comparison with all the others compatible with the same total number of particles and the same total energy. We arrive then at the following expression for the number of particles in the shell s:

$$M_s = Q_s \frac{1}{e^{B+\epsilon_s/kT} + 1}. \tag{59}$$

There is no point in putting for Q_s the value appropriate to radiation-gas, since the Bose formula has already proved adequate for that case. Fermi put the value appropriate to material gases, and obtained:

$$M_s = \frac{2V}{h^3(\pi)^{1/2}} (2\pi m)^{3/2} \frac{(\epsilon_s)^{1/2} d\epsilon_s}{e^{B+\epsilon_s/kT} + 1} = F(\epsilon_s) d\epsilon_s. \tag{60}$$

This formula is the point of departure for the theory of the electron-gas in metals revived and remodelled by Pauli and Sommerfeld, to the experimental tests of which most of the rest of this article will be devoted.

Application of the Fermi Statistics to the Electrons in Metals

We are asked to conceive of a piece of metal as a region populated with "free" electrons, and surrounded by a wall; the electrons being distributed according to the formula of Fermi.

The Fermi distribution-function involves the total number N of the electrons, which is a disposable constant. It also involves the volume V which this assemblage of N particles pervades. For this

we set the volume of the piece of metal—a decision which is tantamount to ignoring the atoms, to supposing the metal a vacuum inhabited by free electrons only. So remarkable an assumption, even though it be made only in approximation, requires some excuse. Its strangeness may be mitigated by recalling, first, that slow electrons may go through atoms (at least through certain kinds of atoms) as imperturbably as if the atoms were not there; and second, that a wave-train may go without being scattered at all through a crowd of particles individually quite able to scatter it, provided that the particles are arranged in a regular lattice having a spacing smaller than the wave-length of the waves. The speeds attributed to electrons in metals are so low and their wave-lengths are so great, that perhaps they do behave in such a way.

The "wall" is the agency which prevents the electrons from escaping; it is commonly imagined as a sharp and sudden gradation of potential at the surface of the metal. Any electron moving towards it from within, with a velocity of which the component normal to the bounding surface may be denoted by u, is supposed to be driven back into the body of the metal if the corresponding "component of kinetic energy" $\frac{1}{2}mu^2$ is less than a certain constant W_a; while if $\frac{1}{2}mu^2 > W_a$ the electron escapes, but with its kinetic energy diminished by W_a. According to newer ideas electrons may sometimes escape even when their values of $\frac{1}{2}mu^2$ are smaller than W_a, and may sometimes fail to escape in the contrary case; but the earlier and simpler conception remains approximately valid, and I will abide by it for a time. The constant W_a may be named the *work-function*.

Like the constant N, the work-function figures as a disposable constant in the theory. It is an ambition of physicists to explain as many as possible of the differences between different metals, by varying only the values of these two constants. Later we shall find it necessary to introduce others, beginning with the one which in the older theories appeared as the mean free path of the electrons; but there are several results of value which can be obtained with no other but these two.

I repeat now from (60) the Fermi formula for the distribution-in-energy of an assemblage of N particles in volume V at temperature T, with two changes made to bring the notation into harmony with that of Sommerfeld:

$$F(\epsilon) = G\frac{2}{(\pi)^{1/2}}\frac{V}{h^3}(2\pi m)^{3/2}(\epsilon)^{1/2}\frac{1}{A^{-1}e^{\epsilon/kT}+1}. \tag{61}$$

Here the symbol $1/A$ replaces e^B, and a factor G to which we shall

assign the value 2 is introduced for a reason which will be stated later. The corresponding distribution-function in the coordinates and momenta is this:

$$f(x, y, z, p_x, p_y, p_z) = \frac{G}{h^3} \frac{1}{A^{-1}e^{\epsilon/kT} + 1}. \tag{62}$$

The first step now is the same as in the classical statistics: to determine the constant A in terms of N by integrating $F(\epsilon)$ over the whole range of energy-values from 0 to ∞, and equating the integral to N:

$$\int_0^\infty F(\epsilon) d\epsilon = N. \tag{63}$$

This was an easy step in the classical statistics, but here it is very hard. The integral of $F(\epsilon)$ is not one of the common well-known functions to be found in mathematical tables, nor a combination of such; and we do not get a simple equation to be solved for A in terms of N. Sommerfeld indeed found it necessary to compromise by deducing two series-expansions for the integral, one being available for values of A smaller than unity, the other for the opposite extreme. By a stroke of luck which seems almost too good to be true, the first one or two terms of one or the other of these expansions form an approximation amply good enough for all the cases where as yet theory and experiment can be compared.

I consider first the approximation which is of *no* importance in the theory of electrons in metals—the one for values of A so very small that the second term in the denominator of $F(\epsilon)$ is negligible by comparison with the first. Then the distribution-law approaches that of Maxwell and Boltzmann, and of necessity the constant A must possess the value which in the limit makes $F(\epsilon)$ identical with the classical expression written in equation (27): to wit, the value:

$$A = \frac{Nh^3}{GV}(2\pi mkT)^{-3/2}. \tag{64}$$

It would however be a great error to suppose that this value of A can be substituted into the function F under all circumstances. This value is acceptable only if A is very small relatively to unity, which is to say, if the quantity to which A is here equated is very small. So the question arises: in any physical case, is the combination on the right-hand side of equation (64) a small fraction of unity, or is it not?

Now for any material gas under any conditions usual in the labora-

tory, A is indeed very small. The new statistics leads to a result indistinguishable from that of the old statistics. To discriminate between the two by experiments on material gases, one would have to work with temperatures so low and densities so high that the gases would probably either be liquefied already, or at least would be in a condition very different from that "ideal" state to which the statistics is tacitly supposed to apply. Perhaps though it is not impossible to make the test with helium or hydrogen.

A this point apparently Fermi stopped. But it occurred to Pauli that if the new statistics were applied to an electron-gas as dense as that which Riecke and Drude had supposed to pervade the interiors of metals, the deviations from the classical distribution would be much more pronounced. For, in the first place, the mass m of the individual electron is smaller by several orders of magnitude than the mass of the atoms or molecules of any material gas. And, in the second place, if the number N of free electrons in a piece of metal is as great as or greater than the number of atoms, then it is thousands of times as great as the number of particles in an equal volume of a material gas. Now the quantity equated to A in equation (64) contains N in the numerator and $m^{3/2}$ in the denominator, and for the hypothetical electron-gas within the metals it is no longer small. The expression (64) for A is then no longer acceptable.

While the statement just made about m is based on a fact of experience, the statement about N is not so firmly grounded. We have no direct knowledge of the number of free electrons in a given volume, say the number n $(= N/V)$ in unit volume, of a metal. This as I said above is a disposable constant of the theory. One of the tests of the theory is whether one can obtain correct numerical values of half-a-dozen properties of a metal by choosing a single value of n for that metal. So long as the classical statistics was applied to the electron-gas, this was impossible. If the value of n was put as high as the number of atoms in unit volume, the predicted value of specific heat (and we may now add, the predicted value of susceptibility) turned out to be too large; if n was lowered sufficiently to avoid this particular discordance, other predictions were impaired. It was however the general impression, that one should put n equal to the number of atoms or a small multiple thereof. I suspect that this decision was largely due to a feeling that since the free electrons are detached from atoms, and since all the atoms are alike, any atom should supply as many free electrons as any other. However that may be, it was natural though not inevitable for Pauli and for Sommerfeld to link the Fermi statistics with the postulate that there are as

many free electrons as there are atoms, and test the combination of these two assumptions.

On putting for m the mass of the electron, for N/V the number of atoms per unit volume of any metal, for T any temperature from zero absolute up to several thousand degrees, and for G any small integer, one finds that the quantity equated to A in (64) is very large. Thus with $N/V = 5.9 \cdot 10^{22}$ (the number of atoms in a cc. of silver), $T = 300°$ K., $G = 2$, Sommerfeld computed:

$$(nh^3/G)(2\pi mkT)^{-3/2} = \text{about } 2400 \tag{65}$$

a result which invalidates equation (64).

We turn then to the other series-expansion of the integral $\int F(\epsilon)d\epsilon$, the one which Sommerfeld proved applicable for large values of A. The first two terms of this expansion are as follows:

$$\int_0^\infty F(\epsilon)d\epsilon = N = \frac{GV}{h^3}\frac{4\pi}{3}(2mkT \log A)^{3/2}\left(1+\frac{\pi^2}{8}(\log A)^{-2}+\cdots\right). \tag{66}$$

Taking the first term only of this expansion and putting the aforesaid values of N/V and T and solving for $\log A$, one finds a very large value indeed ($\log_e A = 325$). Assuredly then we may use the first two terms of this expansion by themselves when we are dealing with the electron-gas in a metal, and indeed the first term will for some purposes be amply sufficient.

We have thus the following first—and second—approximation formulæ for A in terms of n or N/V:

$$\begin{aligned} 2mkT \log A &= h^2(3n/4\pi G)^{2/3} \qquad \text{first approx.} \\ 2mkT \log A &= h^2(3n/4\pi G)^{2/3}\left[1-\frac{(2\pi mkT)^2}{12h^4}\left(\frac{3n}{4\pi G}\right)^{-4/3}\right] \\ &\qquad\qquad \text{second approx.} \end{aligned} \tag{67}$$

(the second approximation being computed by putting the first-approximation value of $\log A$ into the second term of the series expansion).

On substituting one or the other of these into the distribution-functions (61) and (62), we have the postulated distribution of the free electrons expressed to as high a degree of approximation as we require, with no disposable constant except n; and we are ready for the applications.

The Specific Heat

As it was the notorious difficulty with the specific heat which spoiled the old electron-gas theory in which the classical statistics

was coupled with the assumption that there are as many free electrons as atoms, let us first of all find out whether this difficulty remains.

Any distribution-law for an assemblage leads immediately to a formula for the total energy E thereof as function of the temperature, which is:

$$E = \int_0^\infty \epsilon F(\epsilon) d\epsilon. \tag{68}$$

Putting the distribution-in-energy (27) derived from the classical statistics, we find:

$$E = \frac{3}{2} NkT \tag{69}$$

and putting the one just derived from the Fermi statistics with the first-approximation value of A, we find:

$$\begin{aligned} E &= E_0 + \tfrac{1}{2}\gamma V T^2, \\ E_0 &= \frac{2\pi}{5} \frac{VGh^2}{2m} \left(\frac{3n}{4\pi G} \right)^{2/3}, \\ \gamma &= \frac{\pi}{3} mG \frac{(2\pi k)^2}{h^2} \left(\frac{3n}{4\pi G} \right)^{1/3}, \end{aligned} \tag{70}$$

two exceedingly different formulæ.

The derivatives dE/dT of these expressions are the formulæ supplied by the two statistics for the specific heat of the electron-gas. The classical theory predicts for the specific heat a constant value, while the Fermi statistics makes it proportional to the temperature—being thus in harmony with Nernst's heat theorem, while the other is not—and gives it even at room temperatures but a small fraction of the classical value.

Experimentally the specific heat of the electron-gas cannot be measured separately from that of the lattice of atoms, which constitutes the metal—an admission which seems to condemn as vain all hope of testing these formulæ. Nevertheless one can conclude with fair certainty that the classical expression is inadmissible; for the specific heat of an ordinary metal agrees so well with the value attributed by statistical theories both old and new *to the atoms alone*, that there is simply none left over for the electrons—no such great excess, that is to say, as the amount $3nk/2$ which the classical theory requires. The device of reducing n to so low a value that $3nk/2$ would be inappreciable makes trouble in other directions, as I have intimated. But with the new statistics the theoretical value γVT is inappreciable even when n is made as great as the number of the atoms and T as

great as many hundreds of degrees. Probably no one who did not often lament the defeat of the old and so very desirable electron-gas theory by that hard fact about the specific heat will ever quite realize the rejoicing caused by this victory of the new, which by this achievement succeeded *a se faire pardonner* many deficiencies in other fields.

Features of the Fermi Distribution

I will now mention some of the features of the Fermi distribution which has thus justified itself by passing its first test.

The most startling of these may be inferred from the distribution-function (62) or (61), by inserting the first-approximation formula for A presented in equation (67), and a new symbol W_i:

$$f = \frac{G}{h^3} \frac{1}{e^{(\epsilon - W_i)/kT} + 1}; \qquad W_i = \frac{h^2}{2m} \left(\frac{3n}{4\pi G} \right)^{2/3}. \tag{71}$$

At the absolute zero the exponential term is either infinity or zero, according as the variable ϵ is greater or less than W_i. Therefore the density of the electrons in phase-space is constant and equal to G/h^3 for all energy-values less than W_i, zero for all values of energy greater than W_i.

This striking result can easily be deduced from Fermi's basic assumption, without any statistics at all. Absolute zero is by definition the temperature of the state, being in which the assemblage can give away no energy whatever. If not more than one electron may occupy any compartment of the phase-space, absolute zero is attained when there is an electron in every compartment from the origin outwards to a sphere which is centered at the origin, and which has just the volume needful to contain as many compartments as there are electrons. The number of electrons in unit volume of the phase-space is and remains equal to the number of compartments in unit volume, *i.e.* to the reciprocal of the volume of the elementary compartment, from the origin outward to this sphere; there it suddenly sinks to zero, and so continues. Cooling-down of an assemblage is settling-down of the particles into this the most condensed of all permissible arrangements; it is like crystallization upon a lattice, only the lattice is in the phase-space.

The foregoing statements may all be repeated, with the words *phase-space* replaced by *momentum-space*. In the momentum-space, a sphere of radius p, consequently of volume $4\pi p^3/3$, contains $(4\pi p^3/3)/(h^3/V)$ of the elementary compartments. If we set this number equal to the total number of electrons N, and solve the resulting equation for p, we get the radius of the sphere which would just contain all the

electrons if there were one in each compartment. If we set the volume of the sphere equal to Nh^3/VG and solve for p, we get the radius p_m of the sphere which would just contain all the electrons if there were G of them in each compartment. But this is the maximum value of the momentum of the electrons, it is the momentum of the fastest of the electrons. The corresponding speed v_m of the fastest of the electrons is p_m/m, therefore is given by the expression:

$$v_m = \frac{h}{m}\left(\frac{3n}{4\pi}\right)^{1/3} \tag{71a}$$

and the corresponding kinetic energy $\frac{1}{2}mv_m^2$ is the same as W_i.

It is expedient to set down for future reference the *mean values* of speed v and of several integer powers of v, for a gas distributed according to the Fermi law at the absolute zero. The general formula for the mean value of any power v^s of v is this:

$$\overline{v^s} = \frac{1}{n}\int v^s f(v)dv = \frac{1}{n}(4\pi G/m^3h^3)\int_0^{v_m} v^{s+2}dv$$

and in particular:

$$\overline{v^{-1}} = 3/2v_m; \qquad \bar{v} = 3v_m/4; \qquad \overline{v^2} = 3v_m^2/5; \qquad \overline{v^3} = \tfrac{1}{2}v_m^3. \tag{72a}$$

The corresponding values for the Maxwell distribution are these:

$$\begin{aligned} \overline{v^{-1}} &= 2(m/2\pi kT)^{1/2}; \qquad v^2 = (2kT/m); \\ \bar{v} &= 2(2kT/\pi m)^{1/2}; \qquad \overline{v^3} = 4\pi(2kT/\pi m)^{3/2}. \end{aligned} \tag{72b}$$

Plotted as functions of ϵ, the distribution-function f in the coordinates and momenta starts out as an horizontal straight line at a distance G/h^3 from the axis of abscissæ, while the distribution-function F in the energy starts out as a concave-upward parabolic arc; these continue as far as the abscissa $\epsilon = W_i$, and from then on the curves coincide with the axis of abscissæ.

The foregoing statements are valid for absolute zero; what happens as the temperature rises? Sommerfeld has proved that the sharp angles in the distribution-curve are *very gradually and slowly* rounded off, the curve always traversing the midpoint of the vertical arc BC (Fig. 1). The far end of the curve sinks down to the axis of abscissæ in the fashion of the Maxwell law. Even at room-temperature and even far above, however, the distribution departs so little from the absolute-zero form that many phenomena may be interpreted in a qualitative way, simply by imagining the absolute-zero distribution—the completely degenerate distribution, it is called—to persist all

through the observable range of temperatures. Indeed, in calculating electrical resistance and certain other properties of metals, one may use the mean values of the various powers of v which are tabulated in (72*a*). There are however other properties of metals, thermal conductivity for instance, for the estimation of which it is not sufficient to assume that the mean values of the powers of v are always the same as at absolute zero, and one must derive more nearly approximate values for them; for these however I refer the reader to Sommerfeld.

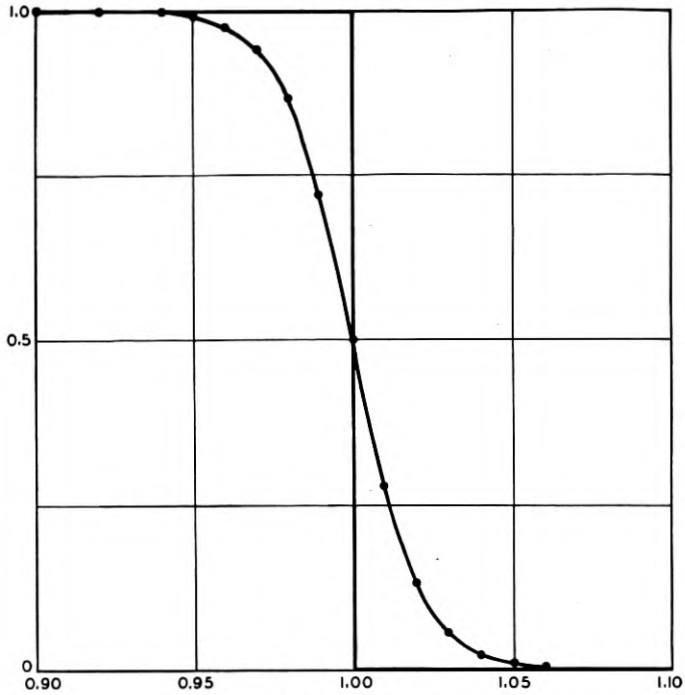


Fig. 1—Graphs of the Fermi distribution-function f plotted against ϵ/W_i as independent variable, for an electron-gas having $6.5 \cdot 10^{22}$ particles per cc., at temperatures zero (rectilinear curve) and 1500° K. (rounded curve). The value of W_i is 6 equivalent volts.

In practice, the values of W_i are rather astonishingly great; no less, for example, than 5.6 equivalent volts for silver, 5.7 for tungsten, 6.0 for platinum. Obviously they depend on the compactness of the lattice, being greater the more closely-packed the atoms are. In potassium and sodium the atoms are relatively widely spaced, and the corresponding values of W_i are about 2.1 and 3.2 in equivalent volts.

The contrast between this and the classical situation is evidently enormous. Where formerly we were asked to think of the electrons in a metal at usual temperatures as being distributed Maxwell-wise about a very modest mean energy, say about 0.02 of an equivalent volt, we are now invited to conceive them as distributed all through a range of energies extending from zero up to as much as half a dozen equivalent volts, and more abundantly the nearer one approaches to the top of this range, abruptly though the distribution ceases when the very top is reached. This is "zero-point energy" with a vengeance!

The pressure of the electron-gas is related to the energy-per-unit-volume by the equation valid also in the classical theory:

$$p = \frac{2}{3}\frac{E}{V},$$

and therefore varies like the total energy—starting from a value absurdly high at the absolute zero (hundreds of thousands of atmospheres) and increasing therefrom very slowly at first, though according to a T^2 law, as the temperature rises. I do not know of any manometer for measuring internal electron-pressures, but if anybody should invent one he had better make it strong.

There is manifest ground for doubting these remarkable proposals: thermionic data seem to show that the work-function which opposes the egress of the electrons from a metal is itself less than half-a-dozen volts (in the usual measure), for some metals less than two—what then keeps these fast electrons confined within the metal? It turns out, however, that in augmenting the *vis viva* of the electrons the new theory also raises the top of the wall which they must overleap. Here indeed we meet with the first of the new experiments which tend to confirm the new theory.

Thermionic Emission

The simplest theory of the thermionic current is, that it consists of all the electrons belonging to the interior electron-gas which fly against the boundary-surface of the metal with velocities such that the component u thereof perpendicular to the boundary-surface is great enough to make the "energy-component" $\frac{1}{2}mu^2$ greater than a constant W_a—the said constant being interpreted as the work-function or the retarding potential-drop at the edge of the metal. The thermionic electrons are those which swim up to the surface with an outward-bound velocity-component so large, that by means of the kinetic energy of their outward motion they can climb over the wall.

Evidently any thermionic emission must distort the distribution of the electron-gas inside the metal, as it is an unbalanced outflow of electrons. The situation in which the efflux is balanced by a corresponding influx from an electron-gas outside the metal is much regarded in thermodynamic theory, but one cannot measure currents in that situation any more than one can measure heat-flow between two bodies at equal temperature. In assuming, then, that the distribution of the internal electrons is that of Fermi or that of Maxwell, we shall probably be invalidating our conclusions except for the limiting case of an infinitesimal emission. It seems probable, however, that with the thermionic currents of practice the approximation is good enough.

The simplest theory of the thermionic current, then, consists entirely of the equation:

$$i = e\left(\frac{m}{h}\right)^3 G \int_{w=u_0}^{\infty} du \int_{-\infty}^{\infty} dv \int_{-\infty}^{\infty} dw \cdot u \, \frac{1}{A^{-1}e^{\epsilon/kT} + 1}. \qquad (73)$$

This is a restatement of the first sentence of this section, *plus* the assertion that even when electrons are leaking out through the wall of the metal the distribution within remains practically that of Fermi. The factor e stands for the electron-charge; the symbol i thus for the thermionic current-density in electrostatic units. The factor m^3 enters because, in conformity with usage, I have translated from the momenta into the velocity-components u, v, w as independent variables. The quantities u_0 and W_a are related by the equation:

$$W_a = \tfrac{1}{2}mu_0^2. \qquad (74)$$

In the integrand we are of course to put $\frac{1}{2}m(u^2 + v^2 + w^2)$ for ϵ.

Setting for A the first-approximation value from (107) with the symbol W_i defined in (71), we obtain:

$$i = e\left(\frac{m}{h}\right)^3 \int_{u_0}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u \, \frac{1}{e^{(\epsilon - W_i)/kT} + 1} \, dudvdw. \qquad (75)$$

The integration is perfectly straightforward if the second term in the denominator may be neglected relatively to the first. This seems unnatural, for we have just been noticing that over part of the energy-range, from $\epsilon = W_i$ downwards, the second term is larger than the first. But if W_a is considerably larger than W_i—and by this I mean, if $(W_a - W_i)/kT$ is positive and considerably larger than unity—then the electrons which escape are those which belong to the extreme

upper part of the energy-range, where the first term is much the larger. Writing then $u \cdot \exp\left[-(\epsilon - W_i)/kT\right]$ for the integrand, we integrate with ease by well-known formulæ, and get:

$$i = \frac{2\pi emG}{h^3}(kT)^2 e^{-(W_a - W_i)/kT}. \tag{76}$$

The experimental test consists in plotting $(\log i - 2 \log T)$ against $1/T$; we should get a straight line provided that W_a does not vary with temperature. The experiments do lead to precisely this result. The slope of the line varies from metal to metal, and depends on the state of the metal surface; identifying it with $(W_a - W_i)/k$, one finds that W_a exceeds W_i by amounts ranging from one equivalent volt upward to five or six; so the approximation just mentioned is abundantly justified, even up to temperatures of incandescence.

The contrast with the predictions of the classical theory is peculiarly interesting. Assuming the Maxwell distribution for the interior electrons, one arrives easily (the reader can do it by substituting the value of A from (64) into (73)) at the formula:

$$i = \frac{en}{(2\pi m)^{1/2}}(kT)^{1/2} e^{-W_a/kT}. \tag{77}$$

On testing this formula by plotting $(\log i - \frac{1}{2} \log T)$ against $1/T$, it is found that the experiments yield lines as beautifully straight as those obtained by plotting $(\log i - 2 \log T)$. Indeed—as everyone knows who has dabbled in thermionics—a function of the type $\exp(-c/T)$ varies so exceedingly rapidly with $1/T$ that it makes no perceptible difference to the graph whether or not the function is multiplied by a constant or by any modest power of T. One cannot then use the graphs to distinguish between the theories, even if one could be sure that W_a is not a function of T. But the classical theory proposes that we identify the slope of the aforesaid line with W_a/k; and it has been the custom so to do.

Now *if the Fermi distribution-function is the right one, physicists have been underestimating the work-function all along.* They have plotted experimental curves which agreed in shape with (76) and (77) and from these they have evaluated the constant figuring in the exponent, a constant which they have denoted usually by $-b/k$; and then they have equated b to W_a, whereas if it is right to apply the new statistics they should have added W_i to b and then equated the sum to W_a. Nor is the alteration slight: for if there are as many free electrons in the metal as there are atoms, then W_i is six volts

or thereabouts, and the quantity to be added to the observed constant b is larger than b itself.

Is there then any other way of determining the work-function than out of this apparently ambiguous current-vs-temperature curve? If there is a direct and independent way, it may serve not only to decide between the two statistics, but also—if it favors the new by yielding a value for W_a greater than the thermionic b—to give an *experimental* value for $(W_a - b) = W_i$ and hence for the disposable constant n which is the one uncertain quantity in the theoretical formula for W_i. It may, that is to say, serve to determine the number of electrons per unit volume of the electron-gas.

Now it seems that the diffraction of electrons by crystals provides an independent and direct way of determining the work-function. For in the phenomena in which negative electricity behaves as a wave-motion, the work-function figures in the index of refraction; and the index of refraction of a metal may be determined from the diffraction-patterns which it forms when irradiated with slow electrons. Ample data concerning one metal—nickel—have already been acquired by Davisson and Germer, and from these it transpires that the work-function is much in excess of the thermionic constant b—so much indeed, that the corresponding value of W_i implies that there are twice as many free electrons as atoms in the metal, or even more.[9] The values deduced for the work-function from the refractive index vary however with the speed of the electrons; and it is evident that much remains to be understood.

The factor which multiples $T^2 \exp\left[-(W_a - W_i)/kT\right]$ in the right-hand member of (76) involves universal constants only (supposing that G is such) and is therefore the same for all metals—a principle derived by Richardson from the first and second laws of thermodynamics twenty years ago, without any assumptions at all about the distribution of the electrons. Its actual value $2\pi k^2 meG/h^3$ differs only by the factor G from the value derived by Dushman, which is numerically equal—in the customary units—to 60.2 amperes per cm.^{2} per degree squared. There are several metals for which the experimental value of this quantity—commonly known as A, a symbol which in this article is monopolized by another meaning—agrees well with 60.2. One might infer that G must be unity, a choice which would demolish the theory of paramagnetism; but there is another recourse; one may suppose that half the electrons which

[9] L. Rosenfeld, E. E. Witmer (*l. c. infra*). From Rupp they cite values of refractive index for six other metals (Al, Cr, Cu, Ag, Au, Pb) and compute values of n. In considering these, however, the reader should assess G. P. Thomson's criticism of Rupp's values.

come up to the bounding surface from within with energy sufficient to escape are nevertheless reflected. A factor $(1 - r)$—r being called the coefficient of reflection, and being put equal to $\frac{1}{2}$—then enters into the formula, and balances out the factor 2 introduced by giving the preferred value to G. Moreover one may explain values of the constant still smaller than 60.2, or between 60.2 and 120.4, by adjusting r accordingly. But there are also recorded values enormously greater than 120; so evidently something remains to be understood.[10]

The methods of wave-mechanics have been applied by Fowler and Nordheim to the problem of evaluating this coefficient of reflection. They have attained some notable results in the fields of thermionics, cold discharge, and photoelectric effect. These however are consequences not of the new statistics only, but of a combination of the new statistics with the new way of considering the transmission and reflection of electron-waves at surfaces. There is not space for me to deal with the latter, beyond indicating its point of departure and its chief results.

Thus far I have been speaking of a metal as an equipotential region surrounded by a surface at which there is a sharp potential-drop, and beyond which there is the equipotential region of outer space. Fowler and Nordheim however, like Schottky and others before them, conceive a metal as an equipotential region surrounded by a surface, beyond which lies a region in which there is a field (or at all events an image-field) the strength of which is a function of the distance from the surface. The quantity W_a appears as the integral of this field-strength, from the surface to infinity. Electrons of a given kinetic energy being supposed to fall against the surface from within, the fraction which fails to pass completely through the region of the field depends upon the kinetic energy of the electrons and upon the *shape* (not solely upon the integral) of the field-strength-vs-distance curve. The average value of this fraction for all the electrons of all speeds coming up to the surface from within—the average being taken with due regard to the relative proportions of the electrons of various speeds, that is to the distribution-in-velocity—is the coefficient r aforesaid. For certain simple shapes of the field-vs-distance curve it may be calculated. One thus arrives at the be-

[10] If the empirical equation is of the form $i = aT^n \exp(-b/T)$, then whatever may be the values of the constants a and b, one can always claim that it agrees with the foregoing theory provided one assumes that W_i or r or both vary with temperature in just the proper way. But, as Fowler puts it: "the variety of possible uncontrollable hypotheses (if such assumptions are to be admitted) becomes too large for profitable discussion."

ginnings of a theory of the effects of surface-conditions on thermionic-emission, and of the cold discharge.

The Cold Discharge

Suppose that a potential-difference is now applied between the metal and a neighboring electrode, such that near the metal the resulting field-strength is very great. Does it penetrate the region which I have just been describing? Assume that it does penetrate as far as the "surface" just defined. Then, everywhere beyond the surface, the actual field is the resultant of this "applied" and the previously-mentioned "intrinsic" field. The shape of the field-strength-vs-distance curve is thus changed, in a way which is calculable if we have postulated some particular original shape; and in certain simple cases it is possible to calculate the consequent change of the coefficient r, and therefore the electron-current—or the additional electron-current—which the applied field causes to emerge from the metal. This is the current known as the "cold discharge."

Nordheim adopted for the field-strength-vs-distance curve, in the absence of applied P. D., the shape which is most commonly proposed —the inverse-square curve, the law of the "image-force" which a charge in the vicinity of a conductor experiences because of its "electrical image" in the conductor. For the current i of the cold discharge he then derived this approximate formula:

$$i = c'SF^2 \exp(-c''/F), \tag{78}$$

in which F stands for the applied field, S for the surface-area of the metal exposed to it, c' and c'' for constants which can be calculated when W_i is known.

There are certain experiments (for an account of which I refer to Nordheim's paper in the Physikalische Zeitschrift) which indicate that the actual relation between the current and the field-strength agrees in form with (78), but that the predicted values of c' and c'' are too large—too large by a factor of the order of ten in the latter case, by several orders of magnitude in the former. However it is possible to explain away these contradictions. One may assume, for instance, that the actual surface concerned in the discharge is a collection of small spots which altogether have but a small fraction of the total area of the metal surface; and that over these small spots, the field-strength is much higher than it would be if the metal were everywhere uniform and smooth. The ratio of the observed to the predicted value of c' then gives the fraction of the total surface which is covered by these "active" spots, and the ratio of the observed to

the predicted value of c'' gives the reciprocal of the factor by which the field-strength must be multiplied.

This explanation has the disadvantage of being not only plausible but much too easy. So long as there is not any independent evidence about the area of the effective spots or the field-strength prevailing over them, the theory simply delivers an equation with two disposable constants, which is not very valuable for testing the underlying assumptions. It appears from the data examined by Nordheim, however, that the ratio of the predicted to the observed value of c'' is always between 10 and 20, and the ratio of the observed to the predicted value of c' is always about 10^{-10} (at least for tungsten). This uniformity of the two quantities which figure in the theory as the disposable constants may be taken as a confirmation of some weight.

Photoelectric Effect

According to the former theory, the elementary process of the photoelectric effect runs thus: a quantum dives into a metal, and gives its whole energy (say E_0) to an electron initially at rest, which then may escape from the metal after suffering a reduction of kinetic energy equal at least to W_a and possibly more (more, that is to say, if the electron loses kinetic energy on its way to the surface). Even if we suppose that the electron originally belonged to an electron-gas conforming to the classical statistics, its initial energy would almost always be quite negligible compared to that which the quantum gives it. But if the electron-gas obeys the new statistics, it comprises electrons with energy-values ranging up to W_i. However if the new statistics is valid, then the reduction of kinetic energy at the boundary is also greater by W_i than we have hitherto supposed. The net result is, that by the new statistics as by the old we derive Einstein's equation for the maximum kinetic energy of the electrons expelled by quanta of frequency ν:

$$E_{\max} = E_0 + W_i - W_a = h\nu + \text{const.} \tag{79}$$

only the additive constant is now $(W_i - W_a)$ instead of $(-W_a)$. This additive constant should as before agree with the thermionic constant b.

The new theory has one marked distinction, probably an advantage, over the old: it implies a sharply definite maximum kinetic energy—that is to say, the distribution-in-energy function of the escaping electrons should jump suddenly from zero to some definitely higher value at the energy-value $E = E_{\max}$ prescribed by (79); the slope of

the curve representing this function should make an acute angle with the axis of E where they intersect at E_{max}. The Maxwellian distribution predicted by the classical statistics for the interior electrons suggests however that the curve in question should approach the axis asymptotically. The new statistics leads also to certain inferences about the shape of the curve for values of E less than E_{max}. A great quantity of data bearing on this subject has been obtained by Ives and his collaborators; but the interpretation is made difficult by the presumption that some allowance must be made for the energy-losses suffered by electrons after they absorb quanta but before they reach the surface, and will require much study.

Paramagnetism of the Electron-Gas

The susceptibility of the electron-gas was calculated by Pauli even before the specific heat was evaluated by Sommerfeld, but as it involves an extra complication I have inverted the historical order.

The complication is due of course to that assumption which is made in order to explain why the electron-gas should be magnetic at all—the assumption that electrons are magnets. Perhaps I am too cautious in referring to it as an assumption, it being so well authenticated by the gyromagnetic effect and by the general usefulness of the "spinning electron" in the explanation of spectra. These phenomena impose a special value on the magnetic moment μ_0 of the electron, to wit, the value,

$$\mu_0 = eh/8\pi m_0 c, \tag{80}$$

m_0 standing for the rest-mass of the electron. Further they require that when the electron is floating in a magnetic field, its moment (considered as a vector) shall be either parallel or anti-parallel to the field. Denote by θ the angle between the moment of the electron and the magnetic field: then θ must be either 0 or π.[11] Now when a magnet of moment M is inclined at an angle θ to a magnetic field H, its "extra magnetic energy" is $-MH\cos\theta$.[12] In dealing with the electron-gas, then, we are in effect assuming that when a field H is applied to it the energy of every electron is either increased or decreased by the amount:

$$\Delta = ehH/8\pi m_0 c. \tag{81}$$

[11] It comes to the same thing, and may on other grounds be preferable, to assign μ_0 twice the value given in (80), and to θ the values 60° and 120°.

[12] The magnetic energy, or energy due to the "interaction between the magnet and the field" is put equal to zero when the magnet is transverse to the field, which is consistent with the picture that the field alters the energy of the magnet by speeding up or slowing down the revolving electricity. The formula here given is a first approximation.

As I mentioned earlier, the idea of compartments in the momentum space may be replaced by the idea of "permitted energy-values," at least in some situations. Let us for convenience return to the latter conception. Then we may say that when a magnetic field is applied to an electron-gas of which the particles are magnets, each of the permitted energy-values is split into two. To any previously-permitted value ϵ correspond a pair of new ones, $\epsilon + \Delta$ and $\epsilon - \Delta$; or let me say $\epsilon + m\Delta$, using m as a symbol which may have only the values $+1$ and -1.

Pauli assumed that the most probable distribution of the electrons among this doubled set of energy-values is to be determined by the new statistics, including Fermi's postulate so modified as to state that not more than one electron may possess any one of the permitted values.

Previously I used the symbol Z_{is} to denote the number of compartments or permitted energy-values which lie in the shell s and are occupied by i electrons apiece; and the symbol Q_s to denote the total number of permitted energy-values in the shell s. Now there are Q_s permitted values which are shifted upward by Δ from the original ones, and Q_s more which are shifted downward by Δ from the original ones. Let Z_{is+1} stand for the number in this upward-shifted group which are occupied by i electrons apiece, and Z_{is-1} for the corresponding number in the downward-shifted groups; Z_{ism} shall be the general symbol for the two.

Now consider the distribution:

$$Z_{ism} = \alpha_{sm} e^{-iB - i(\epsilon_s + m\Delta)/kT}, \qquad \begin{matrix} m = +1, -1, \\ i = \quad 0, \quad 1. \end{matrix} \tag{82}$$

This answers the standard requirements for thermal equilibrium. For if we define W^* in Bose's way, and then say that the entropy is k log W^*, we find that when the energy of the assemblage is varied by δE—the total number of cells and the total number of particles remaining constant—the first variation of the entropy is $\delta E/T$, as it should be. I leave it to the reader to prove this statement as corresponding statements for other distributions (*e.g.* (47)) were earlier proved, and to determine the values of the quantities α_{sm}; on doing which, and substituting the results into this distribution, he should obtain

$$Z_{1s,+1} = \frac{Q_s}{e^{B+(\epsilon_s+\Delta)/kT} + 1}, \qquad Z_{1s,-1} = \frac{Q_s}{e^{B+(\epsilon_s-\Delta)/kT} + 1}, \qquad Q_s = \frac{2}{(\pi)^{1/2}} \frac{V}{h^3} (2\pi m)^{3/2} (\epsilon_s)^{1/2}. \tag{83}$$

It is best, perhaps, to regard these formulæ as the descriptions of two electron-gases, one composed entirely of magnets parallel to the field, the other of anti-parallel magnets; the actual electron-gas is a mixture of the two. The quantity $Z_{1s,\,-1}$, for example, is the number of parallel magnets having energy-values shifted downwards through Δ from the Q_s originally-permitted energy-values which lay in the shell s. True, these magnets are no longer themselves in the shell s, owing to the shift; their energies lie between $(\epsilon_s - \Delta)$ and $(\epsilon_s - \Delta + d\epsilon_s)$, not between ϵ_s and $\epsilon_s + d\epsilon_s$; but for ease of integration it is better to think of them as being associated with the original unshifted energy-values. The total number of magnets comprised in the parallel gas is then given by the equation:

$$N_1 = \frac{2}{(\pi)^{1/2}} \frac{V}{h^3} (2\pi m)^{3/2} \int_0^\infty \frac{(\epsilon)^{1/2} d\epsilon}{e^{B+(\epsilon-\Delta)kT} + 1} \cdot \tag{84}$$

For brevity denote by L the constant before the integral; and use the symbol ϕ for the function defined as follows:

$$\phi(u) = \int_{\epsilon=0}^\infty \frac{(\epsilon)^{1/2} d\epsilon}{e^{\mu+\epsilon/kT} + 1} \cdot \tag{85}$$

Then expanding N_1 as a power-series in the variable Δ/kT which is $\mu_0 H/kT$, we find:

$$N_1 = L \left[\phi(B) - \frac{\mu_0 H}{kT} \phi'(B) + \text{terms of higher order in } H \right] \cdot \tag{86}$$

Similarly one obtains, for the total number N_2 of electron-magnets in the "anti-parallel gas," a formula:

$$N_2 = L \left[\phi(B) + \frac{\mu_0 H}{kT} \phi'(B) + \text{terms of higher order in } H \right] \cdot \tag{87}$$

Now the total magnetic moment of the "parallel gas" is $N_1\mu_0$ in the same sense as the field, and the total magnetic moment of the "anti-parallel gas" is $N_2\mu_0$ in the sense opposed to the field; so that the net magnetic moment of the entire assemblage of electrons is $(N_1 - N_2)\mu_0$. We will carry out the computations only for values of H so low that we may ignore all terms beyond the second in the expansions for N_1 and N_2. The net magnetic moment is then approximately proportional to H; its quotient by H, the *susceptibility* χ of the electron-gas, is constant. It is a fact of experience that with nearly all paramagnetic substances the susceptibility *is* independent of H up to the highest attainable values of this variable. The limi-

tation which we are here accepting will probably therefore not prove serious. Our approximative formula for χ is then as follows:

$$\chi = (N_1 - N_2)\mu_0/H = -\frac{2L\mu_0^2}{kT}\phi'(B) = -\frac{N\mu_0^2}{kT}\frac{\phi'(B)}{\phi(B)}, \quad (88)$$

and all that remains is to make the step made in every previous case —to determine the last remaining unspecified constant, B, in terms of the total number N of the particles of the assemblage.

This number N is the sum of N_1 and N_2. Ignoring the terms of higher order in H, we have:

$$N = 2L\phi(B) \quad (89)$$

and this is substantially the equation which was used to determine Sommerfeld's constant A in terms of N; for e^B and $1/A$ are one and the same. To make this equation identical with (63), or rather to make (63) identical with this one, we must there put $G = 2$, as we did—this is the reason for having introduced that factor G.

The procedure is then as follows: put $-B$ for $\log A$ in the right-hand member of equation (66)—differentiate it with respect to B—insert into the derivative the value of B obtained by equating to N the right-hand member of (66), *i.e.*, the value given in (67)—and substitute into (88). The resulting value for χ is this:

$$\chi = 12\left(\frac{\pi}{3}\right)^{2/3}\mu_0^2 n^{1/3} m_0 h^{-2}. \quad (90)$$

To pass now to the experiments: is it permissible to suppose that the susceptibility of any metal is due entirely to the electron-gas within it? This is the same sort of uncertainty as confuses the question of the specific heat. Here we have every reason to expect that the magnetization of an ordinary paramagnetic metal is a threefold effect, involving not only the orientation of the electrons but also the orientation of the atoms, and finally that alteration of the electron-orbits in the atoms which gives rise to diamagnetism. To disentangle these three contributions to the net magnetic moment seems almost beyond the powers of any theory. With the alkali metals, however there is strong evidence that the second may be absent. Spectroscopic data show quite definitely that the magnetic moment of the alkali-metal ion—the atom minus its valence electron—is zero. If every atom in an alkali metal has surrendered its valence electron to the electron-gas, then there will be no orientation of the ions by the magnetic field, and the number of electrons forming the electron-gas will be

equal to the number of atoms. The values of χ computed with this last assumption should then be not less than the actual susceptibility; they may be somewhat greater, because of the diamagnetic effect which is opposed in sign to the paramagnetic effect and therefore neutralizes it in part. Of this diamagnetic effect we can predict the order of magnitude, and we may expect that it will be greater, the higher the atomic number of the metal.

The value of χ proposed above is the value for absolute zero; for higher temperatures a closer approximation can be obtained by using two terms of the expansion in (66), instead of the first term only. It appears, however, that the alteration is slight. Like the average energy and the pressure, the susceptibility of the electron-gas should be very nearly the same at all temperatures from absolute zero up through room-temperature and far beyond. Now it is a fact that the susceptibility of the alkali-metals is independent of temperature—a fact so surprising, that the desire to explain it seems to have been Pauli's principal incentive in undertaking this research. For if the electron-gas were governed by the classical statistics, and the electrons were as many as the atoms, the susceptibility of a metal would increase as the temperature diminished and attain enormous values near the absolute zero.

When Pauli published the theory to which this section is devoted, the experimental data indicated that the susceptibilities of potassium and sodium were somewhat lower, those of rubidium and cæsium markedly lower than the predicted values—divergences which might be charged to the diamagnetic effect or to faults in the theory. Recent Canadian work, coming out very much *a propos*, has improved the situation remarkably. This tabulation (taken from E. S. Bieler, to whose article I refer for the sources) shows the comparison:

	Na	K	Rb	Cs
Theoretical (Pauli)	0.66	0.52	0.49	0.45
Experimental:				
McLennan et al.	0.61	0.42	0.31	0.42
Lane	0.65	0.54		

(All numerical values to be multiplied by 10^{-6})

Singularly enough, the agreements are too good! one would expect the diamagnetic effect to be more considerable than the very slight discrepancies between the experimental and the theoretical values for sodium, potassium, and cæsium. Perhaps further work on the theory of the diamagnetic effect would now be desirable.

Returning once more to the meaning of G, one sees that the placing of the value 2 for G in equation (61) and all of its descendants amounts to the making of the assumption that the electron-gas is really a mixture of two equally numerous and entirely independent assemblages of particles, each for itself obeying the Fermi statistics. This seems a rather odd idea, but inevitable.

Theory of Conduction

The new theory of conduction developed by Houston and Bloch is based upon the wave-theory of negative electricity, in which the interior of a metal is conceived to be filled not with darting corpuscles, but with stationary waves—as many distinct patterns of loops and nodes, it may be, as in the corpuscle-picture there are free electrons. It is not a consequence of the Fermi statistics alone, but of the Fermi statistics plus the wave-theory. Of course, if we come to decide that the Fermi statistics implies the wave-theory and *vice versa*, this warning will seem superfluous; but it is not superfluous, so long as the new statistics is used with reference to corpuscles. Now the corpuscle-picture of negative electricity is not only familiar, but seems likely to survive as the most convenient for describing most of the phenomena in which electrons figure. I will therefore express as much as possible of the new theory of conduction in the language of corpuscles, although eventually I shall be forced to make an assumption which will come to the same result as converting the corpuscles into waves.

To realize the things to be explained, conceive a slab of metal, having a thickness d measured along the x-axis; suppose a potential-difference V to exist between its faces, so that a field $E = V/d$ directed along the axis of x pervades it.

If the electrons in the metal moved perfectly freely, then any which was introduced without kinetic energy at the negative side of the slab would fall forthwith to the positive side, arriving there with the full kinetic energy eV and the full corresponding velocity of magnitude $(2eV/m)^{1/2}$ directed along the axis of x. Certainly nothing of the sort occurs. When a potential-gradient exists along a wire, for instance, heat is developed uniformly everywhere and there is nothing to suggest that the electrons are moving more rapidly at the positive than at the negative end.

We must then suppose that the free flight of the electron is interrupted at frequent intervals, and that at every interruption it loses the kinetic energy and the component of velocity up the potential gradient which it has acquired from the field since the last one previous. Or at least, the average loss of kinetic energy and of "drift-speed"

at interruptions must be balanced by the average gain between interruptions.

In the corpuscle-theory these interruptions are pictured as actual impacts or collisions of the electrons with the atoms. Evidently, if we could assume that whenever an electron hits an atom it rebounds in some direction perfectly transverse to the field, then we should have a mechanism in which the drift-speed of the electron up the potential-gradient is annulled at every impact. This would be much too artificial. But if we think of both the electrons and the atoms as elastic spheres, the latter being so massive that they never budge when struck, the result is in effect the same. For then, the angle between the direction along which an electron approaches an atom and the direction along which it flies away after collision is *on the average* 90°. The rebound is as likely to be backward as forward; the rebounding sphere retains on the average no memory of its former direction of flight. This I will prove later.

There is a difficulty, which I must not leave unmentioned, although in this place I can do nothing to clear it away. In the development of these ideas we shall in effect assume that at the end of each free path the electron loses not only the forward drift-speed but the whole of the kinetic energy which it acquired while traversing that free path under the influence of the field. But if it collides with infinitely massive spheres it does not lose kinetic energy at all. If it collides with spheres of the mass of an atom, it loses kinetic energy, but does not completely lose its drift-speed. The theory of this latter case has been developed by Compton and Hertz for use in the study of conduction in gases, and might be applied to the problem presented by metals, but probably fits them no better than does the other hypothesis.

With this elastic-sphere model, then, the average interval between impacts is the average interval during which the electron is piling up drift-speed, only to lose it all at the end of the interval and be forced to start afresh from scratch. Denote by t_0 the length of this average interval. Since the acceleration of the electron is eE/m, its drift-speed at the end of the period t_0 is (eEt_0/m), its average drift-speed is half as great. Now I must dispel the impression that the drift-speed is the whole of the speed which the electron has. On the contrary, the mean speed of the thermal agitation—let me call it $\bar{v}$—is immensely larger than the small contribution which any ordinary field (indeed, any not very extraordinary field) can impart to an electron over a distance comparable with the distance between atoms. The field must not be supposed to do more than bend very slightly the rectilinear

paths of the electrons from impact to impact. This statement is true with the classical statistics, *a fortiori* with the new. Denote by l the average distance traversed by an electron between impacts. Then t_0 is $l/\bar{v}$, and the average drift-speed is $\frac{1}{2}(eEl/m\bar{v})$.[13] The corresponding current-density is the product of this by the numer of electrons in unit volume multiplied by the charge of each. So, for the current-density produced by unit-field-strength, which is by definition the *conductivity* σ, we obtain the formula:

$$\sigma = \tfrac{1}{2}ne^2l/m\bar{v}. \tag{91}$$

The constant l, the *mean free path*, is the third disposable constant of the theory of electrons in metals.

I fear that the foregoing passage sounds very old-fashioned; but nevertheless it expresses the corpuscle-theory of conduction. The notion of elastic spheres is only accessory—an image which may or may not be the best to represent the central idea, the idea that the life-history of a corpuscle in a metal pervaded by a field is an alternation of gradual gain and sudden loss. The mean free path is the average distance of uninterrupted gain.

The common test of the formula (91) is the test by the temperature-variation. The result of this, incidentally, was regarded as quite as grave a demerit of the old electron-gas theory as the difficulty with the specific heat.

It is a fact of experience that the resistivity $\rho = 1/\sigma$ of any metal varies rapidly with temperature. For many metals it varies directly as T over quite a wide range; at low temperatures even more swiftly, not to speak of the strange phenomenon of supraconductivity. Now in equation (151) we have ρ set equal to a combination of two universal constants with three quantities $\bar{v}$, n, l between which last the responsibility for these great variations must be divided.

According to the classical statistics $\bar{v}$ is proportional to $T^{1/2}$. This is a variation in the right sense, but not fast enough. To make ρ vary as T we must then make nl vary as $1/T^{1/2}$. With the Fermi statistics the requirement is harder. The mean speed of thermal agitation is almost independent of temperature, and the burden of the whole responsibility for making ρ proportional to T must be loaded upon nl. The first step with the new statistics is a step backward.

Can we reasonably assume n to be the cause of the variation?

[13] It is the mean of the reciprocal of the speed, not the reciprocal of the mean speed, which should figure here; but with so rough a formula the distinction is scarcely worth making.

If so, then it must diminish with rising temperature. It would seem reasonable enough for n to *increase* with rise of temperature, for presumably the free electrons arise from ionization of the atoms, and ionization is promoted by heat; but for n to *decrease* would seem very odd, notwithstanding Waterman's successes in accounting for some of the data by such a theory.

Part of the burden, then, must be cast on the mean free path—indeed the whole of it, if we adopt the new statistics so that $\bar{v}$ is held constant, and suppose in addition that n does not vary with temperature. But the elastic-sphere conception cannot stand the strain. It gives for the mean free path a value independent of temperature,[14] except insofar as the metal dilates with increase of heat. This is pretty nearly checkmate.

If however we might suppose that an electron may sometimes go clear through an atom without being reflected or deflected, and that the chance of such a piercing is relatively smaller and the chance of a rebounding relatively greater, the more violently the atom is vibrating—then by this theory the mean free path would diminish as the temperature rises, which is what is desired. This is an idea proposed long since by Wien.

The new idea is in result the same. The probability of the rebounding, or let me say of the *scattering* of an electron by an atom, is supposed to increase with the vigor of the vibrations of the latter. But for this a new reason is advanced: the reason, that while vibrating the atom is most of the time away from its equilibrium-place in the crystal lattice, and its relations with its companions are distorted. The probability of scattering is made to depend not only on the presence of the atom somewhere along the path down which the electron is rushing, but also on the relative positions of all the other atoms in the crystal.

To make such an assumption is, in effect, to compromise between the corpuscle-theory and the wave-theory. For what is the evidence from which it is inferred that a beam of light falling upon a grating, say, or a beam of X-rays falling upon a crystal, are undulatory? Essentially this: the way in which the beam is scattered or diffracted by the regular array of rulings on the grating or atom-groups in the crystal is different—greatly and strikingly different—from the way in which we know that it would be scattered by a single ruling, or suspect with good reason that it would be scattered by a single atom-

[14] The value $1/N\pi R^2$ familiar in the kinetic theory of gases, N standing for the number of fixed spheres per unit volume, R for the sum of the radii of a fixed and a moving sphere.

group. For instance, there are directions in which no light at all is sent by the regular array, though assuredly light would be scattered in those directions by any member of the array if it were solitary. These facts are explained by invoking interference of waves. The wavelets expanding outwards from the various rulings or scattering particles are supposed to arrive in opposite phases at the "dark fringes" of the diffraction-pattern, so that they cancel each other. But one might also say that the beam of light is a stream of corpuscles which are deflected or scattered by the atom-groups or rulings which they happen to strike, and that the law of scattering of the individual atom-group is altered by the marshalling of the scattering elements into a regular pattern, so that in particular the probability of a deflection towards one of the dark fringes is reduced to zero.

I am not prepared to say that such a compromise is a full alternative for the wave-theory, though modern theoretical physics seems to be tending in that direction. But if we wish to describe with the language of the corpuscle-theory the phenomena of diffraction by a crystal, whether of waves of light or waves of negative electricity: then we must certainly adopt the idea of a probability-of-scattering, of a mean-free-path, which varies with the irregularity of the placing of the atoms.

The principle is especially simple and especially startling, if we deal with a beam of which the wave-length—considering it as a beam of waves—exceeds the spacing of the lattice. Waves of such a magnitude would not be diffracted at all by scattering particles placed exactly at the points of the lattice. Though any particle singly would scatter them, they flow through the lattice intact. If then we wish to interpret the beam as a stream of corpuscles, the probability of deflection of any corpuscle by any atom must sink to zero when the arrangement is made perfect; the mean free path must then be considered infinite.

The resistance of a perfect crystal of an element should then be zero when all the atoms are stationary in their places on the lattice—if they ever are, which apparently they are not; and should increase steadily with increasing temperature, in a way which can be computed if we know two things: the way in which the scattering of waves by particles on a lattice varies with the amplitude of the quiverings of the particles about their lattice-points, and the way in which the amplitudes of the particles vary with the temperature. The second of these questions is the subject of the theory of specific heats of solids, developed principally by Debye. The first has been profoundly studied by Debye and by several other physicists interested chiefly in the scattering of X-rays by crystals. Transferring their results

into the theory of the diffraction of electron-waves, Houston demonstrated that over a wide range of temperatures the resistance of a perfect crystal should vary as the absolute temperature.

To determine not only the law of variation of resistance with temperature but the actual value of the resistivity for any metal, it would be necessary to evaluate the mean free path of the electrons. By the thoroughgoing corpuscular theory, this depends on the size of the atoms from which the corpuscles rebound; by the wave-theory, it depends on the scattering-power of the individual atom, which thus takes the place of the "size of the atom." The problem of computing the scattering-power of an atom for electron-waves belongs to the new mechanics. Houston was able to obtain good numerical agreements for several metals.

Another way of introducing irregularity into a crystal of an element consists in replacing a small fraction of the atoms, chosen at random here and there on the lattice, by atoms of another element. Certain alloys, known as "solid solutions," are of this type; and it is not only known that the resistance of such an alloy is greater than that of the element which is most abundant in it, but it has been shown by Nordheim that the dependence of resistance on percentage of substituted atoms follows the rule to be expected from the diffraction theory of resistance.[15]

Since then the conception of mean-free-path can be reinterpreted in terms of the wave-theory, and since it appears to be possible to deduce from the wave-theory a law of variation of mean-free-path with temperature which can be incorporated intact into the corpuscle-theory it is permissible to return to the corpuscle-picture to set up a theory of conduction of heat and of electricity, and of the thermo-electric effects in crystals.

We shall apply what I may call the method of the *perturbed distribution-function*, developed by Lorentz. The idea is, to begin by deriving a distribution suiting the actual case. The functions which we have hitherto employed, that of Maxwell and that of Fermi, are isotropic; it is only in the combination $(\xi^2 + \eta^2 + \zeta^2)$, hereafter to be called v^2, that the velocity-components ξ, η, ζ appear in them.[16] These "standard" functions may be appropriate to a uniform metal in which the temperature and the potential are uniform. Evidently they are not appropriate to a metal in which there is an electric field,

[15] The idea that the free paths of electrons extend from one irregularity of the crystal to another was propounded before the advent of the wave-theory of negative electricity.

[16] I shall use the velocity-components hereafter in lieu of the momentum components, to conform with the custom.

or a temperature-gradient, or which varies in its chemical nature from place to place, as might an alloy. If in such a case we orient the axis of x parallel to the gradient of the variable quantity—be it electric potential, temperature, or whatever else—we must expect ξ to enter differently into the distribution-function from η and ξ.

Various arguments show that as a rule the departures from the standard function must be rather small. Lorentz therefore postulated that in the presence of a gradient directed parallel to the x-axis, the actual distribution should differ from the standard function $f_0\,(v)$ —this he of course assumed to be Maxwell's—by virtue of a small additive term, a new function of v multiplied by the velocity-component ξ:

$$f = f_0(v) + \xi g(v), \tag{92}$$

and he proceeded to determine the new function g by the condition that f should remain constant in time despite the collisions of the electrons with the atoms. More precisely, he found for each of the three cases with which we shall be concerned a function g, such that the distribution-function obtained by adding ξg to f_0 conforms to that condition. This justifies the procedure.

Much the simplest case of the three is that of a uniform metal at a uniform temperature, subject to an electric field; for here the distribution-function need not vary from place to place. It will be well to go through the reasoning in this instance, though the formula for conductivity in which it leads differs but little from (91).

It is required, to find a function g of the combination $(\xi^2 + \eta^2 + \zeta^2)^{1/2}$ such that if at any moment the distribution $(f_0 + \xi g)$ prevails, it continues unchanged throughout time—the number of particles in any compartment or cell of the velocity-space (the momentum-space of the earlier pages, with its unit of length altered in the ratio $m : 1$) stays constant. Choose a compartment enclosed between planes ξ and $\xi + d\xi$, η and $\eta + d\eta$, ζ and $\zeta + d\zeta$. Call it the compartment C. Its volume is $d\xi d\eta d\zeta$, which to save a profusion of Greek letters I will usually denote by $d\tau$. The number of particles in it is $f \cdot d\tau$. This number must remain unchanged, though individual particles are constantly moving into and out of C in either of two ways—by "drift" and by "collision."

Owing to the field E, all of the particles have a steady acceleration $\alpha = eE/m$, because of which they are continually and continuously drifting from cell to cell. One easily sees that the number which thus drift out of C per unit time (it is best to think of "unit time" not as one second, but as a period small compared with the mean

time between impacts) is equal to first approximation to $\alpha f(\xi + d\xi, \eta, \zeta)d\eta d\zeta$.[17] This loss is partly balanced by an inward drift of particles which are accelerated into the compartment from the one lying beyond the plane ξ; the balance is not perfect, for the number drifting in per unit time is equal to $\alpha f(\xi, \eta, \zeta)d\eta d\xi$ and there is a difference $\alpha(df/d\xi)d\tau$ outstanding.[18] This difference must be balanced by the entrances and exits of particles which undergo collisions.

Let adt represent the number of electrons which, being initially in the compartment C, suffer impacts during unit time and are thus suddenly bumped out of it; and bdt the number which, being initially in other compartments, suffer such impacts during unit time that they suddenly turn up in C. The function g must be so chosen, that the lack of balance between the electrons drifting out and the electrons drifting in is just compensated by the lack of balance between those bumped into the compartment and those which are bumped out:

$$\frac{eE}{m}\frac{df}{d\xi} = b - a. \tag{93}$$

We must therefore evaluate $(b - a)$ in terms of the distribution-function.

We already have a formula ready-made for the number of impacts experienced per unit time by the particles of speed v; it is v/l for each particle, so that:

$$a = (v/l)f d\xi d\eta d\zeta. \tag{94}$$

Since however we have also to compute b, we shall find it expedient to classify these impacts according to the destinations of the particles, so to speak—according to the compartments of velocity-space into which they are bounced. A particle of speed v is located, in the velocity-space, on a sphere of radius v centered at the origin. Collision with an immovable atom changes the direction, but not the magnitude of the velocity; in the velocity-space, the particle suddenly moves to some other point on the same sphere. When electrons jump out of the compartment C because of impacts, they land in the other compartments which with C form a spherical layer around the origin. When electrons jump into C because of impacts, they come from the other compartments of that same layer. We shall derive an expression

[17] During a time dt so short that αdt is small by comparison with $d\xi$, the particles which initially lie between the plane $(\xi + d\xi - \alpha dt)$ and the side $(\xi + d\xi)$ of the compartment move out of it, while the particles which initially lie between the side ξ of the compartment and the plane $(\xi - \alpha dt)$ move into it; these two classes of particles number $f(\xi + d\xi, \eta, \zeta)\alpha dt \cdot d\eta d\zeta$ and $f(\xi, \eta, \zeta)\alpha dt \cdot d\eta d\zeta$ respectively.

[18] This expression figures in the equations as a net loss, but in fact has a negative sign (since $df/d\xi < 0$) and therefore is actually a gain.

for the number of particles leaping from C into any other cell C' of the layer, and an expression for the number leaping reversely. The difference or lack of balance between these numbers, integrated over all the cells C', will be the required quantity $(b - a)$.

We begin by inquiring how many particles make such impacts that their paths (in the coordinate-space, of course—not the velocity-space) are deflected through angles between say θ and $\theta + d\theta$. To be deflected through an angle θ, an electron must strike an atom at a point where its surface is so oriented, that the normal (which is the line of centres at the instant of collision) is inclined to the line of approach of the electron at the angle $\psi = \frac{1}{2}(\pi - \theta)$. Denote by R the radius of the atom, and suppose that the radius of the electron is negligibly small.[19] Think of all the $f \cdot d\xi d\eta d\zeta$ electrons which at some particular moment of time are in unit volume of the metal, and belong to the compartment C of the velocity-space. Imagine each of these to be the centre, in the coordinate-space, of a pair of circles lying in the plane perpendicular to its path, and having radii $R \sin \psi$ and $R(\sin \psi + d \sin \psi) = R(\sin \psi + \cos \psi d\psi)$. As time goes on, let these circles travel in the direction normal to their plane with the speed v. During unit time each pair of circles traces out a pair of cylinders of length v, containing between them a cylindrical sheath of volume $v \cdot 2\pi R^2 \sin \psi \cos \psi d\psi$. Multiplying this by the number $f \cdot d\xi d\eta d\zeta$ of the electrons, we get the total volume included in all of these sheaths. Multiplying this by the number N of atoms in unit volume, we get the number of atoms located with their centres in these sheaths—which is the number of atoms so placed that in unit time, electrons of the stated cell impinge on them at angles between ψ and $\psi + d\psi$—which is the number of impacts per unit time in which electrons are deflected through angles between θ and $\theta + d\theta$, which accordingly is this:

$$N \cdot f d\xi d\eta d\zeta \cdot 2\pi v R^2 \sin \psi \cos \psi d\psi = f d\xi d\eta d\zeta \cdot N \pi R^2 v \cdot \tfrac{1}{2} \sin \theta d\theta. \quad (95)$$

It will be convenient to express this as a fraction of the total number of impacts—call it $Z d\tau$—experienced per unit time by all the electrons in question, which by integrating (95) is found to be:

$$Z d\tau = \pi N R^2 v \cdot f \cdot d\tau, \quad (96)$$

substituting which into (95) we get:

$$Z \cdot 2 \sin \psi \cos \psi d\psi = Z \cdot \tfrac{1}{2} \sin \theta d\theta. \quad (97)$$

[19] The formulæ remain valid even if the diameter of the electron is supposed not negligibly small, provided that we interpret R as the sum of the radii of atom and electron; but the generalization is not, so far as I know, of any practical value.

It will be observed that deflections smaller than 90° are equally numerous with deflections greater than 90°, so that on the average the electrons after impact have no reminiscence of their prior direction of motion, as I mentioned earlier. Also, comparing (96) with (94), one derives the expression for mean free path,

$$l = 1/N\pi R^2, \tag{98}$$

cited already in a footnote. To appreciate the most important feature of the expression (97) we must however return to the velocity-space.

In the velocity-space, the electrons of which the paths in coordinate-space are deflected through angles between 2θ and $2\theta + d2\theta$ execute leaps from the compartment C into other compartments of the spherical layer aforesaid, located on a certain region thereof. These occupy a belt or collar on the sphere, intercepted between two cones drawn with their common apex at the centre, their common axis pointing towards C and their apical semi-angles equal to θ and $\theta + d\theta$ respectively. Now the area of this belt is itself proportional to $\sin \theta d\theta$. This is very important: for it means that the electrons which are bounced out of C by collisions are sprinkled uniformly over all the rest of the sphere. More yet: it means that the electrons which are bounced out of *any* cell of the spherical layer are sprinkled uniformly through all the rest of the layer.

Consider then the interchange of electrons between two cells of the layer, say C at (ξ, η, ζ) with volume $d\tau$, and C' at (ξ', η', ζ') with volume $d\tau'$. The number leaping from C to C' is equal to the total number of impacts occurring in C multiplied by the ratio which the volume of C' bears to the volume V of the layer. The number leaping from C' to C is equal to the total number of impacts occurring in C', multiplied by the ratio which the volume of C bears to the volume of the layer. The excess of the latter over the former is then:

$$Z(\xi', \eta', \zeta')d\tau'(d\tau/V) - Z(\xi, \eta, \zeta)d\tau(d\tau'/V), \tag{99}$$

which with the aid of (96) and (98) may be written thus:

$$\frac{vd\tau}{l}[f(\xi', \eta', \zeta') - f(\xi, \eta, \zeta)]d\tau'. \tag{100}$$

This is the quantity of which the integral with respect to ξ', η', ζ', extended over the spherical layer, is equal to $(b - a)d\tau$—the net rate at which compartment C gains particles through impacts.

Making Lorentz' postulate about the form of the distribution-

function f, and remembering that throughout the spherical layer the combination $(\xi'^2 + \eta'^2 + \zeta'^2)^{1/2}$ is confined within a narrow range of values around v we obtain:

$$b - a = (v/l) \cdot \beta(v) \cdot \int (\xi' - \xi) d\tau' / V. \tag{101}$$

To effect the integration it is expedient to change over to polar coordinates in the velocity-space. Leaving the origin where it was, and directing the polar axis towards C, we make the radial coordinates and the colatitude-angle identical with our v and θ, and for the meridian-angle we use the symbol ϕ. Dividing up the layer into compartments by latitude circles and meridians, we have for any one of them:

$$d\tau' / V = (1/4\pi) \sin \theta d\theta d\phi. \tag{102}$$

Consequently we obtain:[20]

$$b - a = (v/4\pi l) g(v) \cdot \int_0^{\pi} d\theta \sin \theta \int_0^{2\pi} d\phi (\xi' - \xi). \tag{103}$$

Everything therefore hinges on the evaluation of $(\xi' - \xi)$—the change in the x-component of velocity which the electron incurs when it leaps from C' to C—as a function of θ and ϕ. Now it may be shown without much difficulty,[21] that:

$$\xi' - \xi = -2v \cos \psi \cos \omega = -2v \sin \tfrac{1}{2}\theta \cos \omega \tag{104}$$

wherein ψ stands as before for the angle between the line of approach of the electron and the line of centres at the instant of impact, and ω stands for the angle between the line of centres and the axis of x. There is also a standard formula [22] relating $\cos \omega$ to ψ, ϕ and the

[20] Lest someone be disconcerted by the apparent difference between this equation and that given by Lorentz, I remark that I am using θ to designate an angle twice as great as the one which he denoted by θ.

[21] Let $\boldsymbol{v}$, $\boldsymbol{v}'$ represent the vector velocities of the electron before and after impact, $\boldsymbol{c}_1$ the unit vector along the line of centres at the moment of impact. The components of $\boldsymbol{v}$ and $\boldsymbol{v}'$ along the line of centres are equal in magnitude and opposite in direction; the components perpendicular to the line of centres are equal in magnitude and direction. Writing these statements down in vector notation:

$$\boldsymbol{v} \cdot \boldsymbol{c}_1 = \boldsymbol{v}' \cdot \boldsymbol{c}_1; \; \boldsymbol{v} - (\boldsymbol{v} \cdot \boldsymbol{c}_1)\boldsymbol{c}_1 = \boldsymbol{v}' - (\boldsymbol{v}' \cdot \boldsymbol{c}_1)\boldsymbol{c}_1 = \boldsymbol{v}' + (\boldsymbol{v} \cdot \boldsymbol{c}_1)\boldsymbol{c}_1, \text{ hence,}$$
$$\boldsymbol{v}' - \boldsymbol{v} = -2(\boldsymbol{v} \cdot \boldsymbol{c}_1)\boldsymbol{c}_1 = -2v \cos \psi \cdot \boldsymbol{c}_1; \; \xi' - \xi = -2v \cos \psi \cos \omega.$$

[22] Imagine two planes P_1 and P_2 intersecting along a vertical axis, the dihedral angle between them being ϕ. Through a point O on the axis draw a horizontal plane N, and from O draw two lines of unit length OR_1 and OR_2, the former in plane P_1 and inclined at β to the vertical axis, the latter in plane P_2 and inclined at ψ to the vertical axis; ω is the angle between them. The points R_1 and R_2 are at heights $\cos \beta$ and $\cos \psi$ above the plane H. On the vertical line through R_2 put a point R_3 at distance $\cos \beta$ above H. Expressions for the sides of the right-angled triangle $R_1R_2R_3$ are easily obtained, and (214) is derived by applying the theorem of Pythagoras to them.

angle $\beta = \arccos(\xi/v)$ between the initial path of the electron and the axis of x, as follows:

$$\cos\omega = \cos\beta\cos\psi + \sin\beta\sin\psi\cos\phi. \tag{105}$$

We now have everything necessary to do the integration in equation (103), and we find:

$$b - a = (v\xi/l)g(v). \tag{106}$$

Substituting this into equation (93), the condition that the distribution-function $(f_0 + \xi g)$ shall be stable by virtue of perfect balance between the rates at which electrons are shifted from compartment to compartment by the impacts and by the accelerating field, we get:

$$(eE/m)\frac{d}{d\xi}(f_0 + \xi g) = (v\xi/l)g. \tag{107}$$

If the term $\xi g(v)$ is truly small by comparison with the term $f_0(v)$, we may neglect the second term on the left; and since $(df_0/dv) = (df_0/dv)(dv/d\xi) = (\xi/v)(df_0/dv)$, the culmination of all the argument is in the formula:

$$\xi g(v) = \frac{\xi l}{v^2}\frac{eE}{m}\frac{df_0}{dv} \tag{108}$$

for the alteration which the applied electric field imposes on the distribution. Notice that g involves ξ and η and ζ only in the combination v; this justifies the procedure of Lorentz.

Now each electron which during unit time crosses any surface imagined in the metal contributes an amount e to the current through that surface; but the contributions made by electrons crossing in opposite senses are opposite in sign—what we perceive as current is *net* current, the excess of the flow of charge one way over the flow the other. Conceive a plane surface-element of area da, normal to the field, therefore normal to the axis of x. We must classify the electrons which traverse it according to their values of ξ. Let $H(\xi)d\xi da$ represent the number passing through in unit time, and having at the moment of passage x-components of velocity in the range $d\xi$ at ξ. This is equal to the number which at any instant have their x-components of velocity in this range, and are situated in the right prism having da for its base and extending a distance ξ down the direction of x:[23]

$$H(\xi)d\xi da = \xi da d\xi\int_{-\infty}^{\infty}d\eta\int_{-\infty}^{\infty}d\zeta f(\xi,\eta,\zeta). \tag{109}$$

[23] This would be immediately obvious, if all the electrons were moving parallel to the x-axis and made no impacts. Electrons having y and z components of velocity in addition to the x-component will drift obliquely out of the prism, and electrons making impacts will be thrown out of the range $d\xi$; but each electron thus lost will be balanced by another coming in from outside.

Since for electrons crossing in opposite senses ξ is of opposite signs, the integral of this expression over all values of ξ, multiplied by e, gives the net current through da when the proper distribution-function is inserted. Evidently the integral will vanish when f is isotropic; there are enormous flows of charge both ways through da, but they are balanced. Unbalance is brought about by the non-isotropic perturbation-term in the distribution-function. Making the postulate of Lorentz, we obtain for the net current through unit area perpendicular to the field, the current-density J, the expression:

$$J = e \int H(\xi) d\xi = e \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} d\xi d\eta d\zeta \cdot \xi^2 g(v). \tag{110}$$

Set for $g(v)$ the expression in (217), and to effect the integration transform into polar coordinates in the velocity-space, orienting the polar axis along the axis of ξ; integrating over the angles, one obtains:

$$J = \frac{le^2E}{m} \frac{4\pi}{3} \int_0^{\infty} v^2 \frac{df_0}{dv} dv. \tag{111}$$

The final integration is easy if one chooses for f_0 the Maxwell function, not quite so easy if one chooses that of Fermi. A further step will be of some advantage. Integrating by parts, and noticing that v vanishes at the lower and f_0 at the upper limit, we find:

$$J = -(8\pi le^2E/3m) \int_0^{\infty} f_0 v dv \tag{112}$$

and the integral remaining, divided by n the number of electrons per unit volume, is seen to be $1/4\pi$ times the mean value of v^{-1},—the average of the reciprocal of the speed of the electrons, in the absence of the field. Denoting this by $\overline{v^{-1}}$, we may write as the general formula for conductivity:

$$\sigma = \frac{J}{E} = \frac{2}{3} \frac{le^2n}{m} \overline{v^{-1}}. \tag{113}$$

The analogy with (91) is obvious, but we must not be misled into identifying the average of the reciprocal speed with the reciprocal of the average speed; they are not quite equal.

The actual final formulæ obtained by the old and the new statistics —substituting, that is to say, the appropriate values of $\overline{v^{-1}}$ from (72b) and (72a), and putting $G = 2$—are as follows:

$$\sigma = \frac{4}{3} \frac{e^2ln}{(2\pi mkT)^{1/2}} \quad \text{(old)}, \tag{114a}$$

$$\sigma = \frac{8\pi}{3} \frac{e^2l}{h} \left(\frac{3n}{8\pi} \right)^{2/3} \quad \text{(new)}. \tag{114b}$$

As Sommerfeld has shown, all of the reasoning by which (108) was reached remains intact even when it is supposed that l is a function of v; in that case, l remains under the integral sign in (111), and the integral itself is equal to $n/4\pi$ times the mean value of $v^{-2}d(lv^2)/dv$. This generalization may be of some value.

Uniform Metal with a Temperature-Gradient; Thermal Conduction

We have now to find a function g such that the distribution $(f_0 + \xi g)$ is stable in a metal in which there is a constant gradient of temperature along the axis of x. When we find it, we shall be able to evaluate the integral

$$W = \tfrac{1}{2}m \int\int\int d\xi d\eta d\zeta \cdot g(v) v^2 \xi^2, \tag{115}$$

which is like the integral in (219) except for the differently-chosen form of g and the substitution of $\frac{1}{2}mv^2$ for e, and therefore represents the net rate of flow of kinetic energy borne by electrons through unit area perpendicular to the gradient—the contribution of the electrons to the flow of heat, under the circumstances stated.

The standard distribution-function f_0, involving as it does the temperature T, is in this case itself a function of x. One might expect that this dependence of f_0 on x would be sufficient to fulfil the requirements. An isotropic distribution which varies from point to point cannot however be stable; the particles conforming to such a one at any given moment would proceed to drift off down the gradient. A stable distribution cannot be isotropic. We must repeat the process of balancing the rates at which particles enter and leave each compartment of the phase-space through collisions and through drift. I say the phase-space now, instead of the velocity-space; for this case is made more complicated than the previous one by reason of the fact, that we now must make the balance separately for the electrons contained in each of the six-dimensional cells $d\xi d\eta d\zeta dx dy dz$, whereas previously we could make it *en bloc* for all of the particles in the entire metal comprised within any velocity-cell $d\xi d\eta d\zeta$.

Consider then the six-dimensional cell $d\xi d\eta d\zeta dx dy dz = ds$, and the $f(\xi, \eta, \zeta, x, y, z)ds$ electrons in it. The first three factors in ds denote the range of velocity, the last three the range in position, within which an electron must lie if it is to belong to ds. Electrons in the proper range of position are continually entering or leaving the proper range of velocity, because of impacts. The net rate at which ds gains electrons in this way is given, as before, by (106). Electrons in the proper range of velocity are continually drifting into the proper

range of position, coming into $dxdydz$ from the region adjacent to it on the side towards smaller or greater values of x, according as ξ is positive or negative. By the same sort of reasoning as led to the term $\alpha df/d\xi$ in (93), one sees that the net rate at which ds loses electrons in this way is $\xi df/dx$. Equating the two, we have:

$$\xi df/dx = b - a = (\xi v/l)g(v), \tag{116}$$

and when we put f_0 for f as before, on the ground that the term (ξg) in the full expression for f makes but a small contribution to the left-hand member, we have all that is required for computing g from (116) and W from (115) for whatever standard function we elect.

At this point, however, there arises a difficulty. If having adopted this way of determining g we proceed to compute the electric current J in the metal by formula (110) we find that it is not zero. The reasoning has led to the conclusion that wherever there is a net current of heat in a metal, there is also a net current of electricity. This conclusion is not in accord with experiment. Yet there is apparently no other way to circumvent it, than to suppose that when a gradient of temperature is maintained in a metal there arises a spontaneous internal electric field, of just such a magnitude as to counteract the electric current which would otherwise persist. The gradient of temperature calls forth a gradient of potential; the actual distribution-function is the one which is stable under both these gradients combined. In the bookkeeping of the compartment ds, the net gain from impacts $(b - a)$ is balanced against the sum of the net loss through drift in the coordinate-space $(\xi df/dx)$ and the net loss through drift in the velocity-space $(\alpha df/d\xi)$. Putting these statements into the form of equations, and denoting by E the hypothetical electric field and by $\alpha(= eE/m)$ the acceleration which it imparts to each electron, we have:

$$\alpha(df/d\xi) + \xi(df/dx) = b - a = (\xi v/l)g \tag{117}$$

$$J/e = \int\int\int d\xi d\eta d\xi \cdot \xi^2 g = 0, \tag{118}$$

a pair of equations for determining E and the function g.

Lorentz, adopting the Maxwell-Boltzmann function for f_0, solved the equations, and obtained:

$$W = \frac{8}{3}\frac{nlk^2T}{(2\pi mkT)^{1/2}}\left(\frac{dT}{dx}\right) \qquad \alpha = \frac{1}{2}\frac{k}{m}\frac{dT}{dx}. \tag{119}$$

Sommerfeld adopted the Fermi function, and obtained for the degenerate case:[24]

$$W = \frac{8\pi^3}{9}\frac{lk^2T}{h}\left(\frac{3n}{8\pi}\right)^{2/3}\left(\frac{dT}{dx}\right). \tag{120}$$

The coefficient of dT/dx in these expressions for W is by definition the *thermal conductivity*, usually denoted by κ. One notices that these expressions for κ like those for σ, involve the more or less disposable constants n and l. This however is not true of the ratio of the conductivities.

The Wiedemann-Franz Ratio

For the ratio of thermal to electric conductivity, the old statistics and the new supply expressions involving nothing but T and the ratio of the universal constants k and e, and differing only by a slight numerical factor:

$$\kappa/\sigma = 2(k/e)^2T(= 4.2 \times 10^{-11} \text{ at } T = 291^\circ K) \tag{121}$$

by the old statistics, and

$$\kappa/\sigma = \tfrac{1}{3}\pi^2(k/e)^2T(= 7.1 \times 10^{-11} \text{ at } T = 291^\circ K) \tag{122}$$

by the new.

This "Wiedemann-Franz ratio" seems to have been predestined to encourage the devotees of the electron-gas theory. Every other formula offered by the theory contained either n or l or both, and therefore could not serve as an ultimate critical test; for any discrepancy with the data could be removed by adjusting these constants. True, the ensemble of the formulæ provided by the classical theory ran counter to the data in so many different ways, that the net result was quite unfavourable; but one could not point out any single prediction which was certainly wrong. If however the Wiedemann-Franz ratio had departed by an order of magnitude or more from the value of $2(k/e)^2T$, the electron-gas theory could hardly have survived the blow. But in this one case where disagreement would have been fatal, there was agreement; not perfect, but rather too good to be discarded as fortuitous. For many of the familiar metals the ratio, when measured at room-temperature, turned out to be around 6 or $7 \cdot 10^{-11}$. This more than any other one fact was what kept alive the feeling, that in spite of all its difficulties the electron-gas theory must be fundamentally right.

[24] To derive this formula it was necessary to proceed to the second-approximation expression for the Fermi distribution-function; the first approximation merely yielded zero for W.

For the twelve metals Al, Cu, Ag, Au, Ni, Zn, Cd, Pb, Sn, Pt, Pd and Fe, the average of the values of κ/σ at 291° K is $7.11 \cdot 10^{-11}$. The agreement with the prediction of the new statistics is more than good. It is so very good, that it must be partly accidental, especially as the individual values from which the average is formed depart from it by varying amounts. One may still doubt whether it is to be admitted as one of the items which compel the adoption of the new statistics. Drude, be it recalled, obtained the value $6.3 \cdot 10^{-11}$ out of the crude assumption that all of the electrons in any volume-element have the same speed.[25] It used to be regarded as rather amusing that the elaborate calculations of Lorentz merely impaired the agreement which Drude had attained in a naively simple way.

Both theories require that the ratio be proportional to T; this is fairly well satisfied over wide ranges of temperature, but at extreme degrees of cold there is marked divergence, which is inconvenient. It may be desirable to invoke other mechanisms of conduction to supplement the free electrons—as for instance the passing-along of electrons from atom directly to atom to assist in the conduction of electricity, or the transmission of elastic vibrations to aid in the transfer of heat. Indeed, when one reflects that insulators though they lack free electrons yet have some device for the transmission of heat, one wonders why this device should not be available to metals also, and exalt their values of κ and of κ/σ above the predictions of the electron-gas theory.

Intrinsic Potential Difference

We have seen that in a metal where there is no electric current and yet there is a current of heat, an internal electric field must be imagined. We shall now see that in a metal where there is no electric current, but the number of electrons per unit volume varies from point to point, there must also be an internal electric field. This sounds plausible to intuition, for one would expect the electrons to diffuse from regions of higher to regions of lower density unless they were impeded by some force. The equations (117) and (118) enable us to evaluate this force.

Returning to these equations, introduce polar coordinates v, θ, ϕ in the velocity-space as we have formerly done; then $\xi = v \cos \theta$. Multiply both sides of equation (117) by $\cos \theta$; the right-hand member of the new equation is then proportional to $\xi^2 g$. Integrate both

[25] Drude of course could evaluate the ratio k/e without knowing either k or e accurately or at all, since it is the same as the ratio $N_0 k / N_0 e$ — N_0 standing for the number of molecules in a gramme-molecule, the Loschmidt number—and $N_0 k$ is the gas-constant R while $N_0 e$ is the Faraday constant of electrolysis.

members of this new equation over the entire velocity-space. The integral on the right then vanishes by reason of (118), and for the integrals on the left we have:

$$\alpha \int d\tau (df_0/d\xi) \cos\theta + \int d\tau \xi (df_0/dx) \cos\theta = 0. \tag{123}$$

By obvious transformations we get:

$$\alpha \int d\tau \frac{df_0}{dv} \cos^2\theta + \frac{d}{dx} \int d\tau f_0 v \cos^2\theta = 0. \tag{124}$$

Integrating over the angles:

$$\frac{4\pi}{3} \alpha \int \frac{df_0}{dv} v^2 dv^3 + \frac{4\pi}{3} \frac{d}{dx} \int f_0 v^3 dv = 0. \tag{125}$$

Leaving the second term as it is, but integrating the first by parts, we find that as f_0 vanishes (whichever statistics we use) at one limit and v at the other limit of integration, we get:

$$-\frac{2}{3} \alpha \int 4\pi f_0 v^{-1} \cdot v^2 dv + \frac{1}{3} \frac{d}{dx} \int 4\pi f_0 v \cdot v^2 dv = 0. \tag{126}$$

The integrals are written in this curious fashion, to bring out the feature that they are proportional to the mean values of functions—the functions v^{-1} and v, respectively—averaged over the electrons in question; which is to say, all the electrons contained in the compartment *dxdydz* of coordinate-space, to which equation (117) has reference. They are in fact equal to the products of these mean values, to wit the *mean reciprocal speed* and the *mean speed*, by the number *ndxdydz* of the electrons in the compartment *dxdydz*. Rewriting (125) accordingly, with overlinings to signify averages, and dividing out the factors *dxdydz* and 1/3, we get:

$$-2\alpha n \overline{v^{-1}} + d(n\bar{v})/dx = 0, \tag{127}$$

and this is the equation for the acceleration α or the accelerating field $E = m\alpha/e$, required to counteract the electric current which otherwise would be produced in the presence of the gradient $d(n\bar{v})/dx$ of the quantity $n\bar{v}$. This is the gradient which evokes the hypothetical electric field; gradients of temperature or of concentration act indirectly, by making $n\bar{v}$ vary.

With the classical statistics the development is extremely simple, for $\bar{v}$ depends on temperature only, while n may be varied at will.

Heretofore we have tacitly assumed that n remains the same while T and therefore $\bar{v}$ vary along the axis of x, so that:

$$d(n\bar{v})/dx = nd\bar{v}/dx = n(d\bar{v}/dT)(dT/dx), \tag{128}$$

and the reader can verify the expressions for α and W given in (119) by starting from this point. But now we will assume that T and $\bar{v}$ remain the same while n varies along the x-direction with a gradient dn/dx. Then:

$$d(n\bar{v})/dx = \bar{v} \cdot dn/dx. \tag{129}$$

Putting this into (127), and recalling that in the Maxwell distribution the mean values of v and v^{-1} are thus related,

$$\bar{v} = (2kT/m)\overline{v^{-1}}, \tag{130}$$

one perceives that $\bar{v}$ disappears by division from the two sides of the equation, leaving this:

$$\alpha n = (eE/m)n = (kT/m)dn/dx, \tag{131}$$

the desired equation for the necessary electric field. Integrating it, we obtain another of very familiar aspect:

$$n = n_0 \exp (eE/kT)(x - x_0) = n_0 \exp (- [V - V_0]/kT). \tag{132}$$

This is the celebrated equation of Boltzmann embodying the statement that if in an assemblage of particles at uniform temperature there are variations in the number-per-unit-volume from place to place, then there must also be a field of force against which work must be done to move a particle from place to place—and *vice versa*. Specifically: if at any two points P and O the number-per-unit-volume of the particles has values n and n_0, there must be a field of force such that when a particle is moved from O to P its potential energy is increased by $- kT \cdot log(n/n_0)$. If the particles are electrons and the field of force is electric and derived from a potential having values V at P and V_0 at O, then of course this change in potential energy is expressed by $e(V - V_0)$.

Boltzmann's equation is so deeply rooted in modern physics, that it seems strange and suspicious that the new statistics should substitute another but it does. The reason for the innovation stands out very clearly in (127) when the absolute-zero extreme of the Fermi distribution is applied. Owing to the dependence of $\bar{v}$ on n, *owing to the interrelation between average speed and number per-unit-volume* which

distinguishes a system conforming to the new statistics, the second term in (127) is no longer proportional to dn/dx. Instead, we have:

$$n\bar{v} = \frac{3}{4}nv_m = \frac{3hn}{4m}\left(\frac{3n}{4\pi G}\right)^{2/3}; \qquad \overline{v^{-1}} = 3/2v_m, \tag{133}$$

substituting which values into (237), we obtain:

$$eE/m = \alpha = \frac{h^2}{3m^2}\left(\frac{3}{4\pi G}\right)^{2/3} n^{-1/3}dn/dx, \tag{134}$$

and integrating:

$$-E(x - x_0) = +(V - V_0) = \frac{h^2}{2me}\left[\left(\frac{3n}{4\pi G}\right)^{2/3} - \left(\frac{3n_0}{4\pi G}\right)^{2/3}\right]. \tag{135}$$

This is the formula which supplants Boltzmann's equation.

Consulting (71), we see that (135) may be rewritten thus:

$$e(V_1 - V_0) = (W_i)_1 - (W_i)_0, \tag{136}$$

which is to say: if there is equilibrium between two samples of electron-gas, both being at absolute zero and distributed according to the Fermi law, and the fastest electrons of the two having values of kinetic energy W_{i1} and W_{i0} respectively—then there is a potential-difference between the two, such that if the fastest electron of either group were to cross over to the other, its kinetic energy on arrival would be equal to that of the fastest electron of the group which it joins. So stated, the proposition is easy to remember, and one might even come to think it obvious.

Consider now a pair of pieces of different metals, in contact with one another. One may conceive that they are welded together by an alloy in which the proportion of either varies continuously from zero to one hundred per cent, if one feels the need for a mathematical continuity. If the two pieces were separate, the number of electrons per unit volume would probably not be the same for the two; certainly it is not the same if the number of electrons is equal to the number of atoms per unit volume, for this varies from metal to metal. If the process of welding the metals together does not alter the concentration of the electrons in either at points remote from the junction, then a potential-difference given by (132) or (135)—according as the old or the new statistics is the proper one—must arise between the metals. Taking Sommerfeld's example of potassium and silver: if in unit volume of each of these metals there are as many electrons as atoms, and if this state of affairs continued when the two are welded

together, then between the interiors of the metals across the weld there must be a potential-difference of 4.2 volts,[26] potassium being negative. This figure is calculated by the formula (135); the classical formula gives a value considerably lower, about 0.04 volt. This contrast is characteristic. Both the new and the old statistics associate an internal or intrinsic potential-difference with a difference in electron-concentration, and *vice versa;* but the amount of the P. D. associated with a given pair of concentrations is by no means the same by the two theories; and in actual cases, the new statistics gives much the larger amount.

Though it is not actually possible to measure the potential in the interior of a metal, there are phenomena which indicate that between

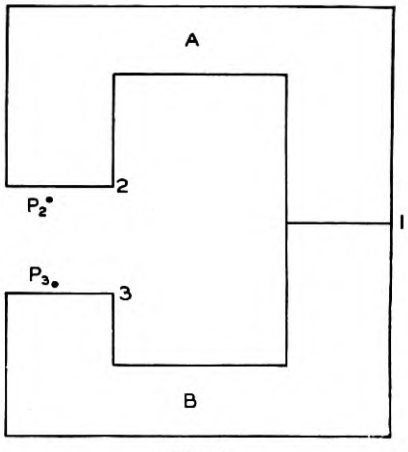


Fig. 2.

two metals touching one another, or between two parts of a metal maintained at different temperatures, there is a difference of potential. These are the *thermoelectric* phenomena—Peltier effect, Thomson effect, thermal electromotive force. The internal potential-gradient reveals itself through the fact that when an electric current is sent through the region where it exists, the rate of generation of heat departs from that which is calculated by Joule's law. We must therefore apply the statistics—this will be the last application which I shall consider—to the problem of evaluating the transport and the generation of heat in an electron-gas, in which the distribution-function is perturbed by an electric field and simultaneously by either of the two other influences—varying temperature, varying concentration of electrons—which we have heretofore considered separately.

[26] Sommerfeld originally computed 5.7 volts, having put $G = 1$; the value 4.2 corresponds to $G = 2$.

Before undertaking this I had better dispel any notion that the "contact" or "Volta" potential-difference between a pair of metals is the measure of the P. D. between their interiors for which we have just been deriving theoretical expressions. It is in fact a measure of something else, as one sees by examining an arrangement like that of Figure 2, where A and B signify pieces of two metals which are in contact at 1, and face one another across a gap between 2 and 3. Consider an electron anywhere inside A, and estimate the potential-barriers which it must cross in order to arrive at the point P_2 just outside of the boundary 2, and also those which it must cross in order to pass through the metal B and reach the point P_3 just outside of the boundary 3. Recalling the symbols and the relations introduced in the section on thermionics, one sees that there is a potential-difference between P_2 and P_3 given by the expression

$$(W_{aA} - W_{aB}) - (W_{iA} - W_{iB}) = b_A - b_B. \qquad (137)$$

This is the contact potential difference; and we see that if the new statistics is correct, it is equal (at the absolute-zero limit) to the difference between the values, for the two metals in question, of that quantity b which appears in Richardson's equation and used to be regarded as the surface work-function. By the old statistics, it differs from $(b_A - b_B)$ by the amount of the internal potential-difference between the metals across the junction 1. Perhaps this difference between the consequences of the two theories could be tested by experiment.

Theory of the Thermoelectric Phenomena

We turn now to the problem of evaluating the rate of generation of heat in a metal through which an electric current is flowing, and in which (according to these theories) there is an intrinsic electric potential-gradient due to a temperature-gradient, or to a gradient of electron-concentration, or both together. The process leads to formulæ which can be tested by experiment, furnishing thus some additional ways of finding out whether these ideas of the new statistics, of the perturbed distribution-function and of the internal electric field are justifiable.

The expression for the rate of generation of heat per unit volume in a conductor traversed by currents of electricity and heat flowing along the axis of x and having current-densities J and W respectively, is $(JE - dW/dx)$. Here E stands for the electric field—not in general for the applied electric field alone, but for the sum of this and the hypothetical internal field. I denote the corresponding acceleration

by α, as before, and the rate at which heat is generated per unit volume by r; then:

$$-r = (m\alpha/e)J - dW/dx. \tag{138}$$

The current-density of heat is given by the formula (115), which I repeat:

$$W = \tfrac{1}{2}m \int v^4 \cos^2\theta g d\tau. \tag{139}$$

Here g stands as always for the non-isotropic perturbation-term in the distribution-function. This and the acceleration α are to be determined from the two equations,

$$\alpha(df_0/d\xi) + \xi(df_0/dx) = (v\xi/l)g = (v^2 \cos\theta/l)g, \tag{140}$$

$$J/e = \int \xi^2 g d\tau = \int v^2 \cos^2\theta g d\tau, \tag{141}$$

which are the same as (117) and (118) except that the electric current is no longer set equal to zero.

Multiply both sides of (140) by $\frac{1}{2}mv^2 \cos\theta$, and integrate over the entire velocity-space. The integral of the right-hand member is W/l; developing the integral of the left-hand member, we find:

$$W = \tfrac{1}{6}ml(-4\alpha n\bar{v} + d(\overline{nv^3})/dx). \tag{142}$$

Multiply both sides of (140) by $\cos\theta$, and integrate over the entire velocity-space; the integral of the right-hand member is J/le; developing the integral of the left-hand member, we get the equation of which (127) was a special case, to wit:

$$-2\alpha\overline{nv^{-1}} + d(n\bar{v})/dx = 3J/el. \tag{143}$$

Evidently these equations suffice to translate (138) into an expression for r in terms of the mean free path, the universal constants, and the averages of various powers of v.

Postulating the Maxwell-Boltzmann distribution-function for f_0; working out the expressions for α and for dW/dx, and importing the formulæ for σ (equation 114a) and κ (equation 119), one finds:

$$\begin{aligned} J(m\alpha/e) &= J \cdot \frac{1}{2}\frac{k}{e}\frac{dT}{dx} - \frac{J^2}{\sigma} \\ dW/dx &= \frac{d}{dx}\left(\kappa \frac{dT}{dx}\right) + 2J\frac{k}{e}\frac{dT}{dx}, \end{aligned} \tag{144}$$

so the value predicted for the rate of generation of heat per unit volume amounts to this:

$$r = +\frac{J^2}{\sigma} + \frac{d}{dx}\left(\kappa \frac{dT}{dx}\right) + \frac{3}{2}\frac{k}{e}\frac{dT}{dx} J. \tag{145}$$

The first term is obviously the Joule heat; the second is not directly a consequence of this current-flow, as it would occur whatever the agency which set up the temperature-distribution in question. It is the third term which concerns us; this is a "reversible heat," proportional to the first power of the current, so that when the current flows in one sense heat is absorbed and when it is reversed heat is evolved. The sign is such, that heat is absorbed when the electrons are flowing towards the hotter part of the metal; the magnitude is such, that as the electrons move onward they acquire just enough energy to raise their temperature to that of the regions which they enter. The coefficient of this term therefore represents the specific heat of the electron-gas, which is the same as that of any other monatomic gas when referred to equal numbers of particles.

Adopting instead the Fermi distribution, and inserting into (142) and (143) the values of $\bar{v}$ and $\overline{v^{-1}}$ and $\overline{v^3}$ prevailing at absolute zero, we find on making the substitutions in the expression (138) for r that the terms containing the first power of J balance one another out. This might have been expected; for we have just seen that these terms form a sum which is proportional to the specific heat of the electron-gas; and if this result may be extended to an electron-gas conforming to the Fermi distribution, then since the specific heat vanishes at zero so also must this "reversible heat." Working through the second approximation, Sommerfeld found that the net coefficient of the term in J in the expression for r is in fact proportional to the specific heat of the electron-gas, being therefore proportional to T, and given by the formula:

$$\frac{2\pi^2}{3}\frac{mk^2}{eh^2}\left(\frac{2\pi G}{3n}\right)^{2/3} T. \tag{146}$$

Now it is a fact of experience that when an electric current flows along a uniform wire of uneven temperature, heat is generated at a rate which involves a term proportional to the current and which changes sign when the current changes sense. This "Thomson heat," like the maximum value which experiments allow us to admit for the specific heat of the electron-gas, has always been much smaller than the value which the classical statistics requires provided that the free

electrons are about as numerous as the atoms. For the Thomson heat as for the specific heat, the new statistics sharply reduces the amounts demanded—to about one per cent of those on which the classical theory insists, at room-temperature that is to say and assuming always that the free electrons are equal in number to the atoms. Agreement in order of magnitude is now attained, and for some metals the advantage is possibly greater; there are indications, too, that the Thomson heat is proportional to T over wide ranges of temperature.

Finally we consider the "Peltier heat"—a term proportional to the current and changing sign when the current changes sense, observed when there is a flow of electricity across a weld or area of contact between two metals. This is clearly to be interpreted as a term in the first power of J, occurring when into the combination $J(m\alpha/e)$—all that remains of the expression (138) for r, when the gradient of temperature is annulled—we substitute the value of α derived from (143) with the assumption that n varies continuously across the weld from the value appropriate to the one metal to the value appropriate to the other. Using the classical statistics, we find that there is such a term; denoting by n and n_0 the electron-concentrations in the two metals, we find for its value:

$$(kT/e) \log (n/n_0)J. \tag{147}$$

Its value for unit current is obviously the intrinsic potential-difference between the metals. Using instead the new statistics, we find that at the absolute zero there is no such term; we must proceed to the next approximation, doing which, Sommerfeld obtained the expression:

$$\frac{2\pi^2}{3}\frac{m(kT)^2}{eh^2}\left[\left(\frac{4\pi G}{3n}\right)^{2/3} - \left(\frac{4\pi G}{3n_0}\right)^{2/3}\right]. \tag{148}$$

Putting the current equal to unity, we find a value very considerably smaller than the intrinsic potential-difference between the metals—a fraction of a millivolt. This is the order of magnitude of the Peltier heat as it is actually observed in many cases. Curiously enough, this fact by itself is in accord with both the theories. By the classical statistics, the intrinsic potential-difference between two metals is generally small, and the Peltier heat for unit current gives its value directly; by the Fermi statistics, the intrinsic potential-difference is generally large, but the Peltier heat for unit current is only a small fraction of it.

Omissions

Among the subjects omitted from this article there are several of much interest, which the reader may trace from the annexed bibliography; in particular:

Sommerfeld's theory of the Hall effect;

Houston's extension of Sommerfeld's theory of intrinsic potential difference, including especially an explanation of the Peltier heat arising when current flows between two differently-oriented crystals of a single substance;

Bloch's use of the new methods of quantum mechanics to make allowance for the influence of the atoms on the conduction-electrons;

Fermi's application of statistical methods to the problem of determining the distribution of electrons in the individual atom;

Fuerth's work on the fluctuations in the new statistics.

The bibliography will indicate other interesting advances in a variety of problems.

BIBLIOGRAPHY

J. M. Adams: Value of W_a for silver. Zeits. f. Physik **52**, p. 882 (1929).
W. Anderson: Zeits. f. Physik **50**, p. 874–877 (1928). Astrophysical applications.
H. M. Barlow: Phil. Mag. (7) **7**, pp. 458–470 (1929). A criticism of the electron theory of metals.
R. S. Bartlett: Proc. Roy. Soc. **A121**, pp. 456–464 (1928). Thermionic and cold discharge.
E. S. Bieler: Journ. Franklin Inst. **206**, pp. 65–82 (1928). General account of new statistics of electron-gas.
H. F. Biggs: Nature **121**, pp. 503–506 (1928). General.
F. Bloch: Zeits. f. Physik **52**, pp. 555–559 (1928). Quantum mechanics of electrons in crystal lattices.
F. Bloch: Susceptibility and change of resistance in magnetic field. Zeits. f. Physik **53**, pp. 216–227 (1929).
Bose: Zeits. f. Physik **26**, pp. 178–181 (1924). First paper on the Bose statistics.
L. Brillouin: C. R. **184**, pp. 589–591 (1927); **188**, pp. 242–244 (1929).
P. A. M. Dirac: Proc. Roy. Soc. **A112**, pp. 661–677 (1926).
J. W. M. DuMond: Proc. Nat. Acad. Sci. **14**, pp. 875–878 (1928); Phys. Rev. (2) **33**, pp. 643–658 (1929). Scattering of x-rays by the free electrons.
C. Eckart: Zeits. f. Physik **47**, pp. 38–42 (1928). Contact P. D.
A. Einstein: Berliner Sitzungsberichte, 1924, pp. 261–267; 1925, pp. 3–17. Application of Bose statistics with modification to material gases.
E. Fermi: Zeits. f. Physik **36**, pp. 902–912 (1926). Fermi statistics.
E. Fermi: Zeits. f. Physik **48**, pp. 73–79 (1928). Application of statistics to electrons within atoms.
E. Fermi: Zeits. f. Physik **49**, pp. 550–554 (1928). Statistical evaluation of the *s*-levels.
E. Fermi: *Quantentheorie und Chemie* (Hirzel, 1928), pp. 95–111. Application of statistics to problems of atomic structure.
V. Fock: Zeits. f. Physik **49**, pp. 339–357 (1928). Dirac's statistical equation.
R. H. Fowler: Proc. Roy. Soc. **A117**, pp. 549–552 (1928). Thermionics.
R. H. Fowler, L. Nordheim: Proc. Roy. Soc. **A119**, pp. 173–181 (1928).
R. H. Fowler: Proc. Roy. Soc. **A120**, pp. 229–232 (1928). Photoeffect.
R. H. Fowler: Proc. Roy. Soc. **A122**, pp. 36–49 (1929). Thermionic constant A.
R. H. Fowler: Monthly Notices Royal Astron. Soc. **87**, pp. 114–122 (1926). Astrophysical application.
J. Frenkel: Zeits. f. Physik **49**, pp. 31–45 (1928). Theory of magnetic and electric properties of metals at absolute zero.
J. Frenkel, N. Mirolubow: Zeits. f. Physik **49**, pp. 885–893 (1928). Theory of metallic conduction by wave mechanics.
J. Frenkel: Zeits. f. Physik **50**, pp. 234–248 (1928). Cohesion.
R. Fuerth: Zeits. f. Physik **48**, pp. 323–339 (1928), and **50**, pp. 310–318 (1929). Fluctuations in the new statistics.

G. E. Gibson, W. Heitler: Zeits. f. Physik **49**, pp. 465–472 (1928). Chemical constant in the new statistics.
E. H. Hall: Proc. Nat. Acad. Sci. **14**, pp. 366–380, 802–811 (1928). Arguments against the electron-gas theory.
W. V. Houston: Zeits. f. Physik **47**, pp. 33–37 (1928). Cold discharge.
W. V. Houston: Zeits. f. Physik **48**, pp. 449–468 (1928). Conduction and thermoelectric effects in metals.
W. V. Houston: Phys. Rev. (2) **33**, pp. 361–363 (1929). Cold discharge.
O. Klein: Zeits. f. Physik **53**, pp. 157–165 (1929). Reflection of electrons at a potential-discontinuity by Dirac's dynamics.
E. Kretschmann: Zeits. f. Physik **48**, pp. 739–744 (1928). Continuation of Sommerfeld's work.
J. E. Lennard-Jones: Proc. Phys. Soc. Lond. **40**, pp. 320–537 (1928). General account.
J. E. Lennard-Jones, H. J. Woods: Proc. Roy. Soc. **A120**, pp. 727–735 (1928). Electron-gas theory with modifications to allow for atoms in metals.
A. C. G. Mitchell: Zeits. f. Physik **50**, pp. 570–576 (1928). Entropy of the electron-gas.
L. Nordheim: Phys. Zeits. **30**, pp. 177–196 (1929). General report on new theory of thermionic emission, cold discharge and photoeffect. Earlier papers: Zeits. f. Physik **46**, pp. 833–855 (1928); Proc. Roy. Soc. **A121**, pp. 626–639 (1928); **A119**, pp. 173–181 (1928) (with R. H. Fowler).
L. Nordheim: Proc. Roy. Soc. **A119**, pp. 689–698 (1928). Alternative way of deriving the distribution-laws.
L. Nordheim: Naturwiss. **16**, pp. 1042–1043 (1928). Conductivity of alloys.
J. R. Oppenheimer: Proc. Nat. Acad. Sci. **14**, pp. 363–365 (1928). Cold discharge.
W. Pauli: Zeits. f. Physik **41**, 81–102 (1927). Paramagnetism.
W. Pauli: *Probleme der modernen Physik* (Hirzel, 1928), pp. 30–45. H-theorem.
F. Rasetti: Zeits. f. Physik **49**, 536–549 (1928). Statistical evaluation of the M-levels.
O. K. Rice: Phys. Rev. (2) **31**, pp. 1051–1059 (1928). Electrocapillarity.
O. W. Richardson: Proc. Roy. Soc. **A117**, pp. 719–730 (1928). Cold discharge.
L. Rosenfeld, E. E. Witmer: Zeits. f. Physik **47**, pp. 517–521 (1928). The radiation-gas.
L. Rosenfeld, E. E. Witmer: Zeits. f. Physik **49**, pp. 534–541 (1928). Refractive indices of metals for electron-waves, values of W_a.
R. Ruedy: Phys. Rev. (2) **32**, pp. 974–978 (1928). Dispersion of metals.
E. Schroedinger: Phys. Zeits. **27**, pp. 95–101 (1927).
A. Sommerfeld: Zeits. f. Physik **47**, pp. 1–32, 43–60 (1928). The fundamental article on the application of the Fermi statistics to electrons in metals. Preliminary note in Naturwiss. **15**, pp. 825–832 (1927).
A. Sommerfeld: Naturwiss. **16**, pp. 374–381 (1928). Extension of the article just listed.
A. T. Waterman: Phil. Mag. (7) **6**, pp. 965–970 (1928). Dependence of conductivity on pressure.
A. T. Waterman: Proc. Roy. Soc. **A121**, pp. 28–41 (1928). Cold discharge.
G. Wentzel: *Probleme der modernen Physik* (Hirzel, 1928), pp. 79–87. Photoeffect.

Physical Properties and Methods of Test For Some Sheet Non-Ferrous Metals

By J. R. TOWNSEND and W. A. STRAW

This paper covers an investigation which was undertaken to secure a simple and reliable method of test for sheet non-ferrous metals. An account of the early development work leading to the adoption of the Rockwell hardness tester for a preliminary inspection of sheet metals and the tensile test as the basic test to be referred to in case the Rockwell test results were near to or outside the established Rockwell limits for a given lot of material was published in 1927.[1] The continuation of this work including establishment of test limits for four grades of brass and for two grades each of nickel silver and phosphor bronze will be published this year.[2]

This present paper describes the development work reported by these two papers.

Considerable attention has been given to the Rockwell tester, which, as a result of this work, has been found satisfactory for use as a specification instrument for brasses 0.020 in. and thicker, when used under standardized methods of test and calibrated with standard test blocks. Other tests such as the bend test, ductility test and other hardness tests have been studied but further development work is necessary.

The Rockwell hardness and tensile strength limits are given for four alloys of brass, and two alloys each of nickel silver and phosphor bronze. The physical properties of the rolling series upon which these limits were based are presented as well as experience data obtained on shipments of commercial material. The limits for brass alloys A, B and E are considered final, but the limits for the other metals are tentative until more complete experience is available.

Grain size limits are given for annealed brass and nickel silver sheet. For inspection purposes the grain size is estimated by comparison with the standard photomicrographs reproduced in the 1929 report of Committee E4 of the A. S. T. M. on Metallography.

Refinements in the calibration of the Rockwell tester are given, as well as the development of testing technique. An experimental model of a motor-driven bend test machine is described.

THE investigation which this paper covers was undertaken because of the need for a simple method of test which would afford a more definite determination of the properties of thin-sheet non-ferrous metals than any which has yet been developed. Large quantities of

[1] "Physical Properties and Methods of Test for Sheet Brass," by H. N. Van Deusen, L. I. Shaw, and C. H. Davis, *Proceedings of the American Society for Testing Materials*, 1927.

In addition to the authors the following took an important part in the investigational work. Messrs. H. N. Van Deusen C. H. Greenall and W. S. Hayford, of the Bell Telephone Laboratories, Messrs. W. H. Bassett, R. M. Tree, Alden Merrill and G. S. Mallett of the American Brass Co., and Messrs. L. I. Shaw, M. D. Helfrick H. F. Culver and G. R. Brown of the Western Electric Co. Messrs. N. E. Newton and W. H. Eastlake of the Northern Electric Co., Montreal, participated in the conferences which were held as the work progressed.

[2] "Physical Properties and Methods of Test for Sheet Non-Ferrous Metals," by J. R. Townsend, W. A. Straw and C. H. Davis, presented before A. S. T. M., June 1929.

brass, nickel silver and phosphor bronze are used in the manufacture of telephone apparatus as structural members, springs, and bearings. Because of space limitations, the parts are necessarily small; many are formed into irregular shapes; spring parts must maintain accurate adjustments and have long fatigue life; certain other parts must resist wear. All requirements are steadily becoming more exacting because of the increasing complexity of telephone systems and it, therefore, became necessary to insure more closely the uniformity of the material used in the apparatus.

The methods of test developed to check the uniformity of material are of equal concern to both consumer and supplier. Both are interested in a test rapid enough for use in inspection work and also so simple that specially trained men are not necessary for the actual testing. Due to the large amount of material that must be inspected it is necessary that any tests used require little time to apply and be adaptable to modern production methods. Furthermore, close agreement in test results is necessary. Since no test method was available fulfilling these requirements, it became necessary either to develop new methods of test or to modify existing methods. In addition, limits were desired that would represent the best quality of material obtainable consistent with commercial mill practices.

As a result of this need a cooperative program of tests was laid out which would lead to the drafting of requirements on thin-sheet metals. Those cooperating in this program were associated in the production and use of such material, and each member had already done preliminary work which evidenced his interest in the problem. The group was limited to one producer and one consumer in order that the work might be expedited, and the results that have been obtained justify the plan.

The results of the investigation are given herein for such value as they may have to the industry. It is hoped that this work may act as a stimulus to others interested in this problem so that ultimately definite standards may result, including the refinements demanded by modern industry.

The results of the study of the methods of test suitable for high and clock brass sheet of the order of 0.020 in. and thicker are presented and it is shown that the Rockwell tester gives the most reliable information of the several hardness testing machines available at the present time. Methods of operation of the Rockwell hardness tester were worked out by the cooperating laboratories and a method of carrying on tension tests was developed. The limitations in the use of the Rockwell hardness tester for sheet metal are pointed out and

the need for the development of a tester capable of higher precision in the harder materials and also capable of making hardness tests on material less than 0.020 in. thick is emphasized.

The tensile strength limits are determined by the following procedure. The tension test results of the rolling series are plotted against the actual percentage of reduction by cold rolling. Two limiting curves are drawn in giving the minimum and maximum tensile strength for all reductions covering the range of commercial anneals. These two curves are based on years of experience with these metals and also on the rolling series. The tensile strength limits were taken from these curves for the theoretically correct reduction for each temper. Rockwell hardness tests are made on the grip ends of the tension test specimens, thus establishing the Rockwell hardness-tensile strength relationship for the material. Having established the tensile strength-reduction relationship, the corresponding hardness values are obtained from the Rockwell hardness-tensile strength curve. These limits are then subjected to trial on a large number of shipments of material.

The Rockwell test is considered a preliminary inspection test and is mainly useful because of its economy of time and material. The tension test, on the other hand, is considered the test upon which the acceptance or rejection of the material is based. In practice, material within the Rockwell hardness limits is accepted unless the hardness reading is near or outside the hardness limits, in which case a tension test is made.

This paper covers high and clock brass sheet and four other alloys of brass. One of these contains less lead than clock brass and is designed for use where a material combining moderate drawing and cutting properties is desired. Another has a nominal composition of 72 per cent of copper and 28 per cent of zinc. Grain size requirements are given for two others mainly used for drawing purposes. One of these consists of nominally 85 per cent of copper and 15 per cent of zinc and the other of 75 per cent of copper and 25 per cent of zinc. This paper also covers two nickel-silver alloys and two phosphor-bronze alloys.

The Rockwell hardness and tensile strength limits given for all of these alloys other than high, clock and alloy E brass may be considered tentative in view of the limited experience had with commercial shipments of materials purchased in accordance with these limits up to the present time. The requirements for high, clock and alloy E brass are considered final. Alloy E brass has a lead content midway between high and clock-brass.

Method of Handling Work

The round-robin tests on the Rockwell hardness tester, standard test blocks and calibration of the machines were made by the American Brass Co. (Waterbury, Buffalo and Kenosha Branches), Bell Telephone Laboratories, and the Western Electric Co. The tension tests were also made by these laboratories. All of the rolling series were manufactured by the American Brass Co. Experience data on shipments of non-ferrous metals were obtained jointly by the American Brass Co. and the Western Electric Co.

Methods of Test

Tension Test:

As stated previously, the tension test is considered the basic mechanical test for cold worked materials. Complete data on tensile properties were therefore obtained on the rolling series. All tests were made in accordance with the standard testing procedure developed for use in the Bell System and given in Appendix II. This method has been in use over 4 years and has been found satisfactory to all concerned. Three tension test specimens were made from each sample. The tensile strength, proportional limit, percentage of elongation in 2 in. and modulus of elasticity were determined for each sample of the high-brass rolling series. The specimen 14 in. in length was used which allowed the use of an 8-in. gage length Anderson [3] extensometer. The shorter specimen with the 2-in. gage length was used for the clock brass and all later work. The tests were made on an Amsler tension testing machine. The machine was carefully calibrated, the load indications being correct to less than $\frac{1}{2}$ of 1 per cent.

The tensile strength and percentage of elongation on the high-brass sheet rolling series are plotted in Fig. 8 against per cent reduction by rolling. Nearly a straight line relation exists between the tensile strength and percentage reduction by rolling and the values are grouped closely about this line. The curve is plotted to show the average result, but separate curves could readily be drawn for each thickness. This test also has little sensitivity in the harder tempers where the curve becomes asymptotic. The proportional limit is difficult to measure with any degree of accuracy on account of the personal factor involved in interpreting curves. The results in this case showed no definite relation to any of the other values.

In order to evaluate the test results, it is necessary to establish control limits to determine whether any variations in the data

[3] H. A. Anderson, "Tension Tests of Thin Gage Metals and Ligh Alloys," *Proc. A. S. T. M.*, Vol. 24, Part II, p. 990 (1924).

are significant. By significant variations are meant variations that can be assigned to definite sources, such as measuring errors, defects in the metal tested, testing errors, etc. For the case under consideration, a method of analysis was employed which allows for small sample numbers. The method used is described elsewhere.[4]

Figure 1 shows an engineering analysis of the tensile strength

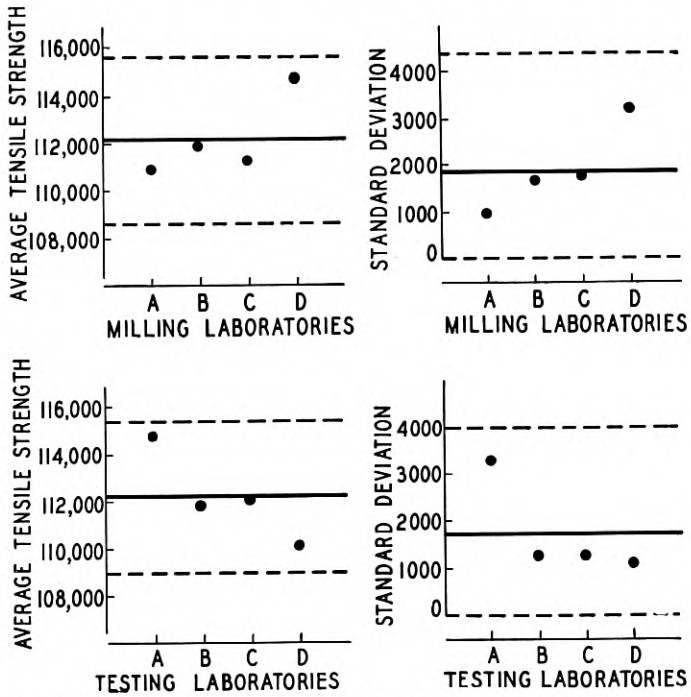


Fig. 1—Statistical Analysis of Round-Robin Tension Tests on Alloy B Nickel Silver.
(*a*) Points are average of all specimens milled at laboratories indicated, and tested by the four laboratories.
(*b*) Points are average of all specimens tested at laboratories indicated, and milled at the several laboratories.

results for alloy B nickel silver plotted in two ways.[5] The data in Fig. 1 (*a*) were obtained by computing, for each of the laboratories, the average tensile strength and the standard deviation [5] of all the specimens milled by that laboratory and tested by each of the four laboratories *A*, *B*, *C* and *D*. The data in Fig. 1 (*b*) were obtained

[4] See Modified Criterion No. 1 in Appendix I to the Report of Sub-Committee XV on Die-Cast Metals and Alloys appended to the report of Committee B-2 on Non-Ferrous Metals and Alloys, presented before the American Society for Testing Materials, June 1929.

[5] The methods used are those described by W. A. Shewhart in a paper on "Quality Control," *The Bell System Technical Journal*, Vol. VI, October 1927.

by computing, for each of the laboratories, the average tensile strength and the standard deviation of the specimens milled in the four laboratories and tested in each of the respective laboratories. In each instance the specimens tested or milled by an individual laboratory are grouped together. The dotted lines on the diagrams represent the control limits within which the data should fall without leaving anything to chance, which means that points falling without these control limits indicate variations due to assignable causes, such as errors in measurements, defects in the material, etc. These diagrams show, therefore, that the tensile strength results between laboratories do not reveal significant or assignable difficulties other than those which could be attributed to chance. In other words, the analysis gave no indication of the presence of assignable variations between the testing laboratories.

These specimens were prepared by cross milling the gage length with a milling cutter shaped to conform to the final shape of the specimen desired. This method, which has been described elsewhere,[6] results in the saving of time and produces specimens of a uniform character.

Scleroscope Hardness Tests:

Previous to this investigation Bell System specifications on nonferrous materials were written in terms of scleroscope hardness, but considerable trouble was encountered between the suppliers and users of metal due to difficulty in checking each other's readings. Because results could not be duplicated on two instruments, it was necessary to allow rather wide limits in each temper of material. This resulted in considerable overlapping in scleroscope limits of the tempers of materials accepted under these specifications. Experience had shown it to be impossible to make correction curves for any two instruments which would hold for any reasonable length of time, and if any replacements such as new hammers were necessary the calibration was changed.

In order to obtain more definite information as to what could be expected from the scleroscope, comparisons were made between four type "C" sclerscopes located in four laboratories using the same samples of materials. Great care was taken in preparing the samples so that they were flat and free from any dirt and roughness. The instruments were in good commercial adjustment and were employed in the usual manner, the magnifier hammer being used. A single

[6] R. L. Templin, "Methods for Determining the Tensile Properties of Thin Sheet Metals," *Proc. A. S. T. M.*, Vol. 27, 1927.

thickness of material was used in each case. The results of one of these comparison tests are shown in Fig. 2. Two other such comparisons were made incidental to the work being done to establish operating technique on the Rockwell tester, but the results shown here are typical. At least five readings were made on each sample

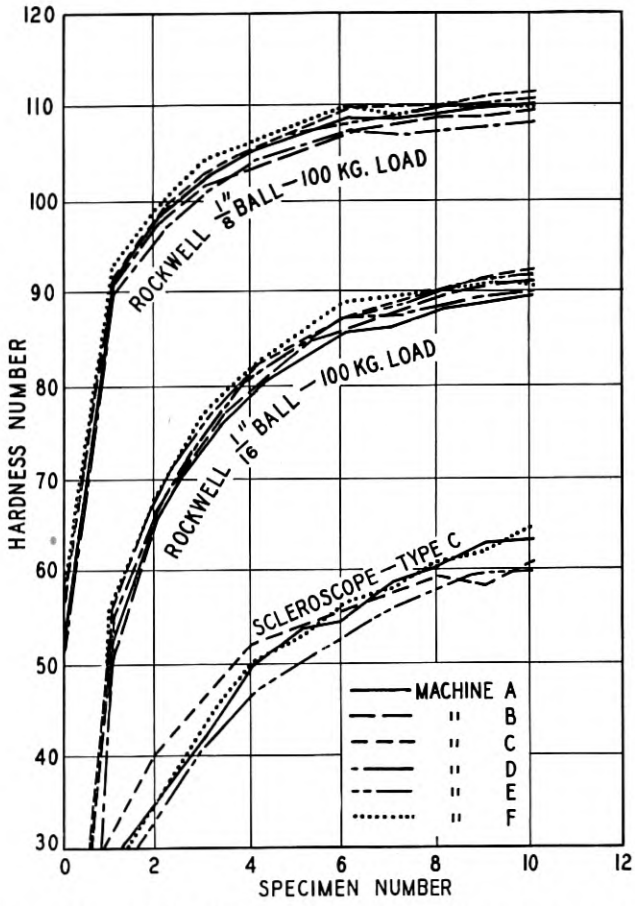


Fig. 2—Brass "Round-Robin" Series. Comparison of Rockwell and Scleroscope Testers.

and the average was used in preparing these curves. It will be noted that the readings on the different scleroscopes varied as much as eight points from each other on cold-rolled materials. The variations in readings between the machines follow no apparent law, one machine reading higher at one part of the scale and lower at another. It will also be noted from the curves that the range of the scleroscope

is much smaller than that of the Rockwell on the same samples. In other words, the sensitivity of the scleroscope is considerably less than that of the Rockwell.

An investigation was made of the type "D" recording scleroscope equipped with a universal hammer to obtain a comparison with the type "C" instruments. The conclusion reached was that there was little difference between the two instruments as regards precision. It was not considered worth while, therefore, to make elaborate comparison tests. Hence, this type of instrument was not used in the round-robin tests.

In view of the difficulties encountered in using the scleroscope, which have also been brought out by previous investigators, no further effort was made to adapt it for use as a specification instrument.

Rockwell Hardness Tests:

While the Rockwell hardness tester has been used quite extensively and with considerable success in testing steels it has been used comparatively little for testing non-ferrous metals. Our experience previous to this investigation indicated that it might prove satisfactory as a specification instrument for non-ferrous metals. Considerable work had already been done by various laboratories to determine the limitations of this machine. Furthermore, the routine testing with the Rockwell of all incoming shipments of material had been instituted for the purpose of accumulating specification data. While each individual machine seemed to give satisfactory results to the user there was little information available as to what agreement could be obtained with machines in other laboratories. Figure 3 gives results of tests on identical samples with five different machines in the laboratories of one of the participating companies before any attempt was made to eliminate mechanical irregularities in the machines. While there was considerable difference between machines, there was a probability that these differences could be reduced by carefully going over the machines and establishing technique of test.

Before making any further comparative tests, careful study was made of each machine and various mechanical irregularities were eliminated. In order to get close comparative results the ball penetrators must be in good condition, the load must be applied without impact and at approximately the same rate in different machines, the value of the load must be the same, and the method of supporting the specimen across the anvil is important. Due to the lever system of applying load and measuring the amount of penetration of the material under this load, a slight amount of friction in the bearings

may result in variations in readings which are not apparent without comparisons with standard blocks. A set of directions for calibrating the Rockwell tester was drawn up which called for a thorough check of these various features of the machine. (See Appendix I.)

Other questions had to be settled, such as the size of ball penetrator to be used and the amount of the load. The possibility of varying the load and size of ball renders the Rockwell machine capable of broad application for testing purposes. Preliminary tests made to determine the most suitable combination of ball and load to be used indicated that the standard "B" scale of the instrument ($\frac{1}{16}$-in. ball and 100-kg. load) would be the most satisfactory for brass. However, it was decided to try also a $\frac{1}{8}$-in. ball with the same load. In our

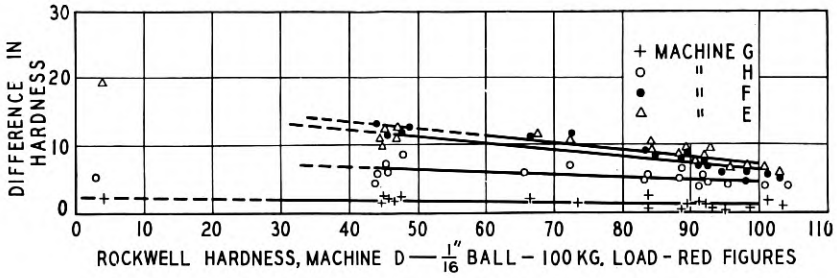


Fig. 3—Preliminary Comparison of Rockwell Testers.

later work on nickel silver and phosphor bronze it was found that the 150-kg. load gave more satisfactory results with those materials. A larger ball is desirable to get readings on annealed brass and copper but with harder materials sensitivity of reading is increased by increasing the load and retaining the $\frac{1}{16}$-in. ball.

The question of the number of thicknesses of material to be tested was also given consideration. On very thin materials there is a distinct anvil effect which can be reduced or eliminated by superposing several samples on each other. However, this method of testing soon proved to be objectionable. The Rockwell tester does not distinguish between actual penetration of the material and descent of the penetrator from any other cause, and since one hardness number of the Rockwell machine corresponds to an 0.00008-in. movement of the penetrator, great care must be taken to insure that movement of the penetrator is due only to actual penetration of the material. Unless the spaces under the penetrator between two or more layers are completely closed by the 10-kg. minor load, low readings will be obtained. With spring materials, which are frequently curved due

to coiling, this condition is often encountered. Furthermore, in very thin materials there is a side flow so that the upper sheets fail to take their share of the load and tend to curl up around the penetrator in such a manner as to prevent the proper measurement of depth. There is also a possibility of including small particles of dirt between the specimens which would result in lower hardness readings. These matters must be very closely watched. An indentation in the anvil just barely visible on a polished surface in strongly reflected light will cause erroneous readings. It is necessary, therefore, that the anvils be sufficiently hard for the purpose. During the past year, harder anvils have been furnished and greatly improved results have been obtained in the testing of thin materials.

A comparison of the methods used by the laboratories in making tests indicated that personal factors might account for some of the differences in readings. One of these personal factors which appeared to be of major importance was the amount of time allowed for the drift of the penetrator after the major load is completely applied. It was the custom of some of the participating laboratories to allow this drift to continue until it had practically ceased while others removed the major load and read the hardness as soon as possible after the major load had been completely applied. This drift is very noticeable in the softer tempers of non-ferrous metals and may amount to as much as 10 Rockwell numbers. In general, it is more noticeable with non-ferrous metals than with steels. In order to eliminate this personal factor, the practice of operating the reset mechanism which removes the major load, immediately upon the full application the major load was adopted.

Three separate rolling series were used in the round-robin tests which were made after the various machines had been checked up. In each case tests were made with both $\frac{1}{16}$ and $\frac{1}{8}$-in. balls using the 100-kg. load. All tests were made on each of six machines located in the six different laboratories. Typical results of one of these series are shown in Fig. 2 which gives the results of a brass rolling series using both the $\frac{1}{8}$ and $\frac{1}{16}$-in. ball. It will be noted that the greatest sensitivity is obtained by using the $\frac{1}{16}$-in. ball, and this was later adopted as a standard ball for testing brass. Much better agreement between Rockwell testers than between scleroscope testers is also apparent.

Rockwell data were obtained on the high-brass and clock-brass rolling series in all thicknesses and tempers using one thickness of material and the standard "B" scale. The plot of the results of these tests on high brass is shown in Fig. 6. These curves appear normal for thicknesses of 0.012 in. and above, but below this thickness the

hardness readings bear no relation to tensile strength due to the anvil effect.

The Rockwell hardness test is carried on in the same manner as given in Appendix I of this paper. All of the Rockwell hardness tests were made on the two grip ends of the tension test specimens. This was done in order to have Rockwell hardness tests and tension tests made on as nearly identical material as possible. Commercial experience with the Rockwell machine in the routine testing of sheet non-ferrous metals has indicated the importance of the following precautions.

Ball Penetrator.—Slight variations in the size and sphericity of the ball penetrators are to be expected and must be guarded against. The results of several hundred measurements made at various points on the "B" scale show that a deviation of less than 0.00002 in. in the diameter of the $\frac{1}{16}$-in. ball has no effect on the hardness readings. Balls showing a greater variation give noticeable error in hardness readings. Penetrator balls have been held within this limit.

Anvil Surface.—It is desirable that the penetrator be perpendicular to the testing surface of the anvil. It is assumed that the penetrator is operating perpendicular to the seating surface and in line with the capstan head. Any lack of perpendicularity would then be due to lack of parallelism of the seating and testing surfaces of the anvil. This is checked by a fixture designed to hold the anvil so that the specimen supporting surface might be checked. A hardened flat, 2 in. square and $\frac{7}{8}$ in. thick, having a $\frac{13}{16}$-in. hole through its center, was lapped so that its two surfaces were flat and parallel to within 0.00005 in. By placing the fixture holding the anvil on the table of an optimeter and sliding the supporting surface of the anvil under a ball-pointed feeler gage, errors in flatness and parallelism may be measured to 0.00005 in. Anvils whose surfaces were found to be plane to within 0.0001 in. were considered satisfactory.

Machine Errors.—Observations on four Rockwell machines at the Western Electric Co., one of which handles approximately 10,000 tests per month and all of which were overhauled and calibrated for their entire range once a week, showed that changes in calibration occurring in this interval were less than one hardness number. Before beginning a series of readings the correction to be applied was determined by taking readings on a standard block. The average change on all three scales was about one hardness number. Variations from calibration were rarely greater than 1.5 numbers unless the penetrator was damaged or the instrument was out of adjustment due to misuse.

The effectiveness with which standard test blocks may be used in bringing Rockwell testers into agreement is shown by the following calibration:

	Approximate Hardness, "B" Scale, 1/16-in. Ball, 100-kg. Load (Red Figures)		
	25	60	80
Before overhauling and checking the entire scale, average difference in hardness readings between three testers	6.0	3.7	3.4
After overhauling and checking the entire scale, average difference in hardness readings between three testers	0.9	1.2	0.9

Calibration of Test Blocks.—It was first thought that a Rockwell hardness machine would hold its adjustment over a long period if carefully maintained and could serve as a standard machine to be used for the calibration of test blocks. More complete experience has shown, however, that wear of the operating parts and slight variations in friction caused the "standard machine" to vary slightly in its readings. Naturally, the calibration of standard test blocks under these conditions is unsatisfactory because it might result in the establishment of several sets of standard blocks, leading to confusion in the application of the Rockwell hardness test limits to commercial material.

A new method of calibrating the Rockwell hardness tester was devised. This method consisted in preparing a set of test blocks covering the range of hardness of the high-brass and clock-brass rolling series. These blocks were checked by the various cooperating laboratories as well as the manufacturers of the Rockwell hardness tester, namely, the Wilson-Maeulen Co. These blocks were considered the basic standards for the Rockwell tests. Sub-standard blocks were calibrated in comparison with these standard blocks and placed in use by the various cooperating laboratories. The high-brass rolling series and the clock-brass rolling series were then retested for Rockwell hardness using the new calibration. It was found that the hardness readings differed one to two points from the values previously determined for several tempers. At the same time experience with commercial shipments of material showed that a shift was necessary in the Rockwell hardness-tensile strength relationship. This verified our conclusions with regard to the adoption of the new method of calibration.

After the establishment of the standard brass test blocks it was

seen that blocks would also be required to cover the nickel-silver and the phosphor-bronze alloys. To prepare new series of blocks especially for these materials would have resulted in a wide variety of blocks and would have been found burdensome to the laboratories. Consequently, a single series of test blocks covering the brass, nickel-silver, and phosphor bronze alloys was selected. This series of test blocks started with the original blocks to which were added a few additional blocks made necessary in order to cover the range of the nickel-silver and phosphor-bronze alloys. The data upon which the calibration of these blocks was based have been submitted to the Section on Indentation Hardness of the A. S. T. M. Committee E-1 on Methods of Testing.

Hardness of Thin Sheet.—The Rockwell hardness test has certain limitations in its application to the testing of thin sheet stock. Material less than 0.020 in. thick gives hardness readings different from thicker sheet of the same temper. Referring to the curves shown hereinafter of Rockwell hardness plotted against tensile strength for materials thinner than No. 24 B. & S. gage the points fall below the curve. This apparently is due to lack of support of the metal about the penetrator and consequently a low reading is given. For still thinner material the penetrator passes nearly through the metal and the hardness reading recorded is inaccurate due to the effect of the supporting anvil in addition to lack of support of the metal.

Cleaning Anvil and Specimen.—Inasmuch as the Rockwell hardness tester measures hardness in terms of penetration, any movement of the penetrator affects the hardness reading. In other words, if the metal has a roughened surface or if the anvil is not polished smooth the metal will flow under the high unit pressure involved and this will cause the Rockwell hardness reading to be lower than its true value. It is considered necessary therefore that the anvil should be polished flat and the material tested should be reasonably free from surface imperfections and oxide film. Table III shows the effect on the hardness readings of polishing the anvil.

In addition to the need of polishing the anvils, if close agreement is to be had, some refinement is also needed in the use of the standard test blocks. It is difficult to obtain test blocks that will not show a variation in hardness. It has been customary, therefore, in calibrating test blocks to take five readings, one in each corner of the test block and one in the center. These readings are made on a machine which has been calibrated with the standard test blocks both before and after calibration of the secondary blocks. Experience in calibrating the Rockwell hardness machine with standard test blocks has empha-

TABLE III—ROCKWELL HARDNESS TESTS ON ALLOY G BRASS SHEET SHOWING EFFECT OF USING UNPOLISHED AND POLISHED ANVIL

Sample			Unpolished Anvil, "B" Scale, 1/16-in. Ball, 100-kg. Load (Red Figures)				Polished Anvil, "B" Scale, 1/16-in. Ball, 100-kg. Load (Red Figures)			
B. & S. Gage	Temper, B. & S. Numbers Hard	Tensile Strength, lb. per sq. in.	Average of 15 Readings	Maximum	Minimum	Range	Average of 15 Readings	Maximum	Minimum	Range
No. 20	2	64,000	73.3	74.0	72.8	1.2	72.9	73.6	71.8	1.8
	4	77,800	83.5	83.8	82.9	0.9	83.5	83.7	83.0	0.7
	6	85,400	87.0	87.6	86.2	1.4	86.7	86.7	86.0	0.7
	8	93,800	90.3	90.0	90.0	0.8	90.7	90.8	90.3	0.5
	10	99,000	92.1	92.7	91.3	1.4	92.7	92.8	92.6	0.2
No. 22	2	65,100	74.2	75.0	73.6	1.4	72.6	73.5	71.3	2.2
	4	79,100	83.4	83.8	83.2	0.5	83.8	84.2	83.0	1.2
	6	88,200	88.3	88.9	88.0	0.8	88.0	88.3	87.8	0.5
	8	99,300	90.0	90.2	89.7	0.5	90.1	90.2	89.8	0.4
	10	98,200	91.8	92.2	91.7	0.5	92.2	92.4	91.8	0.6
No. 24	2	61,000	67.6	68.8	67.1	1.7	63.9	65.3	61.6	3.7
	4	81,600	84.3	84.7	83.7	1.0	82.7	83.0	82.2	0.8
	6	91,400	89.0	89.7	88.7	1.0	87.8	88.0	87.6	0.4
	8	96,600	91.0	91.2	90.7	0.5	89.7	89.9	89.3	0.6
	10	97,800	91.9	92.7	91.2	1.5	91.2	91.5	90.9	0.6
No. 26	2	58,100	64.4	65.9	63.0	2.9	58.4	59.4	56.8	2.6
	4	71,800	77.3	77.8	76.9	0.9	75.1	75.5	74.4	1.1
	6	88,600	87.4	88.0	86.7	1.3	86.2	86.5	85.8	0.7
	8	95,600	90.2	90.7	89.9	0.8	89.5	89.9	89.2	0.7
	10	99,100	91.2	91.7	90.8	0.9	91.0	91.3	90.6	0.7
No. 28	2	61,100	65.4	66.9	64.2	2.7	55.7	57.6	52.3	5.3
	4	71,400	77.0	77.9	76.0	1.7	72.1	73.3	71.1	2.2
	6	83,600	83.2	83.8	82.1	1.7	81.2	81.5	80.7	0.8
	8	94,700	89.6	90.3	89.1	1.2	88.2	88.5	88.0	0.5
	10	100,100	91.5	92.0	90.5	1.5	90.4	90.8	90.1	0.7

sized the need for keeping these blocks clean and free from oxide, grease or other accumulated matter. This applies with equal force to the test specimens. Various methods of cleaning the test blocks have been tried. A successful and convenient method is to rub the surfaces of the test block by hand with chiffon velvet or other lap material dipped in tripoli and water. Test blocks cleaned in this way do not show a greater variation in readings than that of the original test for uniformity.

A variation in hardness must be permitted in these five readings and on the basis of experience with a large number of blocks it is considered that test blocks should have no greater variation than as follows: When tested using the "B" scale, $\frac{1}{16}$-in. ball, 100-kg. load (red figures); for blocks under 40, 3 points variation is allowed; for

blocks from 40 to 60, $2\frac{1}{2}$ points; and for blocks over 60, $1\frac{1}{2}$ points. Acceptance or rejection of blocks is based on these limits, except where a minus reading results and in this case the 60-kg. load is used. Blocks of this uniformity are calibrated for use with the 60, 100 and 150-kg. loads.

Rockwell Scales Employed.—It has been found that the Rockwell "B" scale employing the 100-kg. load, the $\frac{1}{16}$-in. diameter ball and reading the red figures, is satisfactory for rolled brass sheet; but in the case of the nickel-silver and phosphor-bronze alloys it was found that the Rockwell hardness-tensile strength curve became asymptotic, showing very little change in hardness for a large increase in tensile strength in the harder tempers. A load of 150 kg. was substituted for the 100-kg. load and resulted in an improvement since by using the larger load, a greater depth of penetration was obtained and consequently better sensitivity in the higher tempers. The curve, shown by Fig. 13, gives the Rockwell hardness-tensile strength relationship employing the 100 and 150-kg. loads on the Rockwell tester. The 150-kg. load has been adopted for nickel-silver and

TABLE IV

DIAMETER[a] OF AVERAGE GRAIN FOR ANNEALED BRASS AND NICKEL SILVER

Material	Anneal	Average Tensile Strength, lb. per sq. in.	Rockwell Hardness, "B" Scale, $\frac{1}{16}$-in. Ball, 60-kg. Load, Maximum	Diameter of Average Grain m.m.	
				Minimum	Maximum
High-Brass	Light	50,000	80	0.010	0.035
	Drawing	47,000	65	0.035	0.090
	Soft Drawing	43,000	55	0.090	0.200
Alloy C Brass	Light	42,000	76	0.010	0.030
	Drawing	40,000	61	0.030	0.070
	Soft Drawing	39,000	55	0.070	0.150
Alloy D Brass	Light	53,000	82	0.010	0.030
	Drawing	49,000	72	0.030	0.075
	Soft Drawing	45,000	59	0.075	0.200
Alloy E Brass	Light	50,000	80	0.010	0.035
	Drawing	46,000	67	0.035	0.080
	Soft Drawing	44,000	57	0.080	0.160
Alloy A Nickel Silver	Light	55,000	87	0.010	0.025
	Drawing	54,000	80	0.025	0.050
	Soft Drawing	52,500	77	0.040	0.100

[a] The method for determining the grain size of the material is the Jeffries method as described in the note under Section 9 of the Society's Standard Rules Governing the Preparation of Micrographs of Metals and Alloys (E 2–27), see 1927 Book of A. S. T. M. Standards, Part I, p. 778.

phosphor-bronze alloys and the Rockwell hardnesses reported hereinafter for these materials are the values obtained using a 150-kg. load, a $\frac{1}{16}$-in. diameter ball and reading on the "B" scale or red figures. The 60-kg. load is used for testing annealed material. (See Table IV.)

Fatigue Tests:

The endurance limit of each of the metals covered by this paper has been determined for the annealed condition and as rolled four and ten numbers hard. The results of these tests are covered by another paper.[7]

Bend Test Machine:

In connection with the Amsler bend test machine, an improvement on this machine has been developed employing a motor drive and accurately prepared and aligned jaws. Figure 4 shows a model of

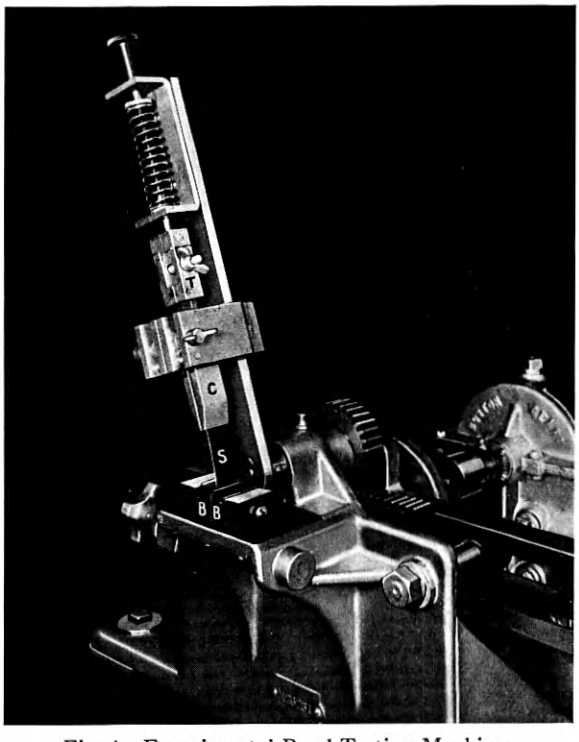


Fig. 4—Experimental Bend Testing Machine.

T is the tensioning grip.
C is the guide clamp.
S is the specimen.
B is the radius block.

[7] J. R. Townsend and C. H. Greenall, "Fatigue Studies of Non-Ferrous Sheet Metals," presented before the American Society for Testing Materials in June, 1929

this machine built for experimental purposes. A strip of metal *S*, $\frac{1}{2}$ in. wide, is clamped between a pair of radius blocks, *B*, selected at random. The upper portion of the specimen is held by a tensioning grip, *T*, and guided by two clamps, *C*. The specimen is bent back and forth over the pair of radius blocks and each bend of 90 deg.

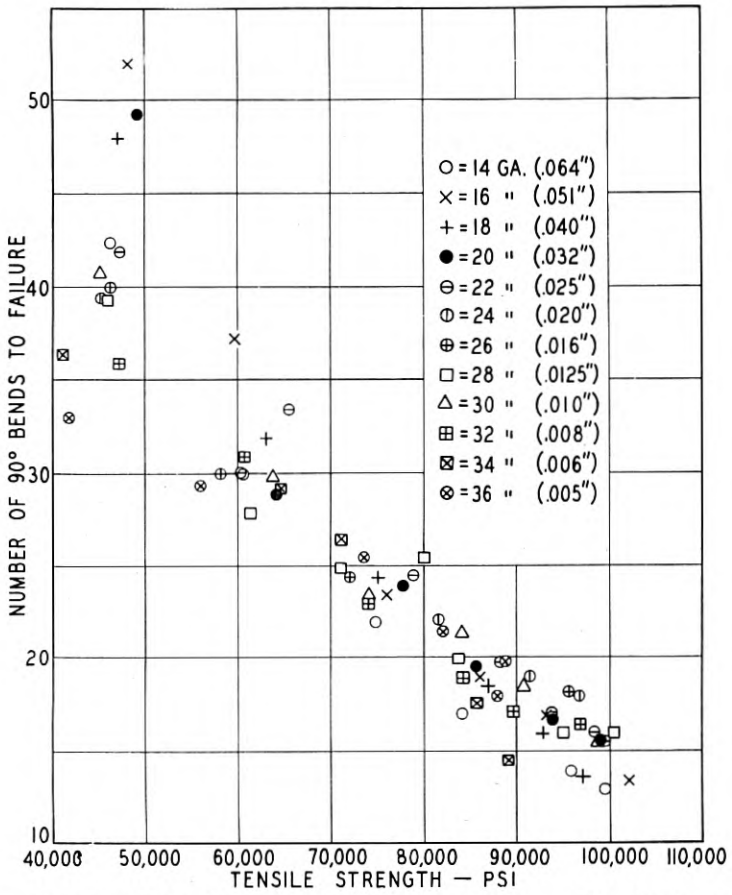


Fig. 5—Relation of Number of 90-deg. Bends to Failure to Tensile Strength of Alloy G Brass. Ratio of Mandrel Radius to Thickness, 8 to 1.

and return is recorded as one bend by a counter. Using another strip of the same metal but a pair of radius blocks of different radii another result is obtained. If the number of bends is plotted on log paper against the ratio of the thickness of the metal to the radius of the blocks, a straight line results. It is possible therefore to select any standard ratio of block radius to thickness of metal and

by two tests employing two different pairs of radius blocks to determine the number of bends for this predetermined ratio by interpolation. The ratio finally selected as most representative was 8 to 1. Figure 5 gives the bend test results for alloy G brass.

The bend test reveals the ductility and toughness of the metal under test. It should show a close correlation to the forming and drawing properties of sheet metal. Further studies will be required to work out such relationships.

Grain Count:

A series of photomicrographs of annealed brass has been adopted covering the grain sizes 0.010 to 0.200 mm. diameter of average grain.[8] Inasmuch as the alpha phase grain structure of nickel silver is similar to that of brass, one set of standards may be employed for these metals. Those standards were carefully counted by the Jeffries method as described in the note to Section 9 of the A. S. T. M. Standard Rules Governing the Preparation of Micrographs of Metals and Alloys (E 2–27) [9] and were photographed at a magnification of 75 diameters. The grain size of a specimen under examination is estimated by comparing the specimen with the standard photomicrographs. For alloy A nickel silver, a magnification of 150 is employed because of the smaller grain size of this material compared with brass. The limits of diameter of average grain of the alloys used for drawing purposes are given in Table IV, namely for high brass sheet, alloys C, D, and E brass, and alloy A nickel silver. (See Table VI for chemical composition.) The grain size limits are based on the range of commercial annealing practice. Inasmuch as these annealed materials are employed only for drawing purposes, no tensile strength limits are necessary. Maximum Rockwell hardness limits are given to exclude cold-worked metal.

Determination of Limits

High and Clock Brass:

Three separate rolling series were used for the round-robin hardness tests. The first of these consisted of a group of materials so arranged as to give a complete range of the hardnesses desired. The second was high brass with a nominal composition of 65 per cent copper and 35 per cent zinc. The third was nickel silver with a nominal composition of 64 per cent copper, 18 per cent zinc, and 18

[8] Report of Committee E-4 on Metallography, presented before the American Society for Testing Materials in June, 1929.

[9] 1927 Book of A. S. T. M. Standards, Part I, p. 778.

per cent nickel. Each of these series included eleven tempers ranging from annealed material to the hardness resulting from 10 B. & S. gage numbers reduction. The samples were approximately 6 in. square and $\frac{1}{8}$ in. thick.

Complete series of samples comprising a fairly wide range of thicknesses were obtained for high-brass sheet in order to thoroughly evaluate the methods of tests. These series were prepared in the following B. & S. gages: Nos. 14, 16, 20, 24, 28, 32 and 36. In each gage the samples had the following hardnesses: Annealed and 2, 4, 6, 8 and 10 B. & S. gage numbers reduction. These rolling series, therefore, included forty-two samples covering the usual commercial range of hardness and the range of thickness of greatest importance in telephone apparatus. These series were made under regular mill conditions with careful supervision to insure representative rolling practice and with complete records of the operations, special attention being given to accurate records of the percentage of reduction. Care was taken to have the metal given approximately a 50 per cent reduction before annealing to insure a uniform grain growth. The temperature of annealing was equivalent to about 600° C. for one-half hour, giving material of about average properties. These series were rolled from four bars of the following compositions:

	Bar No. 1	Bar No. 2	Bar No. 3	Bar No. 4
Copper, per cent	65.19	64.94	65.10	65.13
Zinc, " "	34.77	35.02	34.85	34.82
Lead, " "	0.01	0.02	0.02	0.02
Iron, " "	0.03	0.02	0.03	0.03

The samples were flat strips 10 ft. long and 6 in. wide.

The rolling series on clock brass were made in the following gages and tempers: Nos. 12, 14, 16 and 18 B. & S. gages and 2, 4, 6 and 8 B. & S. gage numbers reduction. The composition was: Copper 61.63 per cent, zinc 36.75 per cent, lead 1.57 per cent, iron 0.05 per cent. This material is so nearly like high-brass sheet in its properties that the data collected when added to that already available were considered sufficient for the preparation of a set of requirements. The tensile strength-percentage of reduction curve for high brass is shown by Figure 10. The physical properties of high and clock brass are shown by Tables I and II.

As mentioned above, Rockwell hardness limits given have been revised to agree with the more recently established Rockwell standards. These limits have been verified by experience data collected on

TABLE I

PHYSICAL TESTS ON HIGH-BRASS SHEET ROLLING SERIES

B. & S. Gage	Thickness, in.	Numbers Hard	Percentage Reduction by Rolling	Proportional Limit, lb. per sq. in.	Tensile Strength, lbs per sq. in.	Elongation in 2 in. per cent	Scleroscope Hardness, Diamond Magnifier Hammer, One Thickness	Rockwell Hardness, "B" Scale, One Thickness	Meyer's Analysis "A," 1/16-in. Ball. 1 mm. in diameter indentation, kg.	Erichsen Ductility, Large Penetrator, mm.	Erichsen Ductility, Small Penetrator, mm.	Olsen Ductility, Large Penetrator, in.	Olsen Ductility, Small Penetrator, in.
No. 36.....	0.0058	Annealed	0	9,500	44,200	40	15	76	80	12.20	3.17	0.462	0.113
No. 36.....	0.0055	2	12.7	33,000	57,200	17	33	76	102	5.64	2.38	0.202	0.064
No. 36.....	0.0058	4	29.25	32,000	67,600	3	41	72	110	4.48	2.17	0.149	0.062
No. 36.....	0.0054	6	46.0	38,000	81,000	1.5	50	76	137	3.27	1.93	0.128	0.062
No. 36	0.0057	8	53.6	39,000	86,000	2	54	74	143	3.35	1.88	0.122	0.048
No. 36.....	0.0060	10	62.5	35,500	90,400	1	58	72	151	2.86	1.86	0.107	0.046
No. 32.....	0.0087	Annealed	0	7,900	44,400	53.5	13	48	83	13.24	3.37	0.488	0.116
No. 32	0.0078	2	22.0	31,600	60,300	16	36	54	107	5.63	2.41	0.209	0.060
No. 32.....	0.0090	6	43.75	29,200	81,200	2	47	59	134	4.30	2.37	0.153	0.061
No. 32.....	0.0085	8	60.8	40,800	91,800	1.5	53	64	160	4.00	2.10	0.135	0.057
No. 28.....	0.0127	Annealed	0	8,300	44,500	60	9	13	62	13.69	3.73	0.542	0.129
No. 28.....	0.0144	2	10.0	38,800	58,000	25.5	24	60	104	7.21	2.94	0.262	0.087
No. 28.....	0.0135	4	37.8	30,500	75,900	5	31	78	132	6.11	2.76	0.209	0.074
No. 28.....	0.0127	6	49.2	40,300	84,400	2	38	82	142	5.16	2.52	0.188	0.059
No. 28.....	0.0134	8	59.1	34,300	89,600	2	49	86	150	4.37	2.27	0.144	0.054
No. 28.....	0.0136	10	66.0	44,300	95,900	1.5	51	89	157	3.94	2.32	0.149	0.050
No. 24.....	0.0221	Annealed	0	12,000	46,600	58	10	16	66	13.62	3.79	0.532	0.119
No. 24.....	0.0204	2	18.4	23,600	58,300	29	25	60	101	8.51	3.22	0.308	0.089
No. 24.....	0.0199	4	39.3	32,000	77,200	6	33	79	133	6.37	2.87	0.233	0.071
No. 24.....	0.0202	6	49.5	31,000	85,500	4	40	83	142	5.44	2.65	0.204	0.058
No. 24.....	0.0210	8	59.4	31,300	90,500	2	44	86	150	4.59	2.35	0.178	0.052
No. 24.....	0.0211	10	67.6	30,000	95,600	2	50	87	155	4.07	2.26	0.163	0.046

TABLE I—*Continued*

B. & S. Gage	Thickness, in.	Numbers Hard	Percentage Reduction by Rolling	Proportional Limit, lb. per sq. in.	Tensile Strength, lbs. per sq. in.	Elongation in 2 in. per cent	Scleroscope Hardness, Diamond Magnifier Hammer, One Thickness	Rockwell Hardness, "B" Scale One Thickness	Meyer's Analysis "A," $\frac{1}{16}$-in. Ball. 1 mm. in diameter indentation, kg.	Erichsen Ductility, Large Penetrator, mm.	Erichsen Ductility, Small Penetrator, mm.	Olsen Ductility, Large Penetrator, in.	Olsen Ductility, Small Penetrator, in.
No. 20.....	0.0334	Annealed	0	6,100	46,300	63	9	10	63.5	14.17	4.10	0.559	0.124
No. 20.....	0.0345	2	13.75	23,200	57,300	32	26	60	103	9.05	3.46	0.348	0.107
No. 20.....	0.0336	4	35.1	23,500	72,100	11	34	77	125	7.52	3.05	0.272	0.076
No. 20.....	0.0333	6	48.9	27,100	83,800	5	48	80	138	6.27	2.48	0.231	0.065
No. 20.....	0.0340	8	58.0	33,300	89,000	4	51	86	147	4.83	1.84	0.175	0.063
No. 20.....	0.0331	10	67.6	33,800	92,700	3.5	56	88	152	3.32	1.90	0.130	0.061
No. 16.....	0.0529	Annealed	0	8,300	45,500	70	10	—4	63	15.00	....	....	....
No. 16.....	0.0524	2	19.6	22,900	59,300	35	36	67	106	10.08	....	....	....
No. 16.....	0.0525	4	35.2	23,100	71,700	9.5	42	79	123	7.50	....	....	....
No. 16.....	0.0523	6	48.7	27,600	80,700	7.5	46	84	137	6.53	....	....	....
No. 16.....	0.0521	8	59.3	30,500	87,900	5.5	51	87	146	4.16	....	....	....
No. 16.....	0.0510	10	68.5	29,600	94,500	5	59	91	155	2.86	....	....	....
No. 14.....	0.0661	Annealed	0	7,900	46,600	63	13	8	66	14.99	....	....	....
No. 14.....	0.0651	2	19.6	23,600	58,900	38	34	68	105	10.05	....	....	....
No. 14.....	0.0662	4	35.1	21,600	71,100	13	36	80	125	8.92	....	....	....
No. 14.....	0.0648	6	49.4	22,000	81,700	8	39	85	137	5.66	....	....	....
No. 14.....	0.0664	8	59.0	30,600	89,300	6.5	47	89	147	2.97	....	....	....
No. 14.....	0.0648	10	67.1	28,300	93,500	6	49	90	154	2.08	....	....	....

TABLE II

PHYSICAL TESTS ON CLOCK-BRASS SHEET ROLLING SERIES

Sample Number	Thickness, in.	Temper	B. & S. Gage Numbers Hard	Reduction by Rolling, per cent	Tensile Strength, lb. per sq. in.	Elongation in 2 in., per cent	Sclero-scope Steel Magnifier Hammer, One Thickness	Rockwell "B" Scale, One Thickness	Meyer's Analysis "A," kg.
502–2	0.081	....	2	20.4	59,250	32	35	68	106
502–4	0.064	....	4	36.6	72,750	9	48	80	125
502–6	0.050	....	6	49.8	83,700	6	54	85	137.5
502–8	0.039	....	8	60.4	90,000	4	58	88	152
340	0.143	Soft	..	....	45,900	62	15	21	67
341	0.081	Quarter Hard	..	....	51,400	40	25	56	86
414	0.040	Half Hard	..	....	61,800	25	35	70	106
412	0.079	"	..	....	64,500	23	40	77	119
342	0.033	"	..	....	70,200	14	39	78	125
190	0.020	"	..	....	64,600	13	..	65	113.5
415	0.040	Hard	..	....	74,300	8	46	82	127
413	0.078	"	..	....	74,800	8.5	45	84	125
191	0.032	"	..	....	74,500	5.5	..	79	123.5

material furnished under the limits by a number of suppliers. Table V shows the Rockwell hardness limits. Figure 6 shows Rockwell hardness plotted against tensile strength for the high-brass rolling

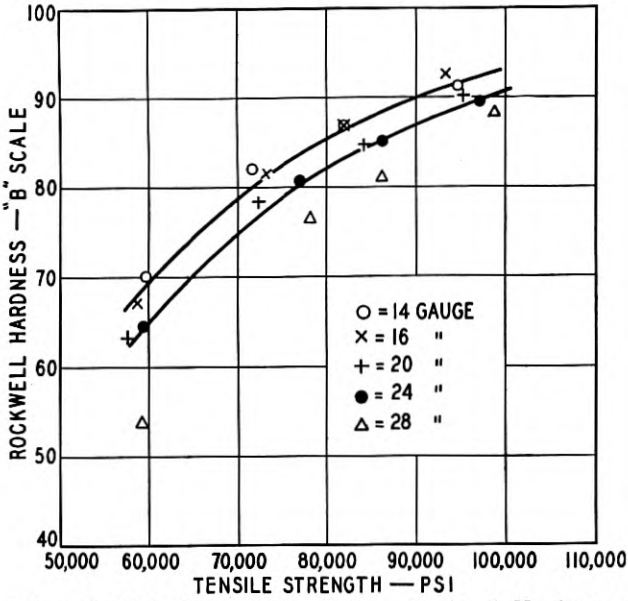


Fig. 6—Relation of High Brass Rolling Series. Rockwell Hardness to Tensile Strength, Using Standard Rockwell Blocks.

TABLE V

REVISED ROCKWELL HARDNESS AND TENSILE STRENGTH LIMITS FOR HIGH, CLOCK AND ALLOY E BRASS SHEET

Thickness	Temper, B. & S. Numbers Hard	Percentage Reduction by Rolling	Tensile Strength,[a] lb. per sq. in.						Rockwell Hardness, "B" Scale, 1/16-in. Ball, 100-kg. Load (Red Figures), for High, Clock and Alloy E Brass Sheet	
			High Brass		Clock Brass		Alloy E Brass			
			Minimum	Maximum	Minimum	Maximum	Minimum	Maximum	Minimum	Maximum
0.040 in. and over	Quarter hard. 1	10.9	46,000	56,000	47,000	57,000	47,000	57,000	30	60
0.020 to 0.040 in.		10.9	46,000	56,000	47,000	57,000	47,000	57,000	30	56
0.040 in. and over	Half hard. 2	20.7	53,500	63,500	54,500	64,500	54,000	64,000	50	73
0.020 to 0.040 in.		20.7	53,500	63,500	54,500	64,500	54,000	64,000	49	71
0.040 in. and over	Three-quarters hard. 3	29.5	61,000	71,000	....	....	....	....	70	80
0.020 to 0.040 in.		29.5	61,000	71,000	....	....	....	....	67	77
0.040 in. and over	Hard. 4	37.1	68,000	78,000	68,000	78,000	68,000	78,000	78	85
0.020 to 0.040 in.		37.1	68,000	78,000	68,000	78,000	68,000	78,000	75	83
0.040 in. and over	Extra hard. 6	50.0	79,000	88,500	79,000	88,500	....	....	85	89
0.020 to 0.040 in.		50.0	79,000	88,500	79,000	88,500	....	....	83	87
0.040 in. and over	Spring. 8	60.5	86,000	95,000	85,000	94,500	....	....	88	92
0.020 to 0.040 in.		60.5	86,000	95,000	85,000	94,500	....	....	85	89
0.040 in. and over	Extra spring. 10	68.7	89,500	98,500	88,000	97,500	....	....	89	93
0.020 to 0.040 in.		68.7	89,500	98,500	88,000	97,500	....	....	86	90

[a] Tensile strength values for high and clock brass sheet are the same as in the previous paper by H. N. Van Deusen, L. I. Shaw and C. H. Davis, "Physical Properties and Methods of Test for Sheet Brass," *Proc. A. S. T. M.*, Vol. 27, 1927.

TABLE VI

CHEMICAL COMPOSITION LIMITS FOR NON-FERROUS METAL SHEET

		High Sheet Brass	Clock Brass	Alloy C Brass	Alloy D Brass	Alloy E Brass	Alloy G Brass	Alloy A Nickel Silver	Alloy B Nickel Silver	Alloy A Phosphor Bronze	Alloy C Phosphor Bronze
Copper, per cent	minimum	64.50	61.00	83.00	73.00	64.00	70.50	70.50	53.50	94.40	91.00
	maximum	67.50	64.00	86.00	76.00	67.00	73.50	73.50	56.50	....	....
Lead, per cent	minimum	0.00	1.25	....	....	0.80	....	....	....	....	....
	maximum	0.30	2.00	0.15	0.25	1.10	0.10	....	0.10	0.05	0.02
Iron, per cent	minimum	0.00	0.00	....	....	....	....	....	....	....	....
	maximum	0.05	0.06	0.05	0.05	0.08	0.05	0.35	0.35	0.10	0.10
Zinc, per cent	minimum	Balance	Balance	Balance	Balance	Balance	Balance	8.50	25.50	....	....
	maximum	Balance	Balance	Balance	Balance	Balance	Balance	11.50	28.50	0.30	0.20
Nickel, per cent	minimum	....	....	....	....	....	....	16.50	16.50	....	....
	maximum	....	....	....	....	....	....	19.50	19.50	....	....
Manganese, per cent	minimum	....	....	....	....	....	....	....	....	....	....
	maximum	....	....	....	....	....	....	....	0.50	....	....
Tin, per cent	minimum	....	....	....	....	....	....	....	....	3.80	7.50
	maximum	....	....	....	....	....	....	....	....	4.80	8.50
Phosphorus, per cent	minimum	....	....	....	....	....	....	....	....	0.05	0.05
	maximum	....	....	....	....	....	....	....	....	0.35	0.25
Antimony, per cent	minimum	....	....	....	....	....	....	....	....	....	....
	maximum	....	....	....	....	....	....	....	....	0.01	0.01
Impurities, per cent	minimum	....	....	....	....	....	....	....	....	....	....
	maximum	0.10	0.10	....	....	0.10	0.10	....	....	trace	trace

series retested using standard blocks. There have been no changes in tensile strength limits. Figure 19 shows Rockwell hardness plotted against tensile strength, for high and clock brass, the individual point representing determinations of Rockwell hardness and tensile strength on commercial shipments of material. The lines drawn show the grouping of the data according to thickness.

Alloys C and D Brass Sheet:

The chemical composition limits for alloys C and D brass sheet are given in Table VI.

Alloy E Brass Sheet:

Four rolling series were made from one bar of metal having the analysis shown in Table VII. These series begin at B. & S. gage Nos.

TABLE VII

CHEMICAL ANALYSES OF ROLLING SERIES

	Alloy E Brass	Alloy G Brass	Alloy A Nickel Silver	Alloy B Nickel Silver	Alloy A Phosphor Bronze	Alloy C Phosphor Bronze
Copper, per cent	66.02	71.73	71.31	55.23	95.35	91.84
Lead, per cent	1.08	0.02	....	0.005	0.01	0.02
Iron, per cent	0.03	0.03	0.12	0.06	0.05	0.03
Zinc, per cent	32.87	28.21	10.74	26.27	0.00	0.00
Nickel, per cent	....	0.01	17.62	18.38	0.00	0.00
Manganese, per cent	....	....	0.12	0.11	....	....
Tin, per cent	....	0.00	....	....	4.48	8.08
Phosphorus, per cent	....	....	....	....	0.08	0.03
Graphite, per cent	....	....	0.00	0.00	....	....
Combined Carbon, per cent	....	....	0.013	0.018	....	....

10, 14, 18 and 22 and were rolled 1, 2 and 4 B. & S. gage numbers from commercial anneals in the 500 to 650° C. range. Figure 7 shows Rockwell hardness plotted against tensile strength for alloy E brass.

In a previous paper [10] it was shown that different Rockwell hardness values were required for material 0.040 in. and thicker, and materials less than 0.040 in. thick. This change in limits is dependent upon thickness and is not sharply defined, but it has been found sufficiently accurate for all practical purposes to consider that the division occurs at 0.040 in. In the case of the alloy E brass rolling series results shown by Fig. 7, only one curve was drawn since not sufficient data were available to show this division. The curve practically agrees with the curves given in a previous paper [10] for the

[10] H. N. Van Deusen, L. I. Shaw and C. H. Davis, *loc. cit.*

high-brass rolling series, and also with the data shown on Fig. 19 for commercial shipments of high and clock brass.

The composition of alloy E brass is midway between that of high and clock-brass, and consequently it was to be expected that the Rockwell hardness limits for this material would be the same as for

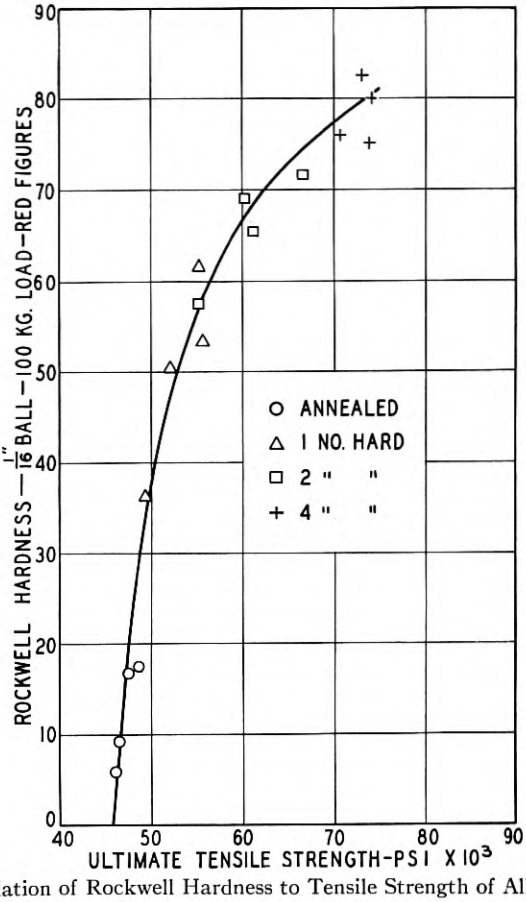


Fig. 7—Relation of Rockwell Hardness to Tensile Strength of Alloy E Brass.

the foregoing. The tension test is more sensitive to such changes in chemical composition and since it is the test upon which acceptance or rejection of material is based, separate tensile strength limits are given. The separate tensile strength and the combined Rockwel hardness limits for these alloys are given in Table V. The physical properties of alloy E brass are given in Table VIII.

TABLE VIII

Physical Properties of Alloy E Brass Sheet

B. & S. Gage	Temper, B. & S. Numbers Hard	Thickness, in.	Percentage Reduction by Rolling		Proportional Limit, lb. per sq. in.	Tensile Strength, lb. per sq. in.	Rockwell Hardness, "B" Scale, 1/16-in. Ball, 100-kg. Load (Red Figures)	Modulus of Elasticity, lb. per sq. in.	Elongation in 2 in., per cent
			At Edges	At Center					
No. 10	0	0.1057	0	0	.....	48,330	17.4	.....	65.8
No. 11	1	0.0903	14.2	15.2	20,050	55,130	61.6	13,740,000	41.0
No. 12	2	0.0817	22.0	23.4	30,417	60,630	71.8	14,600,000	30.7
No. 14	4	0.0657	37.2	38.5	27,230	72,930	82.6	16,540,000	10.8
No. 14	0	0.0656	0	0	.....	46,230	9.1	.....	64.3
No. 15	1	0.0568	10.6	10.6	18,525	52,000	50.3	16,460,000	46.5
No. 16	2	0.0508	23.1	22.8	25,870	60,200	69.2	16,760,000	28.5
No. 18	4	0.0402	38.5	38.5	29,670	74,066	80.1	14,480,000	8.0
No. 18	0	0.0402	0	0	.....	46,130	5.8	.....	67.3
No. 19	1	0.0371	9.2	9.3	18,870	49,600	36.3	.....	52.3
No. 20	2	0.0338	17.4	17.2	20,885	55,230	57.6	15,380,000	37.8
No. 22	4	0.0260	36.7	36.3	23,385	70,530	76.0	17,000,000	11.2
No. 22	0	0.0256	0	0	.....	47,500	16.8	.....	59.0
No. 23	1	0.0225	15.8	14.1	25,410	55,500	53.2	12,340,000	38.7
No. 24	2	0.0205	23.2	21.8	36,460	61,010	65.7	12,300,000	29.0
No. 26	4	0.0169	36.7	35.1	34,090	74,050	75.2	.....	7.5

Note.—Tensile strength and percentage elongation are an average of 3 specimens in each case; Rockwell hardness values are an average of 15 determinations of 5 readings on each of 3 tension test specimens.

It will be noted that the Rockwell hardness and tensile strength limits for alloy E brass are given only for material 1, 2 and 4 numbers hard. This alloy is not normally used in higher tempers as it is especially adapted for use where both forming and machining operations are involved, so that higher leaded or harder materials would not be suitable.

The chemical composition requirements for alloy E brass sheet are given in Table VI.

TABLE IX

PHYSICAL PROPERTIES OF ALLOY G BRASS SHEET

B. & S. Gage	Temper, B. & S. Numbers Hard	Thickness, in.	Actual Percentage Reduction by Rolling	Rockwell Hardness, "B" Scale, 1/16-in. Ball, 100-kg. Load (Red Figures)	Tensile Strength, lb. per sq. in.	Elongation in 2 in., per cent
No. 14	0	0.0655	0	8.5	46,200	68
	2	0.0655	20.6	70.1	60,500	32.5
	4	0.0660	35.0	83.4	74,600	11
	6	0.0658	50.6	88.2	83,800	7.5
	8	0.0655	60.4	92.6	95,600	6.5
	10	0.0655	68.8	93.4	98,600	6.5
No. 16	0	0.0520	0	22.7	48,300	60
	2	0.0522	21.4	69.0	59,700	33
	4	0.0515	37.6	83.7	76,200	9
	6	0.0500	49.9	87.7	86,400	6
	8	0.0503	61.5	91.0	93,000	6.5
	10	0.0512	68.6	93.5	102,000	4.5
No. 18	0	0.0424	0	21.8	48,700	63
	2	0.0409	21.0	73.2	63,400	25.5
	4	0.0412	36.8	81.6	74,900	9.5
	6	0.0408	49.6	87.9	86,800	6
	8	0.0415	59.4	90.1	92,700	4
	10	0.0412	69.6	92.4	96,900	3
No. 20	0	0.0339	0	27.6	48,900	57
	2	0.0333	22.2	73.3	64,000	25
	4	0.0329	37.4	83.5	77,800	7.5
	6	0.0330	49.9	87.0	85,400	5
	8	0.0324	60.7	90.3	93,800	3
	10	0.0320	68.5	92.1	99,000	3
No. 22	0	0.0256	0	15.4	47,100	64
	2	0.0260	23.0	74.2	65,100	22.5
	4	0.0265	39.9	83.4	79,100	7.5
	6	0.0263	50.5	88.3	88,200	2.5
	8	0.0260	60.7	90.0	93,300	3
	10	0.0255	68.8	91.8	98,200	2.5
No. 24	0	0.0201	0	16.0	46,300	61
	2	0.0203	21.1	67.6	61,000	26
	4	0.0204	39.6	84.3	81,600	6
	6	0.0204	53.3	89.0	91,400	3.5
	8	0.0208	59.3	91.0	96,600	2
	10	0.0207	66.6	91.9	97,800	2

TABLE IX—*Continued*

B. & S. Gage	Temper, B. & S. Numbers Hard	Thickness, in.	Actual Percentage Reduction by Rolling	Rockwell Hardness, "B" Scale, 1/16-in. Ball, 100-kg. Load (Red Figures)	Tensile Strength, lb. per sq. in.	Elongation in 2 in., per cent
No. 26	0	0.0169	0	11.0	46,200	58.0
	2	0.0167	17.3	64.4	58,100	26.5
	4	0.0168	34.3	77.3	71,800	9
	6	0.0168	50.9	87.4	88,600	2
	8	0.0165	62.4	90.2	95,600	1.5
	10	0.0172	66.5	91.2	99,100	1.5
No. 28	0	0.0133	0	13.2	45,800	61.5
	2	0.0135	21.1	65.4	61,100	24
	4	0.0134	33.5	77.0	71,400	8
	6	0.0135	49.0	83.2	83,600	2
	8	0.0135	60.1	89.6	94,700	1.5
	10	0.0132	67.7	91.5	100,100	1
No. 30	0	0.0108	0	28.5	45,600	54
	2	0.0103	18.8	55.7	63,600	20
	4	0.0106	36.6	75.7	74,000	4
	6	0.0106	48.7	81.2	84,000	2
	8	0.0108	58.6	86.1	90,600	1.5
	10	0.0108	68.4	90.3	98,600	1
No. 32	1	0.0085	0	....	46,800	51.5
	2	0.0084	17.8	....	60,500	24.5
	4	0.0085	33.6	....	73,900	5
	6	0.0087	46.0	....	84,300	1.5
	8	0.0086	59.4	....	89,700	1
	10	0.0085	63.0	....	96,700	0.5
No. 34	0	0.0071	0	....	41,200	39
	2	0.0065	29.2	....	64,400	20.5
	4	0.0071	39.6	....	70,900	4.5
	6	0.0071	46.9	....	79,800	1.5
	8	0.00705	57.1	....	85,700	1
	10	0.0072	64.5	....	89,200	0.75
No. 36	0	0.0057	0	....	41,600	33
	2	0.0054	16.7	....	55,700	18
	4	0.0056	34.1	....	73,400	4
	6	0.0056	51.5	....	81,900	2
	8	0.0056	60.9	....	87,700	2
	10	0.0052	68.3	....	93,100	1

Note.—Tensile strength and percentage elongation are an average of 3 specimens in each case; Rockwell hardness values are an average of 15 determinations of 5 readings on each of 3 tension test specimens.

Alloy G Brass Sheet:

It has been pointed out in a previous paper [11] that this material, consisting nominally of 72 per cent of copper and 28 per cent of zinc, gives the maximum tensile properties possible with the alpha

[11] W. H. Bassett and C. H. Davis, "Physical Characteristics of Copper and Zinc Alloys," *Proceedings*, Inst. Metals Div., Am. Inst. Mining and Metallurgical Engrs., 1928.

brasses in the cold-rolled condition. A bar of alloy G brass of the chemical composition shown in Table VII was rolled 2, 4, 6, 8 and 10 numbers hard in every even B. & S. gage number from No. 14 to No. 36, inclusive. Table IX gives the physical properties of the

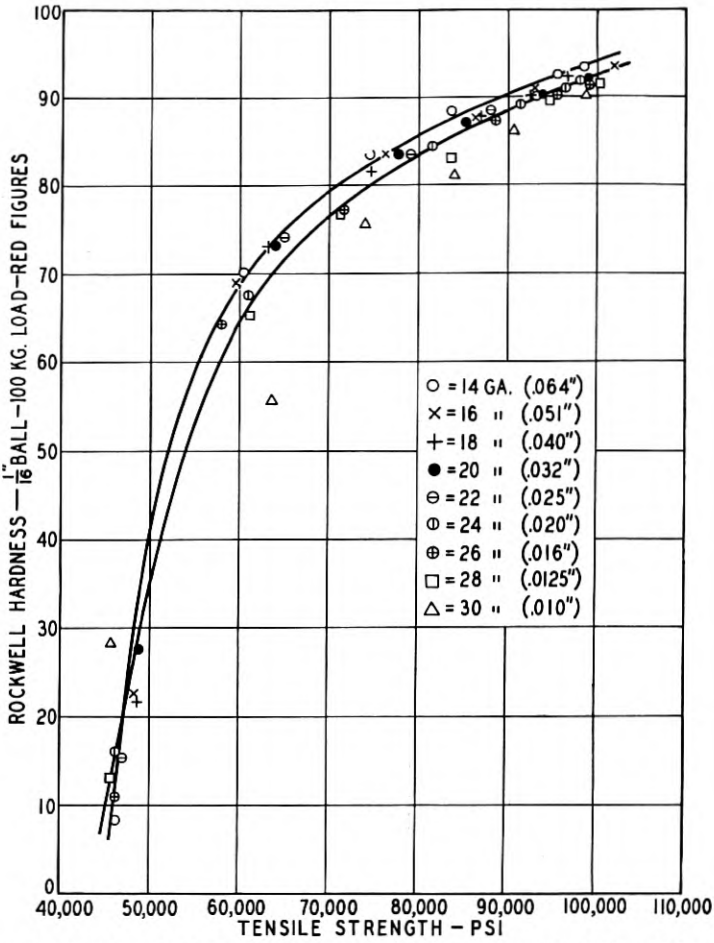


Fig. 8—Relation of Rockwell Hardness to Tensile Strength of Alloy G Brass Rolling Series.

rolling series. The Rockwell hardness-tensile strength curve is shown in Fig. 8 and the tensile strength-reduction curve is shown by Fig. 9. The chemical composition limits are given in Table VI, and the tensile strength and Rockwell hardness limits are given in Table X.

Nickel Silver:

Two nickel-silver alloys were investigated, one containing nominally 72 per cent of copper, 10 per cent of zinc and 18 per cent of nickel, which will be designated "alloy A"; and the other containing 55 per cent of copper, 27 per cent of zinc, and 18 per cent of nickel and designated "alloy B." Alloy A is used mainly for forming and deep drawing purposes, whereas alloy B is used for springs.

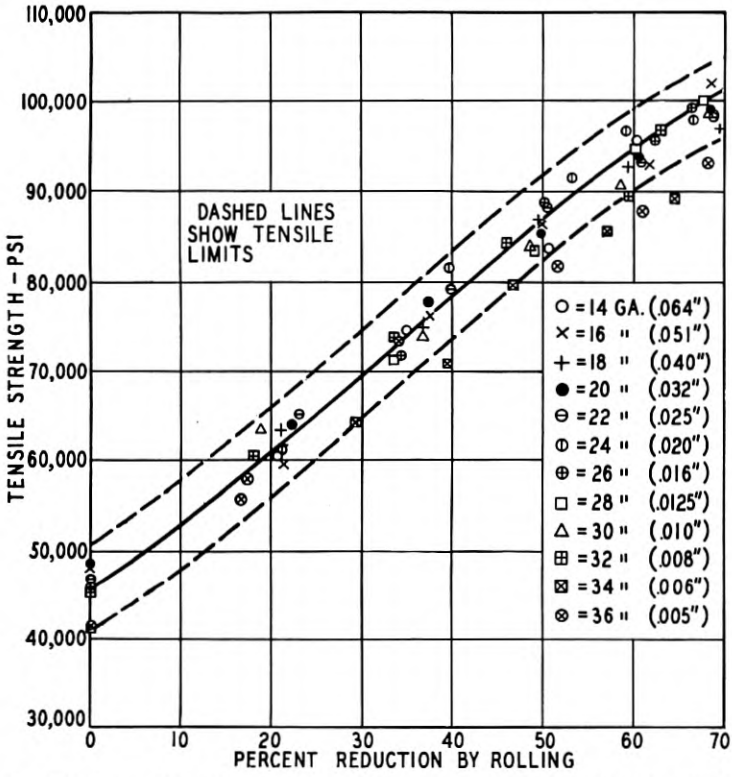


Fig. 9—Relation of Tensile Strength to Percentage of Reduction by Rolling of Alloy G Brass Rolling Series.

Alloy A Nickel-Silver Sheet.—The average composition of three bars of alloy A cast from the same pot of metal are given in Table VII. Table XI shows the physical properties of alloy A. Figure 11 shows the Rockwell hardness-tensile strength relationship and Fig. 12 shows the tensile strength limits plotted against percentage of reduction. The tensile strength and Rockwell hardness limits for this material are given in Table XII and the composition limits in Table VI.

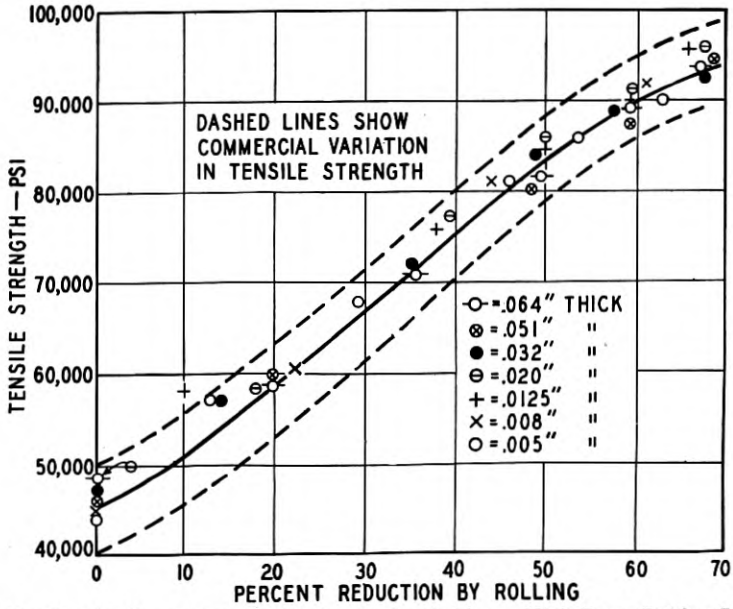


Fig. 10—Tensile Strength *versus* Reduction by Rolling. High-Brass Rolling Series

TABLE X

ROCKWELL HARDNESS AND TENSILE STRENGTH LIMITS FOR ALLOY G BRASS SHEET

Thickness	Temper, B. & S. Numbers Hard	Percentage Reduction by Rolling	Tensile Strength, lb. per sq. in.		Rockwell Hardness, "B" Scale, 1/16-in. Ball, 100-kg. Load (Red Figures)	
			Minimum	Maximum	Minimum	Maximum
0.040 in. and over	Quarter hard 1	11.0	49,000	59,000	35	67
0.020 to 0.040 in.		11.0	49,000	59,000	30	63
0.040 in. and over	Half hard 2	20.7	56,500	66,500	63	76
0.020 to 0.040 in.		20.7	56,500	66,500	57	73
0.040 in. and over	Hard 4	37.1	71,000	81,000	80	86
0.020 to 0.040 in.		37.1	71,000	81,000	77	84
0.040 in. and over	Extra hard 6	50.0	82,500	91,500	87	91
0.020 to 0.040 in.		50.0	82,500	91,500	85	89
0.040 in. and over	Spring 8	60.5	90,500	99,500	90	94
0.020 to 0.040 in.		60.5	90,500	99,500	88	92
0.040 in. and over	Extra spring 10	68.7	95,000	104,000	92	96
0.020 to 0.040 in.		68.7	95,000	104,000	90	94

TABLE XI

PHYSICAL PROPERTIES OF ALLOY A NICKEL-SILVER SHEET

B. & S. Gage	Temper, B. & S. Numbers Hard	Thickness, in.	Actual Percentage Reduction by Rolling	Elongation in 2 in., per cent	Rockwell Hardness, "B" Scale, 1/16-in. Ball (Red Figures)		Tensile Strength, lb. per sq. in.
					100-kg. Load	150-kg. Load	
No. 10.....	0	0.1034	0	37.5	37.5	—13.5	54,500
No. 11.....	1	0.0910	7.3	17.5	73.3	37.5	62,200
No. 10.....	2	0.1022	21.1	8.5	80.1	48.7	70,400
No. 10.....	4	0.1037	36.9	5.5	84.0	55.2	79,500
No. 10.....	6	0.1035	49.8	5.0	86.7	59.7	82,900
No. 16.....	0	0.0517	0	36.5	37.3	—16.7	55,100
No. 17.....	1	0.0460	10.6	21.5	69.1	30.1	61,300
No. 16.....	2	0.0516	22.3	7.5	78.7	45.5	69,500
No. 16.....	4	0.0503	39.5	3.5	84.5	56.2	79,700
No. 16.....	6	0.0512	49.8	4.0	86.3	59.0	82,800
No. 22.....	0	0.0257	0	32.0	36.4	....	54,400
No. 23.....	1	0.0301	7.8	20.0	66.6	26.1	60,600
No. 22.....	2	0.0266	17.2	7.0	75.5	41.1	69,100
No. 22.....	4	0.0263	36.1	2.5	81.4	51.0	79,600
No. 22.....	6	0.0263	48.7	2.0	83.3	55.0	83,700
No. 28.....	0	0.0128	0	35.5	23.0	....	54,500
No. 27.....	1	0.0145	11.5	16.0	65.6	....	62,800
No. 28.....	2	0.0132	21.2	4.5	72.2	....	69,200
No. 28.....	4	0.0137	35.6	1.5	77.1	....	80,700
No. 28.....	6	0.0130	47.4	1.0	81.1	....	84,900
No. 30.....	0	0.0105	0	30.5	....	....	54,600
No. 32.....	0	0.0085	0	32.0	....	....	55,100
No. 34.....	0	0.0059	0	28.5	....	....	55,600
No. 34.....	1	0.0077	12.5 17.6	6.0	....	....	65,200
No. 34.....	2	0.0058	25.0 29.4	2.5	....	....	74,400
No. 34.....	4	0.0062	41.2	1.0	....	....	78,800
No. 34.....	6	0.0060	50.8	1.0	....	....	87,900

Note.—Tensile strength and percentage elongation are an average of 3 specimens in each case; Rockwell hardness values are an average of 15 determinations of 5 readings on each of 3 tension test specimens.

Alloy B Nickel-Silver Sheet.—The five bars of alloy B had the average analysis given in Table VII. The rolling series made from these five bars consisted of the tempers 2, 4, 6, 8 and 10 numbers hard for each of the even B. & S. gage numbers from No. 14 to No. 36, inclusive. The physical properties of this material are given in Table XIII. The Rockwell hardness-tensile strength relationship is shown by Fig. 13 and the tensile strength limits plotted against

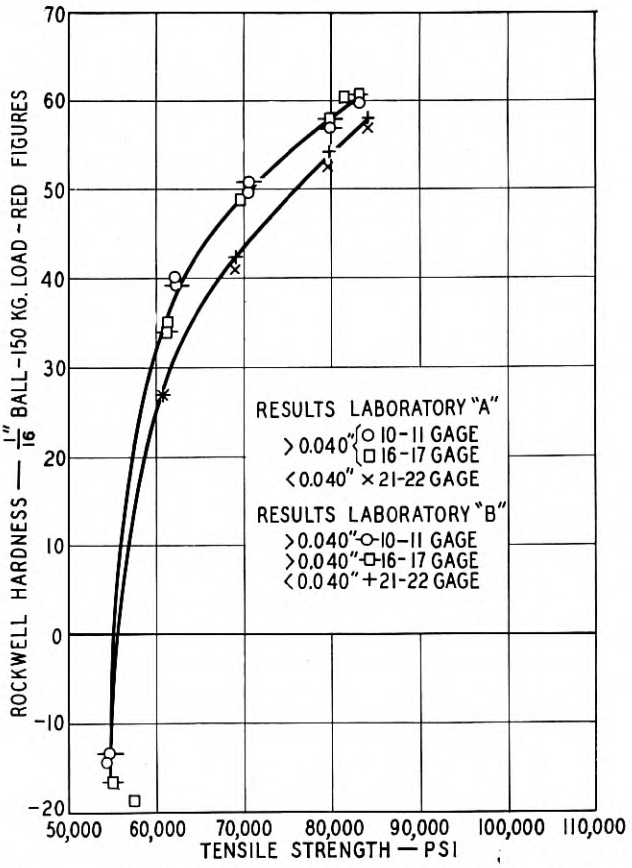


Fig. 11—Relation of Rockwell Hardness to Tensile Strength of Alloy A Nickel Silver Rolling Series.

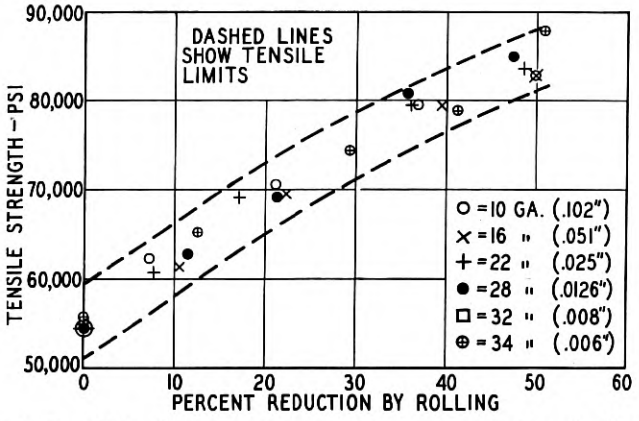


Fig. 12—Relation of Tensile Strength to Percentage of Reduction of Alloy A Nickel Silver Rolling Series.

TABLE XII

ROCKWELL HARDNESS AND TENSILE STRENGTH LIMITS FOR ALLOYS A AND B NICKEL-SILVER SHEET

Thickness	Temper, B. & S. Numbers Hard	Nominal Percentage Reduction by Rolling	Alloy A				Alloy B			
			Tensile Strength, lb. per sq. in.		Rockwell Hardness, "B" Scale, 1/16-in. Ball, 150-kg. Load (Red Figures)		Tensile Strength, lb. per sq. in.		Rockwell Hardness, "B" Scale, 1/16-in. Ball, 150-kg. Load (Red Figures)	
			Minimum	Maximum	Minimum	Maximum	Minimum	Maximum	Minimum	Maximum
0.040 in. and over	Quarter hard 1	11.0	58,500	67,000	26	47	.....	.....	..	..
0.020 to 0.040 in.	Quarter hard 1	11.0	58,500	67,000	19	40	.....	.....	..	..
0.040 in. and over	Half hard 2	20.7	65,500	73,500	44	54	78,000	93,000	54	70
0.020 to 0.040 in.	Half hard 2	20.7	65,500	73,500	38	48	78,000	93,000	48	67
0.040 in. and over	Hard 4	37.1	75,000	82,000	54	61	92,000	106,500	69	77
0.020 to 0.040 in.	Hard 4	37.1	75,000	82,000	49	55	92,000	106,500	66	75
0.040 in. and over	Extra hard 6	50.0	81,000	88,000	59	64	102,000	115,000	75	82
0.020 to 0.040 in.	Extra hard 6	50.0	81,000	88,000	54	60	102,000	115,000	72	79
0.040 in. and over	Spring 8	60.5	.....	.....	..	..	108,000	120,000	78	84
0.020 to 0.040 in.	Spring 8	60.5	.....	.....	..	..	108,000	120,000	75	81
0.040 in. and over	Extra spring 10	68.7	.....	.....	..	..	111,000	123,000	80	85
0.020 to 0.040 in.	Extra spring 10	68.7	.....	.....	..	..	111,000	123,000	77	82

TABLE XIII

PHYSICAL PROPERTIES OF ALLOY B NICKEL-SILVER SHEET

B. & S. Gage	Temper, B. & S. Numbers Hard	Thickness, in.	Actual Percentage Reduction by Rolling	Tensile Strength, lb. per sq. in.	Elongation in 2 in., per cent	Rockwell Hardness, "B" Scale, 1/16-in. Ball (Red Figures)	
						100-kg. Load	150-kg. Load
No. 14.....	0	0.0666	0	66,100	45.5	53.6	9.5
No. 14.....	2	0.0653	22.6	81,300	13.0	87.6	62.2
No. 14.....	4	0.0664	37.2	93,400	5.0	92.8	70.5
No. 14.....	6	0.0670	47.8	103,400	3.0	96.0	76.2
No. 14.....	8	0.0659	58.8	108,600	3.5	97.9	69.0
No. 14.....	10	0.0654	68.3	114,300	3.8	99.1	82.3
No. 16.....	0	0.0513	0	65,100	47.0	51.1	5.3
No. 16.....	2	0.0522	21.6	85,600	12.0	89.6	65.4
No. 16.....	4	0.0506	40.1	97,400	3.0	93.8	70.6
No. 16.....	6	0.0513	51.3	104,500	3.0	96.6	75.3
No. 16.....	8	0.0500	60.8	111,400	3.0	98.9	79.3
No. 16.....	10	0.0510	67.6	114,500	3.0	99.7	80.9
No. 18.....	0	0.0418	0	65,300	45.5	50.6	7.4
No. 18.....	2	0.0413	19.4	81,100	16.0	85.3	59.0
No. 18.....	4	0.0415	37.6	97,400	4.0	93.4	70.7
No. 18.....	6	0.0415	50.9	103,600	2.5	94.9	73.4
No. 18.....	8	0.0398	62.1	110,400	2.0	96.8	77.2
No. 18.....	10	0.0419	66.8	115,000	2.5	98.9	79.1
No. 20.....	0	0.0326	0	67,600	44.0	56.2	14.7
No. 20.....	2	0.0335	19.9	80,700	16.5	85.0	59.6
No. 20.....	4	0.0330	35.4	96,300	3.0	90.9	69.9
No. 20.....	6	0.0325	51.4	107,200	2.0	94.6	74.9
No. 20.....	8	0.0328	61.4	109,500	2.0	95.3	76.3
No. 20.....	10	0.0339	67.8	113,600	2.0	96.7	77.8
No. 22.....	0	0.0286	0	74,600	35.5	67.6	34.5
No. 22.....	2	0.0271	16.5	80,600	19.5	83.4	57.2
No. 22.....	4	0.0259	37.8	98,700	2.0	91.3	70.1
No. 22.....	6	0.0260	49.6	107,000	1.5	94.0	74.2
No. 22.....	8	0.0262	60.6	112,200	1.5	95.5	76.6
No. 22.....	10	0.0264	68.7	113,500	2.0	96.0	77.3
No. 24.....	0	0.0209	0	66,900	42.0	60.6	19.0
No. 24.....	2	0.0210	19.4	94,900	8.5	90.6	68.8
No. 24.....	4	0.0211	34.9	98,700	2.0	90.9	69.0
No. 24.....	6	0.0208	49.1	107,000	1.0	93.4	73.2
No. 24.....	8	0.0211	58.8	112,400	1.5	95.5	76.1
No. 24.....	10	0.0209	68.9	116,200	1.5	96.6	78.5

TABLE XIII—*Continued*

B. & S. Gage	Temper, B. & S. Numbers Hard	Thickness, in.	Actual Percentage Reduction by Rolling	Tensile Strength, lb. per sq. in.	Elongation in 2 in., per cent	Rockwell Hardness, "B" Scale, 1/16-in. Ball (Red Figures)	
						100-kg. Load	150-kg. Load
No. 26.....	0	0.0177	0	65,200	41.0	57.9	2.7
No. 26.....	2	0.0166	21.9	83,000	11.5	86.3	58.7
No. 26.....	4	0.0167	35.8	108,200	2.0	94.2	72.6
No. 26.....	6	0.0172	40.3	106,600	1.5	93.3	71.9
No. 26.....	8	0.0166	60.2	112,600	1.0	95.4	74.8
No. 26.....	10	0.0170	67.4	114,700	1.0	96.2	77.4
No. 28.....	0	0.0139	0	64,500	34.0	59.8	3.3
No. 28.....	2	0.0136	20.3	83,300	12.5	84.9	50.3
No. 28.....	4	0.0126	38.5	102,800	1.5	92.4	63.8
No. 28.....	6	0.0128	51.4	116,300	1.5	96.3	72.7
No. 28.....	8	0.0133	58.4	111,500	1.0	95.4	73.3
No. 28.....	10	0.0133	68.0	116,400	1.0	96.3	74.7
No. 30.....	0	0.0106	0	65,300	39.0	56.2	....
No. 30.....	2	0.0111	18.4	77,500	14.5	80.6	....
No. 30.....	4	0.0112	35.9	95,600	2.5	89.2	....
No. 30.....	6	0.0105	46.4	107,500	1.0	94.0	....
No. 30.....	8	0.0107	59.1	119,500	1.0	97.1	....
No. 30.....	10	0.0106	66.7	115,100	1.0	96.7	....
No. 32.....	0	0.0092	0	67,000	38.0	55.1	....
No. 32.....	2	0.0087	17.0	76,800	16.0	71.5	....
No. 32.....	4	0.0089	34.6	91,200	2.0	83.0	....
No. 32.....	6	0.0092	49.1	102,400	1.0	90.4	....
No. 32.....	8	0.0084	60.5	112,600	1.0	93.4	....
No. 32.....	10	0.0086	66.9	122,200	1.0	97.6	....
No. 34.....	0	0.0070	0	66,200	33.0	66.4	....
No. 34.....	2	0.0068	23.5	86,200	6.5	75.7	....
No. 34.....	4	0.0066	40.0	96,300	2.0	79.1	....
No. 34.....	6	0.0069	50.0	100,500	1.0	81.0	....
No. 34.....	8	0.0068	61.1	109,200	1.0	83.3	....
No. 34.....	10	0.0067	68.3	115,900	1.0	87.2	....
No. 36.....	0	0.0055	0	69,400	34.0	78.9	....
No. 36.....	2	0.0063	23.0	83,200	8.5	81.5	....
No. 36.....	4	0.0058	35.3	96,500	1.5	81.6	....
No. 36.....	6	0.0055	45.0	102,700	1.0	83.1	....
No. 36.....	8	0.0059	57.7	104,100	1.0	82.9	....
No. 36.....	10	0.0057	67.1	111,100	1.0	84.9	....

Note.—Tensile strength values are an average of 3 specimens in each case; Rockwell hardness values are an average of 15 determinations of 5 readings on each of 3 tension test specimens.

percentage reduction are shown in Fig. 14. Table XII shows the Rockwell hardness and tensile strength limits for alloy B. The chemical composition limits are given in Table VI.

Referring to the curve plotted in Fig. 14, showing the tensile strength limits plotted against percentage of reduction, it is seen that one series lies close to the upper curve whereas the remaining rolling

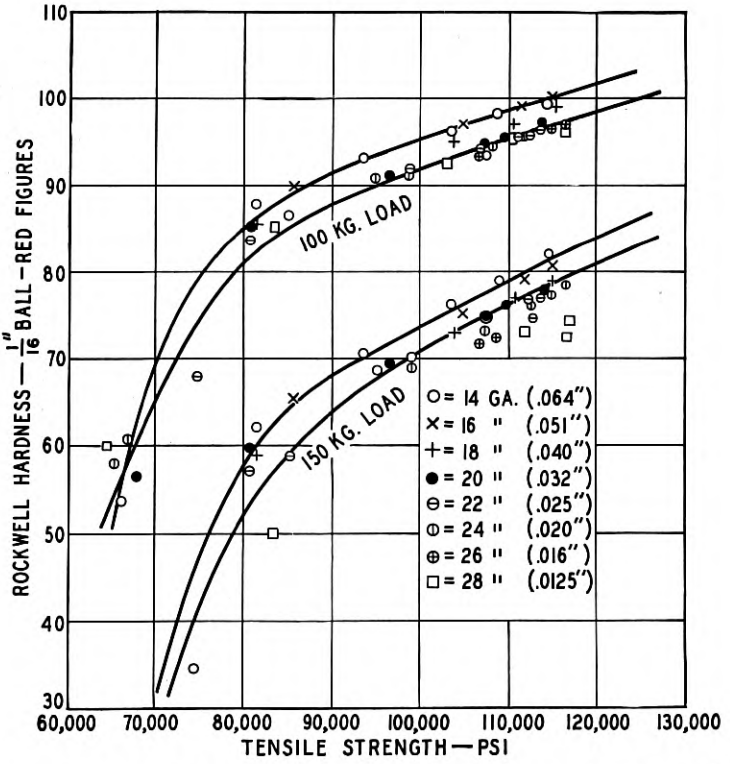


Fig. 13—Relation of Rockwell Hardness to Tensile Strength of Alloy B Nickel-Silver Rolling Series Using 100 and 150-kg. Loads.

series are nearer to the lower curve. This is due to the fact that the bar rolled to No. 22 B. & S. gage sheet had a lighter anneal in the ready-to-finish condition than the other bars, but it was felt that this anneal was close to the commercial range of annealing practice and consequently this series was given some weight when the commercial limits were drawn up.

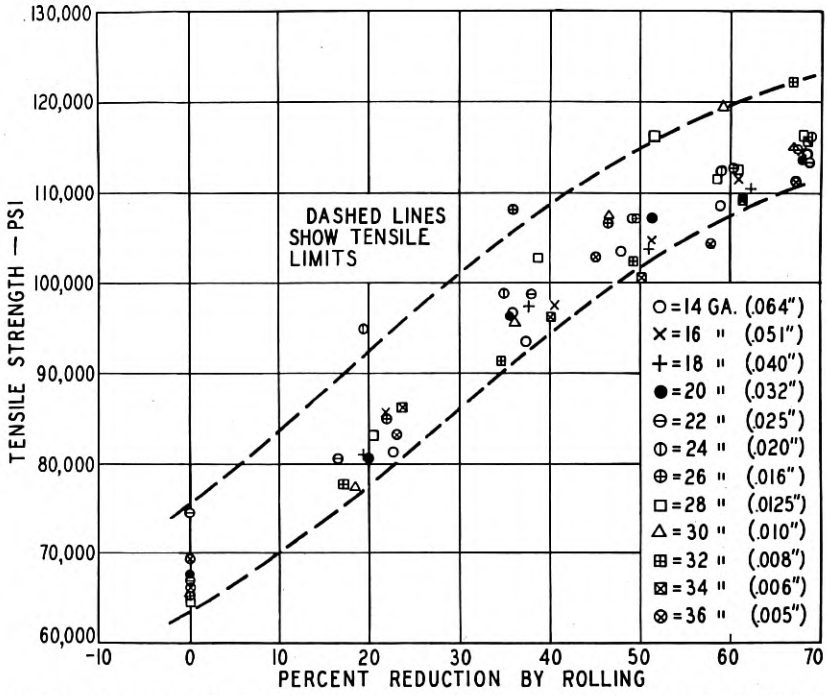


Fig. 14—Relation of Tensile Strength to Percentage of Reduction of Alloy B Nickel-Silver Rolling Series.

TABLE XIV

PHYSICAL PROPERTIES OF ALLOY A PHOSPHOR-BRONZE SHEET

Temper, B. & S. Numbers Hard	B. & S. Gage, in.	Actual Percentage Reduction by Rolling	Tensile Strength, lb. per sq. in.	Elongation in 2 in., per cent	Rockwell Hardness, "B" Scale, 1/16-in. Ball (Red Figures)	
					100-kg. Load	150-kg. Load
Soft.....	0.0636	0	46,900	53	23.8	—6.9
2.....	0.0501	21.2	61,200	21.8	75.9	43.3
4.....	0.0395	37.8	75,150	8.0	84.8	58.3
6.....	0.0321	49.5	85,700	4.0	88.9	65.2
8.....	0.0248	61.0	93,900	2.5	91.4	68.8
10.....	0.0203	68.1	97,450	2.0	92.2	69.7
Soft.....	0.0324	0	47,000	51	26.6	—7.6
2.....	0.0242	25.2	62,500	15.5	75.3	43.2
4.....	0.0199	38.6	73,600	6	80.6	48.0
6.....	0.0166	48.8	83,950	2.5	84.4	55.0
8.....	0.0126	61.0	91,900	1.5	88.1	....
10.....	0.0105	67.6	97,350	1.5	89.3	....
Soft.....	0.0162	0	46,350	38.5	....	....
2.....	0.0129	20.4	61,250	15.5	....	....
4.....	0.0107	33.3	72,800	4.5	....	....
6.....	0.0082	49.2	84,900	1.5	....	....
8.....	0.0062	61.7	94,450	1.3	....	....
10.....	0.0050	68.8	96,400	1	....	....

Note.—Tensile strength and percentage elongation are an average of 3 specimens in each case; Rockwell hardness values are an average of 15 determinations of 5 readings on each of 3 tension test specimens.

Phosphor Bronze:

Two grades of phosphor bronze have been investigated, designated alloys A and C, containing 4 and 8 per cent of tin, respectively.

Alloy A Phosphor-Bronze Sheet.—Three rolling series representative of the entire number investigated were made from the bar of metal of the analysis shown in Table VII. These three series were rolled to all even numbers hard from 2 to 10, B. & S. gage for Nos. 14,

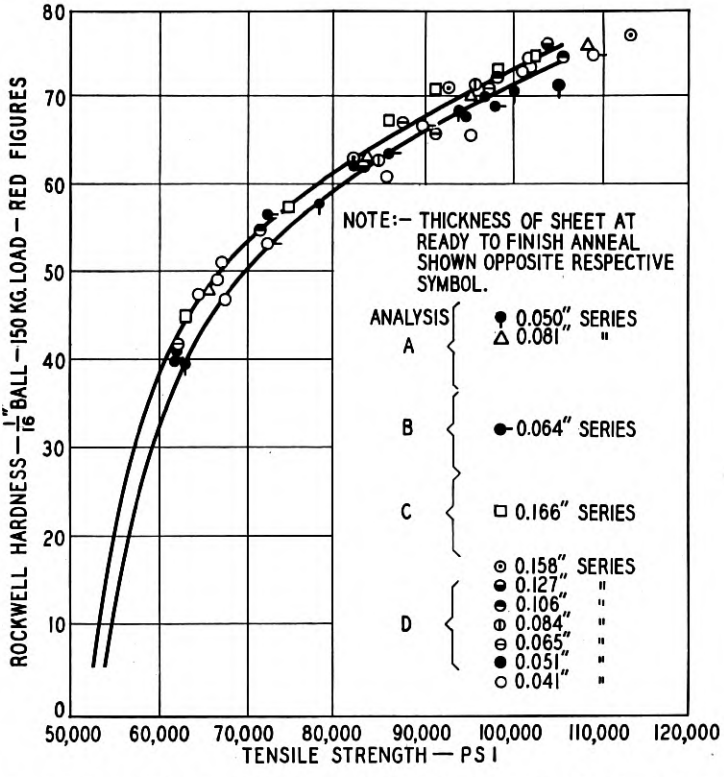


Fig. 15—Relation of Rockwell Hardness to Tensile Strength of Alloy A Phosphor-Bronze Rolling Series.

20 and 28. The physical properties are given in Table XIV. The Rockwell hardness-tensile strength relationship for alloy A is given in Fig. 15 and the tensile strength-reduction relationship is shown in Fig. 16. The Rockwell hardness and tensile strength limits for alloy A are given in Table XV and the chemical composition requirements are given in Table VI.

Alloy C Phosphor-Bronze Sheet.—The alloy C rolling series were made from five bars of metal having the average composition shown in Table VII. This material was rolled to all even hardness numbers from 2 to 10 numbers hard, for all even B. & S. gage numbers from

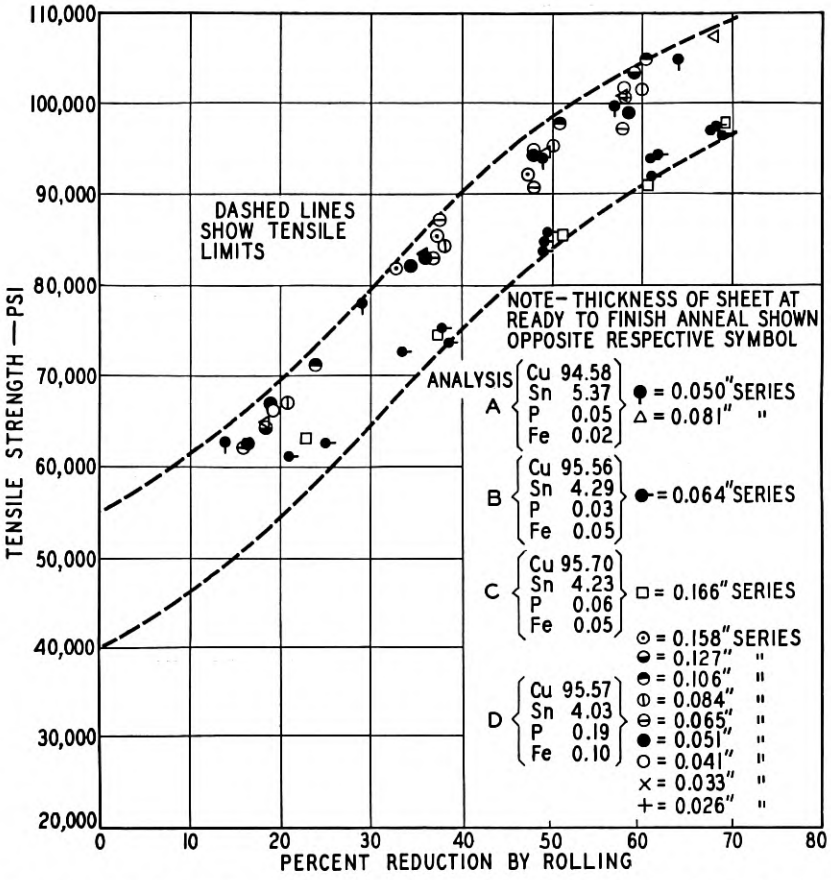


Fig. 16—Relation of Tensile Strength to Percentage of Reduction by Rolling of Alloy A Phosphor-Bronze Rolling Series.

No. 14 to No. 36, inclusive. The physical properties of this material are given in Table XVI. The Rockwell hardness-tensile strength relationship is shown in Fig. 17 and the tensile strength-reduction relationship is shown in Fig. 18. The physical requirements for alloy C are given in Table XV and the chemical requirements in Table VI.

TABLE XV

ROCKWELL HARDNESS AND TENSILE STRENGTH LIMITS FOR ALLOYS A AND C, PHOSPHOR-BRONZE SHEET

Thickness	Temper, B. & S. Numbers Hard	Nominal Percentage Reduction by Rolling	Alloy A Phosphor Bronze				Alloy C Phosphor Bronze			
			Tensile Strength, lb. per sq. in.		Rockwell Hardness, "B" Scale, 1/16-in. Ball, 150-kg. Load (Red Figures)		Tensile Strength, lb. per sq. in.		Rockwell Hardness, "B" Scale, 1/16-in. Ball, 150-kg. Load (Red Figures)	
			Minimum	Maximum	Minimum	Maximum	Minimum	Maximum	Minimum	Maximum
0.040 in. and over	Half hard 8	20.7	55,000	70,000	20	53	69,000	84,000	47	68
0.020 to 0.040 in.	Half hard 8	20.7	55,000	70,000	15	51	69,000	84,000	40	64
0.040 in. and over	Hard 4	37.1	72,000	87,000	55	66	85,000	100,000	69	77
0.020 to 0.040 in.	Hard 4	37.1	72,000	87,000	53	64	85,000	100,000	65	74
0.040 in. and over	Extra hard 6	50.0	84,000	98,500	64	73	97,000	111,500	76	82
0.020 to 0.040 in.	Extra hard 6	50.0	84,000	98,500	62	71	97,000	111,500	73	79
0.040 in. and over	Spring 8	60.5	91,000	105,000	69	75	105,000	118,500	79	85
0.020 to 0.040 in.	Spring 8	60.5	91,000	105,000	67	73	105,000	118,500	76	82
0.040 in. and over	Extra spring 10	68.7	96,000	109,000	72	77	109,500	122,000	81	86
0.020 to 0.040 in.	Extra spring 10	68.7	96,000	109,000	70	75	109,500	122,000	78	83

TABLE XVI

PHYSICAL PROPERTIES OF ALLOY C PHOSPHOR-BRONZE SHEET

B. & S. Gage	Temper, B. & S. Numbers Hard	Thickness, in.	Actual Percentage Reduction by Rolling	Tensile Strength, lb. per sq. in.	Elongation in 2 in., per cent	Rockwell Hardness, "B" Scale, 1/16-in. Ball (Red Figures)	
						100-kg. Load	150-kg. Load
No. 14.....	0	0.0663	0	62,400	68.0	56.3	14.7
No. 14.....	2	0.0640	22.0	72,800	36.0	83.7	56.0
No. 14.....	4	0.0649	37.5	86,900	17.0	93.0	71.3
No. 14.....	6	0.0659	50.5	100,900	10.0	97.3	78.3
No. 14.....	8	0.0671	59.4	110,300	7.5	99.6	81.6
No. 14.....	10	0.0669	67.7	116,300	6.5	101.4	84.7
No. 16.....	0	0.0515	0	54,700	76.0	42.0	—7.3
No. 16.....	2	0.0536	19.4	80,400	33.0	88.4	63.6
No. 16.....	4	0.0539	35.4	83,800	20.0	91.9	67.7
No. 16.....	6	0.0519	49.8	97,100	9.0	96.4	76.2
No. 16.....	8	0.0535	60.2	107,700	7.0	99.2	80.6
No. 16.....	10	0.0549	67.1	115,700	7.0	101.1	83.2
No. 18.....	0	0.0429	0	60,400	71.0	50.7	7.7
No. 18.....	2	0.0412	20.6	71,200	36.0	82.3	55.0
No. 18.....	4	0.0430	34.9	99,200	11.5	96.0	73.4
No. 18.....	6	0.0431	48.2	95,700	10.0	94.5	73.4
No. 18.....	8	0.0420	59.9	106,500	5.5	97.7	78.4
No. 18.....	10	0.0416	69.3	116,800	4.0	100.2	82.6
No. 20.....	0	0.0336	0	61,200	65.5	50.2	6.7
No. 20.....	2	0.0344	18.6	78,300	34.0	85.4	60.3
No. 20.....	4	0.0332	35.6	87,500	15.0	91.0	69.1
No. 20.....	6	0.0334	49.4	114,200	5.0	98.3	80.4
No. 20.....	8	0.0347	57.9	105,900	5.0	96.1	77.1
No. 20.....	10	0.0332	68.5	113,600	3.0	98.6	80.7
No. 22.....	0	0.0264	0	53,700	74.0	45.8	—0.4
No. 22.....	2	0.0267	22.0	80,200	32.0	85.4	60.8
No. 22.....	4	0.0273	35.5	95,400	12.0	92.9	72.8
No. 22.....	6	0.0263	49.5	102,400	7.0	94.6	75.1
No. 22.....	8	0.0260	61.3	121,000	3.0	99.0	82.9
No. 22.....	10	0.0266	67.7	112,100	3.0	97.3	80.3
No. 24.....	0	0.0221	0	59,700	67.5	52.3	11.0
No. 24.....	2	0.0200	22.4	77,300	26.0	85.8	60.2
No. 24.....	4	0.0219	35.2	95,500	14.0	92.2	71.4
No. 24.....	6	0.0221	48.8	108,500	5.0	95.9	77.1
No. 24.....	8	0.0213	55.1	110,700	3.0	96.9	78.8
No. 24.....	10	0.0216	67.4	124,800	2.0	100.0	84.0

TABLE XVI—*Continued*

B. & S. Gage	Temper, B. & S. Numbers Hard	Thickness, in.	Actual Percentage Reduction by Rolling	Tensile Strength, lb. per sq. in.	Elongation in 2 in., per cent	Rockwell Hardness, "B" Scale, 1/16-in. Ball (Red Figures)	
						100-kg. Load	150-kg. Load
No. 26.....	0	0.0174	0	54,300	69.0	57.3	25.0
No. 26.....	2	0.0176	22.6	73,600	38.0	81.5	52.6
No. 26.....	4	0.0178	32.9	86,800	15.0	89.2	65.9
No. 26.....	6	0.0166	49.9	108,500	4.0	96.1	76.3
No. 26.....	8	0.0173	60.1	116,600	2.0	97.3	78.7
No. 26.....	10	0.0168	65.7	115,700	2.0	97.7	80.3
No. 28.....	0	0.0130	0	57,500	58.0	61.6	26.5
No. 28.....	2	0.0129	23.2	73,400	30.0	82.4	46.4
No. 28.....	4	0.0137	40.1	91,500	11.0	90.8	66.1
No. 28.....	6	0.0128	52.9	105,300	4.0	95.1	69.6
No. 28.....	8	0.0150	55.6	107,000	3.0	95.1	74.1
No. 28.....	10	0.0138	66.4	117,600	1.0	99.0	81.1
No. 30.....	0	0.0110	0	55,400	61.0	65.3	43.1
No. 30.....	2	0.0110	17.6	72,400	27.5	85.2	50.3
No. 30.....	4	0.0111	34.2	86,500	12.5	90.3	54.2
No. 30.....	6	0.0109	53.0	102,200	2.5	94.5	63.1
No. 30.....	8	0.0115	58.8	101,900	1.0	96.1	69.7
No. 30.....	10	0.0114	66.7	111,000	1.5	98.3	74.4
No. 32.....	0	0.0086	0	58,200	52.0	....	....
No. 32.....	2	0.0077	28.6	89,000	8.0	....	....
No. 32.....	4	0.0087	28.6	91,400	9.0	....	....
No. 32.....	6	0.0090	47.6	97,200	3.0	....	....
No. 32.....	8	0.0088	61.3	108,200	1.5	....	....
No. 32.....	10	0.0088	67.0	115,300	1.5	....	....
No. 34.....	0	0.0063	0	63,900	50.0	....	....
No. 34.....	2	0.0063	24.0	81,900	17.5	....	....
No. 34.....	4	0.0061	41.9	99,100	7.0	....	....
No. 34.....	6	0.0063	44.5	111,300	2.5	....	....
No. 34.....	8	0.0064	63.4	110,200	1.5	....	....
No. 34.....	10	0.0066	70.0	118,800	1.0	....	....
No. 36.....	0	0.0054	0	61,300	51.0	....	....
No. 36.....	2	0.0054	16.1	77,900	25.0	....	....
No. 36.....	4	0.0049	35.4	97,900	4.5	....	....
No. 36.....	6	0.0053	52.4	103,000	4.0	....	....
No. 36.....	8	0.0054	58.0	114,800	1.5	....	....
No. 36.....	10	0.0059	68.3	110,500	1.5	....	....

Note.—Tensile strength and percentage elongation are an average of 3 specimens in each case; Rockwell hardness values are an average of 15 determinations of 5 readings on each of the 3 tension test specimens.

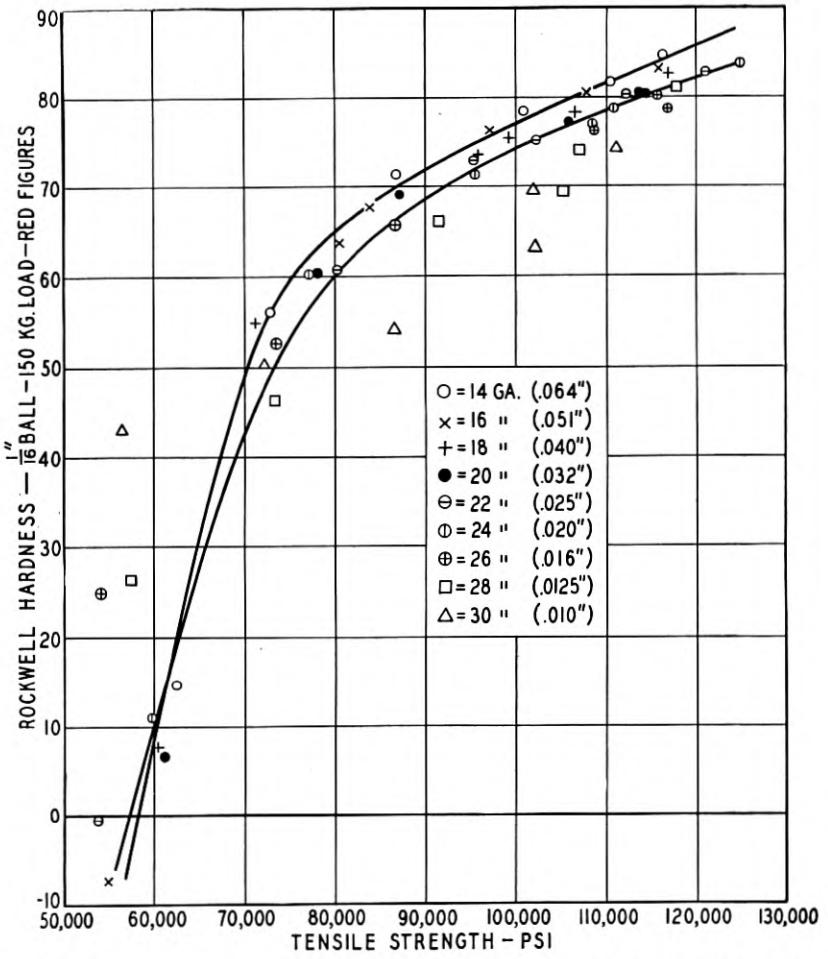


Fig. 17—Relation of Rockwell Hardness to Tensile Strength of Alloy C Phosphor-Bronze Rolling Series.

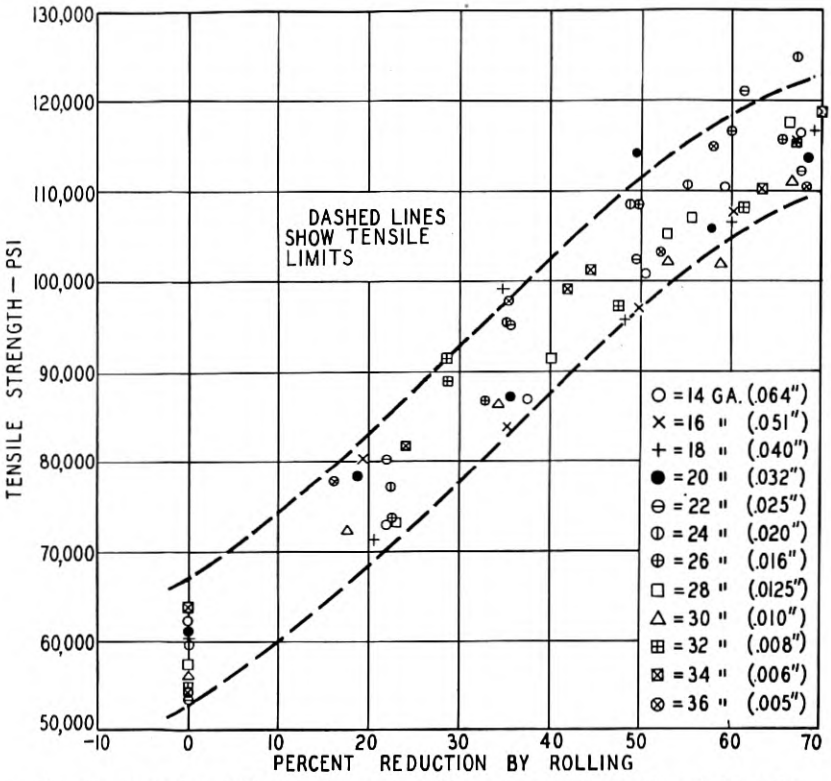


Fig. 18—Relation of Tensile Strength to Percentage of Reduction of Alloy C Phosphor-Bronze Rolling Series.

Experience Data

During and subsequent to the laboratory work covered in this paper, a series of data was obtained on regular commercial shipments of non-ferrous sheet material. About 700 lots of brass, nickel-silver and phosphor-bronze sheet in various tempers and grades were tested on the Rockwell and scleroscope by both the producer and consumer. The lots ranged in size from 100 to about 50,000 lbs. of sheet metal and included material from at least five different mills. The sheet ranged in thickness from 0.010 to 0.500 in. A representative sample about 6 in. long and from 1 to 6 in. wide was taken from each lot. Tests were made on identical samples, five readings being made with each instrument and the average taken. All tests were made by the inspectors who were normally responsible for the control of the material, even though they were not in all cases familiar with the operation of the Rockwell machine. The Rockwell machines were, of

course, put in good mechanical condition and the methods of test outlined in Appendix I were followed. The results obtained constituted an independent verification of the conclusion that the Rockwell machine was more sensitive but much less subject to variation and personal error than the scleroscope.

Figure 19 is representative of the results obtained, showing a

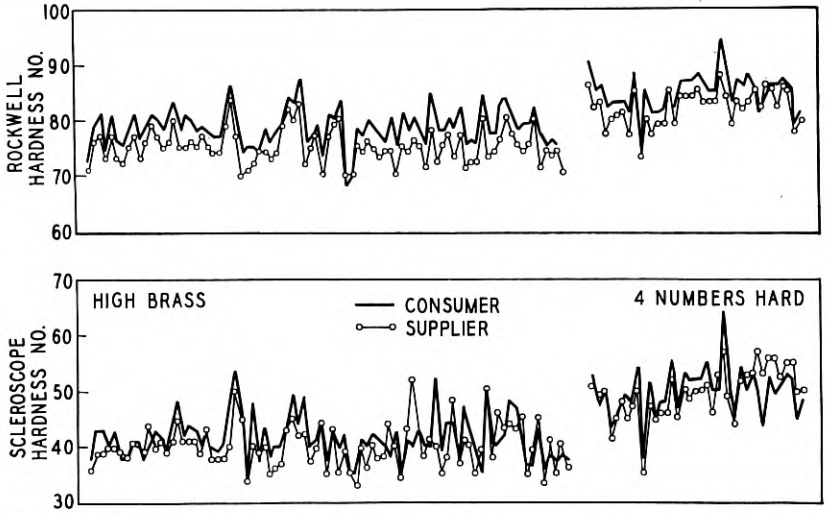


Fig. 19—Typical Comparisons of Rockwell and Scleroscope Machines in Commercial Tests on Brass.

direct comparison of the Rockwell and the scleroscope readings on two tempers of high brass. It illustrates the consistency of the readings of the two Rockwell instruments as compared with the two scleroscopes. Figure 20 represents similar results obtained on nickel

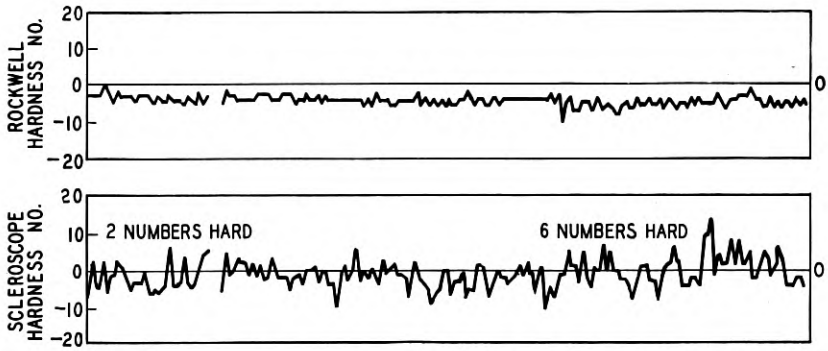


Fig. 20—Typical Results Showing Differences in Rockwell and Scleroscope Machines in Commercial Tests on Nickel Silver.

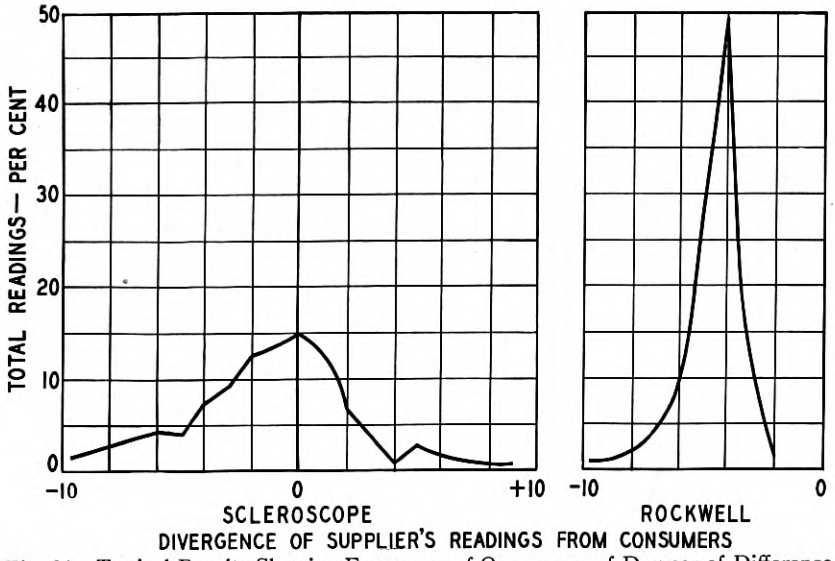


Fig. 21—Typical Results Showing Frequency of Occurrence of Degrees of Difference Between Two Machines.

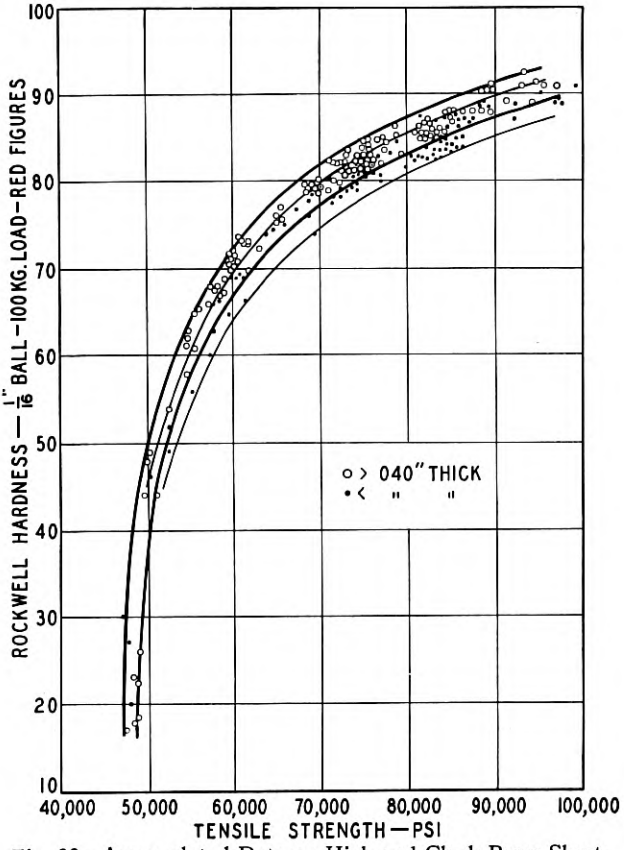


Fig. 22—Accumulated Data on High and Clock-Brass Sheet.

silver plotted to show directly the differences between readings on the two machines of each type, one being assumed as a standard. Figure 21 shows the frequency of occurrence of the various degrees of difference between the two instruments, giving results for both the Rockwell and the scleroscope. One of the instruments was taken as a standard and the variations of the other instrument from it are plotted to left or right as they are negative or positive. The fre-

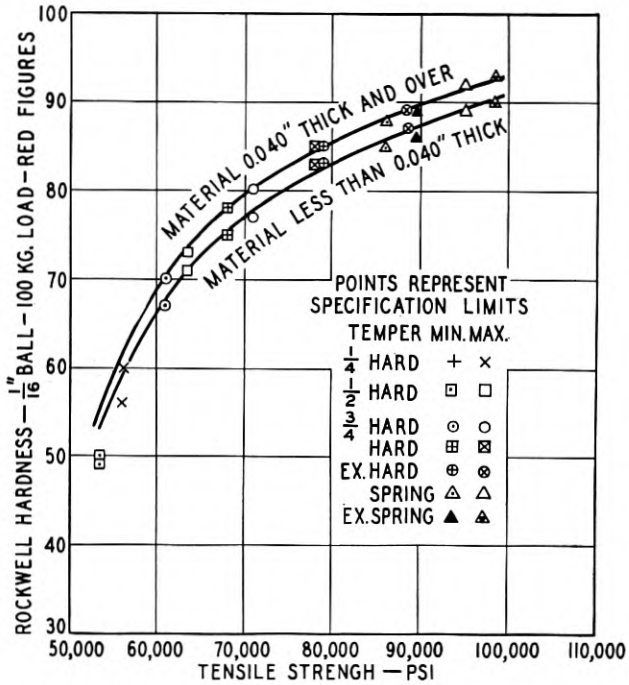


Fig. 23—Comparison of Curves of Results from Experience Data with Limits Adopted, for High-Brass Sheet.

quency of occurrence of differences of the magnitudes represented by the abscissas are plotted as ordinates in per cent of the total readings made on the group of samples for which the curve is drawn. In this typical case it is seen that about 50 per cent of the Rockwell readings from one machine are 4 points below those from the other machine. Application of a conversion factor would, therefore, move the curve to the right 4 points and very close checks between the two machines would result. The scleroscope curve shows that readings

of the two instruments agree on only about 15 per cent of the total, and that in this case the differences are both positive and negative, thus not permitting the application of a conversion factor.

Experience data were also collected for Rockwell hardness and

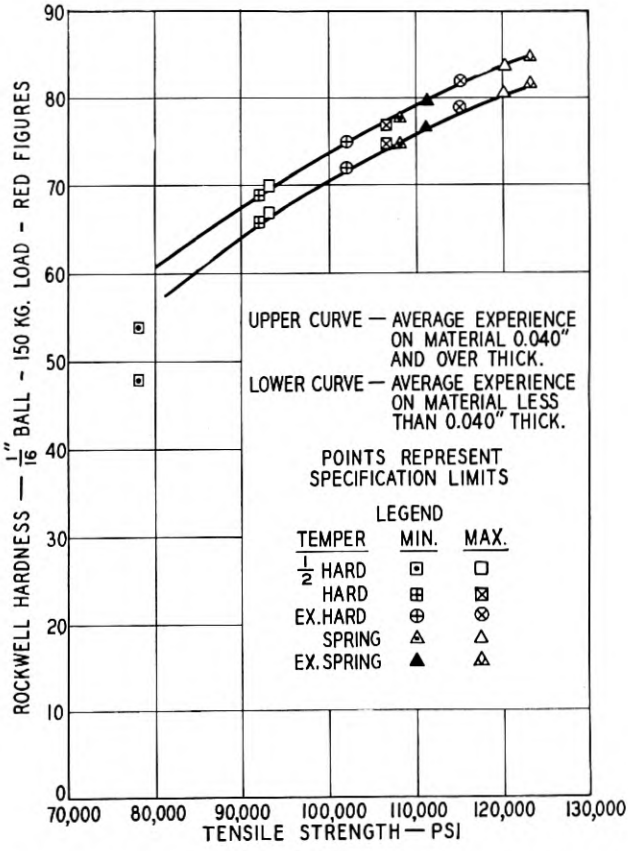


Fig. 24—Comparison of Curves of Results from Experience Data with Limits Adopted, for Alloy B Nickel-Silver Sheet.

tensile strength of the metals, the values being averages of tests on individual shipments.

Figure 22 shows graphically the data obtained on high and clock-brass sheet, and Fig. 23 the experience data which have been averaged and shown as a smooth curve for comparison with the specification

limits shown as points for high-brass sheet. The agreement is close. Figures 24, 25, and 26 show similar smooth curves of experience data for nickel silver and phosphor bronze. In these latter cases the experience data are not sufficiently complete to verify the specification limits in all tempers, but in general the accuracy of the specification limits is verified.

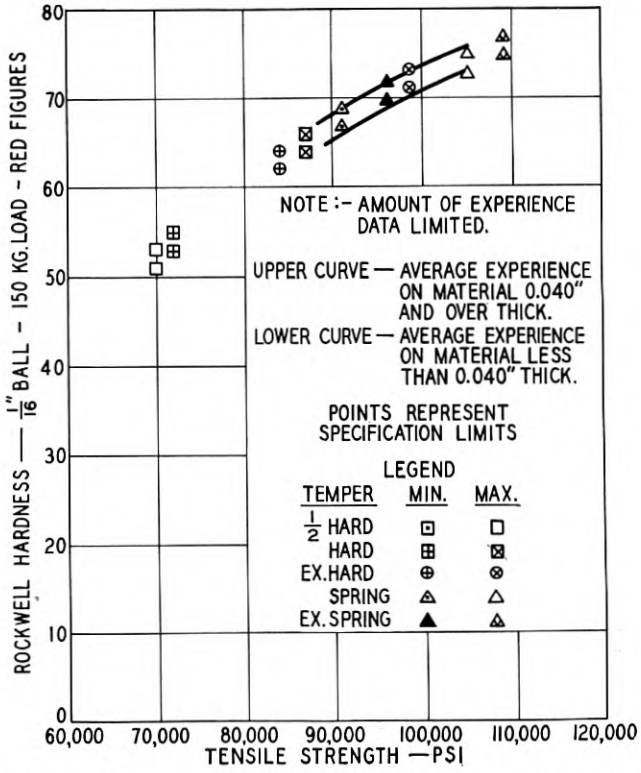


Fig. 25—Comparison of Curves of Results from Experience Data with Limits Adopted, for Alloy A Phosphor-Bronze Sheet.

In addition to the data, the experience of mills supplying material indicates that the limits proposed are satisfactory. Instances where the limits were not met served only to emphasize the advantages of control of the physical properties. Variations from the nominal annealing temperature, composition or percentage reduction in all cases explained the variations from the specification limits.

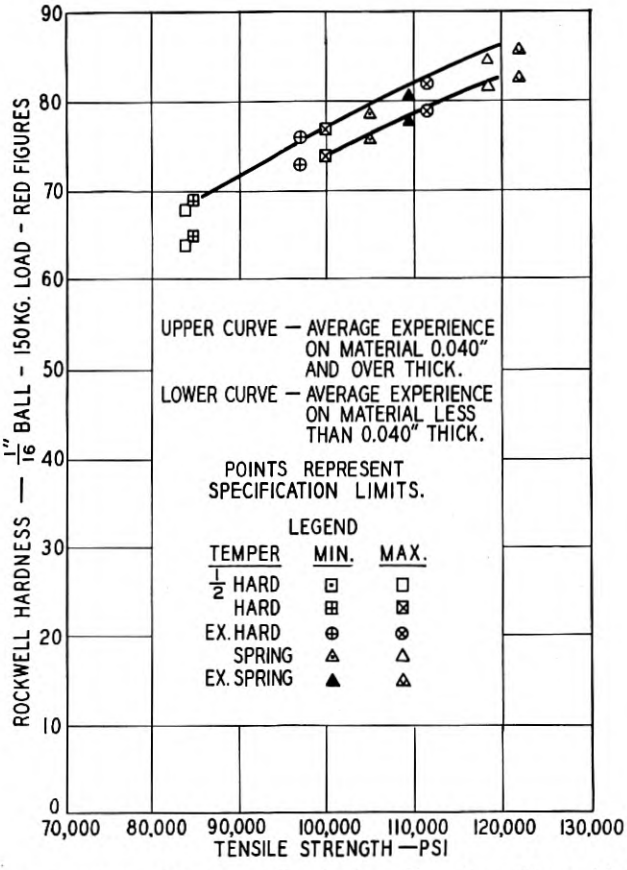


Fig. 26—Comparison of Curves of Results from Experience Data with Limits Adopted, for Alloy C Phosphor-Bronze Sheet.

Conclusion

The purposes stated in this paper have been accomplished, namely, the development of a commercial method of test for sheet non-ferrous metals and limits for use in commercial specifications.

The conclusion reached that the tension test is the best available static test for non-ferrous sheet metal in the hard tempers has been confirmed.

Because of its close correlation to the tension test, the Rockwell hardness test has been adopted for preliminary inspection purposes. Other hardness tests have been studied but this test is the most satisfactory for sheet metals 0.020 in. thick and thicker. It has limited use for material less than 0.020 in. thick. Material near to or outside

of the Rockwell hardness limits is subjected to the tension test. Rejection of material is based solely on the tension test.

Refinements in the application of the Rockwell hardness test as an inspection instrument have been worked out and it is shown that close agreement can be obtained between producer and consumer on commercial shipments of material.

The necessity for the adoption of a series of Rockwell standard blocks covering the range of hardness of certain non-ferrous sheet and the calibration of testing machines to these standards is emphasized.

Grain-size limits are given for annealed brass of four compositions and for one nickel-silver alloy. This is the most satisfactory method of controlling the annealed material.

Tentative tensile strength and Rockwell hardness limits for brass alloy G, nickel-silver and phosphor-bronze alloys have been developed. These limits are considered preliminary until more complete information on actual shipments of material is available. The limits for high and clock brass are considered entirely satisfactory.

A modified bend test has been presented as having considerable value in determining the forming and drawing qualities of sheet metals.

APPENDIX I

STANDARD TEST PROCEDURE FOR ROCKWELL HARDNESS TESTS

Standard Test Blocks

1. Standard test blocks are in the possession of the Bell Laboratories and sub-standard blocks are calibrated from these blocks. These sub-standard blocks are used to calibrate the Rockwell machines. All specifications are written in terms of these standard blocks. Before a Rockwell machine is used for tests on metal supplied to specifications it should be calibrated by the sub-standard block.

Adjustments

2. (*a*) *Dash Pot.*—The dash pot on the Rockwell tester shall be so adjusted that the operating handle completes its travel in from five to ten seconds with no specimen on the machine and with the machine set up to apply a major load of 100 kg.

(*b*) *Index Lever Adjustment.*—As specified in the Rockwell tester instruction book, the following tests (and adjustments, if necessary) should be made.

"Put a piece of material on the anvil and turn the capstan elevating nut to bring the material up against the ball penetrator. Keep turning to elevate the material until the hand feels positive resistance to further turning, which will be felt after the 10-kg. minor load has been picked up and when the major load is encountered. When excessive power would have to be used to raise the work higher, take note of the position of the pointer on the dial. After setting the dial so that C-0 and B-30 are at the top then:

(*1*) If pointer stands between B-50 and B-70 no adjustment is needed.
(*2*) If pointer stands between B-45 and B-50 adjustment is advisable.
(*3*) If it stands anywhere else, adjustment is imperative.

"As the pointer revolves several times when the work is being elevated it is pointed out here that the readings mentioned apply to that revolution of the pointer

which occurs as the reference mark on the gage stem disappears into the sleeve. The object of the adjustment is to see that the elevation of the specimen to pick up the minor load shall not be carried so far as to cause even a partial application of the major load, which to make a proper test, must be applied only through the release mechanism."

"*To Make the Index Lever Adjustment, if Necessary.*—When the test piece is elevated until it starts to pick up the major load, loosen the lock nut of the screw through the index lever which carries at its lower end a small steel plate engaging the ball on the penetrator shaft while using a small screw-driver to firmly hold the screw until after the nut is loose. Then turn the screw very slightly and note result on position of gage pointer which should be at B-60 to B-70 before reclamping the screw with its lock nut."

(*c*) *Protection Against Vibration.*—If the bench or table on which the Rockwell tester is mounted is subject to vibration, such as felt in the vicinity of other machines, the tester should be mounted on a metal plate on sponge rubber at least 1 in. thick or on any type of mounting that will effectually eliminate vibration from the machine. Otherwise the pointer will penetrate farther into the material than when such vibrations are absent.

Cushioning of Latch.—If, when the operating handle is being returned to its normal position, the latch operates with such a snap as to noticeably change the position of the dial pointer, felt or rubber washers should be placed under the trip button in order to cushion this blow. If this snap is severe, difference in reading of several hardness numbers may result.

Testing Methods

3. The anvil used shall have a polished bearing surface for the material of about $\frac{3}{16}$ in. in diameter. It shall be hard enough so that no visible indentation is made when the thinnest material is tested.

Before using the machine it shall be operated several times on a piece of scrap material in order to firmly settle the penetrator, anvil and moving parts of the machine. This should be done every day before the machine is used.

Make a reading on at least one of the standard test blocks. This need not be done every day if it has been found that the machine does not change.

The Rockwell ball should be replaced by a new one occasionally and the operator should be on the lookout for any permanent deformation of the ball which will ordinarily be indicated by high readings on the standard test blocks.

When applying the minor load, the capstan screw should be turned up so that the pointer stops at "O" with a maximum variation of ± five divisions. The last movement of this screw must always be in such a direction as to elevate the specimen.

In applying the major load the operating handle shall be allowed to revolve without interference until the major load is completely applied. This may be observed in two ways, (1) when the pointer suddenly slows down, or (2) watching the weight arm to see that it is completely free from the control of the dash pot. When the major load has thus been completely applied, the operating handle shall be immediately brought back to the latched position without waiting for it to complete its revolution. This should be accomplished in less than two seconds after the major load is completely applied.

The test specimen shall be well supported to prevent errors due to the overhang of the specimen.

All tests shall be made upon one thickness of the material.

APPENDIX II

STANDARD TEST PROCEDURE FOR TENSION TESTING OF THIN SHEET METALS

General

1. (*a*) These specifications cover general methods for testing the tensile strength of thin sheet, strip and flat wire materials for use in the manufacture of telephone apparatus. They shall form part of individual specifications covering sheet metals for specific purposes whenever so provided in the individual specifications.

NOTE.—By thin sheet material is meant material 0.05 in. or less, in thickness. This thickness lies in the range where it becomes necessary to abandon the use of the ordinary micrometer and to use the barrel type micrometer or a still more sensitive device in order to keep the probable error of measurement as low as possible.

(*b*) In case of any difference in the provisions of this and the individual specification, the provisions of the individual specification shall govern.

SPECIMENS FOR TENSION TESTS

2. *Selection of Specimen.*—The test specimens shall be chosen in such numbers and from such locations in each lot of material as to be representative of the quality of this material. The method of selecting specimens will be specified in the individual specifications when found necessary.

3. *Dimensions of Specimens for Tension Test.*—

(*a*) *Materials ½ in. or Less in Width.*—For strip or flat materials less than ½ in. wide, the specimen shall consist of the appropriate length cut from the material as furnished. For gage lengths of 2 and 8 in., the length of the test specimen shall be not less than 8 and 14 in., respectively.

(*b*) *Materials More Than ½ in. in Width.*—For materials over ½ in. wide, two alternative specimens are allowed.

(*1*) *Standard Specimen:*

The specimen which shall be regarded as standard shall have dimensions and be machined as shown in Fig. 27.

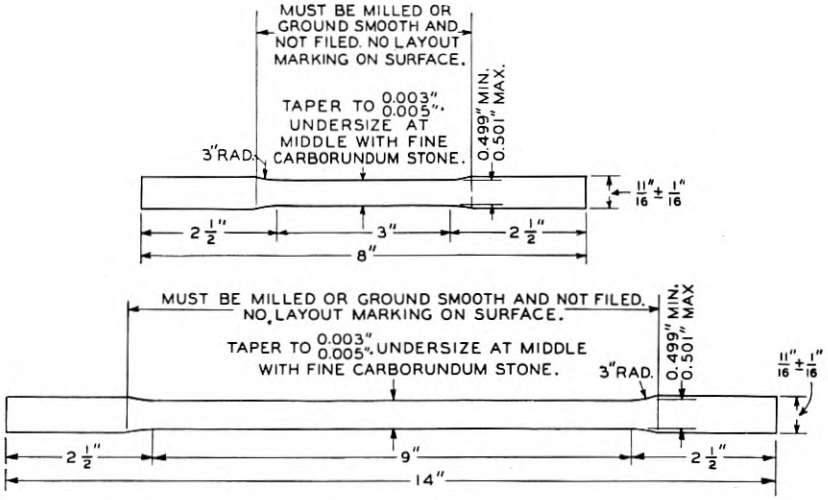


Fig. 27—Tension Test Specimens for Sheet Material Over ½ in. in Width.

(*a*) Specimen for Use in Determining Tensile Strength and Elongation in 2 in.

(*b*) Specimen for Use in Determining Proportional Limit and Modulus of Elasticity.

(*2*) *Alternative Specimen:*

This specimen shall be in the form of a strip up to 1½ in. in width, having parallel sides and having about 0.001 to 0.002 in. of the material removed from the edges of the specimen near the center by grinding with a fine-grained stone. This specimen shall have the lengths given above for materials less than ½ in. wide. It may only be used so long as good breaks (which occur at the point of reduced section and without tearing) are obtained and where the strength of the material does not

approximate the acceptance limits. The milled specimens (see Standard Specimen above) are to be regarded as standard and are to be used where good breaks are not obtained with these wider specimens and as referee test specimens where the strength of the material approximates the acceptance limits.

4. *Cleaning of Specimens.*—Before being measured, each test specimen shall be thoroughly cleaned with carbon tetrachloride applied by rubbing with a soft cloth. When being measured, each specimen shall be wiped clear of dust immediately before being inserted in the measuring instrument.

5. *Measurement of Specimens.*—The dimensions of the specimen shall be measured between flat parallel surfaces of hardened steel, or other appropriate material, approximately $\frac{3}{16}$ in. in diameter which shall be applied in such a manner as to exert on the specimen, while the measurement is being taken, a normal pressure of approximately 16 oz.

6. *Calibration of Measuring Instrument.*—The measuring instrument shall be suitably adjusted and calibrated to indicate the distance between the measuring surfaces of the mandrels when exerting a normal pressure of approximately 16 oz. or 35 lb. per sq. in. of mandrel end area with an error not greater than ± 0.00005 in., *i.e.*, to the nearest 0.0001 in.

Notes

1. When measuring very thin materials, the micrometer or measuring instrument should be adjusted and calibrated with the greatest practicable accuracy.

2. The measuring instrument should be calibrated preferably by means of a hardened steel gage block having flat parallel surfaces and of a thickness approximately equal to that of the material to be measured.

3. In order to obtain an accurate check, it is essential that the mandrel surfaces and faces of the gage be freed from dust immediately before being brought together.

4. If a barrel type micrometer indicating 0.0001 in. per scale division is used, care should be taken to have the instrument in good operating condition. The instrument can then be adjusted to give the first click of the ratchet at a mandrel pressure of 16 oz. ± 4 oz. by suitable adjustment of the ratchet spring, measuring the mandrel pressure by means of a small spring balance, having flat parallel contact surfaces, inserted between the mandrels. The micrometer screw and nut shall be so adjusted that no end play or shake of the spindle may be felt when the spindle is gently pushed back and forth in the direction of the axis of the screw.

5. In calibrating the micrometer against a standard gage block, or in measuring a specimen, the spindle shall be rotated slowly into contact with the work by means of the ratchet until the ratchet clicks once, be retracted about 0.002 in. and again slowly rotated into contact until the ratchet clicks once before reading is taken. In order to avoid exceeding the normal pressure of the mandrels in operating the micrometer, the mandrel shall be brought against the work at a rate not exceeding 0.001 in. per second at both first and second contacts given above. This rate is approximately the maximum rate at which the operator is able to count the half thousandths marks on the barrel as they pass the zero line on the frame.

7. *Method of Computing Cross-Section.*—

(*a*) *Material ½ in. or Less in Width.*—For strip material of flat wire ½ in. or less in width, the width and thickness shall be measured along the major axis of the specimen at the middle point and at distances of approximately 2 and 4 in., respectively, at each side of the middle point. The cross-section of strip and flat wire material shall be assumed to be a rectangle. The area of cross-section to be used in computing the tensile strength of the material shall be the product of the average of the five thickness readings and the average of the five width readings.

(*b*) *Cross-Section of Standard Test Specimen for Material Over ½ in. in Width.*—For the standard test specimen described in Section 3 (*b*), the width shall be measured at the point of minimum width as determined by trial measurements. The thickness shall be measured in a plane normal to the major axis of the specimen at the point of minimum width, at the middle and at each edge of the specimen. When measuring the thickness at the edge of the specimen, the outer edges of the contact surfaces of the measuring instrument shall be approximately $\frac{1}{32}$ in. inside of the edge of the material. The area of cross-section shall be taken as the product of the minimum width and the average of the three thickness readings above specified.

NOTE.—In some cases, it may be found that the thickness of the material near the fillets is so much less than the thickness in the plane of minimum width that the minimum area of cross-section occurs near the fillet. In measuring specimens, the operator should be on the lookout for this condition and, in case it is found, the width of the specimen at the point of minimum width should be further reduced by grinding with a stone until the minimum cross-section is brought to the middle of the specimen.

(*c*) *Cross-Section of Test Specimen Over* ½ *in. in Width.*—For specimens over ½ in. wide as described in Section 3 (*b*), the measurements shall be made as on the standard specimen, as described in Paragraph (*b*) above.

METHOD OF TESTING

8. *Testing Machine.*—

(*a*) *Design.*—Testing machines shall be of a design in which the applied load is balanced by increasing the lever arm of a movable weight or by some other means employing weights without the use of springs.

(*b*) *Method of Mounting.*—The testing machine shall be installed on a solid foundation of such a nature that the machine is maintained in a level position.

(*c*) *Range.*—It is desirable that the testing machine should be capable of applying and measuring a load about 50 per cent greater than the normal anticipated breaking strength of the test specimen.

(*d*) *Sensitivity.*—All bearings and knife edges of the testing machine shall be so proportioned and adjusted that the position of the beam, when balanced at the normal breaking load of the specimen, shall be perceptibly changed by a change of load of 0.1 per cent.

(*e*) *Accuracy.*—The machine shall be calibrated with dead weights so that a load equal to the breaking load of the material tested shall be indicated with an error of less than ± 0.1 per cent.

(*f*) *Method of Applying Load.*—The mechanism for applying load to the test specimen shall be such as to advance the pulling head of the machine at a uniform rate, such as may be obtained by a motor drive operating from a well regulated source of power. The machine shall be capable of being operated at different uniform rates of speed, including the rate of approximately 0.05 in. per minute of the moving head. Unless otherwise specified in individual specifications, the maximum speed of the moving head in making a test shall be 0.025 in. per inch free length of specimen per minute.

(*g*) *Jaws.*—The jaws or grips of the testing machine shall be designed so as to produce and maintain axial alignment of the specimen without producing tearing stresses. The edges of the jaws shall be rounded so as not to noticeably deform the specimen near the borders of the region where the specimen comes in contact with the jaws.

Articulation Testing Methods

By H. FLETCHER and J. C. STEINBERG

This paper is chiefly concerned with the technique of making articulation tests. The construction of a syllabic testing list, the selection of a testing crew, the methods of comparing articulation data for various crews, and the significance of the test as a measure of the speech capabilities of a system are discussed. Various types of lists for different uses are also discussed.

THE transference of thought by means of speech is a very complicated, although common, process. So long as the process runs smoothly, its complications are forgotten. When an auditor fails to understand the speaker, however, inquiry into the reasons for the difficulty begins.

The production, the transmission, and the reception of speech constitute the three important elements of the process. To determine defects in any one of these, it is necessary to have a quantitative means of measuring the recognizability of the speech sounds that the auditor hears. The term "recognizability" as used here refers to correctness with which an auditor identifies a sound as being one, or some combination, of the fundamental speech sounds, when no meaning is associated with the sounds.

During the past few years methods of measuring the recognizability of speech sounds have come into greater and greater use both in this country and abroad. In order to compare the results obtained by various crews in various languages, it is desirable to standardize the methods of test and to set up reference circuits for purposes of calibration. It is the aim of this paper to discuss the methods that have been found the most useful, not only in determining defects in transmission, but defects in the production and reception of speech as well.

One needs only to tabulate the various devices that are used for transmitting speech to realize the importance of a quantitative method of rating their performance. There may be mentioned, for example, the various telephone and radio systems, the phonograph, sound pictures, rooms and auditoriums with various types of acoustic treatment, audiphone sets for the deafened, speaking tubes, etc.

Methods of measuring the recognizability of speech sounds have not been used so extensively for determining the ability of persons to speak properly. Such methods should be of value in the training

of public speakers, actors, students of foreign languages or pupils in deaf schools who are learning to speak.

The rating of auditors by measuring the recognizability of speech sounds which they hear has been used to some extent. For example, such methods have been used to determine the ability of students to interpret a spoken foreign language. Also, the deafness of a person can be determined by such methods. In this case, however, the specialists have usually tried to vary transmission systems between the speaker and listener so as to compensate for the loss of hearing, the amount of such compensation being determined by measuring the recognizability of speech sounds.

The best method of determining the recognizability of the speech sounds naturally depends upon which of the things just enumerated is to be rated. In principle, the method in each case consists in the pronunciation of "selected speech sounds" by a speaker, the transmission of these sounds to an observer's ears, and the recording by the observer of the sounds which he recognizes. Such methods applied to telephone systems have been frequently referred to as articulation tests. The term "articulation" would be more logically used if it were applied only to cases where the speaking abilities of persons are being determined. However, it has been used so frequently in connection with rating transmission systems that it seems convenient to retain it.

The "selected speech sounds" which are ordinarily used in articulation tests are meaningless monosyllables. The percentage of the total number of spoken syllables which are correctly observed is called the syllable articulation. This percentage has frequently been called simply "the articulation."

A syllable is considered to be incorrectly observed, if one or more of the fundamental speech sounds which it contains are mistaken. It is frequently desirable to analyze these mistakes and determine the articulation of the speech sounds. The percentage of the total number of spoken sounds which are correctly observed is called the sound articulation. When the attention is directed toward a specific fundamental sound, such as "b" or "t" or "ā," etc., then the term "individual sound articulation" is used. For example, the individual sound articulation for "b" is the percentage of the number of times that "b" was called that it was observed correctly. Similarly, the terms "consonant articulation" or "vowel articulation" refer to the percentages of the total number of spoken consonant or vowel sounds which are correctly observed.

The articulation values as defined above are taken as the measures

of the recognizabilities of the various phonetic units. English words and short sentences have also been used for testing purposes. When material of this kind is used, a new element enters, namely, the thought or meaning associated with the sentence or word. The criterion for the correct observation of words or sentences is also different from that used in the case of articulation tests. If the thought or meaning of a word or sentence is correctly understood, it is considered to be correctly received, even though the observer may not have correctly recognized each sound that was spoken. The terms "word articulation" or "sentence articulation," therefore, seem inappropriate when referring to the results of such tests. The term "intelligibility" has frequently been used in this sense. Since it has also been used in a more general sense, the terms "discrete word intelligibility" and "discrete sentence intelligibility" will be used when referring to the results obtained by using disconnected words or sentences for the testing material. They are defined as the percentage of the total number of spoken words and sentences, respectively, that are correctly interpreted according to the criterion given above.

Very early in the work of developing the telephone, words and sentences which were chosen in a haphazard way were used for testing purposes. Word lists of various sorts have been worked out and used with some success. Even in very recent years some of these word lists have been used to good advantage. The main objections which have developed, to the continuous use of such lists are: (*a*) it is hard to make the lists equally difficult without resorting to very long lists of words, (*b*) a very large number of lists are necessary in order to avoid memory effects.

Dr. G. A. Campbell [1] was one of the first to propose a system of syllabic speech sounds for testing the transmission characteristics of the telephone system. These syllables had no meaning and were constructed by combining the various initial consonants with the vowel "ee," such as bee, fee, etc. With these lists the consonant articulation was taken as a measure of the system.

Later Dr. I. B. Crandall [2] worked out a system which used both simple and compound consonant forms in a vowel-consonant and consonant-vowel type of syllable. All of the common vowels were used, and the combinations were formed in ways which are usually found in written speech. The sounds occurred with the same frequency as they occur in ordinary written material. As in the Campbell lists, the articulation was based on the consonant sounds alone.

[1] "Telephonic Intelligibility," G. A. Campbell, *Phil. Mag.*, Jan. 1910.
[2] "Composition of Speech," I. B. Crandall, *Phys. Rev.*, 10, p. 74, July 1917.

Several other lists which have not been published were proposed and used, the differences being in the choice of the fundamental speech sounds, in their arrangement into syllables, and in the relative frequency of occurrence, both of the different syllable forms and of the speech sounds in each form. There was a distinct effort to make the lists as nearly like speech as possible by using the syllable forms, and by using the particular combinations of fundamental sounds that occur frequently in English. Difficulties were encountered in testing, however, when this was carried too far in that enough different syllables could not be obtained for continuous testing. On the other hand, when random combinations of sounds were made, without regard to the particular combinations occurring in English, syllables that were very unusual and difficult to pronounce were obtained, unless the combinations were restricted to the simple syllable forms having only two or three sounds. In other words, testing lists must be selected with two things in mind; namely, they must be representative of speech and they must be suitable for making tests. The experience with these various lists also indicated that the results obtained with one system of lists could be calculated approximately from the results obtained with other systems by properly weighting the individual sound articulation values.

This experience led to the adoption by the Laboratories of a system of lists which has been used during the past ten years in studies of the effects of distortion upon the recognition of speech sounds.[3] These lists which have been referred to in the literature as the standard articulation lists were made up of only the con-vow, vow-con and con-vow-con syllable forms. The various fundamental sounds of English were combined at random into the syllables, such that each sound occurred approximately with uniform frequency.

During the past few years it has become evident that still further simplifications in the syllable forms used in the standard articulation lists might be made. Also our methods of making articulation tests and interpreting the results obtained have undergone considerable changes during this time. It is with these new methods that the present paper is chiefly concerned.

In order to distinguish between the old lists and the ones modified as described below, the prefixes "old" and "new" will be placed before the title "Standard Articulation Lists." When there is no chance for confusion, the new lists will be called simply the standard articulation lists, since they are the principal ones now being used in the work at Bell Telephone Laboratories.

[3] "Nature of Speech and Its Interpretation," H. Fletcher, *Journal Franklin Institute*, June, 1922.

New Standard Articulation Lists

In setting up any testing list it is necessary to classify and select a representative group of speech sounds. The National Phonetic Association uses a basic alphabet of 65 different sounds and also uses numerous modifiers which serve to distinguish slight variations in a given sound. Such a system is too complex for testing purposes. The revised scientific alphabet uses 48 simple sounds of which 24 are consonants, 19 vowels, and 5 diphthongs. Besides these fundamental sounds, connected speech contains certain recurrent combinations of them, such as st, ing, etc.

In speech these fundamental sounds are combined into syllables in a large variety of ways, but as mentioned before, in constructing a testing list it is desirable to adhere to very simple syllable forms. More complex forms which include the compound endings are either too few in number or involve unusual speech sound combinations. In either case they are soon memorized by a testing crew working with such lists. In the new lists, therefore, simplifications are made by omitting the con-vow and vow-con types of syllables, leaving only the con-vow-con type. In order to make syllables of this type it is obviously necessary to have the same number of vowels and consonants, provided that each consonant may be used in both the initial and the final position. Some consonants, however, can be used only in the former while others can be used only in the latter position.

With these facts in mind the sounds that are shown in Table I were adopted for these new lists. It will be noticed that all of the consonants are used in both the initial and final positions in the syllable, except h, w, and y, which are used only in the former, and zh, ng, and st, which are used only in the latter position. As was the case in the old standard lists, it will be seen that, in the new lists the vowel variants have been excluded. They occur infrequently in speech and phoneticians do not universely agree on their pronunciation. For this reason they are not included. Also, the diphthongs ī, ou, oi, and ew, which were used in the old lists, were omitted from the new lists. The last two of these diphthong sounds occur very infrequently in speech. Although the diphthongs, ī and ou, do occur quite frequently, it was felt that their essential properties were embraced by the properties of their constituent vowel sounds. By their omission and also by the introduction of the compound st as a final consonant, it is possible to construct any desired number of syllables of the con-vow-con type, from the speech sounds shown in the table.

TABLE I

SPEECH SOUNDS FOR NEW STANDARD TESTING LISTS *

Initial Consonant	I.P.A.	Key Word	Vowel	I.P.A.	Key Word	Final Consonant	I.P.A.
b			a	[ɑː]	father	b	
d			a			d	
f			ā	[eː]	fame	f	
g		go	ā			g	
k			a'	[æ]	fat	k	
l			a'			l	
m			e	[ɛ]	get	m	
n			e			n	
r			ē	[iː]	greet	r	
p			ē			p	
s			i	[ɪ]	tin	s	
sh	[ʃ]	ship	i			sh	
th'	[ð]	this	o	[ʌ]	but	th'	
th	[θ]	thin	o			th	
t			ō	[oː]	go	t	
v			ō			v	
ch	[tʃ]	church	u	[ʊ]	full	ch	
z			u			z	
j	[dʒ]	judge	ū	[uː]	rule	j	
h			ū			zh	[ʒ]
w			o'	[ɔː]	haul	ng	[ŋ]
y	[j]	yawl	o'			st	

Note: Final r and ng are used in the list only when they occur in combination with the following vowels:

a'r (as in carry, paragraph)
ar (as in are, far, tar)
er (as in bury, ferry, verify)
ir (as in spirit)
or (as in her, utter, fir)
o'r (as in for, lore)
ūr (as in your, sure)

a'ng (as in bang, sang)
eng (as in geng, e as in ten)
ing (as in sing, wring)
ong (as in sung, hung)
ung (as in gung, u as in took)
o'ng (as in long, wrong)

* The symbols for the sounds are those of the International Phonetic Association's alphabet. See Pronunciation of Standard English in America, Krapp, Oxford University Press, 1919. See also Revised Scientific Alphabet, Funk and Wagnall's Dictionary.

The testing syllables are formed in the following way. Cards upon which the initial consonants are written are placed in one box; others upon which the vowel sounds are written are placed in a second box; and those upon which the final consonant sounds are written are placed in a third box. A card from each box is drawn at random, thus forming the con-vow-con syllable. By drawing all of the sounds, a list of 22 syllables is formed. This process is repeated three times to obtain a list of 66 syllables which is a unit that has been found convenient to use. A list of syllables of about this length can be used without giving callers and observers a rest period. In such a list each initial consonant occurs three times, each vowel six times, and each final consonant three times.

In forming such syllables only those combinations involving final r and ng that are shown in Table I are included. Much confusion exists concerning the pronunciation of other combinations of these sounds. Syllables that represent slang in English are also omitted. These omissions are made by returning the card upon which the sound in question is written to its box and drawing another card. By combining the sounds at random in this manner any desired number of lists may be made which for practical purposes are all of equal difficulty.

In addition to containing a certain speech sound content, connected speech is characterized by inflection, accent, a rate of utterance, etc. In the earlier articulation studies the test syllables were called singly at intervals of about three seconds. When considered with respect to connected speech this procedure seems somewhat artificial. Comparative tests were made in which the syllables were called singly and as parts of introductory sentences. The tests showed the syllable articulation to be somewhat larger when the introductory sentences were used. The increase was due largely to the greater ease in interpreting the initial consonants of the syllables, when they were inserted in the introductory sentences. The effect was most noticeable for the stop and fricative consonants which have relatively short durations. In order to make the technique more nearly like connected speech the syllables are called in the short introductory sentences. The use of such introductory sentences also helps to insure that any element in the transmission system being tested, whose performance depends particularly upon their immediate past history, will be in the condition in which we are interested for determining speech transmission capabilities.

A list of sentences which is used for this purpose together with a sample record of articulation data is shown in the articulation test record of Table II. For calling purposes, the syllables on the cards are written in the spaces under the columns marked "called" of the test record. These sentences are called uniformly at the rate of 15 per minute. When the syllables in the first column are called, the sentences are repeated using the syllables in the second column and then those in the third column.

The observers are provided with blank articulation test record sheets. They write the sounds which they hear in the corresponding "observed" columns. When the test is completed the observed and called sheets are compared and the various articulations obtained.

For good results it has been found advisable to use a testing crew of ten people—5 men and 5 women. Eight people are ordinarily

TABLE II

ARTICULATION TEST RECORD

DATE 3-16-28 SYLLABLE ARTICULATION 51.5%

TITLE OF TEST PRACTICE TESTS CONDITION TESTED 1500~ LOW PASS FILTER

TEST NO. 10 OBSERVER W.H.S. JR.

LIST NOS. 5-9-37 CALLER E.B.

NO.		OBSERVED	CALLED	OBSERVED	CALLED	OBSERVED	CALLED
1	THE FIRST GROUP IS	máv	náv	póz	po'th	Kōb ✓	Kōb
2	CAN YOU HEAR	pōch ✓	pōch	nēz	nēzh	shēth	siz
3	I WILL NOW SAY	seng ✓	seng	jóch ✓	jóch	fūch ✓	fūch
4	AS THE FOURTH WRITE	chūd ✓	chūd	th'ám ✓	th'ám	thōl ✓	thōl
5	WRITE DOWN	run ✓	run	hab ✓	hab	poth' ✓	poth'
6	DID YOU UNDERSTAND	chiz	Kiz	def	doth	wa'm ✓	wa'm
7	I CONTINUE WITH	foz	fozh	chech	chej	gūm	gūn
8	THESE SOUNDS ARE	lo'l ✓	lo'l	lun	lon	nāsh	nāth
9	TRY THE COMBINATION	jās	zhāth	shāl ✓	shāl	vó'g ✓	vó'g
10	PLEASE RECORD	th'áth	th'ásh	muz ✓	muz	lung	long
11	WRITE THE FOLLOWING	wūr ✓	wūr	led	bēd	diz	dizh
12	NOW TRY	yāp ✓	yāp	wif ✓	wif	KaK	taK
13	THIRTEEN WILL BE	mad	maj	gōst ✓	gōst	th'ár	zhár
14	YOU SHOULD OBSERVE	bēch	bēK	thav	sāv	must ✓	must
15	WRITE CLEARLY	gēm	dēm	Kōf ✓	Kōf	yo'd ✓	yo'd
16	NUMBER 16 IS	th'eb	veb	rá'g ✓	rá'g	jet ✓	jet
17	YOU MAY PERCEIVE	jok	jost	thip ✓	thip	rēp	rēj
18	I AM ABOUT TO SAY	gaf ✓	gaf	yar ✓	yar	th'ēp	hēp
19	TRY TO HEAR	hus ✓	hus	zhūt ✓	zhūt	——	chuv
20	PLEASE WRITE	hiv	thi'th	KūK	tūk	th'ef	th'esh
21	LISTEN CAREFULLY TO	tōg ✓	tōg	fung ✓	fung	bās ✓	bās
22	THE LAST GROUP IS	sho't ✓	sho't	th'ev	vesh	th'o'f	shaf

employed in a test, the remaining two being held for emergencies in order to keep the work going. One member of the crew calls at a time, and the remaining members act as observers. Ordinarily, eight callers are used with four observers recording simultaneously for each caller, although as many as eight or nine observers may be used. The order is arranged such that the various members are equally represented in the test.

Each observer's sheet (Table II) furnishes a value of syllable articulation (the percentage correctly observed), corresponding to a particular caller-observer pair and to 66 syllables or 198 speech sounds called. These values of syllable articulation are recorded in the form shown in Table III. The average of each column gives the average articulation for each observer. The averages of the rows give the callers' articulation.

TABLE III

ARTICULATION TEST RESULT RECORD

FROM 2/29/28 TITLE PRACTICE TESTS

TO 3/21/28 CONDITION 1500 CYCLE LOW PASS FILTER

REFERENCE MM-2186 - 8/27/28 REMARKS

CONDITION															
	OBS. / CALLER	E.B.	W.H.S.	F.S.	C.M.	H.C.	A.H.	E.L.F.	M.W.	P.H.		AVE.			
	E.B.		51.5	51.5	50.0	48.5	57.5	59.0		45.5		51.9			
	W.H.S.	72.5		62.0	69.5	59.0	66.5	60.5		56.0		63.7			
	F.S.	47.0	54.5		56.0	48.5	57.5	50.0		42.5		50.9			
	C.M.	51.5	62.0	53.0		42.5	65.0	57.5		38.0		52.8			
	H.C.	48.5	56.0	48.5	36.5		50.0	54.5	30.5			46.4			
	A.H.	41.0	53.0	50.0	53.0	35.0		47.0	42.5			45.9			
	E.L.F.	51.5	51.5	51.5	45.5	53.0	57.5		41.0			50.2			
	M.W.	50.0	48.5	50.0	36.5	51.5	59.0	63.5				51.3			
	AVE.	51.7	53.9	52.4	48.6	48.3	59.0	56.0	38.0	45.5		51.6			

When the articulation values are near 100 per cent or 0 per cent, then a group of values will not distribute itself symmetrically about the arithmetic mean or average value. For the high values, this is due to the fact that one cannot get a higher value than 100 per cent. To some observers, the 100 per cent mark may be obtained very easily and to others it may be obtained only with considerable effort. This difference in difficulty cannot be registered in the percentages obtained. A similar reason exists for the unsymmetrical grouping for values near zero. Our experiments have shown that this grouping is symmetrical in the range from 20 per cent to 80 per cent. For the range from 80 per cent to 100 per cent, the average value is less than the most frequent value, and for the range from 0 per cent to 20 per cent, the average value is greater than the most frequent value. These differences are of the order of 1 or 2 per cent. From an extensive series of tests, the averaging factor curve of Fig. 1 was constructed which enables the data to be averaged, in a way, such that the average value is approximately equal to the value which would be most frequently observed in a large number of tests. To do this, each observed articulation, based on 66 syllables, is converted into an averaging factor by means of the above curve. These factors are then averaged and the average value reconverted into average

articulation by means of the above curve. The value so obtained is taken as the average syllable articulation for the test.

The articulations for the various sounds are recorded on the Articulation Test Analysis Record of Table IV. In making the analysis the total errors for each caller are counted. The occurrences per caller are the products of the number of times the sounds are spoken by the caller, and the number of observers. One analysis sheet contains the results for a complete test.

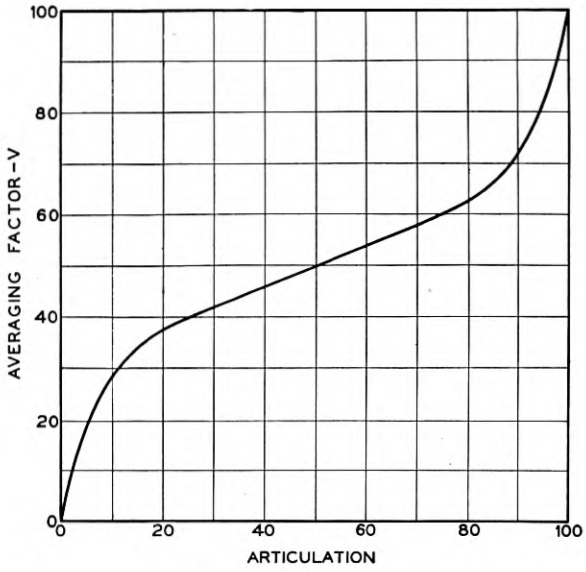


Fig. 1—Averaging factor curve

In dealing with the syllable articulation the unit is 66 called syllables, or the result of one caller and one observer. Hence a number of syllable values are obtained for one test which are averaged by means of the averaging factor curve. When we deal with the articulation of the individual speech sounds, it is advisable to use a larger unit since each sound occurs only six times in a 66 syllable unit. In Table IV the errors are shown for each caller as a unit, and since there were 7 observers per caller, each sound occurs 42 times. This is a unit of sufficient size to qualitatively compare various voices. However, in drawing conclusions as regards the effects of the circuit upon the transmission of the various sounds, it is best to use eight callers as a unit, so that each sound, depending upon the number of observers, occurs of an order of 200 to 300 times. In this case the articulation of an individual sound has a precision that is comparable

with the precision of the syllable articulation when based on 66 called syllables, as will be discussed in a later paragraph.

TABLE IV

ARTICULATION TEST ANALYSIS RECORD

DATE 2/29/28 To 3/21/28 TITLE PRACTICE TESTS

REFERENCE MM-2186 8/27/28 CONDITION 1500~ LOW PASS FILTER

OBSERVERS EB, WHS, FS, CM, HC, AH, ELF, MW, PH SOUND ARTICULATION 79.3 SYLLABLE ARTICULATION 51.6

LETTER	OCCUR. PER CALLER	ERRORS PER CALLER														TOTAL ERRORS	TOTAL OCCUR.		IND. SOUND ART.	
		EB		WHS		FS		CM		HC		AH		ELF		MW				
a	42	3		2		10		1		15		2		2		3		38	336		88.7	
ā	42	1		1		4		0		7		2		0		2		17	336		94.9	
a'	42	2		0		8		7		20		22		27		11		97	336		71.1	
e	42	2		4		4		4		22		19		20		15		90	336		73.2	
ē	42	0		0		2		0		1		4		0		1		8	336		97.6	
i	42	2		0		2		0		9		8		9		3		33	336		90.2	
o	42	12		1		6		5		12		5		4		10		55	336		83.6	
ō	42	1		0		0		0		0		1		0		0		2	336		99.4	
o'	42	0		0		5		2		0		0		0		0		7	336		97.9	
u	42	5		2		19		1		8		3		4		4		46	336		83.6	
ū	42	1		1		1		1		4		4		2		1		15	336		95.5	
TOTAL	462	29		11		61		21		98		70		68		50		408	3696		89.0	
b	42	7		0		11		10		13		1		4		6		52	336		84.5	
ch	42	13		4		11		13		2		8		8		7		66	336		80.4	
d	42	9		3		6		8		8		7		10		14		65	336		80.7	
f	42	6		11		7		12		25		23		21		22		127	336		62.2	
g	42	9		4		8		13		2		3		1		2		42	336		87.5	
h	21	2		0		0		0		2		0		1		1		6	168		96.4	
j	42	17		4		10		9		3		6		3		2		54	336		83.9	
k	42	17		6		6		11		3		12		11		6		72	336		78.6	
l	42	1		0		3		0		0		2		3		0		9	336		97.3	
m	42	1		2		4		6		2		5		7		0		27	336		92.0	
n	42	4		5		0		10		7		6		9		15		56	336		83.3	
ng	21	0		0		2		3		1		1		4		5		16	168		90.5	
p	42	11		11		4		5		19		21		14		8		93	336		72.3	
r	42	3		0		0		2		4		3		4		2		18	336		94.6	
s	42	18		5		16		9		23		22		18		17		128	336		61.9	
sh	42	25		21		28		26		14		25		17		20		176	336		47.6	
st	21	6		0		8		2		7		6		2		1		32	168		81.0	
t	42	15		8		15		13		33		24		25		13		146	336		56.5	
th	42	22		24		17		22		27		16		22		27		177	336		47.3	
th	42	20		24		25		19		17		12		20		15		152	336		54.8	
v	42	17		7		12		10		4		9		5		14		78	336		76.8	
w	21	0		0		0		0		0		0		0		0		0	168		100.0	
y	21	1		0		0		0		1		0		1		0		3	168		98.2	
z	21	6		4		11		4		12		9		9		13		68	168		59.5	
zh	42	33		24		32		35		27		27		23		26		227	336		32.4	
TOTAL	924	263		167		236		242		256		248		242		236		1890	7392		74.4	

E-1248 (6-28)

SELECTION OF TESTING PERSONNEL

It is necessary to set up the technique of testing, such that the values of syllable articulation can be reproduced within acceptable limits. The limits depend upon the control of the auditory and vocal

characteristics of the testing crew, the control of numerous haphazard factors, and the control of the practice or experience that the crew acquires in the testing of circuits.

The departure from normal, in acuity of hearing of prospective crew members, should be measured with a good audiometer. In our laboratories the 2-A audiometer is used for this purpose. Only those individuals whose average hearing loss departs from normal in the speech range of frequencies (100 to 8,000 cycles per second) by less than 5 db (decibels) are selected. Although of normal hearing, some observers of a crew usually obtain higher values of syllable articulation than do others. The averages of the columns of Table III are a typical set of results for nine observers who have passed such a hearing test and who have had a year or more of experience in observing.

Observer A. H. obtained the highest percentage, namely 59, and observer M. W. obtained the lowest percentage, namely 38. In general this order would be preserved in a series of tests, although haphazard variations in a single test might change the order. The spread in observations is of an order of 20 per cent. More extended tests have shown that the spread tends to decrease as the observed percentages approach 0 or 100. In order to make a replacement in the observing personnel from time to time without causing a probable change of more than 2 per cent in the average percentage, it is necessary to use an observing crew of 8 to 10 persons. Our experience has shown that men and women show no characteristic difference when acting as observers.

The ability of prospective crew members to enunciate the sounds in a normal way is determined in the following manner. An extensive series of tests on various voices have yielded data which are arbitrarily used as a basis for determining normalcy. These tests were made with a simplified list consisting of common English words which will be described in a later paragraph (see Table XVII). Tests were made under three conditions; namely, direct transmission through the air in a quiet, well damped room, transmission over a circuit which uniformly transmitted the frequency range from 100–4,500 cycles, transmission over a circuit having a carbon transmitter. A diagram of this latter circuit is shown in Fig. 7. The sounds were observed by a crew of experienced observers. Table V gives the results of tests that were made upon 21 male and 23 female voices, the personnel being selected from various departments of the Laboratories. The average articulations of the simple consonant sounds are shown. The data are given separately for men and women.

TABLE V

NORMAL ENUNCIATION

Speech Sound	Air Transmission		Band 100 to 4,500~		Carbon Transmitter Circuit	
	Av. % Articulation of Sounds		Av. % Articulation of Sounds		Av. % Articulation of Sounds	
	Men	Women	Men	Women	Men	Women
b	98.7	98.0	96.2	90.1	95.0	91.5
ch	100.0	99.3	98.4	98.0	98.7	98.4
d	99.4	100.0	98.7	98.3	91.9	88.6
f	97.5	84.8	96.5	87.7	79.6	65.2
g	100.0	100.0	99.4	94.6	70.6	63.0
h	100.0	98.0	99.4	98.0	94.4	97.8
k	100.0	100.0	99.1	99.7	78.1	82.1
l	100.0	100.0	98.1	97.7	96.0	87.5
m	99.2	91.3	97.5	95.5	86.5	86.0
n	99.4	100.0	99.4	99.1	95.5	94.5
ng	100.0	100.0	99.7	99.4	93.7	100.0
p	96.9	95.8	97.2	96.6	80.0	73.5
r	100.0	100.0	100.0	99.7	96.9	75.5
s	97.5	99.1	95.0	68.0	91.2	67.8
sh	100.0	100.0	100.0	99.8	98.3	93.1
th	90.4	80.6	75.2	56.3	60.1	50.2
th'	97.5	87.0	93.3	87.7	74.1	77.2
t	98.3	100.0	99.4	96.6	92.9	72.1
v	93.3	78.9	96.2	83.6	87.1	75.0
w	100.0	99.3	100.0	98.9	97.5	84.8
z	100.0	100.0	95.9	70.8	91.2	65.2
Aver.	98.2	94.6	96.4	90.2	87.5	80.5

For our work, a prospective crew member is required to call such a list of syllables to a crew of experienced observers. If the observed articulations of the sounds are reasonably close to those indicated in Table V for each circuit, and if no obvious irregularities are noticed in the speech, the prospect is considered satisfactory for testing work. Measurements are also made upon the individual's speech power, but it has not been found necessary to use the information in the process of selection.

Aside from the practical application to the methods of testing, the table is interesting in showing characteristic differences between the voices of men and women. In general, woman's speech is more difficult to interpret than man's, particularly, in the case of the sibilant and fricative consonants. This is probably due to the fact that in woman's speech, these sounds are not only fainter, but occupy higher frequency ranges than in man's speech. The frequency range from 6,000 to 8,000 cycles for the former, is approximately equivalent to the range from 4,000 to 6,000 cycles for the latter. In the case of

the voiced sounds, woman's speech has only one half as many components as man's, which also may cause greater difficulty in interpreting the former.

With respect to the vowel sounds, the crew members are instructed in the correct manner of enunciation. Only those vowels which have definite differences have been included in the testing lists, so that, slight differences in enunciation do not seriously affect the observed results.

The object of the selection process is to determine in a broad but definite way the normalcy in speech of prospective members, and to eliminate those individuals who have speech characteristics which are not readily reproducible should it be necessary to change the testing personnel. The row averages of Table III show a typical set of results for 8 callers who were selected in the above way, and have had a year or more of experience in calling.

The spread in results is of an order of 20 per cent, so that, if a crew of 8 to 10 callers is used, a replacement may be made in the calling personnel without causing a change in the average percentage of more than 2 per cent. Owing to inherent differences in the voices of men and women, they are equally represented on the testing crew. Individuals who have the equivalent of a high school education, and whose ages range from 18 to 23 years, are usually selected for this work.

Control of Haphazard Factors

Haphazard factors arise from various sources, some of which can be controlled reasonably well. The observers work in a sound-proof room, so that extraneous noises will not affect the articulation results. The calling is ordinarily done in a sound-proof booth that has been especially treated with sound absorbing material so as to reduce the reverberation time to an order of a few tenths of a second. Ordinarily the crew does not test more than two to four hours during the day, and the schedule is usually arranged so that this is not done continuously.

The intensity level of each caller is also measured during the test, as small variations in intensity level may cause rather large variations in articulation. Ordinarily the various callers are permitted to call at the intensity level most natural to them, although in some tests the callers all attempt, by watching an indicator, to call at the same level. Various instruments have been used for measuring the intensity levels during tests. The volume indicator [4] has proven quite satis-

[4] This instrument depends for its readings, essentially upon the syllabic powers of the vowel and semi-vowel sounds, so that the reading of the instrument is determined largely by the amplitudes in the frequency range from 100 to 2,000 cycles.

factory and is the instrument ordinarily used for this work. It has the advantage over some of the other instruments that were tried, of being in much more general use on speech circuits.

Control of other haphazard factors of a more or less psychological character, may best be obtained by taking enough data so as to average out their effects. This involves the number of syllables that are called by each speaker and the number of caller-observer pairs that are used in the test. The variability of caller-observer pairs for a calling unit of 66 syllables may be seen from Table III. The probable error [5] in percentage articulation of a single observation (ϵ_s) i.e., one caller-observer pair as taken from the data in the table, is $\pm$ 9. The probable error of the average articulation ($\epsilon_{av.}$) of the 56 caller-observer pairs is $\pm$ 1.2.

It has been found from a large number of tests that the probable error of a number of crews, each consisting of one caller and one observer, is of an order of $\pm$ 12 (per cent articulation) for a 66 syllable unit when the syllable articulation is around 50 per cent. This value tends to decrease with increasing experience in testing, and with increasing or decreasing values of syllable articulation. The use of 36 caller observer pairs obviously reduces the probable error to an order of $\pm$ 2 in percentage articulation, which is about the order of magnitude of the errors involved in maintaining the testing personnel over a period of time.

Since as will be shown in a later paragraph, the syllable articulation is equal to the cube of the sound articulation, the probable error in the sound articulation [6] for one caller and one observer, or a unit of 198 sounds, is of an order of $\pm$ 6 when its value is around 80 per cent. Since each individual sound is called only six times, the probable error for each individual sound for a single caller-observer pair is of an order of $\sqrt{\frac{198}{6}} \times 6 = \pm 35$. If a test comprises 4 observers per caller and 8 callers, each sound is called 192 times, which reduces the probable error for the articulation of each sound to $\pm$ 6. Under

[5] $\varepsilon_s = .67 \frac{\Sigma d^2}{n-1}$ and $\varepsilon_{av.} = \frac{\varepsilon_s}{\sqrt{n}}$; where, n = number of caller-observer pairs; d = difference between the articulation of a caller-observer pair and the average articulation of n caller-observer pairs.

[6] $\varepsilon_s = \frac{\partial S}{\partial L} \varepsilon_L = 3L^2 \varepsilon_L$

$\varepsilon_s = \pm 12$, $S = .5$, $L = S^{1/3}$,

$\varepsilon_L = \frac{\varepsilon_s}{3S^{2/3}} = \pm 6.$

ε_s = prob. error in syl. art. for one caller-observer pair.
ε_L = prob. error in sound art. for one caller-observer pair.

the same circumstances, the probable error for the average syllable articulation is ± 2, and that for the sound articulation is ± 1.

Control of Practice Effects

The third factor entering into the reproducibility of articulation results is practice and experience. The practice effect manifests itself in various ways. An increase in articulation takes place as the observers become more familiar with the vocal characteristics of the speakers. Similar effects are observed as they become more accustomed to a given technique, or to a particular type of distortion. In general, these effects become smaller as the testing crew becomes more experienced.

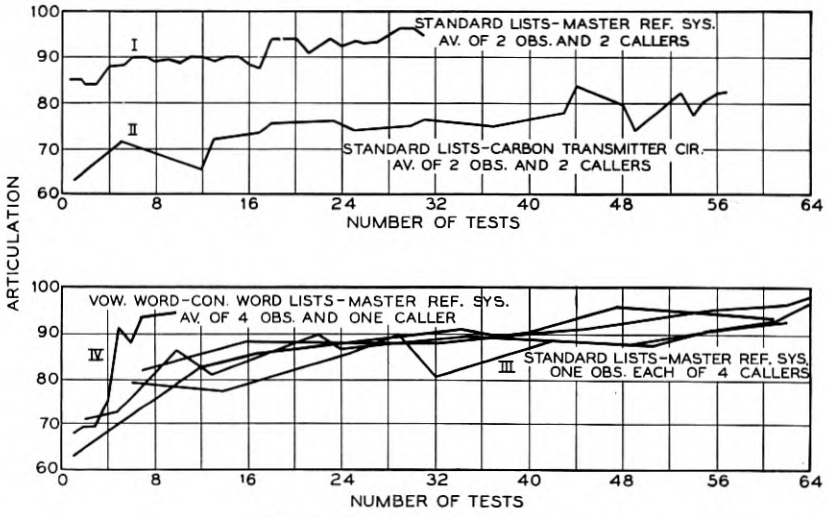


Fig. 2—Typical growth curves

Fig. 2 shows several typical growth curves that were obtained in the process of training new crew members. In this process the new members observe continually on various circuits until the results compare favorably with the results that are obtained by experienced observers. In such tests, experienced speakers are used. The averages for two new observers, of the results that were obtained on a high grade circuit, are shown by Curve I. Two speakers were used in these tests. A limited amount of testing was done by the observers prior to the above tests. Upon the completion of the tests of Curve I about 30 or 40 additional tests were made on various circuits. A

series of tests, in which several speakers were used, were then undertaken on a carbon transmitter circuit. In Curve II the averages for the two observers, of results on two voices, are shown. Three to four weeks' time was spent by the observers in making the various tests mentioned above.

The curves under III show similar data that were taken at a later date by one new observer, for several voices. All of the above tests

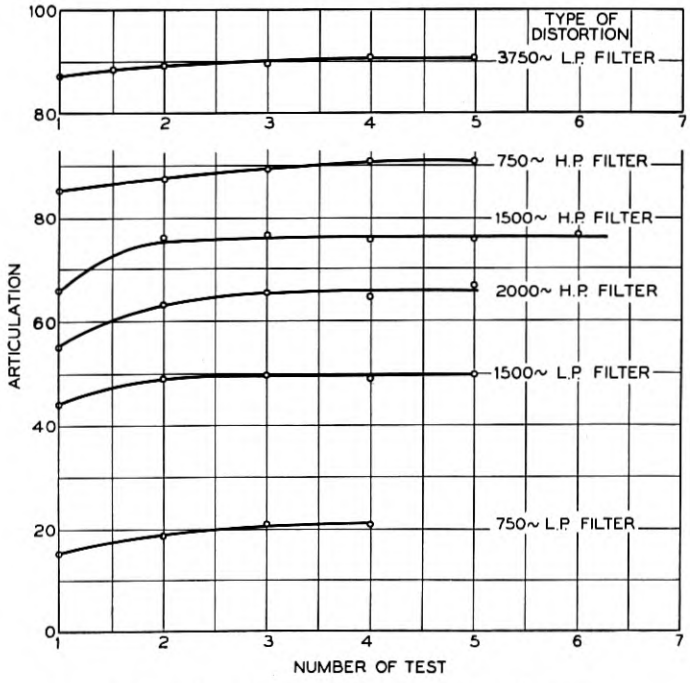


Fig. 3—Practice effects for an experienced crew

were made with the standard lists. In Curve IV, data are shown that were obtained with the vowel and consonant word lists (see Table XVII). In these tests four new observers were used and no preliminary training was given. It is evident that with the word lists, the results reach a state of saturation very quickly.

After a crew has spent several months in testing, its performance becomes largely mechanical. Under such circumstances the practice effects are rather small for types of distorted speech with which it has had experience. When the speech distortion is unusual, however, rather large practice effects may be obtained. Fig. 3 shows such

practice effects for several types of distortion for a crew of eight people. All six of the circuits were tested on each test before going on to the following test. The first three tests were made successively and covered a period of about two months. In each test the types of distortion were interspersed. In other words half of the first test was completed with the filters in one order, and the other half with the filters in the reverse order. The fourth test was made about three

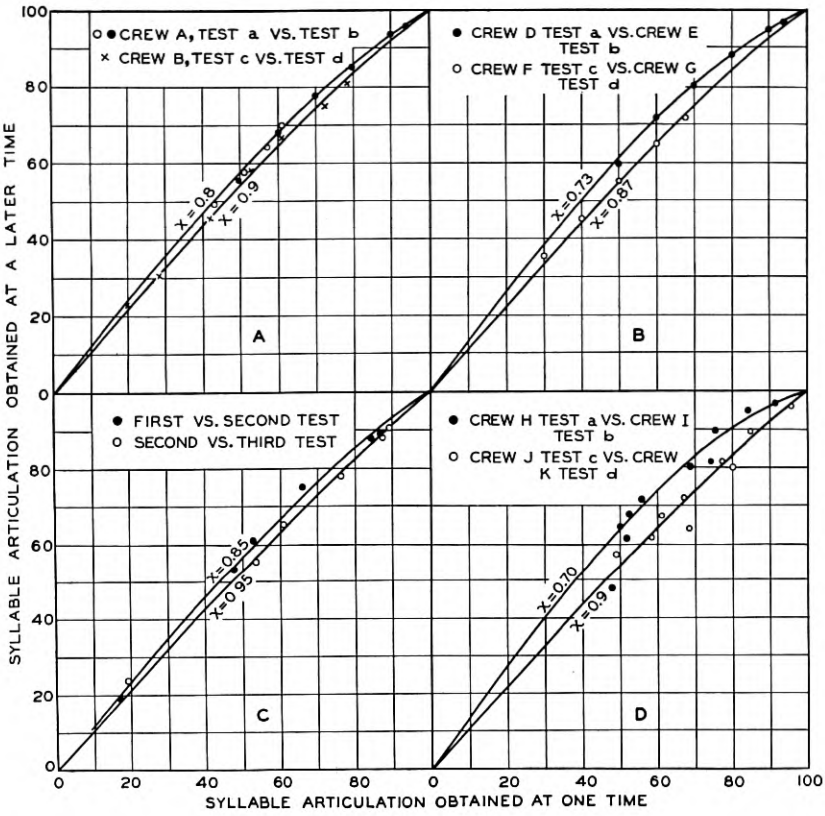


Fig. 4—Practice effects

months later, and the fifth test was made approximately six months later. Although the crew had been testing various circuits for about a year, and were thoroughly accustomed to the routine, these particular circuits had not been previously tested, so that, the crew's experience with these types of speech distortion was small. It is evident that the articulation of an experienced crew reaches saturation very quickly. It is probable that practice effects for a crew of several years' experience with distorted speech would be negligible.

Several procedures are followed in order to correct, in so far as possible, for practice effects. In comparative tests, whenever it is possible, the circuits to be compared are interspersed so as to average out practice effects. If it is desired to compare the articulation of a very new or unusual circuit (from the standpoint of the speech distortion) with one of common experience, several successive tests are made upon the new circuit until no further increase in articulation with practice appears. When it is impossible to intersperse the tests, the data may be corrected to a given state of practice by means of curves which were obtained in the following way. Although as will be seen, this procedure is valid only under certain restrictions, which will be discussed, such a correction will always tend to correct the data to a more comparable basis.

In Fig. 4-*a* a practice curve is shown that was obtained for a crew, from two series of tests that were separated by an interval of three months. The dots represent tests that were made upon a circuit which uniformly transmitted a frequency range from 100 to 5,500 cycles. The circles were obtained from a circuit of the type shown in Fig. 7 involving the carbon transmitter. In both cases the various articulation values correspond to different received speech levels. The crosses represent similar results that were obtained with a different crew on the latter type of circuit.

In Fig. 4-*c* the data of the first three tests in Fig. 3 are shown. In this case the distortion was varied and the received speech level held constant. As previously stated, in so far as was known, the crew had no previous experience with these types of speech distortion so that the practice for the various types of distortion ought to be comparable.

All of the solid curves are graphs of the following equation

$$(1 - S') = (1 - S)^x \cdots, \tag{1}$$

where S' = decimal value of syllable articulation obtained on a given circuit at one stage of a crew's career,

S = the value obtained on the same circuit at a later stage of the crew's career,

x = a number called the practice factor.

The values of the practice factor x that were necessary in order to fit the observed values are shown in the figure. It is impossible to state definitely that a crew has uniform practice with various types of distortion for the reason that experience is cumulative. A crew's experience with one type of distortion may be of aid in the under-

standing of some other type of distorted speech. With this in mind it will be seen that a constant value of x fits the data for the various types of distortion reasonably well. In the case of changing speech levels with a constant type of speech distortion, where the question of uniformity of experience is not so important, the fit is even better.

It is reasonable to suppose that an inexperienced observer must make a greater mental effort than an experienced observer to obtain the same articulation values. In other words the element reflected

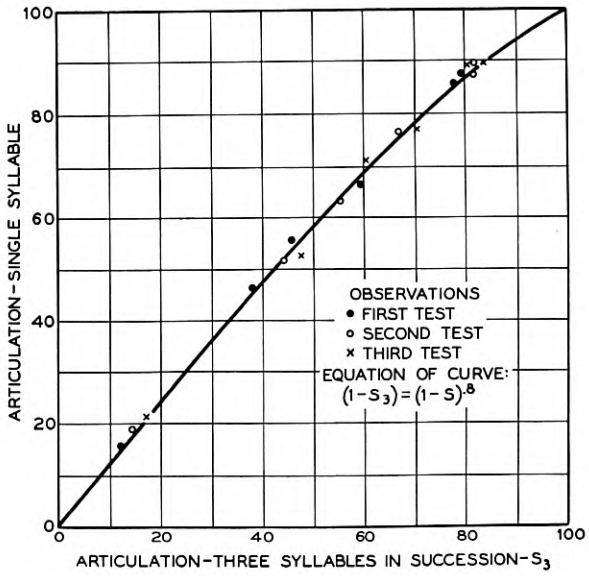


Fig. 5—Relation between techniques

by these curves is closely associated with the burden or strain upon the observer. A somewhat analogous situation obtains when tests are made with two different techniques which differ primarily in the burden imposed on the observer. In making the filter tests described above, two techniques which differed in this respect were used. One was the standard technique, in which one syllable was called with the introductory sentences. In the other, the syllables were called in groups of three (three in succession) with the sentences. The syllables were uttered as nearly in the manner of a three syllable word as was possible.

The results that were obtained with the two techniques are shown in Fig. 5. When the syllables are called in groups of three, the articulation values are smaller than when they are called singly.

It was found, however, that the type of relation shown in Eq. 1 also relates the data obtained with the two techniques. In this case the relation may be expressed as follows:

$$(1 - S_3) = (1 - S)^{\cdot 8 \cdots}, \tag{2}$$

where S_3 = decimal value of syllable articulation when called connectedly,

S = decimal value of syllable articulation when called singly.

The curve of Fig. 5 is a graph of the above equation.

In this case uniformity of experience with the various types of distortion does not enter, as the tests with the two techniques were made simultaneously. The only difference in the techniques was that in the three-syllable case the observer listened to three syllables before writing them down. It seems reasonable to conclude, therefore, that when a crew has the same experience with different types of distortion, then the results obtained by it at one time may be compared with the results obtained by it at some other time by using such a relation. No doubt other types of functions could be found which would also fit the above data. The relation shown here was chosen because it fit both the practice data and the data that were obtained with the two different techniques and is very convenient to use in making such corrections.

It is evident that in order to use the practice curves it is necessary to set up a reference circuit in order to obtain an appropriate value of x. Theoretically, one reference condition should be sufficient, provided that the practice of the crew had the same relative distribution over various types of speech distortion. Since this is usually not the case, it is necessary to use several reference circuits representing various types of speech distortion. When it is desired to correct data for practice effects, the appropriate value of x is determined by making tests upon the reference circuits having types of distortion similar to the circuits for which the corrections are desired. A description of several reference or control circuits which have been found useful with the values of sound and syllable articulation as obtained with the testing crew of five men and five women as previously described, is given below.

(*a*) *Air Transmission. Master Reference System for Telephone Transmission.*—The air transmission tests were made in a quiet, well damped room having a volume of approximately 1,000 cubic feet. The observers faced away and were located at an average distance of 30 inches from the speaker. Sound articulation "L" 99.1 per cent.

Syllable articulation "S" 97.5 per cent. Practically identical results were obtained with the "Master Reference System" [7] with the system set for optimum received speech level, i.e. a sensation level of 70 db, average distance from lips to transmitter 1.5 inches.

(*b*) *Auxiliary Circuit of the Master Reference System.*—The auxiliary circuit of the master reference system consists of networks which are

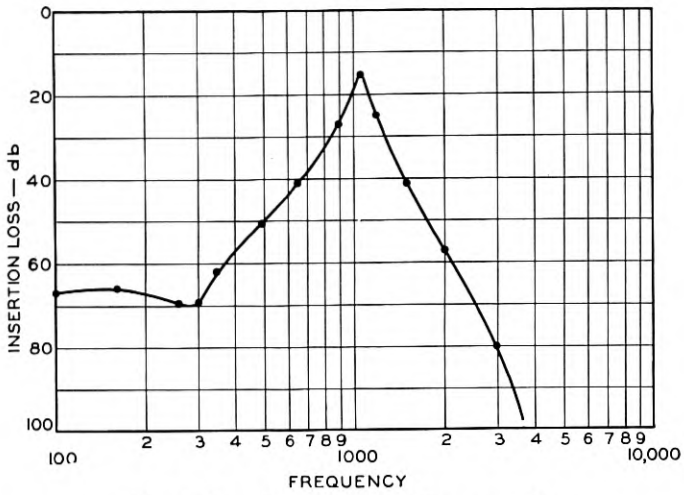


Fig. 6—Insertion loss of auxiliary networks

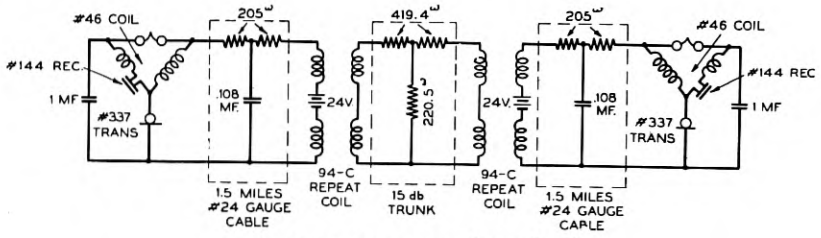


Fig. 7—Carbon transmitter circuit

inserted into the otherwise distortionless reference system, to give it a frequency resonance around 1,100 cycles. The insertion loss of the networks is shown in Fig. 6. This loss is approximately equal to the combined losses of the No. 1 transmitter and receiver distortion networks of the Master reference system. Sensation level 74 db $L = 89.2$ per cent, $S = 72$ per cent.

[7] L. J. Sivian, "A Telephone Transmission Reference System," *Electrical Communication*, 3, Oct. 1924. M. Cohen, "Apparatus Standards of Telephonic Transmission." W. H. Martin and C. H. G. Gray, "Master Reference System," *Bell Tech. Jour.*, July, 1929.

(*c*) *Carbon Transmitter Circuit* (see Fig. 7).—The average values for five transmitters are $L = 93$ per cent, $S = 81$ per cent. In these tests the sensation level of the received speech was 75 db, and the calling level as measured by a volume indicator bridged across the line side of the input repeating coil was -12.5 db.

(*d*) *Master Reference System Plus Filters.*—System set for a sensation level of 70 db.

3,750~ Low Pass Filter	$L = 96.7\%$	$S = 91.0\%$
750~ High " "	$L = 96.7\%$	$S = 91.0\%$
1,500~ Low " "	$L = 77.7\%$	$S = 49.5\%$
1,500~ High " "	$L = 91.0\%$	$S = 76.0\%$

The foregoing discussion has been concerned with methods of correcting the articulation results obtained by a given crew at different times to an arbitrary stage of practice or experience. To do this it is necessary to calibrate the crew for types of distortion that are similar to those of the systems for which the corrections are desired. The method has been described in detail because there are times when it is necessary to make such corrections. However, it has been our experience that such practice effects become negligible with a crew that has been set up in accordance with the methods previously described, when the crew's experience with types of distortion is diversified and when unusual circuits are tested successively until no further increase in articulation with practice occurs.

These methods may also be used to correlate the articulation data of various crews and various techniques, provided that the only essential difference between the crews and techniques is in the demand or burden that is placed upon the observer. This means that the crews must have similar vocal characteristics and similar hearing abilities, and that the testing lists must have similar speech sound content. It has been found, for example, that a crew of women callers obtain a considerably higher articulation than men callers on a circuit which eliminates all frequencies below 1,500 cycles and vice versa on a circuit which eliminates all frequencies above 1,500 cycles. It is obvious, therefore, that the methods described above could not be used to correlate the two crews for such circuits. Similarly, the methods could not be used for comparing two crews, if the hearing level of one is 10 db below the other, or to compare two techniques, one of which is made up entirely of vowel sounds and the other entirely of consonant sounds. As shown in Fig. 4-*b*, data have been obtained with various crews on various circuits which can be correlated very

well by means of the above curves. Fig. 4-*d* gives data that were obtained with other crews which show very poor correlation. In neither case are the characteristics of the crew well enough known to satisfactorily account for the observed differences. At the time the work was done the significance of these factors was not so apparent, so that they were not given the attention they now receive. During the past two years a crew of 10 people has been used almost continuously in testing work. During this time numerous changes in personnel have taken place and only five of the original members are now on the crew. The data obtained during this time appear to be strictly comparable. In some cases it is necessary to use the practice curves. In other cases (circuits that are frequently tested), practically identical results are obtained. For this reason, it is believed that if a similar crew of 10 different people were to be selected as previously described, comparable articulation results would be obtained. It seems reasonable to expect that crews testing in various languages should also obtain comparable results provided that the crews were similar in the sense used here and that the lists were phonetically similar. It seems desirable, therefore, to standardize on the factors which affect the comparison of data, such as, the size and type of crew, the type of list, and the type and number of reference circuits. Best results are likely to be obtained when the crews do not differ by amounts which correspond to values of x less than 0.7. Smaller values indicate that the crews have not had sufficient testing experience, or have speech and hearing characteristics which are essentially different, or that the phonetic content of the testing lists are appreciably different. In the latter case the results may be correlated by means of statistical relations that will be given in a later paragraph.

Relation of Articulation to the Transference of Thought by Speech

The foregoing paragraphs have been concerned with the practical problems of setting up a suitable testing technique and correlating the observed articulation results. The procedure that has been discussed enables us to measure the percentages of the various speech sounds which are correctly recognized when they are spoken in a simple con-vow-con syllable. We desire at this point to consider the broader significance of this measure. In other words, how is the articulation result related to the transference of thought by means of speech? This relationship involves many psychological factors which are difficult to evaluate so it must not be expected that a comprehensive answer can be given here, but it is important to understand

as fully as possible those parts of the problem that can be evaluated. Such a relation involves two questions, (*a*) how do the articulations of the sounds as measured with the testing lists compare with their articulations as they are used in speech, (*b*) how should the articulation values be weighted in order to obtain an index of the speech capabilities of a system.

In the first place, certain fundamental sounds of speech were omitted from the above lists. The most important of these are the consonant compounds. The majority of these sounds may be regarded as the product of a very few combining consonants acting as modifiers to the rest of the consonant alphabet. Since the combining consonants or modifiers occur over and over in combination with various consonants, it might be expected that the interpretation of the compounds would depend primarily upon the interpretation of the various consonants, and not upon the modifiers. In other words, the compounds would be interpreted as simple consonant sounds. The tests discussed below show that this is true on the average, although notable exceptions may occur in individual cases.

The testing lists were made up from the sounds shown in Table VI.

TABLE VI

Consonants		*Vowels*
Initial	*Final*	
b, br,	rb, b,	a'
d, dr,	rd, d,	a
g, gr,	rg, g,	e
p, pr,	rp, p,	i
k, kr,	rk, k,	o
t, tr,	rt, t,	
f, fr,	rf, f,	
th, thr,	rth, th,	
s, sl,	nd, d,	
b, bl,	nj, j,	
g, gl,	nz, z,	
p, pl,	nk, k,	
k, kl,	nt, t,	
f, fl,	ns, s,	
r, l,	n, r,	

These sounds were combined at random into syllables of the con-vow-con-con and con-con-vow-con form. Ten lists of 90 syllables each were made, and a crew of 10 callers, with 5 observers per caller, was used. With this number of tests the probable error in the per cent articulation for each sound is approximately 5 per cent. The tests were made on the auxiliary circuit of the master reference system. The sensation level of the received speech was about 80 db. The results are shown in Table VII.

TABLE VII

ARTICULATION OF CONSONANT COMPOUNDS

Initial				Final			
Sound	% Art.	Sound	% Art.	Sound	% Art.	Sound	% Art.
b	89.0	br	85.3	rb	63.3	b	77.0
d	98.0	dr	88.7	rd	60.0	d	91.0
g	96.3	gr	90.0	rg	57.3	g	88.7
p	78.3	pr	52.0	rp	39.3	p	66.7
k	95.3	kr	93.3	rk	52.7	k	90.7
t	79.3	tr	89.3	rt	58.7	t	88.7
f	56.3	fr	60.7	rf	39.3	f	53.3
th	71.3	thr	88.0	rth	42.0	th	52.0
ave.	83.0	ave.	81.0	ave.	51.5	ave.	76.0
s	57.3	sl	72.7	nd	92.7	d	91.0
b	89.0	bl	95.3	nj	91.3	j	96.0
g	96.3	gl	77.3	nz	82.0	z	76.7
p	78.3	pl	68.0	nk	92.7	k	90.7
k	95.3	kl	86.7	nt	84.7	t	88.7
f	56.3	fl	86.7	ns	72.7	s	44.7
ave.	78.7	ave.	81.1	ave.	86.0	ave.	81.3
Init. Ave.	81.1		81.0	Fin. Ave.	66.3		78.3

Ave. Simple Cons. 79.7
" Comp. " . . . . . 73.5 Exclusive of r () 82.7

The articulation of the consonant compounds as a class does not differ appreciably from the articulation of the corresponding simple consonants. The final r compound is seen to be an exception to this general rule. The errors for combinations containing this sound were caused by the large number of omissions of the modifier. For example, if "barb" were called, "bab" would be recorded. When the final r is combined with a consonant, the tendency is to shorten its duration and to stress it less than is done when it occurs as a simple final. Also, as mentioned before, the r sound materially modifies the vowel preceding it and usually in such a way that the vowel and r sounds are spoken as a vowel. For these two reasons, it escapes detection more readily than when used as a simple final. It will be noticed from the table that f is definitely more difficult to recognize than fl, while p is definitely less difficult to recognize than its compounds pl and pr. There is also a large difference between the results for s and ns. Although these differences are large, some tend to increase and others to decrease the average articulation. It is seen from the table that if the r compounds are omitted, the averages for the simple consonant sounds and for their compounds are approximately equal. Since this class of sounds comprises less than 15 per cent of the speech sounds, the results obtained by using a list in which the consonant

compounds are omitted will be very closely the same as those obtained by lists in which such sounds occur. In view of this, and also because their inclusion would greatly extend the time needed for testing, compound consonants have been omitted.

In conversational or written speech some of the sounds are used much more frequently than others, whereas in the testing lists each sound is used the same number of times. Does this procedure lead to essentially different articulation values, for the various sounds, from those obtained by using the sounds in proportion to their frequencies of occurrence in speech?

TABLE VIII

ARTICULATION OF SOUNDS OF EQUAL VS. UNEQUAL OCCURRENCE

Sound	Equal Occ.		Unequal Occ.	
	No. of Occur.	Art.	No. of Occur.	Art.
a	300	91.0	550	88.0
ā	300	98.3	600	96.5
a'	300	88.3	400	82.3
e	300	91.0	1200	87.4
ē	300	100.0	850	98.8
i	300	96.0	1150	94.5
o	300	86.0	500	90.6
ō	300	98.3	500	97.6
o'	300	97.7	450	96.2
u	300	88.3	150	92.7
ū	300	97.3	250	97.6
Ave.		93.8		92.9
b	300	87.0	400	91.0
ch	300	97.3	150	98.0
d	300	91.0	700	89.1
f	300	73.7	400	67.8
g	300	92.0	450	94.2
h	150	100.0	200	100.0
j	300	98.0	150	97.3
k	300	87.3	900	89.8
l	300	98.7	1200	96.2
m	300	95.3	900	81.3
n	300	93.7	1300	95.5
ng	150	93.3	400	90.0
p	300	75.3	500	71.2
r	300	97.0	1250	97.4
s	300	72.3	850	76.3
sh	300	99.3	300	99.7
st	150	96.0	200	96.5
t	300	83.7	1550	86.6
th	300	60.3	200	60.5
th'	300	70.7	150	68.0
v	300	78.7	250	71.6
w	150	96.7	400	97.8
y	150	99.3	100	94.0
z	300	67.0	250	72.0
zh	150	97.3	50	100.0
Ave.		88.0		87.3

Table VIII gives the results of articulation tests that were made with two such types of lists. In both cases the sounds were combined at random into syllables of the con-vow-con type. The tests were made on the auxiliary circuit of the master reference system.

Realizing that the probable error in the articulation value given for each sound is ± 5, there do not appear to be any outstanding differences in the articulations of the various sounds with the two types of list. The average articulations for the two lists differ by less than the probable error. The test indicate, therefore, that lists having uniform occurrence of sounds give the same individual sound articulation values as lists having the frequencies of occurrence of the sounds proportional to their frequencies of occurrence in speech. At least this is true within the accuracy usually attained in making such tests. The testing advantages of the former type of list have already been pointed out.

It is important to notice that the average sound or the average syllable articulation may not be the same for the two types of lists. even though the articulation for each sound is the same. The averages shown in the table were obtained by assigning equal weights to the articulation for each fundamental sound. If weights which are proportional to frequency of occurrence of the sounds in speech be assigned, the averages obtained will, in general, be slightly different. For the particular circuit corresponding to the data of Table VIII, the averages obtained in the two ways did not differ by more than the observational error. Our data have shown that this is also true for a large class of circuits ordinarily used in telephone work. However, those transmission systems which have a specific effect upon certain consonant or vowel sounds, for example, upon s which occurs 850 times in one list compared to 300 times in the other, would obviously have different values for the sound articulation by using the two methods of obtaining the average.

In speech, certain combinations of sounds occur more frequently than others. In other words, some consonants precede certain vowels more frequently than they do other vowels, and similarly, some consonants follow certain vowels more frequently than others. For example, the combination "es" is used much more frequently than the combination "us" (u as in foot). Since the testing lists are made by random selection, the various con-vow and vow-con combinations occur with uniform frequency. In order to determine how this difference influences the interpretation of the sounds, articulation data on various circuits were examined. Attention was focused first on the final consonant sounds. One hundred errors for each consonant,

or 2,200 consonant errors were selected at random from the articulation data and the number of these 2,200 errors that occurred after each vowel sound ascertained. Similarly, the number of vowel errors, out of a total of 1,100 errors, that occurred after each of the consonant sounds, was determined.

Probability studies indicate that the distribution of these errors as shown in Table IX is of the same order as that to be expected on

TABLE IX

DISTRIBUTION OF VOWEL AND CONSONANT ERRORS

Distribution of Vow. Errors		Distribution of Fin. Con. Errors	
Preceding Con. Sound	No. of Vow. Errors	Preceding Vow. Sound	No. of Fin. Con. Errors
b	58	a	188
ch	73	ā	264
d	52	a'	232
f	49	e	166
g	50	ē	214
h	55	i	172
j	54	o	192
k	60	ō	192
l	45	o'	208
m	49	u	196
n	54	ū	176
p	68		
r	39		
s	50		
sh	53		
th	38		
t	52		
v	40		
w	52		
y	49		
z	60		

the basis that the distribution of errors is due entirely to chance. Since, from the way the lists were constructed, the occurrence distribution is due to chance, it is evident that the errors in recognition of the sounds do not depend upon the particular sounds that they follow. Although the analysis was not made, it would be expected that a similar situation obtains for initial consonant and vowel errors. These data may be interpreted to mean that the consonant articulation, the vowel articulation, the sound articulation, or the syllable articulation, is approximately independent of the particular sound combinations, when a wide variety of combinations are used. The results obtained with these lists, therefore, are as representative of speech as the results that would be obtained with lists employing particular sound combinations in proportion to their frequencies of occurrence

in speech. The analysis was not extensive enough to draw conclusions as to the effects of particular sound combinations upon the articulation of individual speech sounds.

Approximately 40 per cent of the syllables that occur in English are of the con-vow-con type. About 34 per cent are of the con-vow, and vow-con type. The syllables, including the compounds, such as, con-con-vow, vow-con-con, con-vow-con-con, and con-con-vow-con, make up about 16 per cent of the syllables of English. Since, as pointed out above, the interpretation of the consonant compounds depends primarily upon only one of the consonants, the latter syllables may be grouped in the two former classes, which then constitute some 90 per cent of English. Of the remaining syllables, 7 per cent consist of a single vowel, so that the more complex syllable forms constitute only 3 per cent of English.[8] Since 97 per cent of the syllables of English are included in the one, two and three letter forms, there is little reason to include the more complex syllable forms in order to represent speech, when as has been previously stated, they are undesirable from a testing standpoint. As will be shown in a later paragraph, one, two and three-letter syllables all yield equal values of articulation for the various speech sounds. Since the three-letter syllables require a smaller testing time for a given number of called sounds, the other syllable forms were excluded from the testing lists.

Having shown that the standard technique gives, for the various sounds, data that are representative of speech, the question now arises as to the best figure that may be computed from the data obtained with this technique, in order to best represent the speech transmission ability of the system under test. Before discussing this, it is necessary to consider some probability relations existing between the quantities entering into the calculation of such a figure.

Statistical Relations

The syllable articulation S when expressed as the ratio of the number of successes (correct interpretations of the syllables) to the number of trials (syllables called) is the chance of perceiving a syllable correctly. Also, if a similar ratio is used for the sound articulation L, the vowel articulation V, and the consonant articulation C, then these letters represent the probability of perceiving correctly a fundamental sound, a vowel sound or a consonent sound, respectively. If a syllable contains only one fundamental sound, then it is obvious that

$$S = L. \tag{3}$$

[8] These data were obtained from Godfrey Dewey's book "Relative Frequency of English Speech Sounds," Harvard University Press, 1923.

If a syllable has two letter sounds, then the chance of perceiving them both correctly is the same as the chance of perceiving the syllable correctly or

$$S = L^2. \tag{4}$$

Similarly, for a syllable containing m sounds

$$S = L^m. \tag{5}$$

Or if A_1, A_2, A_3, $A_4 \cdots A_m$ give the per cent of syllables in the list containing 1, 2, 3, 4, $\cdots$ m sounds, respectively

$$S = A_1L + A_2L^2 + A_3L^3 + \cdots A_mL^m. \tag{6}$$

Similarly, the chance of perceiving a syllable of the type con-vow or vow-con is VC; of the type con-vow-con, con-con-vow or vow-con-con is VC^2; of the type con-con-vow-con, con-vow-con-con, vow-con-con-con, or con-con-con-vow, is VC^3, etc.

For the old standard articulation lists these formulæ reduced to

$$S = \frac{1}{5}VC + \frac{4}{5}VC^2 = \frac{1}{5}L^2 + \frac{4}{5}L^3. \tag{7}$$

For the new standard articulation lists they reduce to [9]

$$S = VC^2 = L^3. \tag{8}$$

If a list of N syllables is used, then the letter errors and syllable errors will be $3N(1 - L)$ and $N(1 - L^3)$, respectively, or the number of letter errors per mistaken syllable, for the new standard lists, will be

$$m = \frac{3}{1 + L + L^2}. \tag{9}$$

It is seen that m approaches 3 as L becomes small, and unity as L approaches unity. For $L = .30$, $m = 2.06$; for $L = .50$, $m = 1.71$;

[9] When derived from the probability formulæ

$$VC^2 = L^3.$$

However, from the definition of V, C and L,

$$L = (2C + V)/3 \quad \text{so that} \quad VC^2 = L^3.$$

The difference is

$$d = \frac{(V + 8C)(V - C)^2}{27}.$$

Actually V and C are not wholly independent of each other and when values as obtained in tests are substituted in the above equation, the difference turns out to be small.

and for $L = .80$, $m = 1.23$. When observed values of m become consistently greater or less than this theoretical value, it must be concluded that the assumptions underlying this statistical theory are not valid.

All of the above statistical relations are dependent upon the tacit assumption that the chance of perceiving any sound correctly is entirely independent of the other sounds present and also independent of the number of other sounds present. It was shown in the previous section that the articulation of the various sounds is, on the average, independent of the other sounds in the syllables. On the other hand, experiments have indicated that the articulation does depend upon the number of sounds in the syllable. The sound articulation becomes smaller when the number of sounds in the syllables increases beyond three per syllable.

The data from which this conclusion was drawn were taken from three different experiments. In the first, three different transmission systems were tested by using first the standard articulation lists and then the vowel-consonant lists which are described in the last section. When using the vowel list, the vowels only are considered and when using the consonant list the consonants only are considered. These lists together, then may be considered as composed of syllables having only one sound. The syllable and sound articulations are the same when using such lists. The comparison of the results obtained with the two types of lists is shown in Table X. It will be seen that there

TABLE X

ARTICULATION FOR ONE- AND THREE-SOUND SYLLABLES

	Freq. below 1,000~ only			Freq. below 1,950~ only			Freq. above 1,500~ only		
	Vow. Art.	Cons. Art.	Sound Art.	Vow. Art.	Cons. Art.	Sound Art.	Vow. Art.	Cons. Art.	Sound Art.
One-sound Syllables	71.5	62	65	98.5	82.5	88	81	96	91
Three-sound Syllables	69.5	61.5	64	96	83	87	80	96.5	91

is only a slight tendency for the sound articulation to be lower for the three-sound syllable when compared with the one-sound syllable. The differences are within the observational error in testing.

In the second experiment a new list was constructed using the syllable forms con-vow and vow-con. The auxiliary circuit of the master reference system was tested with this list and also with the standard articulation list. The results are shown in Table XI. The

TABLE XI
ARTICULATIONS FOR TWO- AND THREE-SOUND SYLLABLES

	V	*C*	*L*
Two-sound Syllables	95	88	90
Three-sound Syllables	94	87	89

number of syllables used of each type for determining these averages was 1,344. The average sound articulation in each case was determined by giving equal weights to the articulation for each sound. For the three-sound syllables this is done by dividing the number of sounds correctly recognized by the total number of sounds called. For the two-sound syllable the procedure is not so simple. Since each vowel sound occurs twice as often as each consonant sound, it is necessary to obtain an average for the vowels and consonants

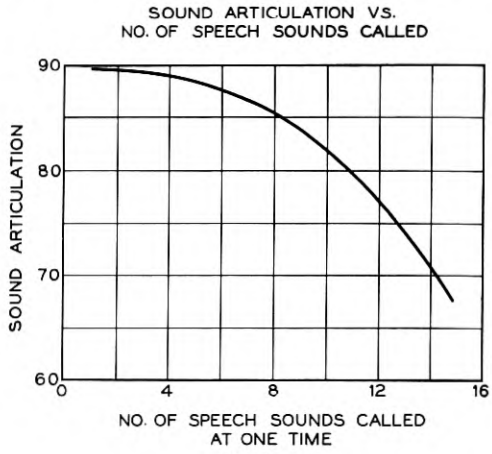


Fig. 8

separately. The final average value for the sound articulation is obtained by assigning weights of 1 and 2 to the vowel and consonant articulations, respectively. It is seen that there is no appreciable difference between the values of L obtained by the two types of lists.

In the third experiment the standard lists were used to test the auxiliary circuit but the syllables were called in groups of 1, 2, 3, 4, or 5 at a time. The results of these tests are shown in Table XII.

TABLE XII

Number of Sounds Called at One Time	Sound Articulation
3	89.0
6	87.5
9	84.0
12	77.0
15	67.0

From these three sets of data and from other available data which could be applied to this problem, the curve shown in Fig. 8 was constructed. It gives the sound articulation which would be obtained for a circuit such as the auxiliary circuit of the master reference system, when the number of sounds that are spoken at a time, that is, before the observer starts writing, is represented by the abscissa. It is evident from the shape of this curve that the assumptions underlying the statistical formulæ are valid for syllables having three or less sounds per syllable, and that they will break down for the more complex types of syllables. These assumptions might be expected to break down also, for certain extreme types of distortion.

Definite relations between the vowel, consonant, sound, and syllable articulations for both the old and the new techniques, have been derived by statistical theory. An experimental relationship between these quantities is shown in Figs. 9 and 10. These were obtained by an analysis of the errors of a large number of tests with widely different types of distortion, the data in Fig. 9 being taken with the old and the data in Fig. 10 with the new technique.

In the figures observed values of sound articulation have been plotted against the corresponding observed syllable articulation values. The solid curves in the two cases were calculated from Equations 7 and 8, respectively. The observed values agree reasonably well with the theoretical curves.

There is very little correspondence between the vowel and syllable or consonant and syllable articulation. The table below shows that

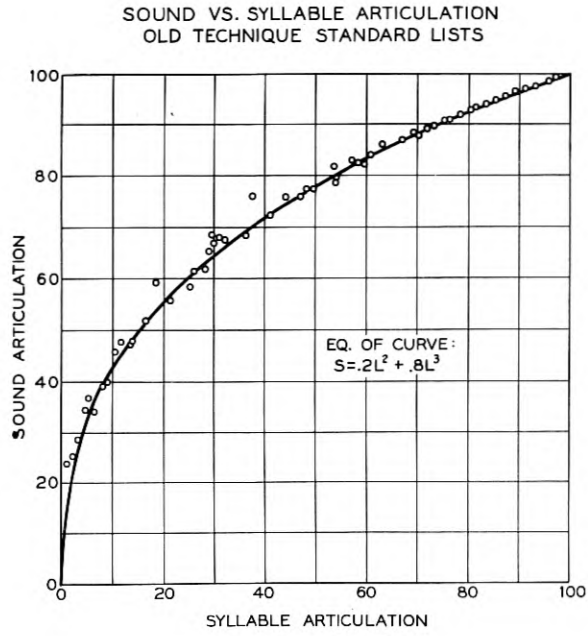
TABLE XIII

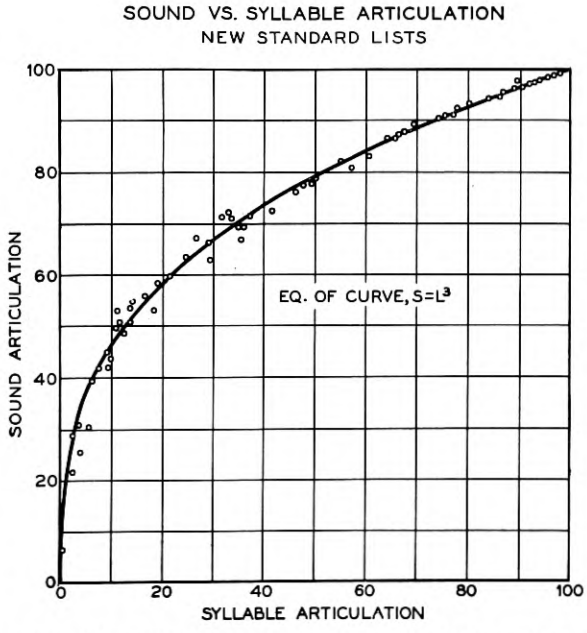
VOWEL, CONSONANT, AND SYLLABLE ARTICULATIONS

	V	C	S	VC^2
3,750 L.P.F.	98.7	95.4	90.8	89.8
750 H.P.F.	93.0	98.6	90.9	90.4
2,850 L.P.F.	98.5	92.9	86.3	85.4
1,000 H.P.F.	89.0	98.4	87.0	86.2

a circuit which discriminates against the vowels may have a syllable articulation equal to another circuit which discriminates against the consonants. However, it is seen that the product VC^2 is equal to S as the statistical theory indicates. If then *the sound, or vowel and consonant articulations are known, it is possible to calculate the syllable articulation*, for the case of two- and three-sound syllables.

We are now in a position to consider the figure which best represents





Figs. 9 and 10—Relation between sound and syllable articulation

the capabilities of systems to transfer thought by means of speech. For giving a complete picture, it is necessary to give the articulation values for each speech sound. Since this involves 36 articulation values, it is difficult to compare various systems. To combine these values into averages raises the question of how such an average shall be taken. At first thought it might seem obvious that the weights assigned to each sound articulation value should be proportional to the frequency of occurrence of that sound in English speech. Many of the most frequently occurring words, however, such as the, of, and, to, in, that, etc., do not carry much of the thought, so that it seems reasonable to exclude the effects of such words in the weighting process. It is evident that many sets of weighting factors could be evolved depending upon how far the exclusion process is carried and depending upon whether written or spoken English is used, in determining the frequencies of occurrence of the sounds. After excluding the twenty or twenty-five most common words, however, further exclusion does not appreciably change the calculated articulation value. The table below gives a set of factors obtained from the frequencies of occurrence used in Table VIII. They are based upon the studies of Messrs. French and Koenig [10] on the frequencies of occurrence of speech sounds in spoken English. The effects of the more common parts of speech, such as, personal pronouns, definite articles, conjunctions, and prepositions have been excluded.

TABLE XIV

Group I	Weight	Group II	Weight	Group III	Weight	Group IV	Weight	Group V	Weight
ā	3.0	i	5.8	r	6.3	d	3.5	z	1.3
ē	4.3	o	2.5	l	6.1	t	7.8	s	4.3
ō	2.5	a'	2.0	ng	2.0	b	2.0	v	1.3
ū	1.3	u	.8	n	6.6	p	2.5	f	2.0
o'	2.3	e	6.0	m	4.5	g	2.3	zh	.3
a	2.8	y	.5			k	4.5	sh	1.5
		w	2.0			j	.8	th'	.8
						ch	.8	th	1.0
						h	1.0	st	1.0
Total Weight	16.2		19.6		25.5		25.2		13.5

It will be noticed that the speech sounds are arranged in five groups. The sounds in each group have very similar characteristics, so instead of dealing with 36 articulation values for a circuit, it is only necessary

[10] "Frequency of Occurrence of Speech Sounds in Spoken English," N. R. French & W. Koenig, *Proc. Acoustical Society of America*, 1929.

to deal with the average value for each of the five groups. The average for the first group is designated V_l, signifying long-vowel index; for the second group V_s, signifying short-vowel index; for the third group C_n, signifying nasalized-consonant index; for the fourth group C_s, signifying stop-consonant index; for the fifth group C_f, signifying fricative-consonant index. If the articulation obtained from any test for each sound be designated by the phonetic symbol for that sound, then,

$$\left.\begin{aligned} V_l &= .19\ \bar{a} + .27\ \bar{e} + .15\ \bar{o} + .08\ \bar{u} + .14\ o' + .17\ a \\ V_s &= .30\ i + .13\ o + .10\ a' + .04\ u + .31\ e + .02\ y + .10\ w \\ C_n &= .25\ r + .24\ l + .08\ ng + .26\ n + .17\ m \\ C_s &= .14\ d + .31\ t + .08\ b + .10\ p + .09\ g + .18\ k + .03\ j \\ &\qquad + .03\ ch + .04\ p \\ C_f &= .10\ z + .32\ s + .10\ v + .15\ f + .02\ zh + .11\ sh \\ &\qquad + .06\ th' + .07\ th + .07\ st. \end{aligned}\right\} \quad (10)$$

The sound index is related to these values by

$$i = .162\ V_l + .196\ V_s + .255\ C_n + .252\ C_s + .135\ C_f. \quad (11)$$

For obtaining the most representative single value for the syllable index I, the equation given below is used.

$$I = .5\ i^2 + .5\ i^3. \quad (12)$$

This equation is based upon the frequency of occurrence of the syllable forms in English speech. As pointed out before, if the compound consonants be considered as simple sounds, then there are less than 10 per cent of syllable forms other than the two- and three-sound type. The frequency of occurrence of these two types is approximately equal.

Similar formulæ to the above may be used to relate articulation results in English to articulation results in a different language. To do this it is necessary to select the fundamental sounds of the different languages that correspond to the 36 fundamental sounds of English, where the correspondence is based on similar phonetic characteristics and similar positions of the vocal organs in producing the sounds. When this is done, the coefficients in Eqs. 10, 11 and 12, must be modified to correspond with the frequencies of occurrence of the sounds and syllables in the language.

Observed values of individual sound articulation are thus reduced

to a single index, or for a more comprehensive picture, to five indices corresponding to the five groups of speech sounds. In order to compare the indices obtained by a given crew with those of a reference crew, it is necessary to correct the data in accordance with Eq. 2 for the effects of practice. To do this, as previously discussed, articulation tests are made upon one or more of the reference circuits by the crew in question. If I' is the syllable index so obtained, the practice factor for the crew is given by the relation

$$x = \frac{\log (1 - I')}{\log (1 - I)}. \tag{13}$$

The practice factors for the other indices may be obtained also, by substituting the appropriate indices for the syllable index in Eq. 13. In Table XV the reference values for the various indices are given for the reference circuits that were previously described.

TABLE XV

REFERENCE VALUES

Circuit	V_l	V_s	C_n	C_s	C_f	i	I
Master Reference System	98.5	98.9	99.6	99.2	98.8	99.3	98.0
Auxiliary Circuit of Master Ref. Sys.	95.0	95.0	96.5	88.5	66.5	90.0	77.0
Carbon Transmitter Circuit	97.0	97.0	96.5	93.5	82.0	94.0	85.5
Master Ref. Sys. plus 3,750- L.P.F.	99.0	99.5	99.5	99.0	86.5	97.6	94.0
" " " " 750- H.P.F.	96.0	92.5	99.0	99.5	98.5	97.1	93.0
" " " " 1,500- L.P.F.	93.5	86.5	90.5	76.0	52.5	80.2	58.0
" " " " 1,500- H.P.F.	85.0	82.5	96.0	97.5	97.0	92.0	81.0

If the values for the sound index be compared with the sound articulation values based on uniform weighting, that were given under the section on practice effects, it will be seen that for these circuits there is very little difference, between the two sets of values. In other words, the average sound articulation is very nearly equal to the average that is obtained when the individual sound articulations are weighted according to the frequencies of occurrence of the sounds in English.

Similar comparisons have been made for a large number of other transmission systems. They showed similar small differences between the weighted and unweighted averages. For this reason we consider it unnecessary to use the weighted average when great accuracy is not required, for example, in a great deal of our routine work where comparisons are being made between circuits which have similar characteristics. This means that when testing an unknown circuit,

having an electrical characteristic similar to one of the reference circuits, the syllable index I can be calculated from the observed syllable articulation S (as obtained with the new standard lists) by means of the equation,

$$I = .5\ S^{2/3} + .5\ S. \tag{14}$$

This value must now be reduced to the reference condition of practice by the methods which have already been described. In such cases it is thus possible to obtain the syllable index from the observed syllable articulation values, and it is unnecessary to analyze the data for the individual sound articulation values.

The weighted average, however, is the more logical way of obtaining a single index and should be used when it is suspected that it might give results which are essentially different from the unweighted average.

It is possible to carry the probability relations a step further and apply them to cases of English words and sentences. In order to do this it is necessary to make assumptions as to how the thought or meaning of the words affects the interpretation of the sounds. These assumptions are not only somewhat uncertain, but owing to psychological factors in testing are difficult to verify experimentally. In general, the meaning associated with words makes them easier to interpret than meaningless words. For single-syllable words, these effects are small. Two-syllable words are easier to interpret than single-syllable words. The interpretation of words containing from three to five syllables, and short sentences, depends almost entirely upon interpreting those parts which are not indicated by the thought or meaning.

Other Testing Methods

For most articulation studies it has been found desirable to use the standard testing technique which has been described, but it is frequently necessary, in special cases, to use other techniques. In such cases it is desirable, if possible to interpret the results in terms of the standard technique. In the course of research work, several different articulation testing methods have been used which give information on the type of correlation between them that may be expected.

The probability relations have been made use of in constructing two other types of lists which are called vowel-consonant and vowel word-consonant word lists. These lists are designed to give the same values of sound articulation as given by the standard lists. The former lists are shown in Table XVI. The various vowels are combined with the same consonant, and the various consonants with the

TABLE XVI

VOWEL LIST

Sound to be Graded	Testing Syllables in the List	
a	at	ta
ā	āt	tā
a'	a't	ta'
e	et	te
ē	ēt	tē
i	it	ti
o	ot	to
ō	ōt	tō
o'	o't	to'
u	ut	tu
ū	ūt	tū

CONSONANT LIST

Sound to be Graded	Testing Syllables in the List					
b	bū	ūb	ba	ab	bē	ēb
d	dū	ūd	da	ad	dē	ēd
f	fū	ūf	fa	af	fē	ēf
g	gū	ūg	ga	ag	gē	ēg
k	kū	ūk	ka	ak	kē	ēk
l	lū	ūl	la	al	lē	ēl
m	mū	ūm	ma	am	mē	ēm
n	nū	ūn	na	an	nē	ēn
r	rū	ūr	ra	ar	rē	ēr
p	pū	ūp	pa	ap	pē	ēp
s	sū	ūs	sa	as	sē	ēs
sh	shū	ūsh	sha	ash	shē	ēsh
th'	th'ū	ūth'	th'a	ath'	th'ē	ēth'
th	thū	ūth	tha	ath	thē	ēth
t	tū	ūt	ta	at	tē	ēt
v	vū	ūv	va	av	vē	ēv
ch	chū	ūch	cha	ach	chē	ēch
z	zū	ūz	za	az	zē	ēz
j	jū	ūj	ja	aj	jē	ēj
h	hū		ha		hē	
w	wū		wa		wē	
y	yū		ya		yē	
zh		ūzh		azh		ēzh
ng		ūng		ang		ēng
st		ūst		ast		ēst

same vowel. The technique of using the list is the same as that previously described, except that the vowel articulation and consonant articulation are measured separately. Only the vowel errors are counted when using the vowel list and only the consonant errors when using the consonant list. These lists have the advantage that they can be used over and over by merely changing the sequence of the syllables.

Table XVII shows two lists similar to the above except that they are made up entirely of common English words. They are designated as vowel word and consonant word lists. This list is used in the same way as the vowel and consonant lists. In using either of these lists

the testing crews should be familiar with the syllables or words in the lists.

TABLE XVII

VOWEL WORD LIST (ENGLISH WORDS)

Sound to be Graded	English Words in the List	
a'	bat	back
ā	bait	bake
e	bet	beck
ē	beat	beak
i	bit	bit
ī	bite	bike
o	but	buck
o'	bought	balk
ō	boat	boat
u	book	book
ū	boot	boot

CONSONANT WORD LIST (ENGLISH WORDS)

Sound to be Graded	English Words in the List	
b	by	by
ch	which	which
d	die	die
f	fie	whiff
g	guy	wig
h	high	high
j		
k	wick	wick
l	lie	will
m	my	whim
n	nigh	win
ng	wing	wing
p	pie	whip
r	wry	wry
s	sigh	sigh
sh	shy	wish
th'	thy	with
th	thigh	thigh
t	tie	wit
v	vie	vie
w	why	why
y		
z	whiz	whiz
st	sty	whist

Note: The h following w is not pronounced in such words as whim, whip, etc.

It usually requires a training period of a month or more for a testing crew to thoroughly master the technique of using the standard lists, that is, to reach a stage where the phonetic symbols are spoken and recorded almost mechanically. The vowel consonant lists require less time, since it is only necessary for the observers to fix their attention on one sound in the syllable. With the word lists this training period is reduced to a minimum. Phonetic symbols are avoided, and attention is given to only one sound in the words. As may be seen from Fig. 2, after a few tests they practically reach a degree of

uniform proficiency. In using the lists, the words are recorded with the English spelling. Only errors in the vowel and consonant sounds of the left-hand column of the above table, are counted. Since only one sound in each syllable is utilized, the above lists require a somewhat greater testing time for a given precision than do the standard lists where all three sounds of the syllables are used.

Table XVIII below, shows data that were obtained with the three types of lists, namely, the standard lists, the vowel consonant lists, and the vowel word consonant word lists. The vowel consonant

TABLE XVIII

ARTICULATION RESULTS WITH VARIOUS LISTS

Circuit	Cons. Vow. List		Cons. Word Vow. Word List		Standard List			Corrected
	C	V	C_w	V_w	S	VC^2	$V_wC_w^2$	$V_wC_w^2$
Air Transmission	99.5	99.0	99.0	99.5	98.0 ± .4	98.0	98.0	97.5
Master Ref. System + 5500 L.P. Fil.	99.0	99.0	99.0	99.5	96.5 ± 1.0	97.0	97.5	97.0
M.R.S. + 3750 L.P. Fil.	95.5	98.5	96.5	99.0	90.5 ± 2.0	90.0	92.0	90.0
M.R.S. + 1950 L.P. Fil.	82.5	98.5	85.5	98.5	67.0 ± 2.0	67.5	72.0	68.0
M.R.S. + 1000 L.P. Fil.	62.0	71.5	67.0	82.0	29.0 ± 3.0	27.5	37.0	33.0
M.R.S. + 1500 H.P. Fil.	96.0	81.0	97.0	86.0	75.0 ± 2.0	75.0	81.0	77.0
Carbon Transmitter Circuit	91.0	96.0	91.5	99.0	81.0 ± 1.5	80.0	83.0	80.0

lists lead to syllable articulations that are essentially the same as obtained with the standard lists. It is evident from the data that the word lists lead to slightly higher values of syllable articulation. The explanation is to be found in the make up of the word list. It was not possible to arrange the lists so that all sounds are equally probable, and still use English words. For instance, an observer would not record "bik" for "beck," nor "wiv" for "with," etc. For this reason, the observers do not make errors which occur frequently in the other two types of lists, where any sound is possible. Hence the observed articulations are somewhat higher with the word lists.

It was found, however, that the word lists may be correlated with the standard lists by means of the relation given by Equation 1. The value of x for this case is 0.9 so that, the word technique may be corrected to the standard technique, by means of the equation.

$$S = 1 - (1 - V_wC_w^2)^{\cdot 9}, \tag{15}$$

where S = syl. art. of standard lists expressed as a ratio,

V_w = vowel art. of vowel word lists expressed as a ratio,

C_w = cons. art. of cons. word list expressed as a ratio.

The corrected values are also shown in the above table.

It is frequently necessary to test very poor systems where the standard lists giving an articulation of a few per cent, are not satisfactory. The vowel consonant lists are somewhat more satisfactory under these circumstances. Lists of sentences have also been found to be very useful for such purposes. The sentences were of the interrogative or imperative form containing a simple idea. They were designed to test the observer's acuteness of perception rather than his intelligence. Tests were made with these sentences and the standard lists on various circuits, involving carbon transmitter circuits and various filter systems. The data are shown in Fig. 11. The

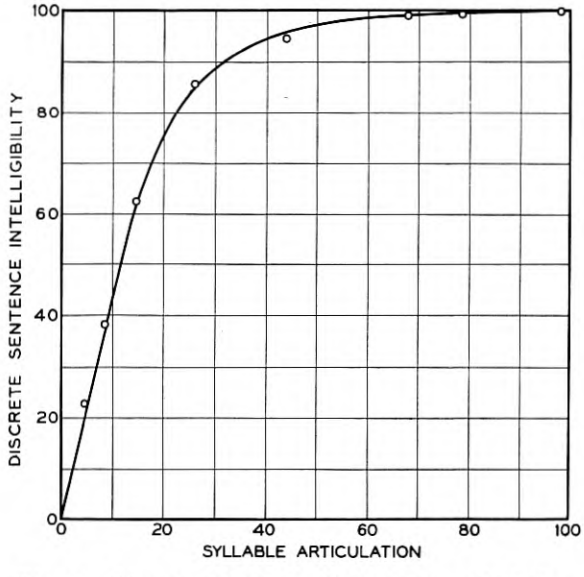


Fig. 11—Discrete sentence intelligibility vs. articulation

sentences were considered to be understood if the observer either recorded the sentence correctly or recorded an intelligent answer. As stated earlier, the percentage correctly observed is called the discrete sentence intelligibility.

It will be seen that for changes in distortion, the changes in the discrete sentence intelligibility will be small for systems having syllable articulations greater than 30 per cent, but very large for systems having syllable articulations below 20 per cent. It is for systems in this latter class that these test sentences are useful. A case in point is the measurement of the degree of secrecy obtained

in sound proofing telephone booths, or in dealing with cross-talk. The sentences have also been found to be useful in making quick qualitative tests of the goodness of an audiphone set for a particular case of deafness.

Because of their general usefulness for these purposes, the complete lists of sentences are given in the appendix. Due to memory effects a set of sentences can be used with the same personnel only a very few times. The psychological factors are also more prominent with sentences than with simple syllable.

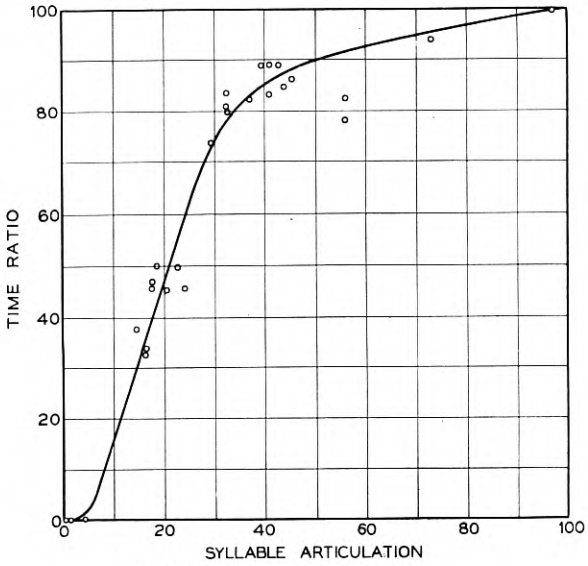


Fig. 12—Articulation vs. time ratio

Sentence lists of the above type have also been used to obtain a notion of how the time taken to transmit an idea correctly over a system depends upon the articulation. To do this, the observer was instructed to reply orally to the question. If the reply indicated that the observer failed to understand, the speaker repeated the question. Both speaker and observer tried to carry out the test in a normal conversational manner. The observer could ask the speaker to repeat, reword or spell out difficult parts of the sentence.

The tests were made on a variety of systems of known syllable articulation. The results that were obtained are shown in Fig. 12. The ordinates of the curve give the ratio of the time required to

transmit correctly one of these test sentences over an ideal system to the time required over the system under test. With the crew used in making these tests, and with an ideal transmission system, it required an average time of 5.2 seconds after the speaker started to pronounce the sentence before the observers grasped the idea. It will be seen from the curve that for systems having approximately 20 per cent articulation, the time required is twice as great. Fig. 11 shows that one out of every four of the sentences is mistaken for this value of articulation. If it is assumed that an observer asks that only sentences which he fails to understand be repeated, it can be shown that this time ratio is equal to the discrete sentence intelligibility.[11]

It is evident from Figs. 11 and 12, that the observed time ratio is appreciably less than the discrete sentence intelligibility. This difference may be taken to indicate that an observer not only asks that sentences which he fails to understand be repeated, but also that sentences about which he is uncertain be repeated. In other words, the time element reflects both factors, the understandability and the uncertainty.

As has been previously mentioned, tests have been made with various types of English word lists. Because of the manner in which the words were selected, and also due to uncertain psychological factors entering into the tests when such words are used repeatedly, it is difficult to compare the results so obtained with syllable articulation results.

However, it was found that if a definite rule were followed in selecting words from a newspaper, consistent results could be obtained with lists containing 500 or more words per list. The method of selection was to take the first word from every third line of a newspaper column. In this selection all proper names and the following six most frequent words of English were excluded, the, of, and, to a, in. When a word was hyphenated from the previous line, the whole word was used. Each of eight callers called a list of 66 words to four observers in the manner of an ordinary standard articulation test. Tests were made with the carbon transmitter circuit and the six circuits indicated in Fig. 3. The data were analyzed to give the discrete word intelligibilities for the one, two, three, four, and five-syllable words occurring in the lists, as well as for the lists as a whole. The lists on the average contained 46.3 per cent one-syllable, 29 per cent two-syllable, 16.8 per cent three-syllable, 6.4 per cent four-

[11] "A Theoretical Study of Articulation and Intelligibility of a Telephone Circuit," John Collard, *Electrical Communication*, 7, page 168, January, 1929.

syllable, and 1.5 per cent five-syllable words, and an average number of two syllables per word. The discrete word intelligibility vs. syllable articulation as obtained with the standard lists is shown in Fig. 13. The dashed curves indicate the relations for the various types of words, and the solid curve for the word lists as a whole. The data for two syllable words practically coincided with the solid curve. Owing to the small amount of data, the curves for the four- and five-syllable words are less reliable than those for the other types. Curves of the above type, both for words and sentences, depend very

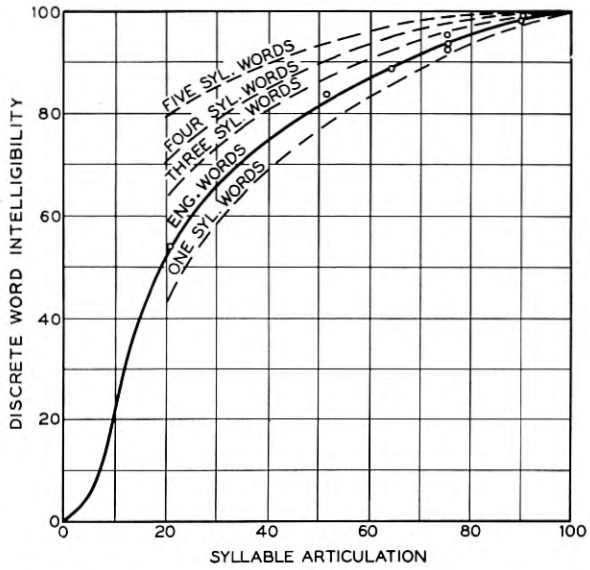


Fig. 13—Discrete word intelligibility vs. syllable articulation

much upon the way the speech material is selected. If, for example, only "different" words had been included in the word lists, appreciably smaller values of discrete word intelligibility would have been obtained.

Tests have also been made with lists made up of the following numbers, 1, 2, 3, 4, 5, 6, 8. These numbers were combined at random into groups of three and called in the manner of an ordinary articulation syllable. The distinguishing characteristic of each of the above numbers is a vowel sound, so that, they are interpreted primarily from recognizing the vowel. Such lists, therefore, do not give a very good picture of the speech capabilities of a system which distorts speech. They are, however, very useful in measuring the deafness

of an observer, for the reason that the number articulation decreases very rapidly as the sounds approach the threshold of hearing. As may be seen from Fig. 14, the number articulation passes from practically 100 per cent to 0 per cent in the short range of 10 or 15 db. It is evident that such lists give a critical measure of the point at which

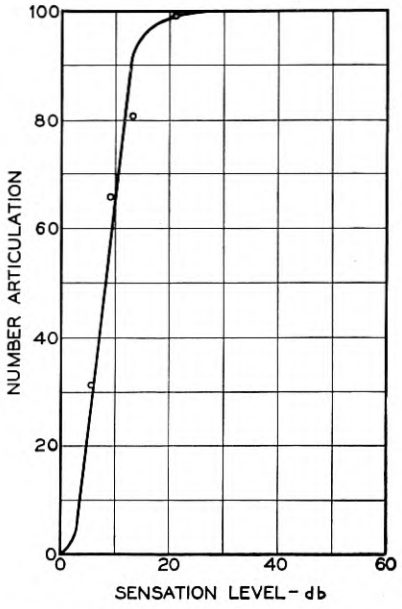


Fig. 14—Number articulation vs. sensation level

an observer fails to hear the sounds. Lists of this type have been used extensively in testing the hearing of school children.

Summary

The standard testing technique is primarily a means of determining the articulation or recognizability of the individual speech sounds when they are spoken in a way that is representative of conversational speech, and in a way which facilitates the carrying out of articulation tests. The articulations of the individual sounds may be converted into an index which indicates the speech capabilities of a system. Other types of lists which yield either the recognizability of speech sounds, or the intelligibility of discrete English words and sentences containing thought, have been described and experimentally correlated with the syllable technique.

It should be emphasized that there may not be a one to one corre-

spondence between all of these measured quantities for all types of speech distortion. The data entering into the sound, vowel, consonant, and syllable articulation curves were very extensive with respect to types of distortion and testing personnel. The theoretical equations relating them seem to rest upon assumptions with few uncertainties. For these reasons, it is felt that these relations can be used with considerable confidence, especially for values of syllable articulation greater than 30 per cent. The curves dealing with English words and sentences are based upon less diversified data and should be regarded as indicative only of the correlation and the type of relation.

During the past few years articulation testing methods have been used more and more, both in this country and abroad. It is felt that in order to compare the results obtained by various crews in various tongues, *it is desirable to use techniques that operate on the same basic principles and to calibrate various crews on similar reference circuits.*

INTELLIGIBILITY LIST

List 1

1. Name a prominent millionaire of the country.
2. How large is the sun compared with the earth?
3. Why are flagpoles surmounted by lightning rods?
4. Give the abbreviation for January and February.
5. Name the tree on which bananas grow.
6. How often does the century plant bloom?
7. What description can you give of the bottom of the ocean?
8. Explain the difference between a hill and a mountain.
9. What is the chief purpose of industrial strikes?
10. Describe the shoes of the native Hollander.
11. Name some uses to which electricity is put.
12. What would cause the air to escape from a bicycle tire?
13. Where is more grain raised, in the East or the West?
14. Tell what is meant by an Indian Reservation.
15. For what invention is Thomas Edison noted?
16. Name a state which has no seacoast.
17. Write the Roman numeral ten.
18. Explain the difference between export and import.
19. Explain why a corked bottle floats.
20. What substance is a good conductor of electricity?
21. Explain why Indians were afraid of firearms.
22. Explain the purpose of fire drills.
23. At what time do ocean waves become dangerous?
24. What medicine would you take to remedy indigestion?
25. What knowledge is covered by the study of astronomy?
26. Name a good restaurant in this vicinity.
27. What is the importance of large windows in stores?
28. Explain why a giraffe eats the foliage of trees.
29. How are the pages of a magazine held together?
30. Explain why the name string-bean is appropriate.
31. Name a nearby city in which there is a shipyard.

55

32. Name a fruit which grows in bunches.
33. Which of our Presidents went to South Africa?
34. Why are wire springs used in beds?
35. Why are books bound in stiff covers?
36. Why did the home people conserve food during the war?
37. Name an insect that has a hard shell.
38. What symbol on the United States money stands for liberty?
39. What weapons did the Indians use in warfare?
40. In what kind of weather does milk sour?
41. What streets in this city have Dutch names?
42. How does turning a ship's wheel steer the ship?
43. What nation aided us in the Revolutionary War?
44. What are some personal characteristics of the people of Japan?
45. What candy is black and good for colds?
46. Name a famous Indian Tribe.
47. Why is this building lighted by reflected light?
48. Why are most lighthouses situated on rocks?
49. Give some ingredients used in soap.
50. Why is a house built of stone superior to others?

Abstracts of Technical Articles From Bell System Sources

Reciprocal Theorems in Radio Communication.[1] JOHN R. CARSON. Two reciprocal theorems, the generalized Rayleigh theorem and the Sommerfeld-Pfrang theorem, are of great theoretical importance in radio communication. A careful analysis of these theorems and their mathematical derivations shows that they are quite distinct and their practical fields of application different. In particular it shows that the Sommerfeld-Pfrang theorem labors under restrictions, implicit in its mathematical derivation, which seriously limit its field of practical applicability.

Telephone Circuits for Program Transmission.[2] F. A. COWAN. Systems of telephone circuits which are extensively used in the transmission of programs to broadcasting stations are described in this paper. Certain stages in the development of these networks are considered and the general requirements for satisfactory transmission at the present time are enumerated. The arrangements of the networks as well as the procedures used in setting up and maintaining them are discussed.

Correlation of Directional Observations of Atmospherics with Weather Phenomena.[3] S. W. DEAN. This paper analyzes some data on the direction of arrival of static at Houlton, Maine, obtained by means of a recorder and a cathode ray radio direction finder; and points out that in certain cases there is a relation between the direction of static and the location of storm centers. Two cases are discussed in which day by day bearings showed static sources in the direction of moving storm centers.

A Direct-Current Amplifier for Measuring Small Currents.[4] J. M. EGLIN. A direct-current amplifier consisting essentially of a Wheatstone bridge, having the amplifying tube in one arm and a balancing tube in another, has been described by P. I. Wold and by C. E. Wynn-Williams. This circuit has now been developed to give a constant amplification for currents in either direction up to 10,000 times the lowest measurable value. The amplification and the lowest

[1] *Proceedings of the Institute of Radio Engineers*, Vol. 10, June, 1929, pp. 952–956.
[2] *A. I. E. E. Journal*, July, 1929, pp. 538–542 (abridgment).
[3] *Proceedings of the Institute of Radio Engineers*, Vol. 17, July, 1929, pp. 1185–1191.
[4] *Journal of the Optical Society of America and Review of Scientific Instruments*, Vol. 18, May, 1929, pp. 393–402.

measurable current are alterable together by changing the resistance introduced between the grid and filament of the amplifying tube. With tubes of high insulation, the amplification can be made as large as 10^6; and the measurable current as low as 10^{-14} ampere. Some improvements of the circuit are: (1) the insertion of a resistance in series with the tube in one arm of the bridge to "compensate" for variations in plate and grid battery voltages; (2) the suspension of the tubes to protect them from mechanical vibrations; (3) the use of tubes with pure tungsten filaments to avoid changes in contact potentials, and with plates enclosing the filaments completely to lower the effects of wall charges. In a "null" method of using the circuit the values of the grid resistance and an auxiliary potential introduced in the grid-filament circuit are sufficient to determine the measured current.

Meeting Long Distance Telephone Problems.[5] H. R. FRITZ and H. P. LAWTHER, JR. There have been written many papers describing various technical and apparatus developments of value in providing long distance telephone service. Several papers have also appeared covering specific transmission or operating problems, or dealing with the advance planning of the telephone plant. Feeling that it might be of interest, particularly to the young engineering graduates, the writers have prepared this over-all sketch of the general problem of actually providing, year by year, the extensions and additions to a comprehensive network of communication channels necessary to keep pace with a growing public demand for long distance service. Since the writers are most familiar with the area served by the Southwestern Bell Telephone Company, the discussion is restricted to that territory.

Some Measurements on the Directional Distribution of Static.[6] A. E. HARPER. The utility of directional data on static is shown, and two types of apparatus devised for such a directional investigation are compared. It is shown that a method which gives the direction of individual crashes is superior to integrating methods. The distribution of thunderstorms over the world is discussed, and comparisons are drawn between this distribution and the observed directional distribution of static. Probable geographical locations are assigned to the sources, based upon thunderstorm data and directional observations.

[5] *A. I. E. E. Journal*, July, 1929, pp. 547–550 (abridgment).
[6] *Proceedings of the Institute of Radio Engineers*, Vol. 17, July, 1929, pp. 1214–1224.

Maximum Excursion of the Photoelectric Long Wave Limit of the Alkali Metals.[7] HERBERT E. IVES and A. R. OLPIN. Earlier experiments have shown that the long wave limit of photoelectric action in the case of thin films of the alkali metals varies with the thickness of the film. A maximum value is attained greater than that for the metal in bulk, which for the majority of the alkali metals lies in the infra-red. The wave-length of the maximum excursion of the long wave limit was first studied for Na, K, Rb and Cs. In each case it was found to coincide with the first line of the principal series, i.e., the resonance potential. If this relation holds for lithium, its maximum long wave limit should be greater than that of sodium. This was tested and confirmed by experiments in which red-sensitive lithium films were prepared, sensitive to 0.6708 μ. It is suggested that photoelectric emission is caused when sufficient energy is given to the atom, to produce its first stage of excitation. The identity of photoelectric and thermionic work functions suggests that atomic excitation is the initial process in thermionic emission as well.

Magnetic Testing Furnace for Toroidal Cores.[8] G. A. KELSALL. When making magnetic tests at high temperatures trouble is often experienced in maintaining the insulation between turns of the magnetizing and exploring windings and between the windings and the test sample.

This paper describes a magnetic testing furnace for toroidal specimens which eliminates these difficulties. By means of this furnace the test sample may be passed through a definite temperature cycle and the variation in magnetization for a constant magnetizing force determined or the temperature may be held constant while measurements are made for the B-H curve or for a hysteresis loop.

Electrical Wave Analyzers for Power and Telephone Systems.[9] R. G. MCCURDY and P. W. BLYE. This paper describes two types of electrical analyzers which have been developed for the direct measurement of harmonic components of voltage and current on power and telephone systems. These devices are assembled mechanically in a form suitable for use either in the laboratory or in the field. Both instruments, which differ chiefly with respect to sensitivity and input circuit arrangement, employ multistage vacuum tube amplifiers and two duplicate interstage selective circuits.

[7] *Physical Review*, Vol. 34, July 1, 1929, pp. 117–128.
[8] *Journal of the Optical Society of America and Review of Scientific Instruments*, Vol. 19, July, 1929, pp. 47–49.
[9] *A. I. E. E. Journal*, Vol. 48, June, 1929, pp. 461–464 (abridgment).

The power circuit analyzer is designed to measure harmonic voltages in the frequency range from 75 to 3000 cycles and over a voltage range from 0.5 millivolt to 50 volts. The telephone circuit analyzer operates over the same frequency range and measures harmonic currents as low as 0.05 microampere and voltages as small as 0.005 millivolt. Both analyzers are adapted to measure small harmonic voltages and currents in the presence of the fundamental component and other harmonics relatively large in magnitude.

A number of devices are described which have been adopted for eliminating various sources of error. The paper presents in detail the characteristics of both instruments with respect to selectivity, sensitivity, linearity, balance of input with respect to ground, generation of harmonics, and susceptiveness to stray fields.

Solution to a Problem in Diffusion in Employing a Non-Orthogonal Sine Series.[10] R. L. PEEK, JR. In this paper there are developed the equations applicable to diffusion through a membrane between a chamber in which a constant pressure of the diffusing material is maintained and a second closed chamber, initially evacuated, into which the material diffuses. Assuming Fick's law to apply, a solution is obtained in the form of an infinite series of a type similar to those applying to other problems in diffusion. The sine series to which the solution reduces at zero time is non-orthogonal, but it is shown that by a modification of Fourier's method the coefficients of the terms may be directly determined. There is included a proof of the convergence of the series considered.

Telephone Transmission Networks.[11] *Types and Problems of Design.* T. E. SHEA and C. E. LANE. In this paper is given a brief résumé of the nature of telephonic signals showing how the qualities of wave composition which distinguish signals from other electrical waves set the requirements on networks and provide a basis for their design. The principal functions of wave filters, equalizers, telephone transformers, line balancing networks, and artificial lines are outlined. In order that these networks may be used in conjunction with other apparatus in the telephone system they must provide efficient transmission, low distortion, good impedance balance, stoppage of longitudinal currents, stable characteristics with current variations, low external coupling, and low reflection coefficient. In addition to these

[10] *Annals of Math.*, 2d Series, Vol. 30, April, 1929, pp. 265–269.

[11] Presented at the Regional Meeting of the South West District No. 7, of the A. I. E. E., Dallas, Texas, May 7–9, 1929. Abridgment in *A. I. E. E. Journal* of August, 1929, pp. 624–628.

requirements the network must not cross-talk into associated circuits and must have desirable impedance characteristics in the attenuation range of frequencies as well as throughout the transmission range. An illustration of the use of transmission networks in a typical three-channel carrier telephone system is given describing the functions of the line filter sets, the directional filter sets, band filters, and equalizers. Some of the engineering limitations on the design and construction of networks are discussed.

Recent Developments in Telephone Construction Practices.[12] B. S. WAGNER and A. C. BURROWAY. In this paper are described some recent developments in telephone cable installation and maintenance practices. The paper is divided into three sections: (1) Gas pressure testing for detecting and locating sheath defects before they result in failure. (2) Methods for reducing bowing and other movements of cable which in time cause fracture of the sheath. (3) Catenary construction for long spans, such as at river crossings.

[12] *A. I. E. E. Journal*, Vol. 48, May, 1929, pp. 366–369 (abridgment).

Contributors to this Issue

KARL K. DARROW, B.S., University of Chicago, 1911, University of Paris, 1911–12, University of Berlin, 1912; Ph.D. in Physics and Mathematics, University of Chicago, 1917; Engineering Department, Western Electric Company, 1917–25; Bell Telephone Laboratories, 1925–. Dr. Darrow has been engaged largely in writing studies and analyses of various fields of physics and the allied sciences. Some of his earlier articles on Contemporary Physics form the nucleus of a recently published book entitled "Introduction to Contemporary Physics" (D. Van Nostrand Company).

H. F. DODGE, S.B., Massachusetts Institute of Technology, 1916; Instructor, Electrical Engineering, 1916–17; A.M., Columbia University, 1922; Engineering Department, Western Electric Company, 1917–25; Bell Telephone Laboratories, 1925–. Mr. Dodge was earlier associated with the development of telephone instruments and allied devices, and is now engaged in development work relating to the application of statistical methods to inspection engineering.

HARVEY FLETCHER, B.Sc., Brigham Young, 1907; Ph.D., Chicago, 1911; Instructor of Physics, Brigham Young, 1907–08 and Chicago, 1909–10; Professor, Brigham Young, 1911–16; Engineering Department, Western Electric Company, 1916–25; Bell Telephone Laboratories, 1925–. During recent years, Dr. Fletcher has conducted extensive investigations in the fields of speech and audition.

EDWARD C. MOLINA, Engineering Department of the American Telephone and Telegraph Company, 1901–19; Department of Development and Research, 1919–. Mr. Molina has been closely associated with the application of the mathematical theory of probabilities to trunking problems and has taken out several important patents relating to machine switching.

H. G. ROMIG, A.B., Pacific University, 1921; University of Washington, 1922; A.M., University of California, 1923; Teaching Fellow in Physics, University of California, 1922–24; Instructor in Mathematics and Physics, San Jose State Teachers College, 1924–26; Bell Telephone Laboratories, 1926–. Mr. Romig is engaged in development work relating to the application of statistical methods to inspection engineering.

L. J. SIVIAN, A.B., Cornell University, 1916; Engineering Department, Western Electric Company, 1917–1919 and 1920–1925; Bell Telephone Laboratories, 1925–. Mr. Sivian's work is in acoustics, chiefly in connection with methods of electroacoustic measurements.

J. C. STEINBERG, B.Sc., M.Sc., Coe College, 1916, 1917; U. S. Air Service, 1917–19; Ph.D., Iowa, 1922; Engineering Dept., Western Electric Company, 1922–25; Bell Telephone Laboratories, 1925–. Dr. Steinberg's work since coming with the Bell System has related largely to speech and hearing.

R. I. WILKINSON, B.Sc., Iowa State College, 1924; Western Electric Company, 1920–21; American Telephone and Telegraph Company, Department of Development and Research, 1924–. Mr. Wilkinson has studied principally the application to telephone problems of the mathematical theory of probability, including sampling and statistical analysis.

W. HOWARD WISE, B.S., Montana State College, 1921; M.A., University of Oregon, 1923; Ph.D., California Institute of Technology, 1926; American Telephone and Telegraph Company, Department of Development and Research, 1926–. Since coming to the American Telephone and Telegraph Company, Dr. Wise has been engaged in various theoretical investigations.