The Bell System Technical Journal

Vol. XXV *July, 1946* *No. 3*

Some Recent Contributions to Synthetic Rubber Research*

By C. S. FULLER

INTRODUCTION

WHEN the war put an end to shipments of natural rubber from the Far East, it became evident that synthetic chemistry would be called upon to fill the gap in our supply of this strategic material. We know now how effectively the emergency was met. In less than three years the production of Buna S type synthetic rubber alone had risen to exceed our total prewar comsumption of natural rubber. Few, however, realize the magnitude of the effort and the extent of the cooperation between groups of experts that was essential for the achievement of this success.

Rubber companies in this country had been experimenting with synthetic substitutes for natural rubber for some time before the present war began. None of these products, however, was sufficiently advanced either from the stand-point of raw materials or in regard to the knowledge of its properties, to warrant production on a large scale as a substitute for natural rubber during the emergency. In 1942, following the advice of the Baruch Committee, we decided to place chief reliance on Buna S, the butadiene-styrene synthetic rubber which the Germans developed about 1934. In addition, considerable support was given to the domestic synthetics, Neoprene, Thiokol and Butyl. The latter rubbers, however, were not considered as useful for tires as Buna S.

Making Buna S in this country and fabricating it were not simple, however. The Germans had kept the details of the process secret and restricted shipments of the product. Besides, as we have since found out, the German chemists did not have any too complete control of the process themselves and the type of rubber made by them, as shown by samples obtained indirectly, was not satisfactory for use on American processing machinery. Our engineers and research men were therefore faced with the problem of setting up a process on an enormous scale to turn out a product which could be used in our tire plants and which would give satisfactory service on the road. Fortunately for us, a few companies had acquired enough knowledge

* The investigations described in this article were carried out under the sponsorship of the Reconstruction Finance Corporation, Office of Rubber Reserve, in connection with the Government synthetic rubber program.

both from German sources and from their own researches to warrant taking the gamble.

As a part of the large program laid out under the auspices of the Government in 1942, provision was made for a cooperative research and development effort to parallel and to contribute to the constructional program designed to provide the much needed rubber. A large number of company laboratories as well as universities contributed to this research. The Bell Telephone Laboratories because of its past contributions in this field of synthetic polymers was asked to participate in this program. The present discussion is intended to describe part of the Bell Laboratories investigations directed toward the improvement of Buna S type rubber, particularly work relating to the characterization and control of the final copolymer.

In order to present the material in a logical and understandable form to readers unfamiliar with the subject-matter, brief mention will be made of the history of the synthetic rubber problem and of progress in the knowledge of polymeric substances during recent years.

The Problem of Synthetic Rubber

The problem of synthesizing natural rubber is almost as old as man's curiosity about the nature of rubber itself which began when Faraday in 1826 first showed it to be a hydrocarbon having the formula $C_{10}H_{16}$. Experiments done by Williams in 1860, in which he obtained isoprene from natural rubber and by Bouchardat in 1879, who showed that isoprene could be polymerized to a rubber-like material, represent about as close as we have come to synthesizing natural rubber in spite of many subsequent efforts. In 1910 particularly, when the price of natural rubber reached $3 per pound, considerable pressure was exerted to bring about this synthesis. Although the chemist failed in this quest his very failure, analyzed in the light of more recent studies on other polymers as well as rubber, has had its virtues. It has emphasized the importance of chemical structure, that is the precise organization of the atoms composing the rubber molecules (in addition to simply the nature of these atoms) in determining the ultimate properties of a polymer.

Although natural rubber eluded synthesis, the early organic chemical work nevertheless laid the basis for our present synthetic rubber. Curiously, much of this pioneering research on synthetic rubber was done in England with the support of strong proponents of natural rubber. However, Germany and Russia were also active contributors. The United States later achieved fame by bringing forth two of the most promising rubbers yet produced, Neoprene and Butyl. The early foreign synthetics were based on the polymerization of hydrocarbons such as 1-methyl butadiene and 2,3 dimethyl butadiene and butadiene itself. They were undoubtedly "rubbers" of a sort but there could be no question about their inferiority to the natural product. Even today the Russians persist in making their synthetic rubber from butadiene and, although there have been improvements, the polymer is still subject inherently to the same fundamental difficulties of structure that existed when it was first synthesized by Lebedev in 1911.

The deficiencies in the early synthetic rubbers and the difficulty of synthesizing natural rubber were appreciated in Germany where in the period 1935–39 several plants were constructed to manufacture synthetic rubber, including Buna S, on a large scale. By polymerizing together butadiene and styrene instead of butadiene alone they achieved several advantages over previous synthetic rubbers. The fact that the best opinion in this country decided in favor of imitating German Buna S, shows that progress in Germany was indeed substantial. As we have already indicated, however, improvements were necessary in both the German product and process if it was to be satisfactory for our use. The product developed in this country and now being currently produced at the rate of nearly 700,000 tons per year, although prepared from the same starting materials as German Buna S, therefore differs from the latter in many important respects. The name Government Rubber-Styrene, abbreviated GR-S, has been given to this product.

History of the Development of Ideas of Composition and Structure of Polimers

All rubbers, both natural and synthetic, as well as all organic plastics and fibers belong to a class of substances called polymers. We now know that they are constructed of large molecules, in turn built up of simple atomic patterns (repeating units) joined end to end. Surprisingly, it was not until about fifteen years ago that this idea gained general acceptance among chemists. Since that time truly remarkable research progress on polymers has been made. It is not our object to present a full account of this work here. Most of it was carried on independently of its application to the synthetic rubber problem but nevertheless has had a profound effect upon it. A brief review of the growth of the present concepts of natural and synthetic polymers will, however, help to emphasize the significance of the more recent researches on synthetic rubber.

For a long time chemists believed that naturally occurring polymers like natural rubber, cellulose and silk were indefinite chemical compounds in which the arrangement of the atoms was so complex as to defy analysis. As has been mentioned, Faraday had shown in the case of natural rubber that carbon and hydrogen atoms were present in the ratio of 16 hydrogens for every 10 carbons. It was not until much later that it was postulated that rubber, inasmuch as it had the same hydrogen to carbon ratio as isoprene obtainable from it, was a compound in which many isoprene groups were in some manner combined together. Thus, Harries about 1904 was inclined to regard rubber as a sort of association complex representing a combination of relatively small ring molecules held together by van der Waals' attractions[1]. This same view of polymers as associations of small molecules was also applied to cellulose by well-known carbohydrate chemists both in England and in Germany.

The influence of the contemporary colloid chemists helped to promote this idea. Even the term "micelle", applied by them to soap and other aggregates, which are in fact van der Waal's or ionic associations, was unfortunately adopted to describe the structure of many of the organic polymers. In addition, early x-ray studies on natural polymers, because of a misinterpretation of the diffraction patterns, lent further support to these views. For some reason or other it was not appreciated by workers in the field that the x-ray unit cell did not necessarily mark the boundaries of the organic molecule. Hence, since the unit cells appeared to be small, many erroneously concluded that the molecules were small also. It is to Sponsler and Dore[2], working in this country in 1926 on the x-ray structure of cellulose fibers, that we must give thanks for being the first to realize the incorrectness of the older x-ray deductions and to postulate a long primary valence chain structure for cellulose.

The realization that natural organic polymers really consisted of very long chains of primary valence bound atoms, in the strictly organic chemical sense, came surprisingly slowly. Staudinger in Germany beginning about 1926 was most insistent on this view[3], although others including Meyer and Mark were developing the same conception. As early as 1910 Pickle in England had conceived of such a chain type of molecule for natural rubber but unfortunately did not follow it up. As the idea of molecules of large size grew, it became more and more popular to try to measure them. Also there was much effort given to working out the details of the "crystal structure" of the natural products insofar as they could be regarded as crystalline. Here again was an opportunity for argument which is still going on today: just what do we mean by the term "crystalline" when applied to these substances? The answer seems to be that we have all degrees of organization of the molecules, or more correctly parts of molecules, in polymers from the completely chaotic or amorphous in some to highly ordered or what may be called crystalline arrangement in others. We shall have occasion to come back to this subject in our later discussion.

It was logical that the interest of scientists in the constitution and structure of polymers should be lavished on naturally occurring high polymers rather than on the synthetic ones. But strangely enough it has been the synthetic polymers which have really led us to a more complete understanding of the natural substances and particularly to the explanation of why polymers have the properties they do.

The early work on synthetic polymers, as we have seen, centered around the constitution of natural rubber and efforts to duplicate it. Soon, however, organic chemists found they could make better products from other dienes than they could from isoprene which seemed to be the progenitor of natural rubber. The approach was necessarily empirical—one of trying out a variety of reaction conditions on the chemical compound to be polymerized and studying the properties of the final product as compared to natural rubber. Nearly always the comparison was disappointing. Following this procedure the Germans and the Russians developed their respective competitors for natural rubber from 1910 to the present time. The organic chemistry of polymerization, the reactions whereby the simple unsaturated compounds join up into longer molecules, was, however, very imperfectly understood in 1910 and still is not clear today.

Perhaps it was for this reason that some organic chemists decided to build large molecules by methods in which they had acquired great confidence in regard to how the atoms come together. Emil Fischer, the first of this group, succeeded in synthesizing a polypeptide molecule of known composition and known organic structure which, although smaller in size than the natural proteins, nevertheless was very large compared to the usual organic molecules. This was in 1906. About 20 years later the matter was again opened up in a more general way by Staudinger and his collaborators who synthesized chains built up of alternate carbon and oxygen atoms, the polyoxymethylenes, and showed how such large molecules could give rise to a pseudo-crystalline type of crystal lattice. Then came the simple and beautiful work of W. H. Carothers and his collaborators beginning in 1928, which led to the development of nylon. These compounds and the linear polyesters, which Carothers had (by improvement of the methods of Vorländer[4] and others) prepared, because they were known to contain long chain molecules of definite structure and composition, were ideal compounds to examine in order to determine what factors were truly responsible for observed polymer behaviors. In this way it was hoped to explain the outstanding toughness, high tensile strength, rubberiness, peculiar softening and flow properties and a host of other characteristics of polymers which make these materials so important in life processes and technology. Researches along these lines have indeed shown that the way the various units are combined and the regularity of the atomic arrangements in the units themselves have a profound effect on properties.

This work has also emphasized the importance of size and linearity of the chain molecules on polymer behavior. For example, the length of the molecules which are present in a polymer is of critical importance to certain properties such as mechanical strength. These facts, as well as the necessity for order in the arrangement of the molecular units along the chains were not appreciated by the early organic workers. That Carothers realized what many of the older organic chemists did not realize is indicated by his statement made in 1934 that the problem of physically characterizing polymers in significant numerical units is of the utmost importance and that it should receive more attention jointly from physicists and chemists.

Some Physico-Chemical Features of Polymers

We have seen very briefly how the quest for the origin of properties of rubbers and polymeric substances in general led of necessity to a study of the intimate details of chain molecule structure on the one hand and a study of the general characteristics of large molecules on the other. Before taking up the specific researches on GR-S synthetic rubber, however, it will be helpful to pursue somewhat further the ideas on the formation and constitution of polymers.

There are two general chemical processes by which polymer molecules are formed, namely polymerization and polycondensation. Chemists, at times, use the first term to represent all processes leading to the formation of large molecules but it is more convenient to distinguish two processes even though the difference between them is academic in some cases. In polymerization, chemical molecules called the monomers, become "activated" either by heat energy or by means of special chemical compounds. In this state they spontaneously grow at the expense of their unactivated neighbors until the growth of the chains is abruptly terminated, either by active chains coming together or by a transfer of energy to other, often foreign, molecules. The entire growth reaction for any given chain usually takes but a fraction of a second for completion. When two or more different monomers capable of polymerization enter together into the same chain molecule formation the process is referred to as "copolymerization".

In polycondensation, identical or non-identical molecules react to give large molecules just as in polymerization. The difference is that in the former reaction a molecule of water (or other substance) is evolved each time a new molecule is added to the growing chain system. Also the reaction resulting in chain growth is step-wise in the sense that each added molecule follows the same steps in reacting that are followed by any other. No special type of activation on the end of the growing chain is necessary. Finally, since there is no activated growth, the phenomena of termination in the sense used above in connection with polymerization do not exist.

Both kinds of polymers are important technically. Thus polystyrene is a polymerization type of polymer. Nylon on the other hand is a polycondensation polymer. Buna S type synthetic rubber is a polymerization copolymer because it is formed by polymerizing together styrene and butadiene monomers.

One of the important characteristics about reactions leading to the formation of polymers is that they result not in molecules of the same size but in a statistical distribution or mixture of molecules of various sizes. These molecular weight distributions, as they are called, in special cases can be

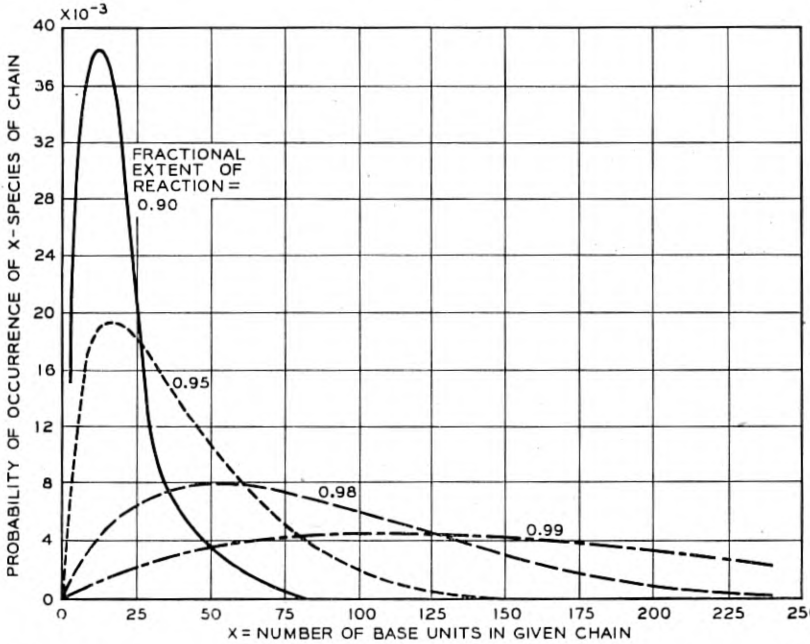


Fig. 1.—Curves showing frequency distribution of molecule species of different chain lengths for linear polyesters (Flory—reference 5).

calculated from the nature of the reaction. In other instances this is not possible, although experimentally it is often possible to arrive at an approximate curve representing a given polymer distribution. Figure 1 shows a series of curves for a linear polyester in which the reaction conditions are such that the calculated curves[5] represent very closely the actual distribution of molecules present. The curves represent the weight fraction of each molecular species present in the mixture at the extent of reaction shown on each curve.

It is customary to speak of an "average molecular weight" therefore in characterizing these polymer mixtures. Several different types of averages are used for convenience. The two most frequently employed are what are termed the "number average" and the "weight average". If osmotic pressure measurements are made on solutions of the polymer and extrapolated to zero concentration these will lead to a number average molecular weight figure. This average represents what we would obtain if we sorted out the molecules according to molecular weight and counted them. Multiplying each molecular weight (m_1) by the number present (n_1) and dividing by the total number of molecules we obtain the number average molecular weight ($\overline{M}_n$) or stated mathematically

$$\overline{M}_n = \frac{\Sigma m_i n_i}{\Sigma n_i} = \frac{1}{\Sigma \frac{f_i}{m_i}} \tag{1}$$

where f_i is the weight fraction of species of molecular weight, m_i.

Usually the osmotic pressure measurements are difficult to carry out and a simpler measurement, that of dilute solution viscosity (DSV) is performed. This determination consists in measuring the relative viscosity of a solution of the polymer at one given low concentration and calculating

$$(DSV) = \frac{\ln \eta_r}{c} \tag{2}$$

where η_r is relative viscosity and c is the concentration in grams per 100 ml. of solution. A more fundamental quantity usually differing little from the DSV value is the so-called intrinsic viscosity. This is defined as $[\eta] = \frac{\ln \eta_r}{c_{c \to o}}$. Measurements are made at several concentrations and extrapolated to zero concentration just as for osmotic pressure. From this value a molecular weight, which may be referred to as a viscosity average molecular weight, can be calculated from the empirical expression $[\eta] = K(\overline{M}_v)^a$ where both K and a are constants over a fairly wide range, and which must be independently determined. In some polymer distributions this viscosity average is very close to the weight average defined by

$$\overline{M}_w = \frac{\Sigma w_i m_i}{\Sigma n_i m_i} = \Sigma f_i m_i \tag{3}$$

where m_i is again the molecular weight of each species and f_i is the weight fraction in which it is present in the mixture. In the example of Fig. 1 the number averages are indicated by the maxima of the various curves. Here the viscosity and weight averages are identical.

In polymers an equally important consideration with molecular size distribution is chain molecule structure. It is convenient to distinguish between micro-chain structure and macro-chain structure. By micro-chain structure we mean the detailed architecture of the chain molecule over distances of the order of length of the repeating unit. The kind of atoms involved in the unit and their spatial arrangement in regard to atoms in the same chain as well as in the neighboring chains are included in this definition. It is the micro-chain structure which determines entirely the chemical properties of the polymer and to a large extent the physical properties as well. Thus, the influence of solvents, oxidizability, hardness at a given temperature, softening point, ability to crystallize are determined largely by the micro-structure of the polymer.

Macro-chain structure on the other hand refers to the long range form of the chain molecule. It ignores composition and concerns itself with the nature of the molecule as a whole and with its interconnections to other molecules.

Certain terminology has grown up in this connection which can be conveniently defined at this time. We speak of "linear" polymers when primary valence bonds can be traced through the molecules from one end to the other without passing over the same atoms twice. We say "branched" molecules are present when the process of tracing leads us into one or more offshoots from the main chain. When the degree of branching becomes excessive the molecules may become insoluble in good solvents for the linear or slightly branched molecules. When nteworks of molecules are present we say the polymer is "netted" or "cross-linked". In this instance closed paths may be traced and the smaller the paths, the "tighter" or more "intense" is the netting. Netted chain molecule systems are invariably insoluble. Insoluble polymers whether because of intense branching or netting are called "gel". We speak of micro-gel when the gel particles (molecules) are microscopic or smaller in size (say less than 1μ) and of macro-gel when the particles are large.[6] Usually macro-gel as well as the micro-gel is associated with soluble molecule species. These latter are referred to as "sol" and represent the linear or the less branched molecular components of the mixture. The complete description of every molecule present in a polymer mixture is thus a very difficult if not impossible task. We are thus forced to employ a statistical treatment.

In the case of copolymers, still other considerations arise. There is the probability that the reacting components will not react with one another at the same rates they do with themselves. When this occurs the composition of the molecules in the mixture varies, some containing more of one component than others do. Also the order in which the components are arranged along the chains may vary molecule to molecule. Such circumstances of course give rise to varying properties in the copolymer mixture. We shall have occasion to consider these questions below in connection with the development of GR-S.

Early Status of GR-S Synthetic Rubber

The process by which GR-S type synthetic rubber is made is known as the emulsion polymerization process. In it, butadiene and styrene in the proper proportions are emulsified in water with small amounts of catalysts and substances called modifiers which serve to control the plasticity of the polymer. During the reaction period of from ten to twenty hours about three-fourths of the butadiene and styrene are converted into the synthetic rubber. The reaction occurs in such a way that very minute particles are formed and the resulting synthetic latex is suggestive of natural latex. To obtain the rubber itself the latex is coagulated with acid and sodium chloride or with aluminum sulfate and the coagulum washed. After drying the rubber crumbs are baled and shipped to the fabricating factories. The above brief sketch of course does not provide an idea of the many complexities which arise in practice nor of the many process variations which can be used to control the final properties of the rubber. A complete treatment of this subject falls outside the scope of this paper.

When the Baruch Committee advised "bulling through" the synthetic program on the basis of Buna S type rubber, it fixed the chemical composition of the product to a very great extent. We knew then, or shortly afterward, that we would be required to use approximately 690,000 tons of butadiene and 197,500 tons of styrene per year to produce the coplymer rubber. Whatever other components might be employed would be available in only insignificant quantities by comparison. One element of choice remained as far as chemical composition was concerned, namely the proportions in which the two components might be used. German Buna S is supposed to consist of 75 parts by weight of butadiene to 25 parts of styrene but, as we shall see later, this ratio does not determine the ratio actually present in the final copolymer which is a function of reaction variables as well as the initial ratio of the ingredients. Consequently it was necessary to examine the composition of the final copolymer and to control it at the proper ratio of butadiene to styrene. The chemical composition was not the only factor to be controlled, however, since as we have seen, the properties of polymers unlike ordinary chemical compounds depend as much if not more on the chain structure. This is of course not only dependent on the nature of the starting ingredients but also on the manner in which they are combined into the chain.

At the time intensive work was undertaken in this country on Buna S type synthetic rubber little attention had been given to its characterization by physico-chemical means. The usual physical testing procedures involving the preparation of compounds by mixing in pigments and vulcanizing were of course being employed to supply useful information about the copolymer produced and the vulcanization properties possessed by it. What was needed, however, were more precise and revealing tests, and tests which could be carried out directly on the copolymer itself. No ordinary chemical methods such as are applicable to the usual type of synthetic chemicals apply, for reasons which should be evident from our previous discussion. New methods of characterization designed to insure uniformity and satisfactory quality in the GR-S copolymer were required.

The precise and early control of the copolymer was of utmost importance. Non-uniformity in the product may cause serious troubles in fabricating operations such as are employed in tire plants, wire coating factories, adhesives manufacture, etc. Furthermore, with a varying product it often cannot be determined whether the trouble, when it occurs, is in the copolymer or in the method of fabrication being used.

What are the characteristics which must be controlled to insure a satisfactory product? To answer this question it was necessary to investigate a variety of GR-S copolymers and to conduct service tests on them in order to determine their practical performance. Some of these tests, particularly those on tires, have been very extensive. Some of the characteristics of the copolymer which experience has taught should be measured and controlled are:

1. The over-all or average styrene content in the butadiene-styrene copolymer.

   This necessitates (1) a method of separating the pure copolymer (which is the rubber-like component) from non-rubber components such as soap, salts, insoluble matter etc., and (2) a suitable method for determining the styrene content of the purified copolymer.
2. The percentage soap, fatty acids and low molecular butadiene-styrene compounds in the rubber.
3. The amount of "gel" fraction, if present, and the swelling volume of the gel.[7]
4. The average molecular size of the "sol" or soluble fraction of the copolymer.
5. The degree of branching of the sol molecules.
6. The molecular weight distribution of the sol.

In addition to the above tests on the final copolymer, control tests which can be used during the polymerization to tell when the reaction has progressed to the proper point were needed. In the following paragraphs we will take up in some detail the problem of characterization and attempt to show the basis on which methods have been evolved to control some of the quantities listed above.

Composition of GR-S and Its Determination

Given a piece of GR-S synthetic rubber, our first task from the standpoint of determining its chemical composition is to separate the pure copolymer which is responsible for the rubber-like properties from the non-rubber constituents. The latter comprise soaps or other emulsifying agents, fatty acids, salts, antioxidant and low molecular weight, non-rubbery butadiene-styrene products to the extent of several percent. Some of these minor ingredients, like the antioxidant, are essential whereas others play no important role subsequent to polymerization. All, however, must be separated from the copolymer before it can be properly evaluated. The analysis for the non-rubber components after separation is fairly straightforward and standard and will not be gone into here.

It has been found that the azeotrope of toluene and ethyl alcohol which consists of approximately 30 parts by volume of toluene to 70 parts by volume of alcohol is an excellent extractant for the non-rubber compounds and hence may be used to effect a separation[8, 9]. The procedure for isolating the copolymer is simply to place a quantity, say 10 grams, of the GR-S in an extraction thimble supported in an extraction flask as shown in Fig. 2. Another, more rapid, procedure is to reflux the azeotrope over the rubber for two hours, when extraction has been found to be essentially complete. This method is now used in the Standard Specification for all GR-S. The pure copolymer, left as residue, is the product to which we now turn our attention.

As has been mentioned, the ratio in which butadiene and styrene are employed in the starting mixture does not determine either the ratio in the whole copolymer at a given stage of reaction or the ratio present in any given chain molecule of the copolymer. Therefore the starting ratio cannot be relied upon to control the composition of the final copolymer. Experiments show that under certain process conditions large differences in composition between different fractions of the copolymer do occur. Even under the best conditions theoretical considerations predict that variations must occur between molecules since the ratio of the reactants is continuously changing during the reaction.

Let us examine the chemistry of the process for a moment to try better to understand why these variations are possible. When styrene (S) reacts with itself polystyrene (S_x) is formed. Analogously polybutadiene (B_y) is formed in the case of butadiene (B). In GR-S both styrene and butadiene react to give a copolymer.

When a quantity of styrene undergoes polymerization, a distribution consisting of various numbers of long chain molecules of various lengths is formed. Thus, if we start with N_1 molecules of styrene, S, the polymerization reaction results in the formation of molecules of polystyrene by the addition of S to S in chain fashion. The result may be expressed as follows:

$$N_1S \rightarrow a_1S_1 + a_2S_2 + a_3S_3 + \cdots + a_nS_n \qquad (4)$$

where each term represents a group of styrene molecules containing $1, 2 \cdots n$ styrene units, n assuming values up to several thousand depending on the

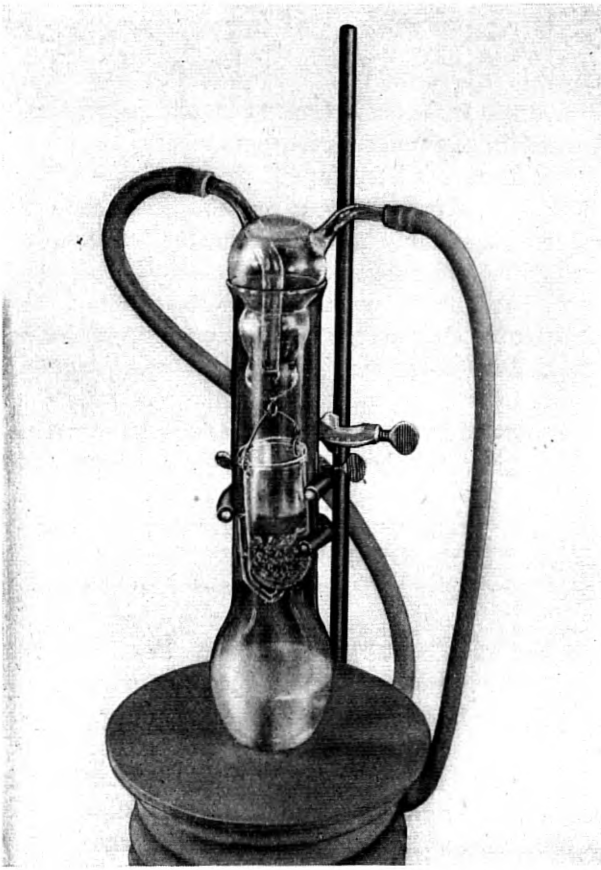


Fig. 2.—Apparatus for extracting non-rubber components from GR-S.

reaction conditions. If N_1 is very large there are of course many molecules, a_n, formed of the length corresponding to each value of n. In fact, $a_1 + 2\,a_2 + 3\,a_3 + \cdots + na_n = N_1$. The first term in (4) allows for the molecules which do not react, the second represents the dimers, the third the trimers, etc.

In an analogous way we may consider N_2 molecules of butadiene, B, to polymerize into chain molecules of various lengths:

$$N_2B \rightarrow b_1B_1 + b_2B_2 + b_3B_3 + \cdots + b_nB_m . \qquad (5)$$

Now if styrene and butadiene molecules react together, as in the production of GR-S, we can represent their copolymerization as the insertion of the styrene chains (or portions of them) of (4) at random points in the butadiene chains of (5) to form chains S_jB_k. That is

$$N_1S + N_2B \rightarrow \Sigma \;\; a_iS_jB_k \qquad (6)$$

where j and k take on a variety of integral values and in any particular chain the arrangement of S and B units is probably random.

In practice, in the reaction represented by (6), N_2/N_1 has the value of approximately 6 since 75 parts by weight of butadiene are employed to 25 parts of styrene. Each chain molecule therefore would be expected to contain about 6 butadiene residues to each one of styrene. It is actually found, however, as indicated above that the starting ratio is not adhered to throughout the reaction, the molecules formed early being richer in butadiene and those formed later being poorer in butadiene than the starting ratio of 6 to 1. But, not only is the ratio B_k/S_j a variable from molecule to molecule of the copolymer formed but also their sequence along the chain is variable. Thus, in equation (6), even when equal numbers of styrene and butadiene molecules are present, a strict alternation is apparently not maintained but "strings" of one pure component or the other, form.

In the GR-S reaction the *weight* ratio of butadiene to styrene in the first molecules formed may be as high as 4:1 or more from a starting charge of ratio 3:1. Thus, the average weight percentage of styrene in the GR-S copolymer first formed is about 8% below that in the original charge (25%) and increases with conversion so that at the point where the reaction is stopped the copolymer forming contains about 29% styrene. Analogously there is evidence to show that in GR-S no regular sequence of butadiene and styrene along the chain molecules exists but rather a more or less random entrance of the two residues into the molecules with a frequency approximating the 6 to 1 ratio, as the extent of combination (percentage conversion) of the two ingredients approaches completion where obviously the two must become equal. Figure 3 illustrates this behavior for a typical sample prepared in the laboratory. An integral curve showing the cumulative percentage styrene and a differential curve representing the percentage styrene in the increment of the copolymer are illustrated.

It must be left to future research to determine how important the molecule to molecule variations in styrene content are in terms of useful properties and to devise ways of eliminating them if necessary. For the present, we are perhaps justified in assuming that these variations can be neglected. The control considered here therefore relates to the over-all or average composition of the copolymer.

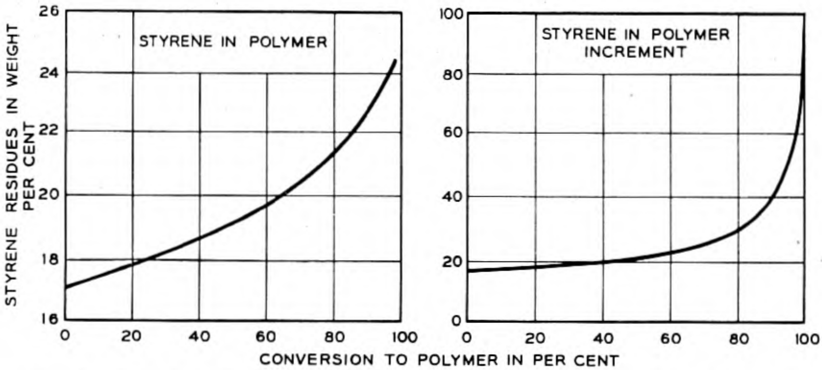


Fig. 3.—*Left:* Cumulative percentage by weight of styrene in the copolymer as a function of percentage conversion for an initial 25 percent styrene charge. *Right:* Percentage by weight of styrene in the polymer forming at any instant as a function of conversion for an initial 25 percent styrene charge.

Determination of Styrene Content

Many suggestions involving both chemical and physico-chemical methods for measuring the average styrene content of GR-S copolymers have been proposed. Physico-chemical methods when applicable have an advantage in speed and precision over straight chemical methods and therefore have been more carefully examined. Both ultra-violet absorption[10] and refraction[8] have been shown to be applicable but since the absorption method is much more sensitive to impurities, the refraction method has proven the most general. It has the advantage also that it can be employed with polymers containing considerable gel fraction.

The refraction method is based on the fact that the styrene residues in the copolymer provide a greater contribution to the refraction of light passing through the solid or a solution of the solid than do the butadiene residues. Early work at the Bell Laboratories showed that the determination of the refractive index of the solid *unpurified* copolymer led to errors. In addition, the determination of the refractive index even of purified polymers was not precise if much gel was present, as frequently was the case with the early synthetic product. As a consequence a method, based on the use of the interferometer, was developed[8, 10]. The procedure is to disperse 2.4 grams of the pure copolymer in benzene, transfer the contents to the interferometer cell and make a reading of the change in refraction compared to the pure benzene. This value, with the help of a curve relating styrene content to refraction, enables the true styrene content to be determined. The curve of refraction as a function of styrene content must be constructed beforehand and is shown in Fig. 4. This curve is obtained by measuring the refraction of pure polystyrene on the one hand and polybutadiene on the other. Checks also were made by independent methods of estimating composition in the range of the usual Buna S-type synthetic rubber.

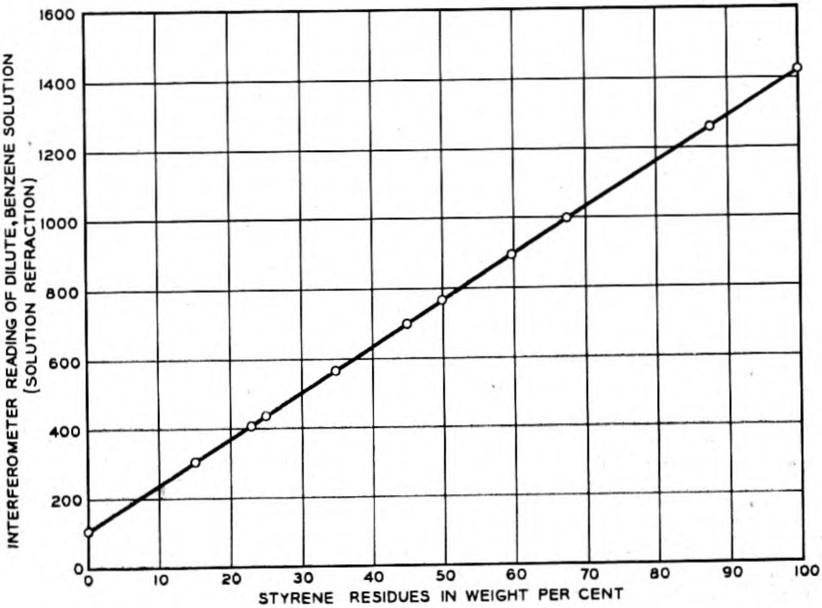


Fig. 4.—Refraction as a function of styrene content for solutions in benzene of polymers containing known percentages of styrene.

Through the use of this method it has been possible to control the styrene content of the copolymer to about ± 0.2 weight percent styrene residues, which is amply close for all purposes. Figure 5 shows the apparatus employed in this determination, the interferometer. More recently, it has been possible to employ a simpler procedure where a milling of the copolymer is introduced to remedy difficulties early encountered in the determination of the refractive index directly on the solid[11]. Although not as precise as the interferometer method, this method is shorter and as a consequence is finding application in process control. It is safe to say that today, with these methods, the control of the average composition of GR-S procuced in this country is now entirely adequate for all purposes.

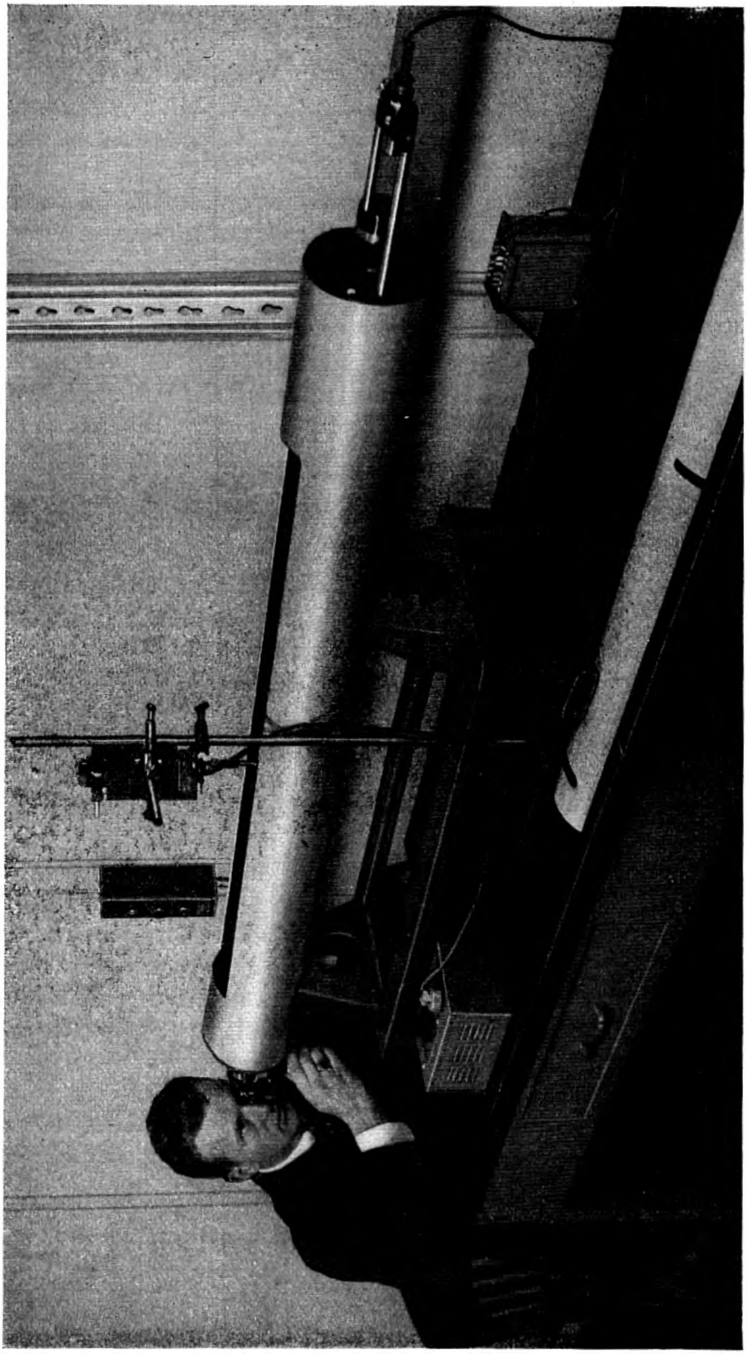


Fig. 5.—Interferometer used in the determination of styrene content of synthetic rubber from refraction.

Molecular Weight Distribution in GR-S

Unlike the linear polyesters whose molecular weight distributions can be calculated from simple assumptions (Fig. 1), the distribution of molecular sizes present in polymerization polymers cannot, at the present state of our knowledge at least, be accurately predicted. With linear polymers of uniform composition it is possible to determine experimentally the approximate molecular weight distribution by fractional precipitation of the dissolved polymer from dilute solution. This procedure, to yield good results, must be carried out under very careful control, and requires considerable time. The usual procedure is to prepare a solution of the polymer to be studied and add to it portions of a precipitant. The successive fractions of the whole polymer precipitated are then examined for average molecular weight by some suitable method. This procedure can give only a crude separation but often furnishes useful information. More accurate results require the use of very dilute solutions and the precipitation is best carried out by lowering the temperature to produce insolubility at each step. The experimental distribution curve is then obtained by plotting as ordinate the weight fraction and as abscissa the average molecular weight (weight average or number average) corresponding to each fraction. In this way an integral curve is obtained which on differentiation gives differential curves of the type shown in Fig. 1.

In GR-S, such a fractionation procedure is complicated by the fact that all of the molecules of the copolymer are not of the same type. For as we have seen we may encounter differences not only in structure between molecules but also in composition either of which alone will, independently of molecular size *per se*, influence solubility.

In fact experiments have shown that fractions separated from GR-S actually do exhibit differences in styrene content attesting to the special complications of determining molecular distributions in copolymers by this method. In spite of this, fractionations of GR-S have been made which no doubt have qualitative value. As a result of such experiments it has been found that molecular size distribution in GR-S is highly dependent on impurities present during the reaction as well as on other factors. When, however, the process and raw materials are suitably controlled it is likely that the shape of the curve does not vary greatly. Under these circumstances the number average molecular weight determined by osmotic pressure furnishes a measure of molecule size.

If the molecules are not too highly branched, we may employ viscosity measurements to furnish a "viscosity average" molecular weight. Since the latter measurements are the simplest to make they are generally employed[12], although care must be used to insure proper interpretation of results. In general, the average molecular weight given by the viscosity will fall nearer to the low molecular weight end of the distribution curve than does the true weight average molecular weight. Only for a homogeneous system does it coincide with the number average value. Hence, the difference between the two can be used as a rough measure of the broadness of the distribution.

Light scattering from solutions offers possibly an absolute way of getting the true weight average value. If the molecules are small compared to the

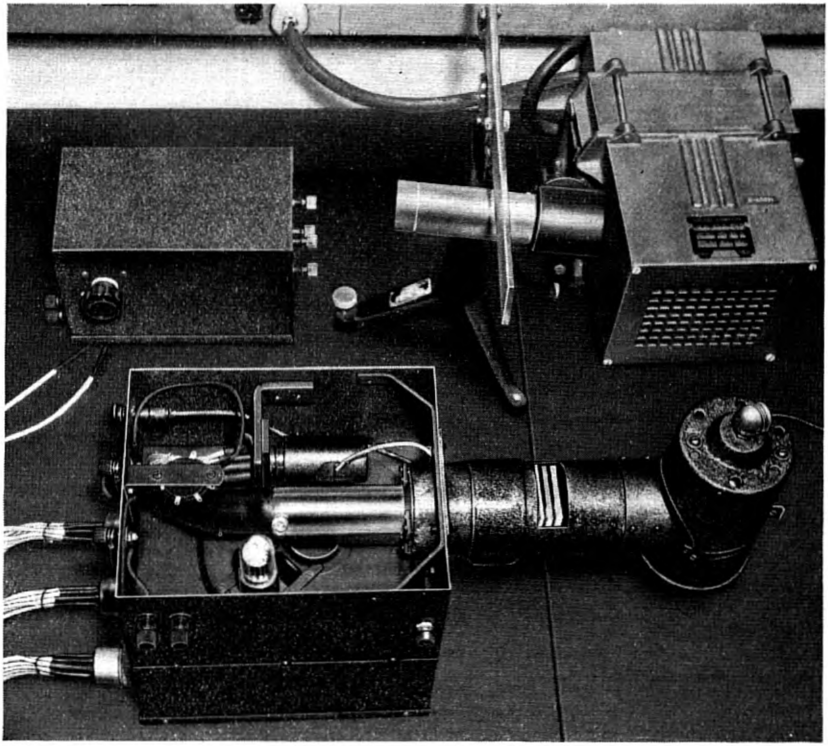


Fig. 6.—Apparatus for measuring the intensity of scattered light from solutions of polymers

wavelength of the light used and solutions of various dilutions are employed, measurements of turbidity τ, i.e. the fraction of the total light scattered per cm. of path, allow the weight average molecular weight $\overline{M}_w$ to be calculated according to

$$\overline{M}_w = \frac{1}{H(c/\tau)_o}$$

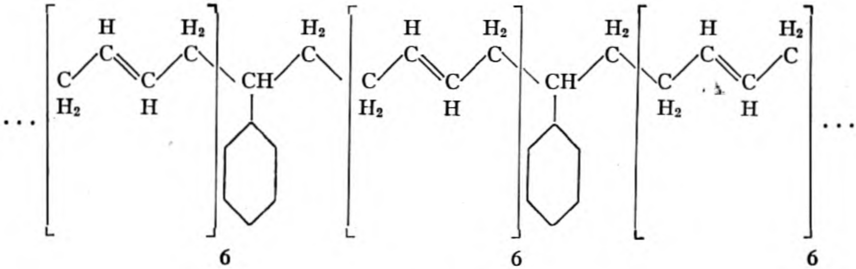
where H is a constant, c is the concentration and $(c/\tau)_o$ is the value found by extrapolation to zero concentration[13]. Further study of this method is required before direct results can be obtained on GR-S. An apparatus employing electron multiplier tubes for measurement of the intensity of the scattered light, which was developed in the Laboratories, is being used to study this new technique. Its original form is illustrated in Fig. 6. Likewise, photographic determination of scattered light has established good correlation with independent molecular weight evaluation of certain other polymers[14].

The distribution of molecule sizes in GR-S has a profound influence on its properties. It is controversial still as to whether a uniform or non-uniform distribution is desirable for all considerations. The presence of low molecular material favors ease of processing but depreciates properties. High molecular material behaves the opposite. It is customary to regard the viscosity, either of the rubber itself or the dilute solution viscosity, as a measure of the average molecular weight. While this assumption is not wholly true, it is partly justified because the shape of the distribution curve as commonly measured for GR-S is roughly constant. When osmotic measurements can be made sufficiently accurately, the number average molecular weight together with the viscosity average provides a more precise measure of the distributions present.

Chain Structure of GR-S and Its Characterization

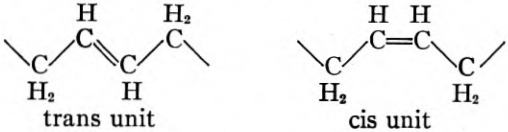
As for simple polymers, the chain structure of GR-S is best considered from two points of view: the micro-structure and the macro-structure. The micro-structure, which has already been briefly discussed, is concerned with the kinds of atoms forming the chain, their arrangement in space and the manner in which they pack with the atoms of neighboring chain molecules. It is this structure which is all-important in determining the nature of the forces between molecules and, in turn, the intrinsic rubber-like properties of the polymer. The micro-structure also determines the chemical properties of the compound. The macro-structure, on the other hand, is not dependent on the kind of atoms in the polymer or their immediate relation to each other but with the length of the chain molecules, their general shape and the extent to which they are joined with the other molecules (netted) or what were originally other molecules in the material. It is the macro-structure which plays the chief role in plasticity and viscosity of the rubber during processing, its smoothness or roughness during extrusion, the extent to which it elongates or creeps on stretching and the extent to which it swells in solvents.

Let us approach the problem of the micro-chain structure of GR-S by considering the possibilities from the organic structural point of view. In the formation of GR-S about 6 butadiene molecules combine with each styrene molecule. In (9) butadiene is shown in brackets and styrene residues are between them. There are at least four important ways in which the chain structure of GR-S can deviate from the simplest structure, namely



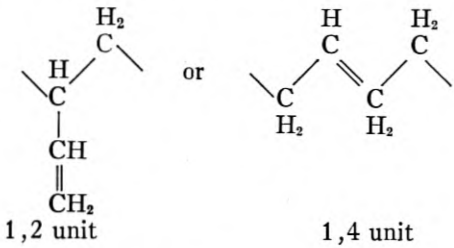
(9)

which would arise from a regular addition of butadiene and styrene. In the first place, as noted before, the styrene and butadiene units can be badly mixed up in the chain and not arranged in any special order. Secondly, the butadiene or the styrene units may be reversed end for end in the chains. This will make no difference in the case of the butadiene provided the molecule has a center of symmetry but this is probably not the case. Again, the butadiene residue may assume either the well-known *cis* or the *trans* configurations shown in (10) because of the double bond present in it.



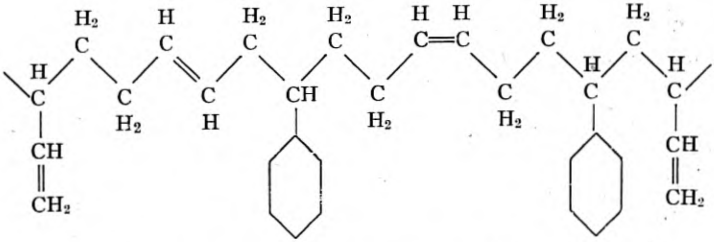
(10)

Finally, the butadiene unit may be combined into the chain as a 1,2 or as a 1,4 unit. In the former case a vinyl group is appended to the chain molecule whereas in the latter it is absent:



(11)

If these possibilities are considered, one representation of the chain molecule over a distance we should consider within the definition of micro is as follows:



(12)

Obviously many other combinations are possible which are even more involved.

There is considerable chemical as well as physical evidence to support the presence of all of these possibilities in the GR-S molecule. It is probable that the butadiene and styrene units enter the chain in an irregular manner, although, as we have seen, one molecule may acquire more total styrene or butadiene than another. The occurrence of *cis* and *trans* forms and head-to-tail arrangements is also irregular. The 1,2 and 1,4 butadiene structures likewise may occur randomly although the amount of 1,2 structure appears to vary somewhat depending on the type of reaction. It is not possible to review here the detailed evidence for the randomness and for the occurrence of these various features. The fact that x-rays when diffracted from stretched or cooled samples of GR-S fail to show evidence of crystalline or even of imperfectly crystalline material is proof that a disordered chain structure exists. X-rays, however, do not specify the cause of this disorder.

Work on synthetic linear polymers of known composition has demonstrated that relatively minute structural changes are able to cause marked disorder in polymer systems[16, 17]. It is not surprising, therefore, to find that GR-S copolymer is disordered. The important question is: what effect has the disorder on the properties and, if it is deleterious, what can be done about improving the chain structure? Without going into detailed arguments there is good reason to believe that an ordered chain structure is desirable for the best properties in a rubber. Only then is it possible for the chain molecules to pack together into crystalline-like regions on stretching and thus provide the resistance to tearing and breaking that are required. Natural rubber possesses this characteristic to an outstanding degree and polychloroprene and polyisobutylene when vulcanized also show considerable crystalline behavior on stretching. Other factors, such as the rate at which crystalline regions develop, are likewise important[17]. But the crucial requirement for toughness is the development of the crystalline type of forces on stressing.

It must be admitted that no great progress in reducing the chain disorder of GR-S has been attained as yet. Obviously, complete order because of the hybrid nature of the polymer is impossible. This was realized at the outset of the research program and for that reason emphasis was placed on improving the macro-structure where obvious changes could be effected. We shall consider this phase of the work next.

We have already seen how the chain molecules of GR-S vary in size and in composition. They may vary also in over-all shape. Branching and cross-linking leading eventually to net-work formation may result during the chain growth or termination reactions. In this way variously shaped molecules may arise. Obviously the situation may become very complex and in reality we may have to do with mixtures where all types of molecular species are present at once.

What influence on the properties of the final compounded and vulcanized rubber do these various branched and netted chain structures have? It was not recognified at first that the gel part of GR-S was particularly different from the sol in its effect on ultimate properties. This was because no reliable measurements of sol or gel had been made and because sol and gel behaved differently during the compounding and processing steps[7, 12]. Some workers also did not appreciate that natural rubber and GR-S behave very differently in regard to the effect of processing on their ultimate properties.

It has since been established that the sol-gel properties are of importance both in the processing and in the final properties of GR-S synthetic rubber. It turns out that the amount of the sol and its molecular weight distribution and the amount of the gel and its swelling volume, which is a measure of the intensity of netting, enables us to make predictions as to what properties a given sample of rubber will exhibit during processing and in the final product[18]. This does not mean that other features of the sol and gel are unimportant. For example, methods of estimating the degree of branching (by means of concentrated solution viscosity)[6, 12] of the soluble portion have been worked out which undoubtedly will be useful if a more refined control proves desirable.

It is possible to make GR-S type rubber which is completely soluble. Such a product requires to be characterized only as to molecular weight distribution, composition and perhaps degree of branching. If the distribution of sol is such that there is an excess of low molecular material, the copolymer besides being soft and difficult to handle, provides cured stocks which have low tensile strength, poor tear and abrasion resistance, poor resistance to the growth of cracks and high hysteresis loss. If, on the other hand, an excess of high molecular material is present in the sol the copolymer is very stiff[18] and cannot be handled in the subsequent compounding and processing procedures. Aside from this difficulty its ultimate properties seem to be superior the higher the average molecular weight. When all considerations of properties and processing requirements are taken into account a copolymer containing as nearly linear molecules as possible and having neither an excess of high or low molecular fraction is probably preferable.

Gel GR-S, depending on its swelling volume (see below), is a tough material totally lacking in plasticity. Swelling volumes as low as 10 are hardly distinguishable from vulcanized gum GR-S and in fact resemble it structurally because vulcanization is actually a special kind of gel formation. Ordinarily the swelling volumes of gel in GR-S range between 20–150[17, 18].

Commercial GR-S may contain both sol and gel, although the trend is to eliminate gel altogether. When large amounts of gel of moderate swelling volume are present the product is hard to mix, although it may extrude smoothly, and after processing, particularly if done hot, it is likely to give products which have higher modulus than copolymer free from gel, and to show poor resistance to cutting and crack growth—properties of great significance in tires and other applications.[18] It is therefore important that we should be able to determine sol and gel in the presence of each other. This need is particularly great in the case of characterization of copolymers after they have been subjected to processing and compounding[18]—treatments which often are responsible for profound changes in its molecular structure.

Methods of Characterization of Sol and Gel

Considerable work has been done at the Laboratories on methods for determining the sol-gel properties of polymers and in investigating the effects of various after treatments of the copolymers on their sol-gel characteristics[6, 18]. Figure 7 shows the type of apparatus employed for effecting the sol-gel separation[7]. The weighed copolymer sample is thoroughly dried, cut into small pieces and distributed on stainless steel screens contained in the bulb of the apparatus. About the 100 ml. of benzene is added and the parts assembled. After 24 hours or more standing without disturbance, the benzene containing the soluble part of the copolymer is carefully withdrawn by opening the stop-cock very slightly. The weight of the swollen gel left on the screens is obtained from the difference between the weight of the assembly after draining off the solution and its original weight. This divided by the original weight of the unswollen gel gives the swelling volume (SV) of the material. The slight density correction can be neglected.

The dilute solution viscosity is determined directly on 5 cc. of the solution withdrawn from the vessel and is calculated from equation (2). The concentration c is determined by evaporating a known volume of the solution and weighing the solid left after evaporation of the benzene. Figure 8 shows the viscometers and bath employed for the measurements of the relative viscosity. A variation of the dilute viscosity method adopted for

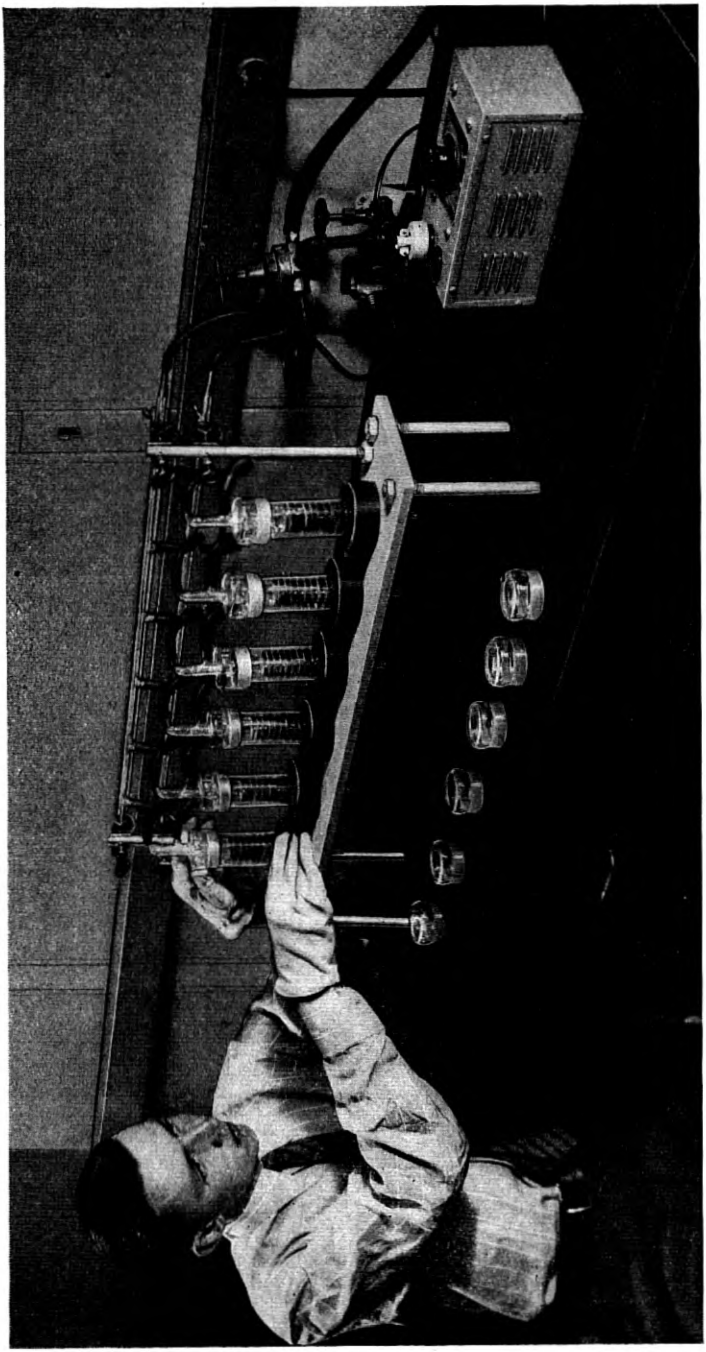


Fig. 7.—Apparatus employed in the determination of the sol-gel content of synthetic rubber.

use directly on latex has been employed as a control during the synthesis of GR-S[19]. This test, referred to as the vistex test, consists in adding 1 ml. of the latex sample to be examined to 100 ml. of a solvent having both hydrophobic and hydrophilic properties, such as a mixture of 70 parts (by volume) of xylene with 30 of pyridine, or 60 of benzene and 40 of *t*-butanol. The clear solution is run through the viscometer in the usual manner and the relative viscosity used as a measure of extent of reaction. The test has the

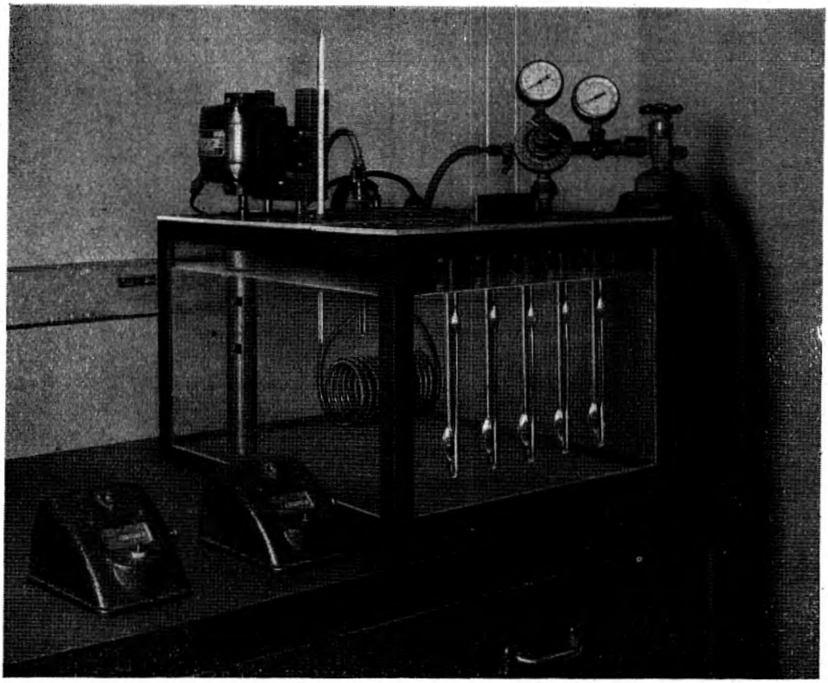


Fig. 8.—Viscometers and bath used for the determination of dilute solution viscosity of polymer solutions.

advantage of great speed, thus providing control of the reaction, step by step. Figure 9 shows the apparatus employed in the determination of concentrated solution viscosity (*CSV*).[20] In this measurement a 15 percent solution of the copolymer in xylene is made by weighing the required quantity of GR-S into a test-tube adding the precise volume of xylene and stoppering. The solution is homogenized by moving a steel armature through it in the test-tube by means of a strong electro-magnet. A trace of acetic acid is added to eliminate thixotropic effects. After complete dispersion has been effected the viscosity is determined by the falling ball method. Branched copolymers show inordinately high concentrated solution viscosities. The latter may therefore be employed as a measure of degree of branching or approach to gelation when supplemented by dilute solution viscosity measurements. Furthermore, the power required to maintain the armature stationary as measured by the current passing through the magnet furnishes data useful in predicting how a given copolymer sample will process.

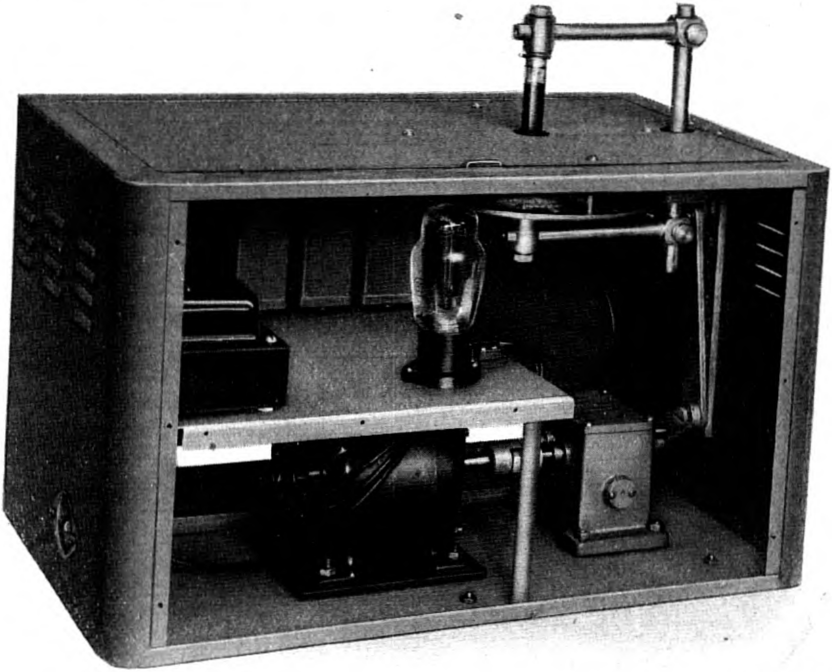


Fig. 9.—Apparatus employed to effect the solution of synthetic rubber prior to the determination of concentrated solution viscosity.

Application of Sol-Gel Methods to Control Processing

In addition to their application to the control of synthetic rubber in production, the sol-gel methods of characterizing the copolymer which have been briefly described above are of very great use in elucidating what happens during the processing of the rubber[18]. By the term "processing" is meant the operations which are carried out on the copolymer subsequent to its manufacture and prior to its vulcanization into its final form. These operations involve working the rubber on machinery (plastication) in order to render it soft and satisfactory for mixing in pigments and for extrusion

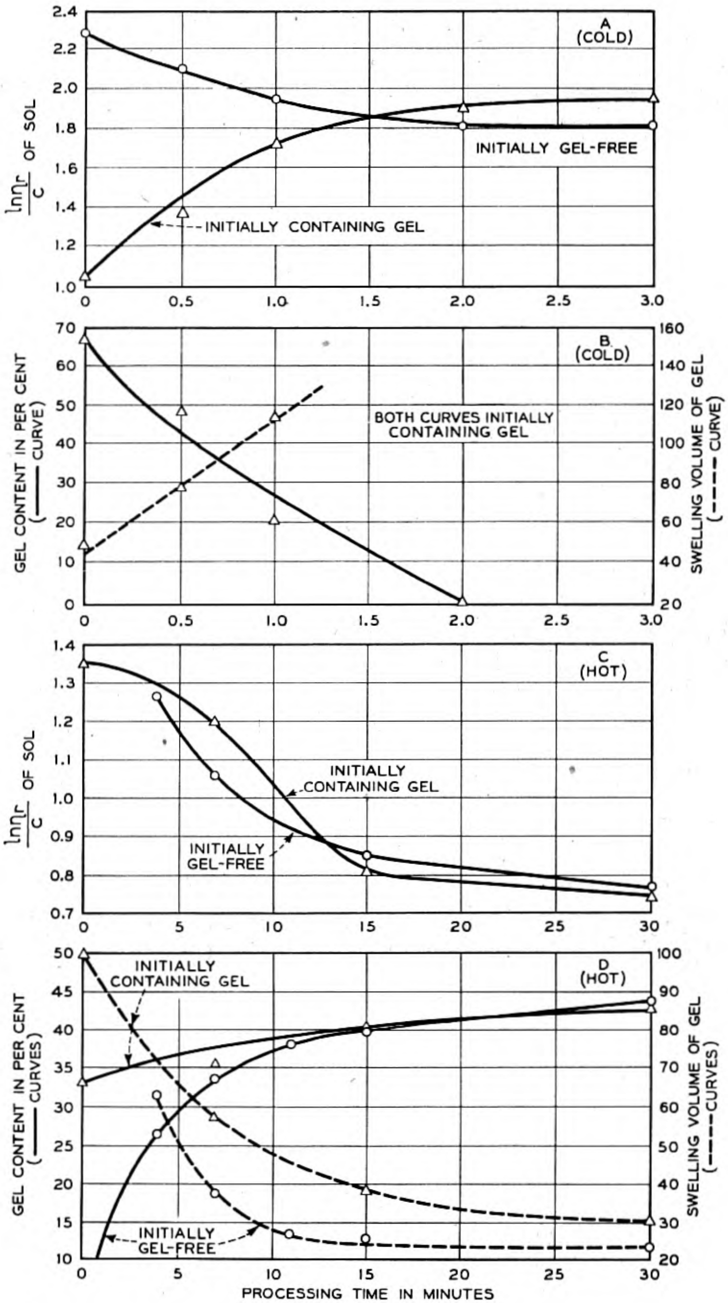


Fig. 10.—Curves showing change in solution viscosity $\left(\frac{\ln \eta_r}{c}\right)$ of sol, gel content, and swelling volume of gel with time of milling. Both cold (A and B) and hot (C and D) milling are shown for samples of synthetic rubber containing no gel and gel of low swelling volume.

and molding. Several different types of machines are employed including mills, calenders, Banbury mixers and extrusion machines. At every stage and particularly in Banbury mixing, where carbon black is generally mixed in, the copolymer undergoes changes which affect its performance in the finished product. This is especially true if, as nearly always happens in practice, considerable heat is developed during the operation. Indeed, it is frequently true that processing operations have more to do with the ultimate rubber properties than do factors in the production of the copolymer itself[18]. It is only fairly recently that this point has been sufficiently emphasized and considerable progress made in controlling the processing steps to the same extent as the polymerization is now controlled.

To see what happens during processing of the copolymer let us assume we subject two extreme types of GR-S, one containing no gel and one containing gel of rather low swelling volume, to a typical hot processing treatment consisting of hot mastication and Banbury mixing in of carbon black.[18] In addition, in order to exhibit differences in processing let us consider the effects of cold processing on the same two samples. As is evident from Fig. 10 which summarizes the results,[18] hot processing tends to build up gel and decrease its swelling volume in both of the rubber samples. The dilute solution viscosity on the other hand falls. This behavior although of advantage to subsequent extrusion and calendering operations is definitely opposed to securing the best mechanical properties in the final rubber. Cold processing has the opposite effect on the samples. Thus, the copolymer not containing gel is little affected, whereas the gel in the other is broken down and gives rise to a higher dilute solution viscosity.

In processing, therefore, important changes in the chain structure of the copolymer are brought about. Under certain circumstances, these are beneficial, but since most processing involves considerable heat development the changes are usually detrimental. It is of the utmost importance therefore that a type of copolymer is produced which is compatible with the type of processing machinery already installed in industry. In addition, uniformity of the copolymer is of very great significance if control of processing operations is to be achieved, for such control cannot be attained with a variable starting material.

The processing step which involves the mixing in of carbon black (or other pigment) is perhaps the most important. Unfortunately, the presence of the carbon black makes it impossible to employ the usual sol-gel analysis because a new kind of insolubility enters.[18, 21] In addition to the primary valence gel discussed up to this point, a secondary valence combination involving the carbon black and the large sol molecules forms. This is not immediately distinguishable from the first type of gel unless we have other reason to know that the latter is absent and does not form during the mixing operation. This phenomenon of the insolubilization of natural rubber by carbon black has been known for some time.[22] Only recently, however, has its relation to the structural features of GR-S copolymer become apparent.[21] Work is now underway to allow an estimation of both types of gel in the presence of one another. When this is achieved the analysis of the reactions occurring during compounding will be further facilitated.

Chain Structure and Polymer Properties

We have now reviewed some of the molecular complexities which are involved in the synthesis of GR-S synthetic rubber. It remains to discuss more in detail the influence of chain structure on the properties we associate with rubber-like behavior. We might begin by asking ourselves two questions: (1) What makes a polymer exhibit rubber-like properties? (2) What composition and chain structure are desirable in a rubber? The first question involves a discussion of the theory of rubber-like elasticity. The answer to the second involves an inquiry into the specific use to which the material is to be applied. Since most of our rubber is employed in tires let us consider the special requirements for that use.

Taking up the first question, we fall immediately into the pit of having to define what a rubber is and how it differs from a plastic. Originally rubber meant "natural rubber". When synthetics with rubber-like properties appeared we adopted the term "synthetic rubber" to describe them. Some have objected (unsuccessfully) to the use of this term because it implies synthetic natural rubber and have proposed the word "elastomer" instead. Others have gone still further and suggested other terms (usually ending in mer) for various plastics and rubber-like materials.

All of these new names seem unnecessary. Polymer is the inclusive term. The term rubber simply has come to mean a polymer which at ordinary temperatures has properties like natural rubber. A plastic is a polymer which at ordinary temperatures is hard and which usually becomes soft and deformable at higher temperatures. Such terms as rubber-like plastic or glass-like plastic are frequently employed. This kind of terminology is admittedly loose but it often tells just as much in familiar words as does the newly proposed nomenclature.

The significant fact is that there is a perfectly consistent and orderly relationship between the properties of polymers and their chemical composition and structure. Fundamentally, the major factor which determines whether a long chain polymer will be a rubber or a plastic is the magnitude of the forces acting between chain molecules. If the forces between polymer molecules are low, the polymer is a rubber; if they are high, it is a plastic. And obviously since these forces can be regulated nicely there are all gradations from the hardest to softest polymer. A rubber therefore may be regarded simply as a soft plastic—one in which the forces between chains are very low—with one important distinction, namely that soft plastics to show rubber-like properties must be "vulcanized" i.e. a few very strong inter-chain linkages must be established to prevent slippage. Some plastics can be vulcanized, too, but here the inter-chain forces are high anyway and the few additional strong bonds are not essential. However, the usual inter-chain forces in plastics being of the van der Waals' type are very susceptible to temperature. Consequently, if we weaken them by raising the temperature we can, provided the plastic is "vulcanized", cause it to acquire rubber-like properties at the higher temperature. So the distinction between rubbers and plastics is in the last analysis slight.

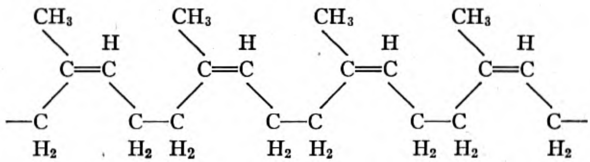
We still have not answered the question as to what causes a polymer to exhibit rubber-like properties. In fact it was only during the last 10 years or so[23] that the answer has been known, which is surprising, because it is simply "temperature". Contrary to previous views, forces between atoms in the same chain have little to do with the long range retraction phenomenon shown by rubbers, at least at elongations up to about 200%. The stretched polymer returns because of the thermal heat motion in the mass which seeks to restore the elongated chain molecules to their more stable, kinked-up configurations. The molecules, through their vulcanization points, communicate their retraction behavior to the entire mass. Thus theory agrees with experience that the chemical constitution of the polymer is of secondary significance. As long as chain molecules are present which are capable of kinking-up by rotations about chemical bonds, as long as the forces between molecules are not large compared to the thermal energy, and as long as the molecules are interconnected at points so as not to slip, we shall have a rubber-like substance whether we call it a rubber or a plastic or an elastomer.

Coming to our second question, it might now take the form: what composition and chain structure are desirable in a rubber for tires? We shall see that the qualitative views expressed above must be altered if we are to explain the more intimate properties of rubber involved in this application. We have already seen from the sol-gel discussion what some of these refinements are. There are certain differences between GR-S type synthetic rubber and natural rubber, however, which go back even farther and involve the manner in which the molecules pack and slip over one another during deformation. These have geen discussed above under micro-structure.

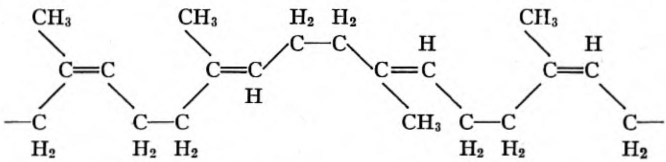
Man learns largely by imitation and our knowledge of rubber-like behavior has been no exception. Natural rubber possesses amazing qualities which no synthetic product has yet been able entirely to duplicate, although we have found in many cases ways to overcome weaknesses in synthetic rubbers by round-about means. For example, the hysteresis loss in vulcanized natural rubber, i.e. the heat generated per cycle of reversible stretching, is less than any synthetic rubber of comparable gum tensile strength. The resistance to crack growth of vulcanized natural rubber, particularly in gum, i.e. unfilled form, is better than any synthetic of comparable low-temperature flexibility.

What are the causes of these differences? We can make synthetic rubbers from hydrocarbons (in fact from isoprene itself) which judged simply from the composition should have the same interchain forces acting as in natural rubber. That this is not so is shown by the fact that polyisoprene in all the properties which would fit it for a tire rubber is much inferior to natural rubber. We are forced to conclude that it is the form of the individual isoprene units (*cis* or *trans*) and the way in which they are placed in the chain that determines.

In stretched natural rubber we are convinced that units are orderly arranged in the *cis* configuration as follows:



In polyisoprene on the other hand they are probably in random disorder:



We might expect then that the interchain forces, which change so critically with distance, are less of a match for the thermal energy in synthetic polyisoprene than in natural rubber. The high degree of molecular uniformity in natural rubber, as X-rays show, gives rise to a crystallization on stretching which is entirely absent from polyisoprene (see Fig. 11). Even before crystallization as such has progressed far (it starts in nuclei and these multiply throughout the mass) the interchain forces have prepared for and compensated in an effective way for the increased effect of stress tearing the chains apart and the increased thermal energy tending to weaken the interchain forces. We may look upon natural rubber as a substance which progressively and automatically transforms itself to a plastic as it is elongated. It is these crystallization phenomena which are responsible for high gum tensile strength[17] and outstanding resistance to the growth of cracks.

To return to our question and ask once more what structure we desire in a rubber for tires we see that although we cannot quite write an order in terms of a chemical formula we can state general requirements. We want a polymer in which the interchain forces are as low as possible to give us low hysteresis. At the same time we want regularity of chain molecules to provide a minimum loss of cohesion with rising temperature and rising elongation. To a chemist this sounds like an order for natural rubber and the design of Buna S which we were forced to imitate in the emergency seems wrong. If research can iron out some of this irregularity, a further improvement in our product perhaps can be achieved. The chemist by the clever trick of adding styrene to butadiene has provided himself a way he can

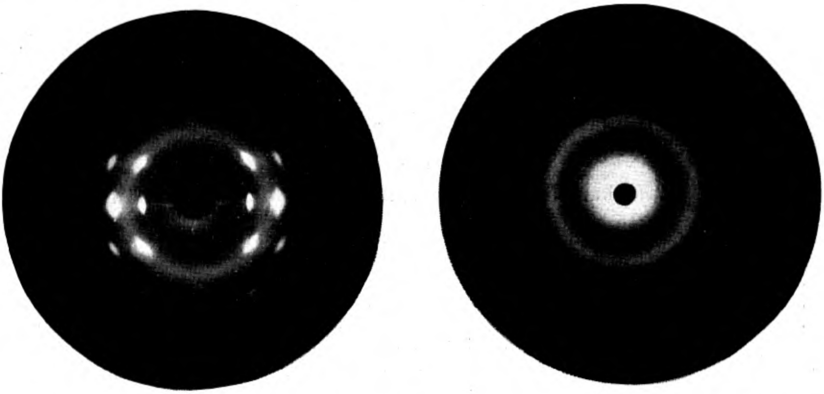


Fig. 11.—X-Ray photographs of natural rubber, stretched (left) and synthetic polyisoprene, stretched (right).

regulate the interchain forces and therefore the degree of rubberiness of Buna S. He is able to make it harder and stronger at will by increasing the amount of styrene, something nature is unable to do. But his task is not finished until he can control also the order and the packing of his molecules or devise some equally clever way of getting the interchain forces to behave.

Conclusion

We have attempted to review some of the problems arising out of the effort to achieve the best possible Buna S type rubber for our war emergency and to show how they have been attacked. We have also tried to give a simple account of some of the theories underlying the behavior of polymers. The story of synthetic rubber is of course much broader both in theory and practice than we have indicated.

Future synthetic polymers will be devised to meet the intimate requirements of many diverse applications. Engineering will be more precise and control of our materials will be based more on scientific methods. It is romantic to read from a recent popular book "you see him in his shirt sleeves cutting off a piece of rubber with his knife, smelling it, biting it and stretching it. Then he either looks satisfied or worried. Laboratory reports give him a complete report on the sample but a prodigious memory and a sixth sense born of years at his job often tell him whether the rubber will make a good tire." But this is hardly the way the future engineer will judge. It is hoped that the present account has helped to point out how scientific methods are being applied and how research can supply a safe guide to the wise control and application of synthetic materials.

Acknowledgement

The author wishes to express his thanks to Dr. W. O. Baker for valuable comments on the manuscript and for assistance in preparing the figures.

References

1. Harries, "Untersuchungen über die Natürlichen und "Künstlichen Kautschukarten"" Berlin (1919).
2. Sponsler and Dore, *Colloid Symposium Monograph, 4*, 174 (1926).
3. Staudinger, *Ber., 59*, 3019 (1926).
4. Vorländer, *Ann., 280*, 167 (1894)
5. Flory, *Jour. Am. Chem. Soc., 58*, 1877 (1936).
6. Baker and Mullen, Private communication, Bell Telephone Laboratories to Office of Rubber Reserve, March 23, 1943; *Ibid.*, June 24, 1943.
7. Baker and Mullen, Private communication, Bell Telephone Laboratories to Office of Rubber Reserve, Sept. 13, 1943.
8. Baker and Heiss, Private communication, Bell Telephone Laboratories to Office of Rubber Reserve, March 23, 1943; *Ibid.*, Sept. 14, 1944.
9. Baker and Heiss, Private communication, Bell Telephone Laboratories to Office of Rubber Reserve, Sept. 13, 1943.
10. Baker and Heiss, Private communication, Bell Telephone Laboratories to Office of Rubber Reserve, June 24, 1943.
11. Baker and Heiss, Private communication, Bell Telephone Laboratories to Office of Rubber Reserve, May 11, 1944.
12. Baker and Mullen, Private communication, Bell Telephone Laboratories to Office of Rubber Reserve, Nov. 23, 1943.
13. Debye, *Jour. Appl. Phys., 15*, 338 (1944).
14. Doty, Zimm and Mark, *Jour. Chem. Phys. 12*, 144 (1944); *13*, 159, (1945).
15. Baker and Pape, Private communication, Bell Telephone Laboratories to Office of Rubber Reserve, March 14, 1944.
16. Baker and Fuller, *Jour. Am. Chem. Soc., 64*, 2399 (1942).
17. Baker, *Bell Lab. Record, 23*, 97 (1945).
18. Baker and Mullen, Private communication, Bell Telephone Laboratories to Office of Rubber Reserve, June 12, 1944.
19. Baker, Mullen and Heiss, Private communication, Bell Telephone Laboratories to Office of Rubber Reserve, Jan. 21, 1944.
20. Baker and Mullen, Private communication, Bell Telephone Laboratories to Office of Rubber Reserve, Nov. 30, 1943; Baker, Walker and Pape, Private communication, Bell Telephone Laboratories to Office of Rubber Reserve, March 14, 1944.
21. Baker and Walker, Private communication, Bell Telephone Laboratories to Office of Rubber Reserve, Sept. 14, 1944.
22. Stamberger, *Kolloid. Z., 42*, 295 (1927).
23. Whitby, *Jour. Phys. Chem., 36*, 198 (1932).

Characteristics of Vacuum Tubes for Radar Intermediate Frequency Amplifiers

By G. T. FORD

1. Introduction

THE desired characteristics for vacuum tubes for use in broad-band intermediate frequency amplifiers are primarily high transconductance, low capacitances, high input resistance, and good noise figure. These characteristics determine the frequency bandwidth, amplification, and signal-to-noise ratio attainable with such an amplifier. The maximum operating frequency is generally limited by the input resistance of the tube which decreases as the frequency is increased, and, in some cases, by the tube noise which increases with increasing frequency. Three other characteristics which are also important are small physical size, low power consumption, and ruggedness. The present paper describes how these characteristics are related to the performance requirements for intermediate frequency (IF) amplifiers used in radar systems and shows how the requirements were met in the design of the Western Electric 6AK5 Vacuum Tube.

In a coaxial cable carrier telephone system of the type which was initially installed between Stevens Point, Wisconsin and Minneapolis, Minnesota[1] the upper frequency of the useful band is of the order of three megacycles per second (mc) and the bandwidth is of the same order. The Western Electric 386A tube was developed a number of years ago for amplifiers such as those used in this system. It is characterized by high transconductance and low capacitances. In radar receiving systems similar but more exacting requirements must be met for the IF amplifier. It is desirable to operate in many cases at a mid-band frequency of 60 mc, to have a bandwidth of the order of 2–10 mc, and to have as close to the ideal noise figure as possible. Additional considerations of great practical importance are low power consumption, small size, and ruggedness.

The choice of the mid-band frequency for the IF amplifier is influenced by considerations, a detailed discussion of which is outside the scope of this paper. For example, the characteristics of the beating oscillator and ite relation to the operation of the automatic frequency control (AFC) system, when AFC is used, are involved. The usual practice has been to

[1] "Stevens Point and Minneapolis Linked by Coaxial System," K. C. Black, *Bell Laboratories Record*, January 1942, pp. 127–132.
"Television Transmission Over Wire Lines," M. E. Strieby and J. F. Wentz, *Bell System Technical Journal*, January 1941, Vol. 20, p. 62.

standardize on a mid-band frequency of 30 mc or 60 mc. The latter frequency has been used in most radars developed at the Bell Telephone Laboratories.

In pulsed radar systems the transmitter is turned on and off by the modulator in such a way that radio frequency energy is generated for a pulse duration τ seconds at a repetition rate which is usually several hundred per second. When the transmitter is modulated by a pulse which approaches that shown in Fig. 1(a), the energy-frequency distribution in the transmitted signal is as shown in Fig. 1(b). Although the maximum of the distribution

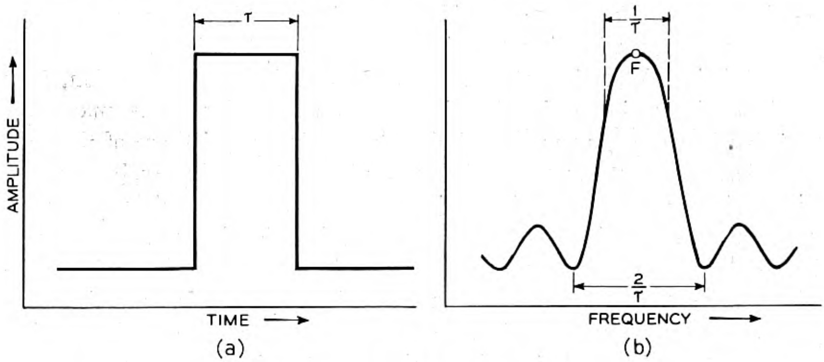


Fig. 1a—Modulating pulse
Fig. 1b—Amplitude—frequency distribution in transmitted RF pulse.

curve lies at the transmitter frequency F, there is considerable energy in the range $F \pm \frac{1}{2\tau}$ and the receiver usually has a bandwidth of at least $\frac{1}{\tau}$ in order to make as efficient use as possible of the energy in the echoes reflected by the target. For many radar applications, this requirement and the problems of transmitter and beating oscillator frequency stability result in the use of a bandwidth of as much as 10 mc for the IF amplifier.

The Western Electric 386A tube, Fig. 2(a), had characteristics which approached those needed to meet the IF amplifier requirements discussed above. It was therefore slightly modified in physical form for convenience of use and recoded the Western Electric 717A tube, Fig. 2(b). The 717A tube was used extensively in IF amplifiers in several radar systems. As the emphasis on small size and light weight for airborne radars increased, and, as the need for better characteristics became more pressing, further development was undertaken which resulted in the Western Electric 6AK5 tube, Fig. 3.

The importance of size and weight for radar systems to be used in airplanes is obvious. The use of miniature tubes in airborne equipment has

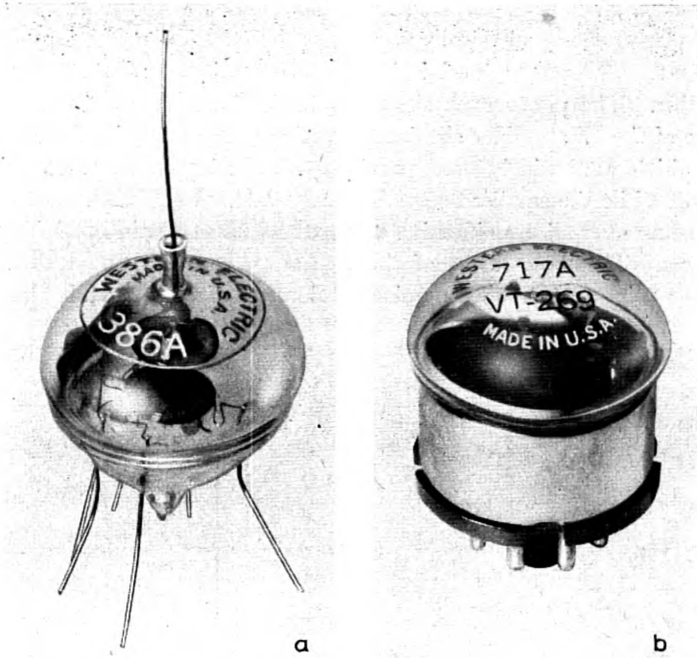


Fig. 2a—Western Electric 386A vacuum tube (full size).
Fig. 2b—Western Electric 717A vacuum tube (full size).

Fig. 3—Western Electric 6AK5 vacuum tube (full size).

been a consequence wherever their power handling capabilities are adequate to meet the performance requirements. The IF amplifier offers an ideal opportunity to effect substantial savings in both size and weight by using small tubes. When the IF frequency is as high as 30 mc or 60 mc, the circuit elements are physically small, so that the tube size is relatively important. The power level is low enough so that miniature tubes are applicable. The photograph shown in Fig. 4 illustrates the reduction in the size of the IF amplifier obtained by using 6AK5 tubes in place of 6AC7 tubes which were widely used previously. The larger amplifier weighs 2 lbs. 4 oz. while the smaller one weighs only 9 oz. Each of these amplifiers provides an over-all amplification of about 95 db at a mid-band frequency of 60 mc. The bandwidths of the two amplifiers are comparable. The amplifier using 6AC7 tubes requires 31.3 watts of power, while the 6AK5

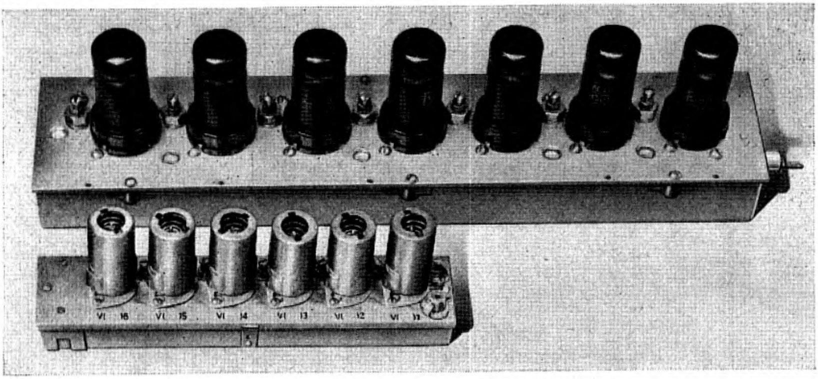


Fig. 4—60 Megacycle IF amplifiers (¼ full size).

amplifier requires 14.4 watts. The power savings possible with the 6AK5 tubes are not particularly important from the power economy standpoint, but rather because of the easier heat dissipation problem. In compact equipment such as airborne radar, the problem of keeping the operating temperatures of the various components within safe limits is formidable.

In the later years of the war the 6AK5 tube became the standard IF amplifier tube for radar systems and because of its superior properties was used for other applications in many other radio equipments.

2. Amplification and Bandwidth

The amplification that can be obtained with a given number of stages, and the useful bandwidth, are closely related. Within certain limits, one can be increased at the expense of the other. In fact, the product of the amplification per stage and the bandwidth is one important measure of the goodness of a particular tube and circuit design. A simple case will illustrate how this comes about. Assume the band-pass interstage shown in Fig. 5. L is the inductance of the coil, R is the shunt resistance equivalent to the load resistor and any loading effects due to high-frequency losses in the tubes or circuit elements, and C is the total shunt capacitance of the tubes, the circuit elements and the wiring. If it is assumed that R, L and C are

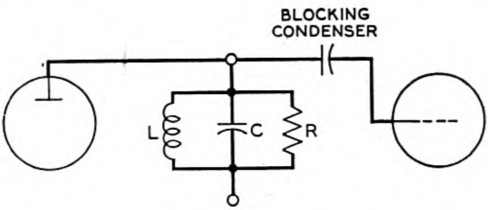


Fig. 5—Band-pass interstage.

independent of frequency, the magnitude of the impedance of this network is

$$|Z| = \frac{1}{\left[\left(\frac{1}{R}\right)^2 + \left(\omega C - \frac{1}{\omega L}\right)^2\right]^{1/2}} \quad \text{(1)}$$

When $\omega = \omega_0 = \frac{1}{\sqrt{LC}}$ the impedance is a maximum and the frequency $f_0 = \frac{\omega_0}{2\pi}$ is the resonant frequency. The maximum value of the impedance is $|Z_0| = R$. If this type of interstage is used in a grounded-cathode type of circuit employing a pentode tube, the small-signal voltage amplification A_v from the control grid of the tube to the plate is

$$A_v = G_m |Z| \quad \text{(2)}$$

where G_m is the grid-plate transconductance of the tube. It is assumed that $|Z|$ is small compared to the plate resistance of the pentode and that there is no feedback. The voltage amplification A_{v0} at the resonant frequency is then

$$A_{v0} = G_m R \quad \text{(3)}$$

If the bandwidth is defined as $\Delta F = f_1 - f_2$, where f_1 and f_2 are the two frequencies where $A_v = \frac{A_{v0}\sqrt{2}}{2}$, it can be shown that

$$\Delta F = f_1 - f_2 = \frac{1}{2\pi RC} \quad \text{(4)}$$

The product of the mid-band voltage gain times the bandwidth can be called the band merit, B_0, and we have

$$B_0 = (\Delta F)(A_{v0}) = \frac{1}{2\pi RC}(G_m R) = \frac{G_m}{2\pi C} \quad \text{(5)}$$

This expression for band merit has been derived from the product of the bandwidth and the voltage amplification for a simple band-pass interstage. Higher band merits can be realized with a given tube by using more complicated coupling networks.

Another interpretation of band merit is to say that it is the frequency at which the voltage amplification is unity. This is the frequency at which the product of the transconductance and the reactance of the shunt capacitance is unity.

From the foregoing expression for band merit it is evident that, in general, the higher the band merit the fewer is the number of stages that are required to obtain a given gain and bandwidth. It is highly desirable to keep the number of stages small in order to save space, weight and power consumption and to avoid the use of unnecessary components which reduce the

TABLE I

Type	Heater Power	Plate Current	Total Power Consumption	Nominal Transconductance	Transconductance per Unit Plate Current	Band Merit
	watts	*ma*	*watts*	*umhos*	*umhos/ma*	*mc*
6AC7	2.84	10	4.7	9,000	900	89.5
6AG7	4.10	30	9.6	11,000	367	85.3
6AG5	1.89	7.2	3.0	5,100	708	90.0
717A	1.10	7.5	2.3	4,000	533	71.4
6AK5	1.10	7.5	2.3	5,000	667	117.

reliability in operation. In practice the total amplification in the receiver is made high enough so that the system noise, with no signal, produces a fair indication on the output device when the gain control is set for maximum gain. For a bandwidth of 5 mc, the equivalent *RF* input noise power level is of the order of 2×10^{-13} watt and the power level necessary for a suitable oscilloscope presentation in a radar is about 20 milliwatts. The net over-all gain needed is then about 110 *db*. Making an allowance of, say, 15 *db* for losses in the detectors and elsewhere, a total of about 125 *db* gain is required. A minimum of about 110 *db* of this is usually in the *IF* part of the receiver. With this amount of gain as a requirement it is easy to see the importance of high band merit, small size, and low power consumption in the *IF* tubes.

Table I, above, compares the salient characteristics of the 6AK5 with those of other similar types of tubes.

The tube design factors which determine the band merit will now be examined by considering an idealized case. For an idealized plane parallel triode structure in which edge effects are assumed to be negligible, the plate current can be related to the tube geometry and the applied voltages approximately as follows:[2]

[2] "Fundamentals of Engineering Electronics," W. G. Dow, pp. 44, 102, et seq.

$$I_b = \frac{2.33 \times 10^{-6} A \left(E_{c1} + \dfrac{E_b}{\mu}\right)^{3/2}}{a^2 \left(1 + \dfrac{1}{\mu}\dfrac{a + b}{b}\right)^{3/2}} \quad \text{(6)}$$

It is assumed that the electrons leave the cathode with zero initial velocities. The failure to take into account the effects of initial velocities makes this equation only a fair approximation for close-spaced tubes, but it is still instructive to assume it is approximately correct for the purposes of the present discussion. "A" is the active area of the structure, "a" is the distance from the cathode to the plane of the grid, "b" is the distance from the plane of the grid to the plate, E_b is the plate voltage, E_{c1} is the effective grid voltage (including the effect of contact potential) and μ is the amplification constant. Dimensions are in cms. This stipulation is unnecessary for equation (6) but is necessary for some of the later equations. For a plane parallel tetrode or pentode, equation (6) is a good approximation if E_b is replaced by the screen voltage E_{c2}, μ is replaced by the "triode mu" μ_{12} (*mu* of control grid with respect to screen grid) and the coefficient M introduced, where M is the ratio of the plate current to the cathode current. The reason this approximation is good is that the field at the cathode, and therefore the cathode current, is determined almost entirely by the potentials of the first two grids. Because of the screening effect of these grids the potential of the plate has little effect on the cathode current. We have then

$$I_b = \frac{2.33 \times 10^{-6} MA \left(E_{c1} + \dfrac{E_{c2}}{\mu_{12}}\right)^{3/2}}{a^2 \left(1 + \dfrac{1}{\mu_{12}}\dfrac{a + b}{a}\right)^{3/2}} \quad \text{(7)}$$

The distance "b" is now the distance from the plane of the grid to the plane of the screen. This expression can be differentiated with respect to E_{c1} to get

$$G_m = \frac{dI_b}{dE_{c1}} = \frac{3}{2}\,\frac{2.33 \times 10^{-6} MA \left(E_{c1} + \dfrac{E_{c2}}{\mu_{12}}\right)^{1/2}}{a^2 \left(1 + \dfrac{1}{\mu_{12}}\dfrac{a + b}{a}\right)^{3/2}} \quad \text{(8)}$$

It is assumed that μ_{12} and M are independent of E_{c1}. If we let I_0 be the cathode current density, then substitute MI_0A for I_b in (7), and eliminate the expression $\left(E_{c1} + \dfrac{E_{c2}}{\mu_{12}}\right)$ between (7) and (8), we have

$$G_m = \frac{3}{2}\,\frac{(2.33 \times 10^{-6})^{2/3} MAI_0^{1/3}}{a^{4/3}\left(1 + \dfrac{1}{\mu_{12}}\dfrac{a + b}{a}\right)} \quad \text{(9)}$$

Neglecting the stray capacitances between the lead wires and also the edge effects, the greater proportion of the cold capacitance (input plus output) in Farads can be approximated by

$$C_0 = .0885A\left(\frac{1}{a} + \frac{1}{b} + \frac{1}{c}\right) \times 10^{-12} \qquad (10)$$

where "a" and "b" have the same meaning as in (7) and "c" is the spacing between the plate and the suppressor (or screen in the case of a tetrode). This is of course a highly idealized case. The cathode, the grids, and the plate, are each assumed to be plane conductors of infinitesimal thickness, each having an area equal to the active area of the structure.* The band merit then becomes.

$$B_0 = \frac{G_m}{2\pi C_0} = \frac{4.74\, M\, I_0^{1/3} \times 10^8}{a^{4/3}\left(\frac{1}{a} + \frac{1}{b} + \frac{1}{c}\right)\left(1 + \frac{1}{\mu_{12}}\frac{a+b}{a}\right)} \qquad (11)$$

With a given cathode current, the factor M increases as the screen current is reduced. The use of small wires for the screen grid is an important factor in obtaining minimum screen current.

I_0, the useful cathode current density, is limited by the emission capabilities of the cathode. In practice it is necessary to operate in a region considerably below the maximum available emission to avoid excessive changes in transconductance which would result from variations in cathode activity with time. Also, the shot noise will begin to rise when the region of temperature-limited operation is approached. It should be noted that B_0 is independent of the area A.

Taking reasonable values for M, I_0 and μ_{12}, ($M = 0.75$, $I_0 = 50\ ma/cm^2$, $\mu_{12} = 25$), equation (11) becomes

$$B_0 = \frac{1.31 \times 10^8}{a^{4/3}\left(\frac{1}{a} + \frac{1}{b} + \frac{1}{c}\right)\left(1 + .04\,\frac{a+b}{a}\right)} \qquad (12)$$

The curves in Figs. 6, 7 and 8 show how B_0 varies with each of the variables "a", "b" and "c" when a constant value is assigned to the other two. The band merit shown on these curves is considerably greater than that which can be realized in an actual circuit of the simple band-pass type assumed because the stray capacitances in the tube, the socket, the circuit elements, and the wiring increase the total interstage capacitance substantially. Also, the grid-cathode capacitance is substantially higher under normal operating

* This assumption is obviously not true, particularly in the case of the suppressor grid, but, by simply regarding the effective suppressor-plate spacing as somewhat greater than the actual spacing, the assumption becomes useful.

conditions, because of the presence of space charge, than when the tube is cold, as assumed in deriving equation (11). It has been assumed in making

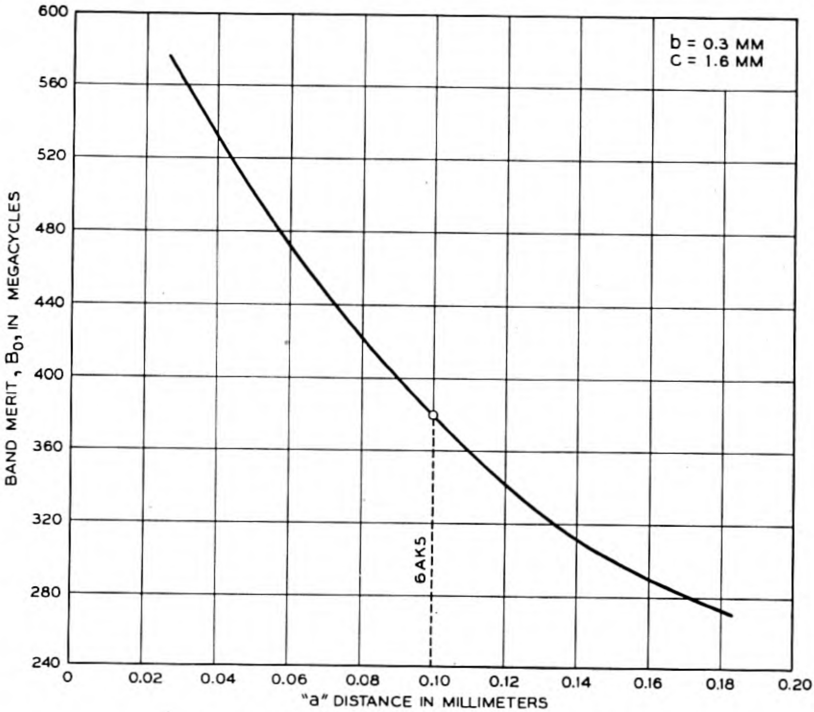


Fig. 6—Band merit vs. grid-cathode spacing.

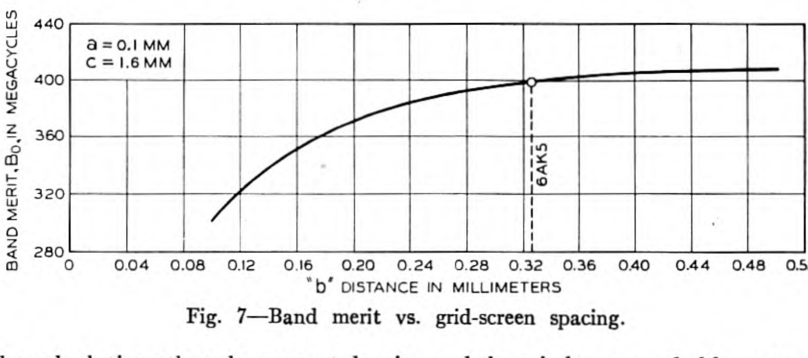


Fig. 7—Band merit vs. grid-screen spacing.

the calculations that the current density and the triode *mu* are held constant while "*a*", "*b*" and "*c*" are varied. In the cases of "*a*" and "*b*", this requires that the control grid and/or screen voltage be varied in such a fashion as to hold I_0 constant. The current density is nearly independent of "*c*" over a reasonable range provided "c" is not so large as to cause the formation of a virtual cathode in the screen-plate space.

Figure 6 shows that B_0 rises as "a" is reduced. Reducing "a" also has the advantage that a lower screen voltage is required with a given grid bias. The improvement in B_0 as "a" is reduced is quite rapid, but of course the mechanical difficulties involved in reducing "a" below about 0.10 mm become very great.

The curve in Fig. 7 shows that B_0 does not increase much if "b" is increased above about 0.30 mm, and increasing "b" has the disadvantage that higher screen voltage for a given grid bias is required.

The curve in Fig. 8 shows that B_0 increases very slowly if "c" is increased beyond about 1.50 mm, and increasing "c" has the disadvantage that the external dimensions of the structure become greater. Increasing "c" also means that the plate comes closer to any shield which is placed around the

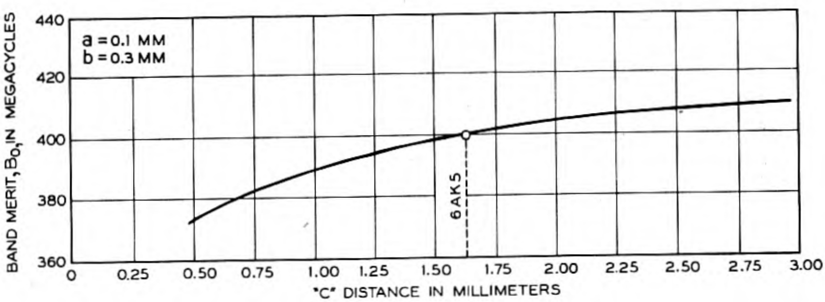


Fig. 8—Band merit vs. plate-suppressor spacing.

bulb of the tube, thus increasing the output capacitance and tending to cancel out any improvement obtained in the structure itself. Also, if "c" is made too large a virtual cathode may be formed in the screen-plate space under some conditions. This would interfere with the normal operation of the tube.

In order to take full advantage of the close grid-cathode spacing, the control-grid pitch should be no greater than about 1.5 times the spacing and the grid wires should be as small as possible. If the pitch is too large, the parts of the cathode directly opposite the grid wires will be cut off while space current is flowing from the sections opposite the spaces between grid wires. This state of affairs shows up in the characteristics as excessive variation of the triode amplification factor, μ_{12}, as the grid bias is varied, and results in a reduction of the transconductance. That is, when the grid is made more negative the amplification factor μ_{12} decreases so that the plate current does not decrease as much as it would if μ_{12} remained constant.

When the grid is made more positive the plate current rise is reduced because μ_{12} increases. The grid wires should be as small as possible so as to block off no more of the area of the structure than is necessary. An ideal grid would be an infinitesimally thin conducting plane which offered no resistance to the passage of electrons through it except that due to its electrostatic potential (sometimes called a "physicist's grid"). In the 6AK5 tube the grid-cathode clearance is 0.089 mm (0.0035 inch), the control-grid pitch is 0.0127 mm (0.0050 inch), and the wire size is 0.0010 inch. Experiments have shown that substantially higher transconductance could have been realized, with the same spacing, if smaller wires and smaller pitch had been used, but the mechanical difficulties would have been much greater.

The way to achieve a high band merit from the tube design standpoint is thus to use as close grid-cathode clearance as practicable, to operate the tube at as high a current density as the emission capabilities of the cathode will permit, and to keep the stray capacitances as low as possible. It was noted above from (11) that B_0 is independent of A. However, if it were possible to maintain the same grid-cathode clearance with a large tube as it is with a small one, the larger tube would have the advantage that the stray capacitances in the tube would be a smaller fraction of the total capacitance so that the band merit for the tube would be higher. It would also be closer to what can be realized when the tube is used in an actual circuit because the capacitances added by the socket, the wiring and the circuit elements would be less important. However, the practical mechanical limitations controlling the minimum grid-cathode clearance have been such that the band merit is roughly independent of the tube size over a moderate range of sizes of high transconductance receiving tubes.

3. Input Conductance

Two factors tend to make the input conductance of tubes higher at high frequencies than at low frequencies. One is the effect of lead inductances and the other is the effect of transit time. If the loading produced by these effects is no more than that required to get the desired bandwidth, it may be no particular disadvantage for stages other than the first one in the amplifier. As will be seen later, however, this effect in the input tube increases the noise figure. The practical result of a consideration of these effects is that the leads are made as short as possible and that small tubes are used in order to use close grid-cathode clearances when the frequency at which the tubes are to be used is above about 10 mc. The expression derived by North[3] for

[3] "Analysis of the Effects of Space Charge on Grid Impedance," D. O. North, I. R. E. Proceedings, Vol. 24, No. 1, January, 1936.

the input conductance of a tetrode or pentode can be re-written, neglecting the higher order terms, to give the approximate expression

$$G_{in} = \frac{5.0 \times 10^{-3} a^2 f^2 G_m}{V_1} \left[1 + 3.3 \frac{b}{a\left(1 + \sqrt{\dfrac{V_2}{V_1}}\right)} \right] \quad \ldots . (13)$$

where G_m is the triode-connected transconductance, a is grid-cathode spacing in cms, b is grid-screen spacing in cms, V_1 and V_2 are the effective grid-plane and screen-plane potentials in volts, and f is the frequency in mc. It is assumed that there are no lead inductances and that there is no potential minimum in the grid-cathode region. For a given transconductance, (13) shows that close spacings are necessary for minimum grid conductance.

If the lead inductances between the external circuit and the tube elements are appreciable, the input loading may be excessive even though the transit time through the tube structure is negligibly small. The general case taking account of the mutual and self inductances of all the leads of a pentode has been treated by Strutt and van der Ziel[4]. The equations are cumbersome even though only the first order terms in frequency are retained. If all of the lead inductances except that in series with the cathode are neglected, and transit time is assumed to be negligible, the input conductance of a pentode becomes approximately[5]

$$G_{in} = \omega^2 G_m L_k C_k \ldots\ldots\ldots\ldots\ldots\ldots\ldots (14)$$

where L_k is the cathode lead inductance and C_k is the grid-cathode capacitance. It is further assumed that the plate-grid capacitance is negligible.

Work done at the Naval Research Laboratory includes data on the input conductance of 6AK5 tubes in the frequency range from 100 mc to 300 mc. Through the courtesy of the Naval Research Laboratory some of the data are reproduced in Fig. 9. It is of interest to check a point on this curve against equation (14). For the 6AK5, 0.02 micro-henry is the estimated* cathode lead inductance, $G_m = 5000 \times 10^{-6}$ mhos, and $C_k = 4 \times 10^{-12}$ farad. At a frequency of 250 mc we have a calculated conductance of 990 micromhos, which checks roughly with the value of 1110 micromhos from the curve in Fig. 9. Equation (13) can be used to obtain an approximate value for the loading due to transit time. Taking $a = 0.0089$ cm., $b = 0.032$ cm., $G_m = 6.7 \times 10^{-3}$ mhos, $f = 250$ mc, $V_1 = +2.3$ volts, and $V_2 = +120$ volts, a calculation gives $G_{in} = 177$ micromhos. These results

[4] "The Causes for the Increase of the Admittance of Modern High-Frequency Amplifier Tubes," M. J. O. Strutt and A. van der Ziel, *I. R. E. Proceedings*, Vol. 26, No. 8, August 1938.

[5] "Hyper and Ultra-High Frequency Engineering," Sarbacher and Edson, p. 435.

* This has been checked roughly by Q-meter measurements.

show that the performance of the 6AK5 near the upper end of its present useful frequency range is probably limited to a considerable degree by the lead inductances rather than by transit time effects in the structure.

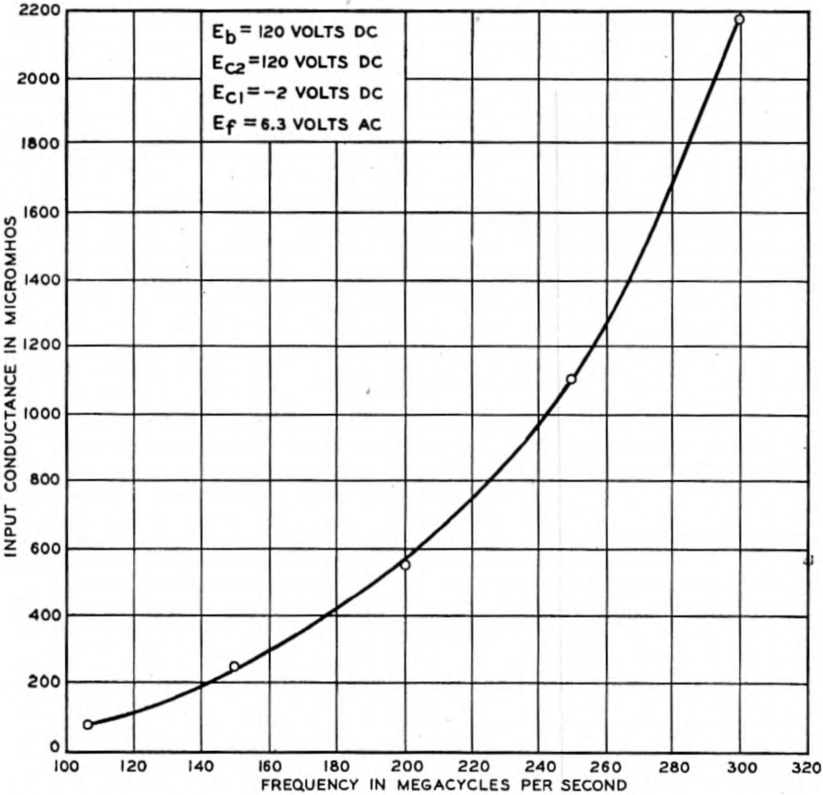


Fig. 9—Average input conductance vs. frequency for six 6AK5 tubes—courtesy of Naval Research Laboratory.

4. Noise

Although it may be possible to employ enough stages of IF amplification to provide the necessary gain and band-width we may still have a relatively insensitive receiver for weak signals. This comes about because there are inherent electrical disturbances in vacuum tubes and passive networks which give rise to random voltages. Since these disturbances may be of the order of the strength of the signal, they must be kept to a minimum in order to maintain a high signal-to-noise ratio. In a receiving system in which no RF amplification is used ahead of the first detector, the signal-to-noise ratio is limited to a large extent by the noisiness of the first detector and the first

IF tube. The noise performance of the first IF stage will be discussed in some detail.**

A convenient method of expressing the departure from ideal performance is the use of the "noise figure" proposed by Friis.[6] The noise figure of a network or amplifier may be defined as follows:

$$NF = \frac{\text{Available output noise power}}{GKT\Delta f} \qquad (15)$$

where G is the "available gain" which is defined as the ratio of the available signal power at the output terminals of the network or amplifier to the available signal power at the terminals of the signal generator. $KT\Delta f$ is the available noise power from a passive resistance, where K is Boltzmann's constant, T is absolute temperature and Δf is the incremental bandwidth. This follows from consideration of a noise generator of resistance R_0 working into a load resistance R_0. This is the condition for maximum power into the load, half of the noise voltage appearing across the source and half across the load. The open-circuit noise voltage appearing across the terminals of a resistance R_0 is

$$V^2 = 4KTR_0\Delta f \qquad (16)$$

The noise power delivered to the load by the source resistance will be

$$P_n = \frac{V^2}{4R_0} = KT\Delta f \qquad (17)$$

If there were no source of noise other than that of the resistance of the signal source itself, the noise figure would be unity. If the signal source works directly into a matched load resistance at the same temperature, the noise figure is 2 since the available output noise power is $KT\Delta f$ and the available gain is one-half.

The importance of the noise figure of the radar receiver is obvious since a reduction in the noise figure is equivalent to the same percentage increase in transmitter power. In recent radar systems the noise arising in the first IF stage constituted a substantial part of the total receiver noise.

If the first IF tube provides at least 15 db gain, noise introduced by its plate load impedance, and by anyother sources in the rest of the amplifier, is usually negligible. The departure of the noise figure of the IF amplifier from unity is then due to noise arising in the first tube and in its input circuit.

In the very high frequency (VHF) range, essentially all of the noise arising

** It will be assumed in all of the discussion about noise that there is no noise due to flicker effect, emission of positive ions, microphonics, sputter, ionization, secondary emission, or reflection of electrons from charged insulators.

[6] "Noise Figure of Radio Receivers," H. T. Friis, *I. R. E. Proceedings*, July 1944.

in the tube itself is due to random fluctuations in the emission of electrons from the cathode. For a parallel plane diode in a circuit such as that shown in Fig. 10(*a*), the mean square noise current is[7]

$$\overline{i_k^2} = 2\Gamma^2 e I_k \Delta f \qquad (18)$$

where e is the electronic charge, I_k is the d-c cathode current, Δf is the incremental bandwidth, and Γ is a factor which takes into account the

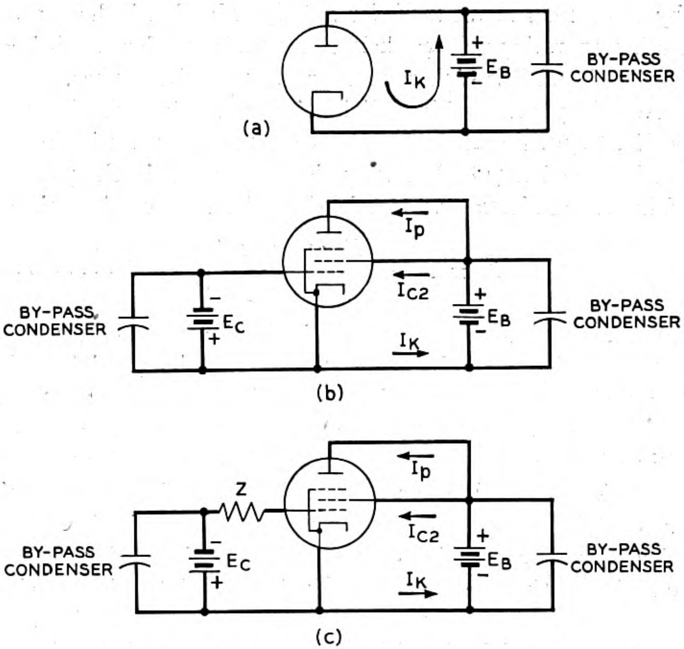


Fig. 10a—Diode, negligible circuit impedance.
Fig. 10b—Pentode, negligible circuit impedance.
Fig. 10c—Pentode, impedance in grid circuit.

"cushioning effect" of space charge. For anode-cathode spacings and operating conditions such that the transit time is not too large a faction of the period of the frequency involved, Γ is of the order of 0.20 when the zero-field emission is several times larger than I_k. Under temperature-limited conditions Γ is unity. That is, under favorable space-charge-limited conditions, the fluctuation noise component in I_k is only about 20% of the value for the same cathode current under temperature-limited conditions. Although the introduction of grids between the cathode and anode

[7] "Fluctuations in Space-Charge-Limited Currents at Moderately High Frequencies," B. J. Thompson, D. O. North, W. A. Harris, *R. C. A. Review*, October 1940, Vol. 5, p. 244.

of the simple diode complicates the noise problem, the factor Γ remains of great importance. It is therefore desirable to control the tube processing in manufacture so as to insure adequate available emission in every tube. An "activity test" is made on completed tubes for this purpose. This test consists of reducing the heater voltage by an arbitrary amount (usually 10%) and observing the change in one of the tube characteristics which is sensitive to changes in available emission. The characteristic used for tubes like the 717A and 6AK5 is the transconductance. If the transconductance decreases by more than about 20% for a 10% reduction in heater voltage, with the other operating voltages held constant, insufficient available emission is indicated and Γ is higher than for more "active" tubes.

In the case of a pentode with negligible impedance in each of its leads, as shown in Fig. 10(b), the fluctuation in the cathode current is the same as that given in equation (18), but the noise component in the plate lead is larger. Thompson, North, and Harris[7] showed that it can be written as

$$\overline{i_p^2} = 2eI_p\Delta f\left[\frac{\Gamma^2 I_p + I_{c2}}{I_k}\right] \quad (19)$$

where I_p is the d-c plate current and I_{c2} is the d-c screen current. It was mentioned above that Γ can be made as low as 0.20 by providing adequate available emission. From the design standpoint, equation (19) also shows that the screen current should be as small as possible. For normal operating conditions in a pentode, the screen current is influenced by the screening fraction (fraction of area blocked off by grid wires) of the screen grid and by the amount of space current turned back to the screen in the screen-plate region. The way to get a minimum screening fraction and still obtain the desired function of the screen grid of reducing the plate-grid capacitance is to use wire of as small diameter as possible. The presence of a suppressor grid, at cathode potential, placed between the screen and the plate to prevent interchange of secondary electrons, causes a certain proportion of the space current which would otherwise go to the plate to be turned back to the screen. Here again, this effect is minimized by using as fine wire as possible for the suppressor grid. It was pointed out in an earlier section that fine wires are desired for the control grid. The ideal for each of the three grids in an IF pentode would be a conducting plane which offers no resistance to the passage of electrons other than the influence of its potential.

When impedances are connected in the various leads of a pentode the noise components discussed above will, in general, be different due to the influence of fluctuation voltages developed between the elements of the tube. In particular it is found experimentally in the VHF range of frequencies that when an impedance Z is introduced between the grid and cathode, as shown in Fig. 10(c), the noise component in the plate lead rises more than would be expected due to thermal noise from Z, because of the effect of grid noise. North and Ferris[8] showed that the grid noise can be taken into account by assuming that the input resistance of the tube is a resistance noise source whose absolute temperature is about 4.8 times ambient, if the input loading is due to transit time effects alone. One consequence of this input loading is that at high frequencies the best signal-to-noise ratio is usually obtained with an input circuit of lower impedance than that which would be used at low frequencies.

Actually, as was brought out in an earlier section, the loading in tubes like the 6AK5 is probably due largely to lead inductances between the active tube elements and the external circuit components up to a few hundred megacycles. According to Pierce[9] the effect of the cathode lead inductance feedback is to reduce the signal component in the output current while leaving the noise current due to screen interception noise unaffected. Input loading may be a limiting factor in tubes like the 717A and 6AK5 when the frequency is of the order of 100 mc or higher, both because of its effect on gain in some cases and because of its adverse effect on the signal-to-noise ratio for early stage use. There is good evidence that the 6AK5 structure would be useful at much higher frequencies than is the case at present if the circuit connections to the tube elements were improved by more advantageous mounting of the structure and the use of more suitable sockets or external connectors.

Noise measurements made by a number of workers at Bell Telephone Laboratories[10] indicate that an average noise figure of about 2.8 can be obtained with the 6AK5 at a midband frequency of 60 mc, with a well-designed input circuit, and bandwidths up to 10 mc. At a mid-band frequency of 30 mc the noise figure is about 2.4. At 100 mc it is about 3.6.

5. Description of the Design of the Western Electric 6AK5 Tube

In order that the reader may have a full appreciation of the dimensions and other requirements of design to meet the characteristics discussed above, a detailed description of the 6AK5 tube follows.

5.1 *Mechanical Description*

The 6AK5 tube is an indirectly heated cathode type pentode employing the 7-pin button stem and the T-5-1/2 size miniature bulb. The outline dimensions are shown in Fig. 11. A photograph of a mount ready to be sealed into a bulb is shown in Fig. 12. Figure 13 is a photograph of a transverse section through the tube at the middle of the structure in a plane

[8] "Fluctuations Induced in Vacuum Tube Grids at High Frequencies," D. O. North and W. R. Ferris, *I. R. E. Proceedings*, Vol. 29, No. 2, February 1941.

[9] Unpublished Technical Memorandum, J. R. Pierce.

[10] S. E. Miller, V. C. Rideout, R. S. Julian.

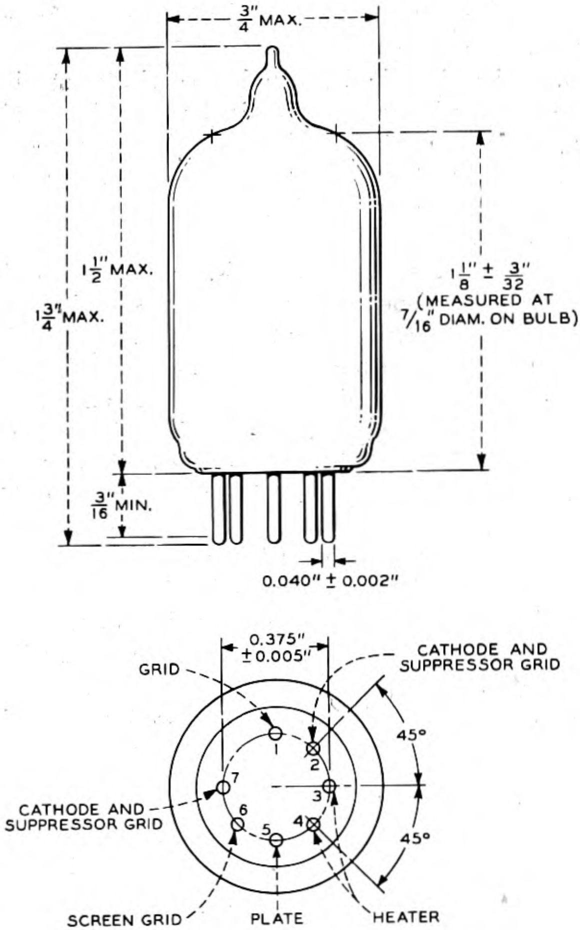


Fig. 11—6AK5 outline dimensions.

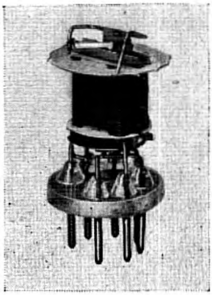


Fig. 12—6AK5 mount structure (full size).

parallel to the stem or "base" of the tube. The magnification in Fig. 13 as reproduced is about 5½ times.

The heater is a conventional folded type with eight legs of coated tungsten wire. The wire diameter is 0.0014 inches and the unfolded length is 3 inches. The insulating coating consists of fired aluminum oxide and is about 0.0025 inch thick.

The cathode is of oval cross-section with the contours of the longer sides shaped to conform to the shape of the control grid. The major axis of the

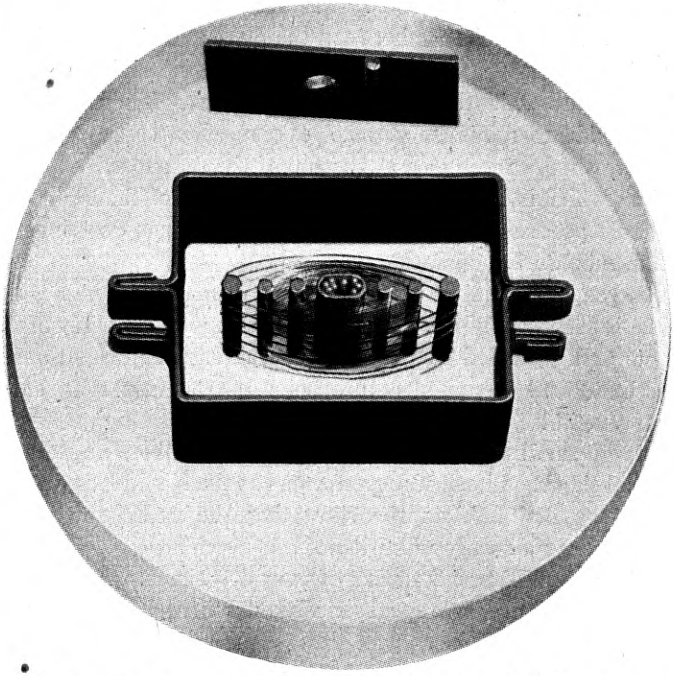


Fig. 13—Transverse section of 6AK5 tube (5½ times).

cross-section of the cathode sleeve is 0.048 inch. The minor axis is 0.025 inch. The length of the sleeve is 0.47 inch and it is coated over a centralized section which extends 0.28 inch along its length. The coating is the usual mixture of oxides of barium, calcium, and strontium. Before the tube is processed, the coating thickness is about 0.002 inch. After processing, it is about 0.001 inch thick.

The grids are oval-shaped and are wound with small diameter wires. The control-grid and the screen-grid lateral wires are 0.001 inch diameter tungsten. The suppressor-grid lateral wires are 0.002 inch diameter molybdenum. The control-grid wires are gold-plated in order to minimize primary emission of electrons from this grid.

The plate has a rectangular transverse cross-section and is made of 0.005 inch thickness carbonized nickel. The carbonization increases the thermal emissivity of the plate surface so that a reasonable amount of power dissipation in the plate can be tolerated.

The end shields are made of nickel-plated iron. The reason for using this material instead of nickel, as is more often the case, is to prevent overheating of these shields during the exhaust process. At the temperatures used, the iron shields pick up less energy in the induction field during the out-gassing of the plate than do nickel shields. The function of the end shields is to minimize the stray capacitance between the plate and the control grid.

The insulators or spacers which hold the tube elements in proper disposition with respect to each other are of high grade mica. They are coated with magnesium oxide to minimize surface leakage effects.

The getter, which can be seen at the top of the structure in Fig. 12, contains barium which is flashed onto the inside surface of the bulb at the end of the exhaust process in order to take up residual gas evolving from the parts of the tube during operation.

Although the individual parts are extremely small and fragile, the completed tube is surprisingly rugged. The short supporting wires in the stem and the support provided by the bulb-contacting top insulator result in the stem, bulb, and mount structure being a relatively rigid unit. The small parts assembled into the mount are very light in weight and therefore exert relatively small forces on their supporting members under conditions of mechanical shock. Shock tests performed at the Naval Research Laboratory and at the Bell Laboratories show that the 6AK5 tube stands up satisfactorily under a steady vibration of rms acceleration 2.5 times gravity and withstands 1 millisecond shocks of over 300 times gravity.

The most important single geometrical factor in the tube is the spacing between the cathode and the control grid. In the 6AK5 tube this clearance is 0.0035 inch after processing. Before processing it is 0.0025 inch. The manufacturing difficulties involved in assembling the structure and maintaining such small clearances are obviously very great. However, as was seen from the discussion above, this close spacing is essential in order to obtain the desired high-frequency performance.

It can be observed that the close mounting of the structure on the stem provides very short lead lengths between the tube elements and the external pins. This is of importance at frequencies where the inductances of the lead wires become comparable in magnitude to the circuit reactances.

5.2 *Low Frequency Electrical Charatceristics*

The usual static characteristics are shown by the curves in Fig. 14. The ratings, nominal characteristics, and cold capacitances are given in Table II

5.3 *Fixed Tuning*

It is highly desirable to design the IF amplifier with fixed tuning in order to minimize the number of adjustments that need to be made when tubes

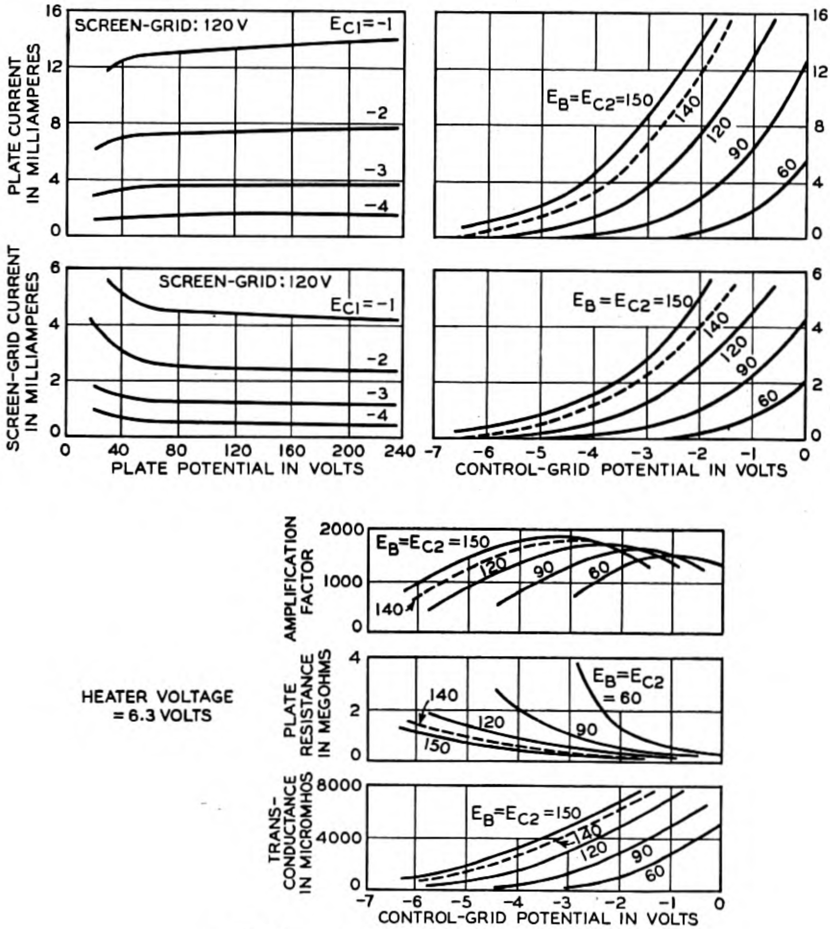


Fig. 14—Average 6AK5 characteristics.

are changed. The difficulty immediately arises that, since the tube capacitances are usually a large fraction of the total shunt capacitance of each coupling network, the variations in capacitance from one tube to another become very important. It is essential that, after the amplifier is lined up at the factory with standard tubes, any stock tubes taken at random shall give satisfactory performance in the amplifier. This requires close control of the capacitances. Not only must the capacitances be held within the limits of the test specification, but the product averages must be kept close to the design center values particularly in the cases of the input and output capacitances.

TABLE II

HEATER RATING			
Heater voltage			6.3 volts, a-c or d-c
Nominal heater current			0.175 ampere
Maximum Ratings (Design-center values)			
Maximum plate voltage			180 volts
Maximum screen voltage			140 volts
Maximum plate dissipation			1.7 watts
Maximum screen dissipation			0.5 watt
Maximum cathode current			18 milliamperes
Maximum heater-cathode voltage			90 volts
Maximum bulb temperature			120°C
OPERATING CONDITIONS AND CHARACTERISTICS			
Plate voltage	120	150	180 volts
Screen voltage	120	140	120 volts
Cathode-bias resistor	200	330	200 ohms
Plate current	7.5	7.0	7.7 milliamperes
Screen current	2.5	2.2	2.4 milliamperes
Amplification factor	1700	1800	3500
Plate resistance	0.34	0.42	0.69 megohms
Transconductance	5000	4300	5100 micromhos
INTERELECTRODE CAPACITANCES (*With JAN 1A No. 314 shield connected to cathode*)			
Control grid to heater, cathode, screen grid and suppressor grid			3.90 $\mu\mu$f
Plate to control grid			0.01 $\mu\mu$f
Plate to heater, cathode, screen grid and suppressor grid			2.85 $\mu\mu$f

6. CONCLUSION

The important factors to be considered in evaluating the suitability of a vacuum tube for broad band IF use in the VHF range are as follows:

1. Band merit
2. Noise figure
3. Input conductance
4. Power consumption
5. Physical size
6. Control of capacitances

We have seen that the tube design features which make a tube good on the basis of these requirements are close grid-cathode spacing, fine grid wires, short lead wires, and small elements. An important consideration from the manufacturing standpoint is the control of cathode emission for low noise. It has been possible to extend the useful frequency range of conventional receiving tubes up through the VHF range and somewhat higher by this line of attack.

The 6AK5 tube is an outgrowth of the many years of development in this general field by the Electronics group of the Bell Telephone Laboratories. Contributions to the success of the development have been made by chemists, physicists, mechanical engineers and electrical engineers too numerous to mention individually and to whom the author is indebted for much of the material presented.

High Q Resonant Cavities for Microwave Testing

By I. G. WILSON, C. W. SCHRAMM and J. P. KINZER

Formulas and charts are given which aid the design of right circular cylinder cavity resonators operating in the TE 01n mode, which yields the highest Q for a given volume. The application of these to the design of an echo box radar test set is shown, and practical considerations arising in the construction of a tunable cavity are discussed.

Introduction

A TUNABLE high Q resonant cavity is a particularly useful tool for determining the over-all performance of a radar quickly and easily[1].* Further, since it uses the radar transmitter as its only source of power, it can be made quite portable. When a high Q cavity is provided with two couplings, one for the radar pickup and the other to an attenuating device, crystal rectifier and meter which serve for tuning the cavity not only can an indication of over-all performance be obtained but other useful information as well. For example, the transmitter frequency can be measured; calibration of the crystal affords a rough measure of the transmitter power; and an analysis of the spectrum can be made by plotting frequency versus crystal current. This information is of particular importance in radar maintenance.

The Q required for this purpose is quite high, comparable to that obtained from quartz crystals in the video range. For this reason, such cavities have many additional possibilities for use in microwave testing equipment and microwave systems. For example, they may form component parts of a narrow band filter, or be used as discriminators for an oscillator frequency control.

Resonant cavities are of two general types—tuned and untuned. A tuned cavity is designed to resonate in a single mode adjustable over the radar frequency range. An untuned cavity is of a size sufficient to support a very large number of modes within the working range. Both are useful, but the tuned variety can give more information about the radar and hence has been more widely used.

While a tuned cavity may be a cylinder, parallelepiped or sphere (or even other shapes), the first of these has been most thoroughly explored by us. It offers the possibility of utilizing the anomalous circular mode, described by Southworth[2] in his work on wave guides, which permits the attainment of high Q's in quite a small size. In addition, it is easier to construct a variable length cylinder than a variable sized sphere.

* Superscripts refer to bibliography.

Due to their interesting properties the history of resonators of the cavity type in which a dielectric space is enclosed by a conducting material, goes back many years. In 1893 J. J. Thompson[3] derived expressions for resonant frequencies of the transverse electric modes in a cylinder. Lord Rayleigh[4] published a paper in 1897 dealing with such resonant modes. The early work was almost entirely theoretical but some experiments were carried out in 1902 by Becker[5] at 5 and 10 centimeters. In recent years, the subject has been fairly thoroughly investigated (at least theoretically) for several simple shapes.

However, many of the presentations are highly mathematical with considerable space devoted to proofs; the results which would be most useful to an engineer are thus sometimes obscured. The purpose of this paper is to present certain engineering results together with information upon the application of the tunable cylindrical cavity to radar testing.

Definitions and Fundamental Formulas

Modes

By fundamental and general considerations, every cavity resonator, regardless of its shape, has a series of resonant frequencies, infinite in number and more closely spaced as the frequency increases. The total number N of these having a resonant frequency less than f is given approximately by:[6]

$$N = \frac{8\pi}{3c^3} V f^3 \tag{1}$$

in which

V = volume of cavity in cubic meters.

c = velocity of electromagnetic waves in the dielectric in meters per second.

f = frequency in cycles per second.

With each resonance there is associated a particular standing wave pattern of the electromagnetic fields, which is identified by the term "mode."

In right cylinders (ends perpendicular to axis) the modes fall naturally into two groups, the transverse electric (TE) and the transverse magnetic (TM). In the TE modes, the electric lines everywhere lie in planes perpendicular to the cylinder axis, and in the TM modes, the magnetic lines so lie. Further identification of a specific mode is accomplished by the use of indices.

The MS Factor

With the cylinder further restricted to a loss-free dielectric and a nonmagnetic surface, there is associated with each mode a value of Q (quality factor)[7] which depends on the conductivity of the metallic surface, on the frequency and on the shape of the cylinder, e.g. whether it is circular or elliptical, and whether it is slender or stubby. The quantity $\frac{Q\delta}{\lambda}$, however, depends only upon the mode and shape of the cylinder and has been referred to as the mode-shape (MS) factor. In this formula, δ refers to skin depth as customarily defined[8], and λ is wavelength in the dielectric, as given by $\lambda = \frac{c}{f}$; both δ and λ are in meters.

Fundamental Formulas

Expressions for standing wave patterns and $Q\frac{\delta}{\lambda}$ are given in Table I, for right rectangular, circular and full coaxial cylinders*. The table is virtually self-explanatory, but a few remarks on mode designation are needed. The mode indices are l, m, n following the notation of Barrow and Mieher.[9] In the rectangular prism they denote the number of half-wavelengths along the coordinate axes. For the other two cases they have an analagous physical significance with l related to the angular coordinate, m to the radial and n to the axial.

In the elliptical cylinder, a further index is needed to distinguish between modes which differ only in their orientation with respect to the major and minor axes; these paired modes are termed even and odd, and have slightly different resonant frequencies.[10] In the circular cylinder they have the same frequency, a condition which is referred to as a degeneracy (in this case, double); that is, in the circular cylinder, odd and even modes are distinguishable only by a difference in their orientation within the cylinder with reference to the origin of the angular coordinate. In Table I, the field expressions are given for the even modes; those for the odd modes are obtained by changing $\cos l\theta$ to $\sin l\theta$ and $\sin l\theta$ to $\cos l\theta$ everywhere.

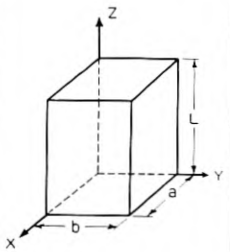
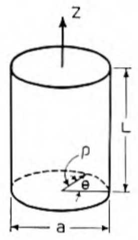
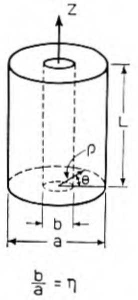
The value of N in the table is based on counting this degeneracy as a single mode; counting even and odd modes as distinct will nearly double the value of N, thus bringing it into agreement with the general equation (1). The distinction between even and odd modes is of limited importance in practical applications, and will not be further mentioned.

In Table I, the *mks* system of units is implied. The notation is in general accordance with that used in prior developments of the subject. For engineering applications, it is advantageous to reduce the results to units in ordinary use and to change the notation wherever this leads to a more obvious association of ideas. For these reasons, in what follows attention

* The elliptic cylinder (closely allied to the circular cylinder of which it is a generalization) is omitted as the necessary functions are not widely known or easily available.

This table belongs between pages 410 and 411 of the article HIGH Q RESONANT CAVITIES FOR MICROWAVE TESTING by I. G. Wilson, C. W. Schramm and J. P. Kinzer in the Bell System Technical Journal, July, 1946. Through an error, it was omitted from the printed journal.

TABLE 1.—*Formulas for Cavity Resonators—Fields, Resonant Frequencies and Mode Shape Factors for Rectangular Prism, Circular Cylinder and Full Coaxial*

TYPE OF CAVITY & CO-ORDINATE SYSTEM	MODE	FIELD EQUATIONS *	DEFINITIONS	RESTRICTIONS ON ℓ, m, n
RECTANGULAR PRISM 	TM	$E_x = \sqrt{\frac{\mu}{\epsilon}}\,\frac{k_1k_3}{k^2}\cos k_1x\ \sin k_2y\ \sin k_3z$ $E_y = \sqrt{\frac{\mu}{\epsilon}}\,\frac{k_2k_3}{k^2}\sin k_1x\ \cos k_2y\ \sin k_3z$ $E_z = -\sqrt{\frac{\mu}{\epsilon}}\,\frac{k_1^2+k_2^2}{k^2}\sin k_1x\ \sin k_2y\ \cos k_3z$ $H_x = -\frac{k_2}{k}\sin k_1x\ \cos k_2y\ \cos k_3z$ $H_y = \frac{k_1}{k}\cos k_1x\ \sin k_2y\ \cos k_3z$ $H_z = 0$	$k_1 = \frac{\ell\pi}{a}$ $k_2 = \frac{m\pi}{b}$ $k_3 = \frac{n\pi}{L}$ $k^2 = k_1^2+k_2^2+k_3^2$ $\lambda = \frac{2\pi}{k}$ ℓ, m, n = INTEGRAL INDICES IDENTIFYING THE MODES. MAY ASSUME THE VALUE ZERO, SUBJECT TO RESTRICTIONS GIVEN IN ADJOINING COLUMN	$\ell > 0$ $m > 0$
	TE	$E_x = -\sqrt{\frac{\mu}{\epsilon}}\,\frac{k_2}{k}\cos k_1x\ \sin k_2y\ \sin k_3z$ $E_y = \sqrt{\frac{\mu}{\epsilon}}\,\frac{k_1}{k}\sin k_1x\ \cos k_2y\ \sin k_3z$ $E_z = 0$ $H_x = \frac{k_1k_3}{k^2}\sin k_1x\ \cos k_2y\ \cos k_3z$ $H_y = \frac{k_2k_3}{k^2}\cos k_1x\ \sin k_2y\ \cos k_3z$ $H_z = -\frac{k_1^2+k_2^2}{k^2}\cos k_1x\ \cos k_2y\ \sin k_3z$		$\ell + m > 0$ $n > 0$
CIRCULAR CYLINDER 	TM	$E_\rho = -\sqrt{\frac{\mu}{\epsilon}}\,\frac{k_3}{k}J_\ell'(k_1\rho)\cos\ell\theta\ \sin k_3z$ $E_\theta = \sqrt{\frac{\mu}{\epsilon}}\,\ell\,\frac{k_3}{k}\,\frac{J_\ell(k_1\rho)}{k_1\rho}\sin\ell\theta\ \sin k_3z$ $E_z = \sqrt{\frac{\mu}{\epsilon}}\,\frac{k_1}{k}J_\ell(k_1\rho)\cos\ell\theta\ \cos k_3z$ $H_\rho = -\ell\,\frac{J_\ell(k_1\rho)}{(k_1\rho)}\sin\ell\theta\ \cos k_3z$ $H_\theta = -J_\ell'(k_1\rho)\cos\ell\theta\ \cos k_3z$ $H_z = 0$	$k_1 = \frac{2r_{\ell m}}{a}$ $k_3 = \frac{n\pi}{L}$ $k^2 = k_1^2+k_3^2$ $\lambda = \frac{2\pi}{k}$ ℓ, m, n = DEFINED AS FOR RECTANGULAR PRISM $r_{\ell m}$ = m^{th} ZERO OF $J_\ell(x)$ FOR TM MODES $r_{\ell m}$ = m^{th} ZERO OF $J_\ell'(x)$ FOR TE MODES	$m > 0$
	TE	$E_\rho = -\sqrt{\frac{\mu}{\epsilon}}\,\ell\,\frac{J_\ell(k_1\rho)}{k_1\rho}\sin\ell\theta\ \sin k_3z$ $E_\theta = -\sqrt{\frac{\mu}{\epsilon}}\,J_\ell'(k_1\rho)\cos\ell\theta\ \sin k_3z$ $E_z = 0$ $H_\rho = \frac{k_3}{k}J_\ell'(k_1\rho)\cos\ell\theta\ \cos k_3z$ $H_\theta = -\ell\,\frac{k_3}{k}\,\frac{J_\ell(k_1\rho)}{k_1\rho}\sin\ell\theta\ \cos k_3z$ $H_z = \frac{k_1}{k}J_\ell(k_1\rho)\cos\ell\theta\ \sin k_3z$		$m > 0$ $n > 0$
FULL COAXIAL  $\frac{b}{a} = \eta$	TM	SAME AS FOR CIRCULAR CYLINDER, BUT SUBSTITUTE: $Z_\ell(k_1\rho)$ FOR $J_\ell(k_1\rho)$ $Z_\ell'(k_1\rho)$ FOR $J_\ell'(k_1\rho)$	SAME AS CIRCULAR CYLINDER, EXCEPT: $r_{\ell m}$ = m^{th} ZERO OF $[J_\ell(\eta x)\,Y_\ell(x) - J_\ell(x)\,Y_\ell(\eta x)]$ $A = \frac{J_\ell(r_{\ell m})}{Y_\ell(r_{\ell m})}$ FOR TM MODES	$m > 0$ SPECIAL CASE OF TM 0,0,n MODE, WITH $r_{\ell m} = 0$
	TE	WHERE $Z_\ell(k_1\rho) = J_\ell(k_1\rho) - AY_\ell(k_1\rho)$ $Z_\ell'(k_1\rho) = J_\ell'(k_1\rho) - AY_\ell'(k_1\rho)$	$r_{\ell m}$ = m^{th} ZERO OF $[J_\ell'(\eta x)\,Y_\ell'(x) - J_\ell'(x)\,Y_\ell'(\eta x)]$ $A = \frac{J_\ell'(r_{\ell m})}{Y_\ell'(r_{\ell m})}$ FOR TE MODES	$m > 0$ $n > 0$

SOURCES: HANSEN, JNL. APP. PHYS., 9, P. 654 BORGNIS, HOCHF: TECH u. ELEK: AKUS., 56, P. 47

* THE TIME FACTOR HAS BEEN OMITTED;

CAVITY	MODE	NORMAL WAVELENGTHS	APPROXIMATION FOR TOTAL NUMBER OF MODES (TE & TM) HAVING $\lambda > \lambda_0$	FORMULAS FOR $Q\frac{\delta}{\lambda}$		DEFINITIONS
RECTANGULAR PRISM	TM	$\lambda = \frac{2}{\sqrt{\left(\frac{\ell}{a}\right)^2+\left(\frac{m}{b}\right)^2+\left(\frac{n}{L}\right)^2}}$	$N = 8.38\frac{V}{\lambda_0^3} - \frac{P}{\lambda_0}$ $V = abL$ $P = a+b+L$	$\frac{abL}{4}\cdot\frac{(p^2+q^2)(p^2+q^2+r^2)^{\frac{1}{2}}}{p^2b(a+L)+q^2a(b+L)}$	$n > 0$	$p = \frac{\ell}{a}$ $q = \frac{m}{b}$ $r = \frac{n}{L}$
				$\frac{abL}{2}\cdot\frac{(p^2+q^2)^{\frac{3}{2}}}{p^2b(a+2L)+q^2a(b+2L)}$	$n = 0$	
	TE	SAME AS TM MODES		$\frac{abL}{4}\cdot\frac{(p^2+q^2)(p^2+q^2+r^2)^{\frac{3}{2}}}{aL[p^2r^2+(p^2+q^2)^2]+bL[q^2r^2+(p^2+q^2)^2]+abr^2(p^2+q^2)}$	$\ell m > 0$	
				$\frac{abL}{2}\cdot\frac{(q^2+r^2)^{\frac{3}{2}}}{q^2L(b+2a)+r^2b(L+2a)}$	$\ell = 0$	
				$\frac{abL}{2}\cdot\frac{(p^2+r^2)^{\frac{3}{2}}}{p^2L(a+2b)+r^2a(L+2b)}$	$m = 0$	
CIRCULAR CYLINDER	TM	$\lambda = \frac{2}{\sqrt{\left(\frac{2r_{\ell m}}{\pi a}\right)^2+\left(\frac{n}{L}\right)^2}}$ $(fa)^2 = \left(\frac{cr_{\ell m}}{\pi}\right)^2+\left(\frac{cn}{2}\right)^2\left(\frac{a}{L}\right)^2$ $c = \sqrt{\mu\epsilon}$ = VELOCITY OF ELECTROMAGNETIC WAVES IN DIELECTRIC f = FREQUENCY	$N = 4.38\frac{V}{\lambda_0^3}+0.09\frac{S}{\lambda_0}$ $V = \frac{\pi a^2L}{4}$ $S = \pi aL$	$\frac{r_{\ell m}}{2\pi}\left[1+p^2R^2\right]^{\frac{1}{2}}\cdot\frac{1}{1+R}$	$n > 0$	$R = \frac{a}{L}$ $p = \frac{n\pi}{2r_{\ell m}}$
				$\frac{r_{\ell m}}{\pi}\cdot\frac{1}{2+R}$	$n = 0$	
	TE			$\frac{r_{\ell m}}{2\pi}\left[1+p^2R^2\right]^{\frac{3}{2}}\frac{1-\left(\frac{\ell}{r_{\ell m}}\right)^2}{1+p^2R^3+p^2(1-R)R^2\left(\frac{\ell}{r_{\ell m}}\right)^2}$		
FULL COAXIAL	TM	SAME FORM AS FOR CYLINDER $r_{\ell m}$ HAS DIFFERENT VALUES	$N \approx 4.4\frac{V}{\lambda_0^3}$ WITH SOME DOUBT AS TO VALUE OF THE COEFFICIENT	$\frac{r_{\ell m}}{2\pi}\left[1+p^2R^2\right]^{\frac{1}{2}}\frac{(1-\eta^2H')}{(1+\eta H')+R(1-\eta^2H')}$	$n > 0$	$R = \frac{a}{L}$ $p = \frac{n\pi}{2r_{\ell m}}$ $H' = \left[\frac{Z_\ell'(\eta r_{\ell m})}{Z_\ell'(r_{\ell m})}\right]^2$ $H = \left[\frac{Z_\ell(\eta r_{\ell m})}{Z_\ell(r_{\ell m})}\right]^2$
				$\frac{r_{\ell m}}{\pi}\cdot\frac{(1-\eta^2H')}{2(1+\eta H')+R(1-\eta^2H')}$ THESE EXPRESSIONS ARE NOT VALID FOR SMALL η WHEN $\ell = 0$	$n = 0$	
	TE			$\frac{r_{\ell m}}{2\pi}\cdot\frac{\left[1+p^2R^2\right]^{\frac{3}{2}}M}{(1+\eta H)+p^2R^2\frac{\ell^2}{r_{\ell m}^2}\left(1+\frac{H}{\eta}\right)+p^2R^3M}$ WHERE $M = \left(1-\frac{\ell^2}{r_{\ell m}^2}\right)-\eta^2H\left(1-\frac{\ell^2}{\eta^2r_{\ell m}^2}\right)$		

is confined to the circular cylinder, with changes in units and notation as specified later.

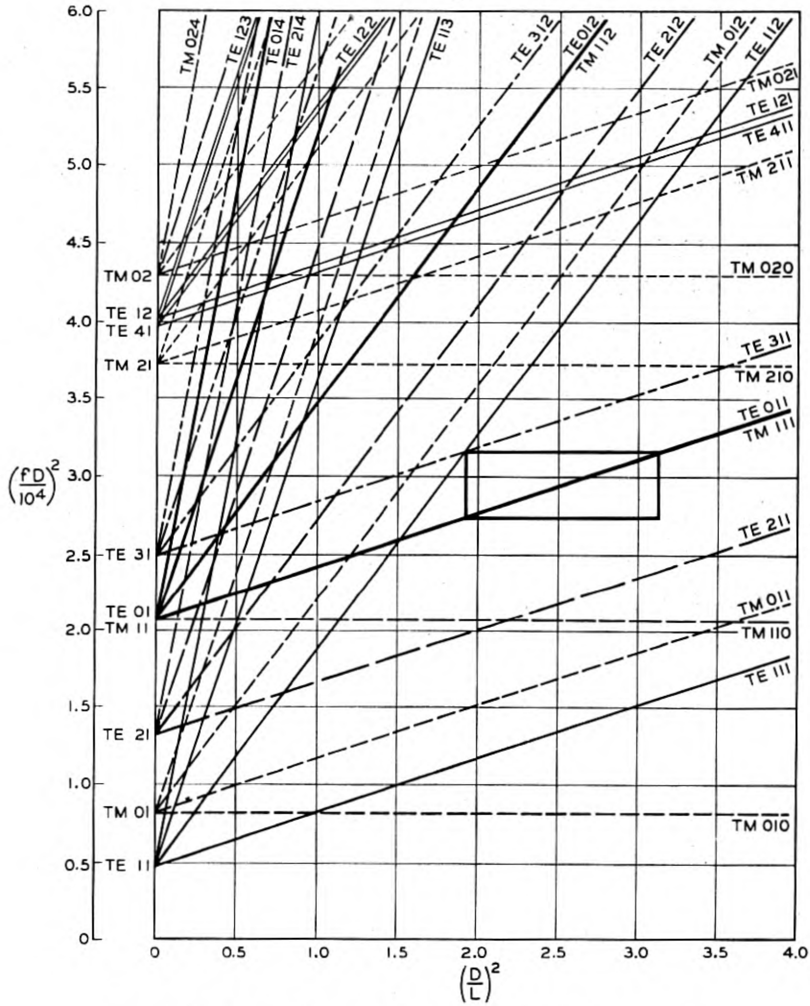


Fig. 1—Mode chart for right circular cylinder resonant cavity.

The Mode Chart

The formula relating the resonant frequency to the mode, shape and dimensions of the cylinder is of prime interest. It may be simply written as

$$(fD)^2 = A + Bn^2 \left(\frac{D}{L}\right)^2 \tag{2}$$

where

f = frequency in megacycles per second.
D = diameter of cavity in inches.
L = length of cavity in inches.
A = a constant depending upon the mode. Values of A are given in Table II for the lowest 30 modes. Values of Bessel function roots are given in Table III for the first 180 modes.
B = a constant depending upon the velocity of electromagnetic waves in the dielectric. For air at 25°C and 60% relative humidity, $B = 0.34799 \times 10^8$.
n = third index defining the mode, i.e., the number of half wavelengths along the cylinder axis.

TABLE II.—*Constants for Use in Computing the Resonant Frequencies of Circular Cylinders*

$$(fD)^2 = \left(\frac{cr}{\pi}\right)^2 + \left(\frac{cn}{2}\right)^2 \left(\frac{D}{L}\right)^2 = A + B\,n^2 \left(\frac{D}{L}\right)^2$$

$B = 0.34799 \times 10^8$ $\qquad c = 1.17981 \times 10^{10}$ inches/second

Mode	r	A
TM 01	2.40483	0.81563×10^8
02	5.52008	4.2975
03	8.65373	10.5617
11	3.83171	2.0707
12	7.01559	6.9415
13	10.17347	14.5970
21	5.13562	3.7197
22	8.41724	9.9923
31	6.38016	5.7410
32	9.76102	13.4374
41	7.58834	8.1212
51	8.77148	10.8511
61	9.93611	13.9238
TE 01	3.83171	2.0707
02	7.01559	6.9415
03	10.17347	14.5970
11	1.84118	0.47810
12	5.33144	4.0088
13	8.53632	10.2770
21	3.05424	1.3156
22	6.70613	6.3426
23	9.96947	14.0175
31	4.20119	2.4893
32	8.01524	9.0606
41	5.31755	3.9879
42	9.28240	12.1520
51	6.41562	5.8050
61	7.50127	7.9359
71	8.57784	10.3772
81	9.64742	13.1265

Value of c is for air at 25°C. and 60% relative humidity. D and L in inches; f in megacycles.

Formula (2) represents a family of straight lines, when $(D/L)^2$ and $(fD)^2$ are used as coordinates, and leads directly to the easily constructed and highly useful "Mode Chart" of Fig. 1.

It will be noted from Table II that the *TE* 0*mn* and the *TM* 1*mn* modes have the same frequency of resonance. This is a highly important case of degeneracy. In the design of practical cavities it is necessary to take measures to eliminate this degeneracy, as the *TM* mode (usually referred to as the companion of its associated *TE* mode) introduces undesirable effects. This is discussed more at length later.

Choice of Operating Mode

Turning now to the expressions for $Q\frac{\delta}{\lambda}$ these are seen to be of a rather complicated nature. For some of the lower order modes, their values are plotted in Figs. 2, 3 and 4. Examination of these leads to the question of which mode has the highest Q for a given volume. It is desirable to keep the volume a minimum, since, as shown by (1), the total number of resonances is a function of the volume. Analysis of the problem is somewhat involved, but leads to the conclusion that operation in the *TE* 01*n* mode* gives the smallest volume for an assigned Q, and also leads to specific values of n and D/L which give this result. In fact, for maximum Q/V in the *TE* 01*n* mode,

$$(fD)^2\left(\frac{D}{L}\right) \doteq 3.11 \times 10^8 \tag{3}$$

which permits easy plotting on a mode chart of the locus of the operating points for best Q/V ratio.

Extraneous or Unwanted Modes

In echo boxes for radar testing, where high Q is of the utmost importance, the *TE* 01*n* mode has been used. The values of n vary from 1 at frequencies around 1 *kmc* to 50 at about 25 *kmc*.

All other modes are then regarded as unwanted or extraneous. The great utility of the mode chart lies in that it permits a quick determination of the most favorable operating area. We consider this now in detail.

Figure 5 shows a portion of a mode chart. It is clear that the sensible way to construct a tunable cylinder is to keep the diameter fixed and vary the length. With fixed diameter, the coordinates of the mode chart are essentially f^2 and $\left(\frac{1}{L}\right)^2$ and it is convenient to refer to them loosely as fre-

* Unimportant exceptions occur for values of $Q\frac{\delta}{\lambda} < 1.2$.

quency and length. With this understanding, if the frequency band to be covered by the tunable cavity extends from f_1 to f_2 then the length must be adjustable from L_1 to L_2. Responses to frequencies within the band, but

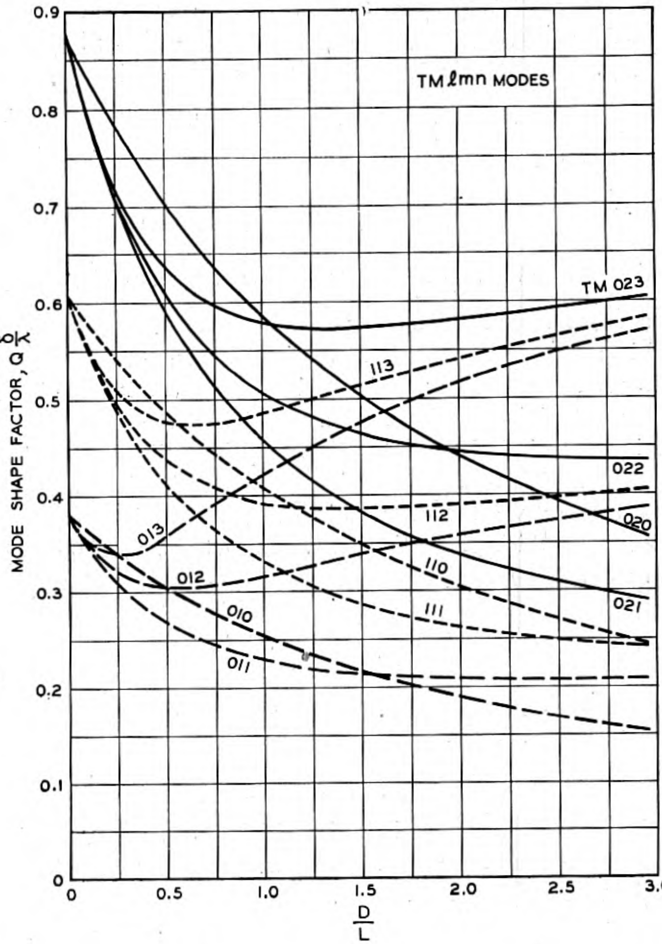


Fig. 2—Mode-shape factors $\left(Q\frac{\delta}{\lambda}\right)$ as a function of diameter to length ratio $\left(\frac{D}{L}\right)$ for circular cylinder resonator—*TM* modes.

which demand lengths outside the range L_1 to L_2 are of little interest because mechanical stops prohibit other lengths. On the assumption that the applied frequency will always lie within the operating band f_1 to f_2, responses to frequencies below f_1 or above f_2 are likewise of little interest. Therefore, major consideration need be given only to those modes which lie within the rectangle of which the desired TE 01n mode forms the diagonal.

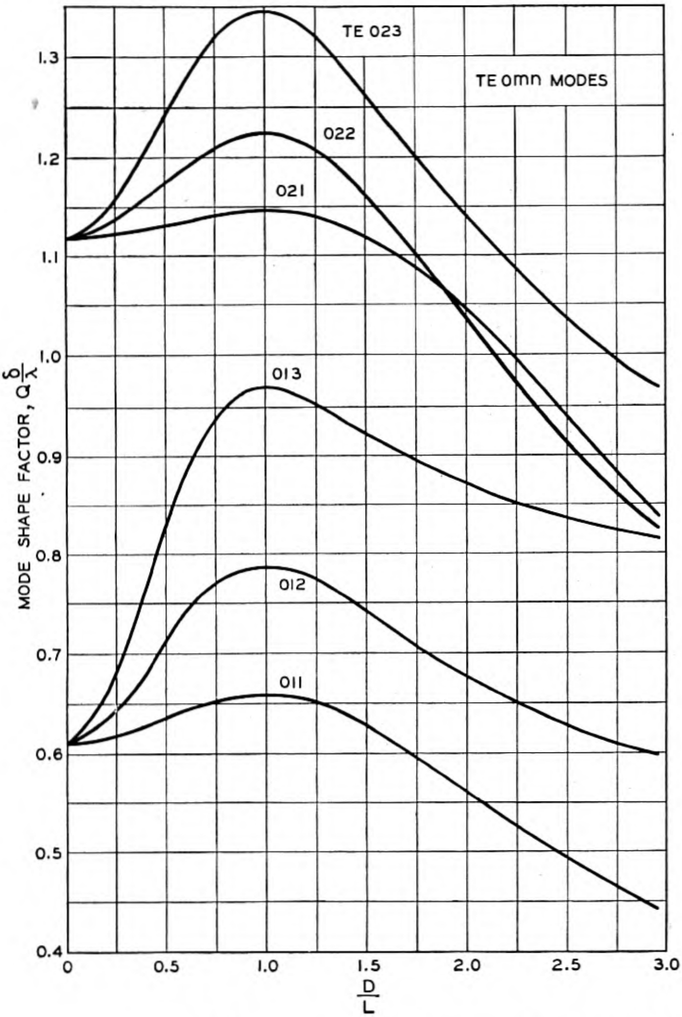


Fig. 3—Mode-shape factors $\left(Q\frac{\delta}{\lambda}\right)$ as a function of diameter to length ratio $\left(\frac{D}{L}\right)$ for circular cylinder resonator — TE modes with $l = 0$.

Any nondiagonal mode is an extraneous mode. Those which do not cross the desired mode within the frame are called interfering modes. They act to give responses at more than one tuning point when the applied frequency is held fixed, or alternatively to give responses at more than one frequency when the tuning is held fixed. In either event they lead to ambiguity and confusion, and their effects must be reduced to the point where this cannot occur.

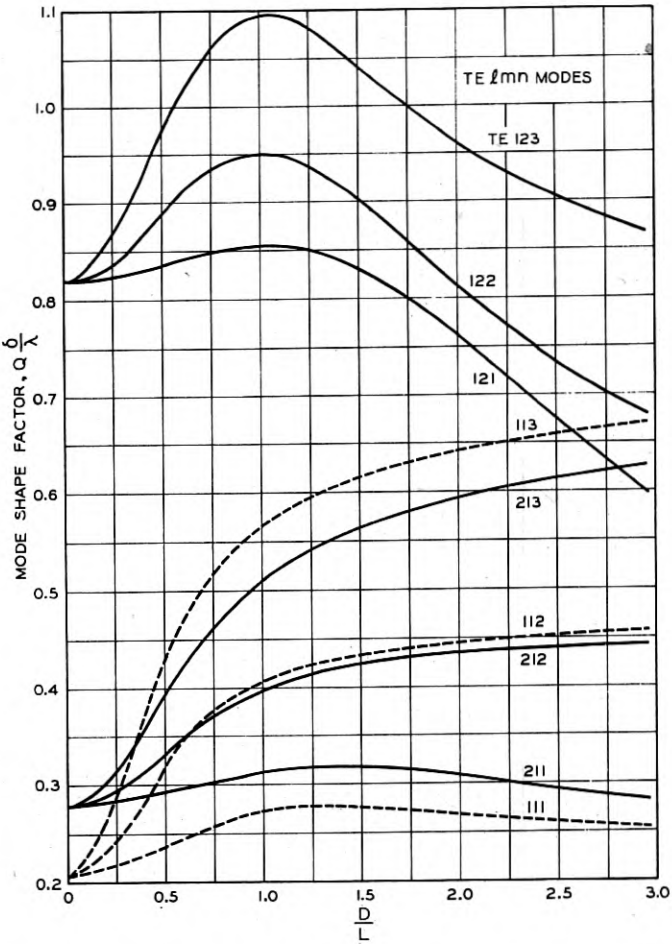


Fig. 4—Mode-shape factors $\left(Q\frac{\delta}{\lambda}\right)$ as a function of diameter to length ratio $\left(\frac{D}{L}\right)$ for circular cylinder resonator — TE modes with $l > 0$.

A special type of interfering mode is the $TE\ 01(n + 1)$ mode. In the nature of things, it is virtually impossible to suppress this mode without likewise suppressing the desired $TE\ 01n$ mode. The width of the operating band is thus strictly limited, if ambiguity is to be avoided. This effect, termed self-interference, becomes an important factor as n increases, since

$$\frac{f_2}{f_1}\ (\text{maximum}) = \frac{n+1}{n}$$

This maximum value cannot be realized because it is incompatible with other requirements.

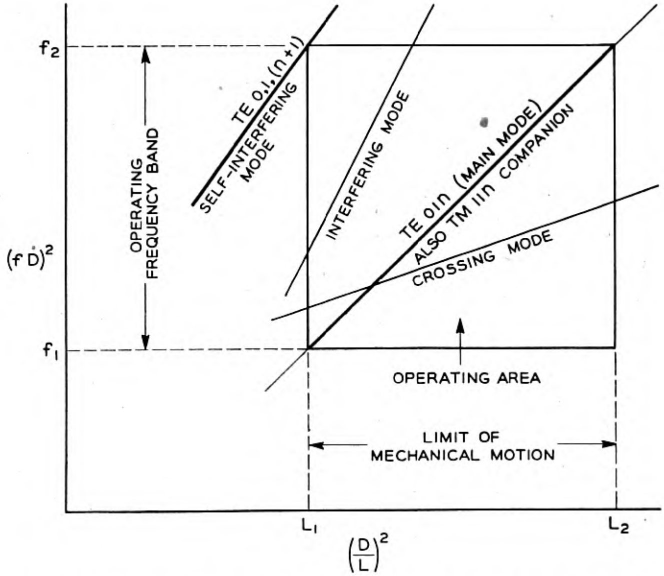


Fig. 5—Mode chart illustrating meaning of crossing and interfering modes and of operating area.

When an undesired mode crosses the main mode within the rectangle it is called a crossing mode. Except in a region close to the crossing point, it acts only to cause ambiguity as already discussed. In the immediate region of the crossing point, however, the cavity is simultaneously resonant in both modes. Violent interaction effects which may seriously degrade the cavity Q frequently occur at such a crossing.

Methods of Minimizing Effects of Unwanted Modes

A major problem in the design of a high Q cavity for radar test purposes, in which the Q and frequency range are set by radar system considerations, is to reduce the effects of all the undesired modes without seriously degrading the main *TE* 01n mode. Among those to be suppressed is the companion *TM* 11n mode which is inherently of the same frequency.

TABLE III.—*Values of the Bessel Function Zero (r_{lm}) for the First 180 Modes in a Circular Cylinder Resonator*

	r_{lm}	Mode*		r_{lm}	Mode*		r_{lm}	Mode*		r_{lm}	Mode*
1	1.8412	E 1-1	46	13.0152	M 3-3	91	18.4335	M 10-2	136	22.6716	E 2-
2	2.4048	M 0-1	47	13.1704	E 2-4	92	18.6374	E 6-4	137	22.7601	M 1-
3	3.0542	E 2-1	48	13.3237	M 1-4	93	18.7451	E 12-2	138	22.7601	E 0-
4	3.8317	M 1-1	49	13.3237	E 0-4	94	18.9000	M 14-1	139	22.9452	M 8-
5	3.8317	E 0-1	50	13.3543	M 9-1	95	18.9801	M 5-4	140	23.1158	M 14-
6	4.2012	E 3-1	51	13.5893	M 6-2	96	19.0046	E 9-3	141	23.2548	E 21-
7	5.1356	M 2-1	52	13.8788	E 12-1	97	19.1045	E 17-1	142	23.2568	M 18-
8	5.3176	E 4-1	53	13.9872	E 5-3	98	19.1960	E 4-5	143	23.2643	E 16-
9	5.3314	E 1-2	54	14.1155	E 8-2	99	19.4094	M 3-5	144	23.2681	E 7-
10	5.5201	M 0-2	55	14.3725	M 4-3	100	19.5129	E 2-6	145	23.2759	M 11-
11	6.3802	M 3-1	56	14.4755	M 10-1	101	19.5545	M 8-3	146	23.5861	M 6-
12	6.4156	E 5-1	57	14.5858	E 3-4	102	19.6159	M 1-6	147	23.7607	E 10-
13	6.7061	E 2-2	58	14.7960	M 2-4	103	19.6159	E 0-6	148	23.8036	E 5-
14	7.0156	M 1-2	59	14.8213	M 7-2	104	19.6160	M 11-2	149	23.8194	E 13-
15	7.0156	E 0-2	60	14.8636	E 1-5	105	19.8832	E 13-2	150	24.0190	M 4-
16	7.5013	E 6-1	61	14.9284	E 13-1	106	19.9419	E 7-4	151	24.1449	E 3-
17	7.5883	M 4-1	62	14.9309	M 0-5	107	19.9944	M 15-1	152	24.2339	M 9-
18	8.0152	E 3-2	63	15.2682	E 6-3	108	20.1441	E 18-1	153	24.2692	M 15-
19	8.4172	M 2-2	64	15.2867	E 9-2	109	20.2230	E 10-3	154	24.2701	M 2-
20	8.5363	E 1-3	65	15.5898	M 11-1	110	20.3208	M 6-4	155	24.2894	E 22-
21	8.5778	E 7-1	66	15.7002	M 5-3	111	20.5755	E 5-5	156	24.3113	E 1-
22	8.6537	M 0-3	67	15.9641	E 4-4	112	20.7899	M 12-2	157	24.3382	M 19-
23	8.7715	M 5-1	68	15.9754	E 14-1	113	20.8070	M 9-3	158	24.3525	M 0-
24	9.2824	E 4-2	69	16.0378	M 8-2	114	20.8269	M 4-5	159	24.3819	E 17-
25	9.6474	E 8-1	70	16.2235	M 3-4	115	20.9725	E 3-6	160	24.4949	M 12-
26	9.7610	M 3-2	71	16.3475	E 2-5	116	21.0154	E 14-2	161	24.5872	E 8-
27	9.9361	M 6-1	72	16.4479	E 10-2	117	21.0851	M 16-1	162	24.9349	M 7-
28	9.9695	E 2-3	73	16.4706	M 1-5	118	21.1170	M 2-6	163	25.0020	E 14-
29	10.1735	M 1-3	74	16.4706	E 0-5	119	21.1644	E 1-7	164	25.0085	E 11-
30	10.1735	E 0-3	75	16.5294	E 7-3	120	21.1823	E 19-1	165	25.1839	E 6-
31	10.5199	E 5-2	76	16.6982	M 12-1	121	21.2116	M 0-7	166	25.3229	E 23-1
32	10.7114	E 9-1	77	17.0038	M 6-3	122	21.2291	E 8-4	167	25.4170	M 16-
33	11.0647	M 4-2	78	17.0203	E 15-1	123	21.4309	E 11-3	168	25.4171	M 20-1
34	11.0864	M 7-1	79	17.2412	M 9-2	124	21.6415	M 7-4	169	25.4303	M 5-
35	11.3459	E 3-3	80	17.3128	E 5-4	125	21.9317	E 6-5	170	25.4956	E 18-
36	11.6198	M 2-3	81	17.6003	E 11-2	126	21.9562	M 13-2	171	25.5094	M 10-
37	11.7060	E 1-4	82	17.6160	M 4-4	127	22.0470	M 10-3	172	25.5898	E 4-
38	11.7349	E 6-2	83	17.7740	E 8-3	128	22.1422	E 15-2	173	25.7051	M 13-
39	11.7709	E 10-1	84	17.7887	E 3-5	129	22.1725	M 17-1	174	25.7482	M 3-
40	11.7915	M 0-4	85	17.8014	M 13-1	130	22.2178	M 5-5	175	25.8260	E 2-
41	12.2251	M 8-1	86	17.9598	M 2-5	131	22.2191	E 20-1	176	25.8912	E 9-
42	12.3386	M 5-2	87	18.0155	E 1-6	132	22.4010	E 4-6	177	25.9037	M 1-
43	12.6819	E 4-3	88	18.0633	E 16-1	133	22.5014	E 9-4	178	25.9037	E 0-
44	12.8265	E 11-1	89	18.0711	M 0-6	134	22.5827	M 3-6	179	26.1778	E 15-
45	12.9324	E 7-2	90	18.2876	M 7-3	135	22.6293	E 12-3	180	26.2460	E 12-

* Nomenclature after Barrow & Mieher, "Natural Oscillations of Electrical Cavity Resonator IRE Proceedings, April 1940, p. 184. M modes take zeros of $J_l(x)$; E modes take zeros of $J_l'($ Number directly following E or M is l; number after hyphen is number of root.

Values less than 16.0 are abridged from six-place values and are believed to be correct; valu more than 16.0 are abridged from five-place values and may be in error by one unit in fourth decin place. **5** in fourth place indicates that higher value is to be used in rounding off to fewer decima

One solution, of limited application, is to choose an operating rectangle free from extraneous responses. An alternative solution is to design the cavity in a manner such that the undesired responses are reduced or suppressed to a point where their presence does not interfere with the normal operation of the cavity. In this latter case, the amount of suppression is naturally dependent upon the use to which the cavity is to be put, and is conceivably different for a high Q cavity used as a frequency meter, for example, and one used as a selective filter.

Experience has shown that certain families of modes are much more difficult to suppress than others, and are to be avoided, if at all possible. The feasibility of doing this can be determined by sliding the operating area (a suitable opening in a sheet of paper) around on a large mode chart until the most favorable operating region, consistent with other requirements, has been found.

Once a suitable operating area has been chosen, the cavity diameter is fixed and length and frequency scales added to the mode chart make it read directly in quantities readily measured.

Cavity Couplings

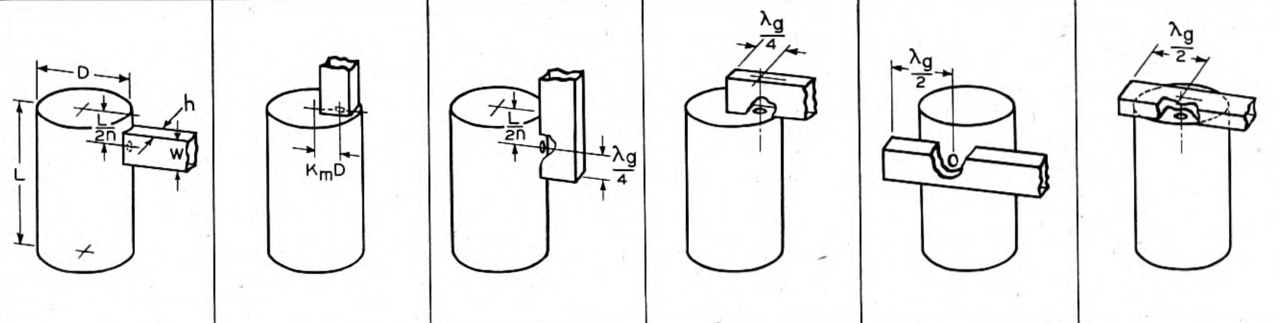
To be useful the cavity must be coupled to external circuits. The problem here is to get the correct coupling to the main mode and as little coupling as possible to all others. Since the electric field is zero everywhere at the boundary surface of the cavity for the TE $01n$ mode, coupling to it must be magnetic. This may be obtained either by a loop at the end of a coaxial line or by an orifice connecting the cavity with a wave guide.

The location for maximum coupling to the main mode is on the side of the cavity, an odd number of quarter-guide wavelengths from the end, or on the end about halfway (48%) out from the center to the edge. Correct orientation of loop or wave guide is achieved when the magnetic fields are parallel. This requires the axis of the loop to be parallel to the axis of the cylinder for side wall feed and to be perpendicular to the cylinder axis for end feed. Wave guide orientation is shown in Table IV.

The theory of coupling loops and orifices is not at present precise enough to yield more than approximate dimensions. Exact sizes of loops and holes have therefore been obtained experimentally for all designs.

On the basis of rather severely limiting assumptions,[11] coupling formulas for a round hole connecting a rectangular wave guide and a TE $01n$ cavity are given in Table IV. The assumptions are that the orifice is in a wall of negligible thickness, its diameter is small compared to the wavelength, it is not near any surface discontinuity, and that the wave guide propagates only its principal (gravest) mode and is perfectly terminated. In echo box applications, this theory leads to a computed diameter that is somewhat smaller than experiment shows to be correct.

The coupling to other modes can be analyzed, at least qualitatively, from the field expressions of Table I. This has been of value in making final

	CASE					
	1A	1B	2A	2B	3A	3B
COUPLING METHOD						
CIRCULAR ORIFICE	$\frac{\Delta f}{f} = -K_c \frac{\lambda^2 d^3}{D^4 L}$ $W_a = K_w \frac{\lambda^2 d^6}{\lambda_g w h D^4 L}$	$\frac{\Delta f}{f} = -K_c \frac{n^2 \lambda^2 d^3}{D^2 L^3}$ $W_a = K_w \frac{n^2 \lambda^2 d^6}{\lambda_g w h D^2 L^3}$	$\frac{\Delta f}{f}$ SAME AS 1A $W_a = K_w \frac{\lambda_g \lambda^2 d^6}{w^3 h D^4 L}$	$\frac{\Delta f}{f}$ SAME AS 1B $W_a = K_w \frac{n^2 \lambda_g \lambda^2 d^6}{w^3 h D^2 L^3}$	$\frac{\Delta f}{f}$ SAME AS 1A W_a SAME AS 1A	$\frac{\Delta f}{f}$ SAME AS 1B W_a SAME AS 1B

CONSTANTS

	1A		1B			2A	2B
	K_c	K_w	K_c	K_w	K_m	K_w	K_w
TE 01n	0.316	1.322	0.1107	0.464	0.2403	0.331	0.1159
TE 02n	1.058	4.43	0.1995	0.836	0.1312	1.108	0.2089
TE 03n	2.225	9.32	0.288	1.207	0.0905	2.330	0.302

NOTATION: λ = FREE SPACE WAVELENGTH OF CAVITY RESONANCE
λ_g = GUIDE WAVELENGTH
d = DIAMETER OF ORIFICE
$W_a = \frac{1}{Q_a}$ = CAVITY LOADING

NOTE: FOR FEED LIKE CASES 2 AND 3, BUT WITH WAVEGUIDE TERMINATED IN BOTH DIRECTIONS, DIVIDE W_a BY 2

TABLE IV.—*Orifice Coupling of Wave Guide (TE 10 Mode) to Cylindrical Cavity (TE 0mn Mode)* small relocations of the coupling points to discriminate against a residual undesired mode.

Principle of Similitude

One other theorem generally applicable to all cavities has been useful in design. It is the principle of similitude, which may be stated as follows[12]:

> A reduction in all the linear dimensions of a cavity resonator by a factor $1/m$ (if accompanied by an increase in the conductivity of the walls by a factor m) will reduce the wavelengths of the modes by a factor $1/m$.

The condition given in parentheses is necessary for strict validity; for high Q cavities, it need not be considered.

Application of Theory

An illustration of an engineering application of the basic information just presented, is the design of an echo box test set for use in radar maintenance.

The test set has a number of components, but only the cavity proper and its couplings will be considered at this time.

Design Requirements

The basic design requirements of the cavity set by the radar are: (1) the working decrement, and (2) the tunable frequency limits (f_1 and f_2). Service use of these test sets in the 3 *kmc* and 9 *kmc* bands has shown that a working decrement of about 3 *db* per microsecond gives satisfactory results.

As seen in the discussion above, Q is a more useful design parameter for the resonant cavity than decrement, d. Hence the conversion

$$Q = \frac{27.3f}{d} \tag{4}$$

where d is expressed in *db* per microsecond, and f is in megacycles, gives the loaded or working Q to be realized.

Determining the Theoretical Q Required

For design purposes, however, there are several factors which dictate the use of a value for theoretical Q which is somewhat higher than the loaded Q just computed. The input and output couplings reduce the theoretical Q of the cavity due to their loadings. The coupling factor, s, expressed as the ratio of loaded to unloaded Q, has values for echo boxes of about 0.90 for the input coupling and from 0.90 to 0.99 for the output coupling. In addition, other factors such as the means used for mode suppression may degrade the Q. But, in simple designs, these may be negligible. Therefore, it is expedient to design for a theoretical Q of about 15 to 25 per cent in excess of the working Q to be realized.

The working Q's of a number of echo box designs are cited here to indicate the order of magnitude required at several frequencies:

1 *kmc*	70,000*
3	40,000
9	100,000
25	200,000

Finding the Cavity Dimensions

With the frequency and theoretical Q known, the dimensions of the cavity can be evaluated but the formulas of Table I require some simplification for engineering use.

The mode-shape (MS) factor, $Q\frac{\delta}{\lambda}$, may also be termed its selectivity which for the $TE\ 01n$ modes may be expressed as follows:

$$Q\,\frac{\delta}{\lambda} = 0.610\,\frac{\left[1 + 0.168\left(\frac{D}{L}\right)^2 n^2\right]^{3/2}}{1 + 0.168\left(\frac{D}{L}\right)^3 n^2} \tag{5}$$

where:

δ = the skin depth in cm. $= \dfrac{1}{2\pi}\sqrt{\dfrac{10^9\rho}{f}}$

ρ = the resistivity in ohm-cm.

The skin depth is a factor which recognizes the dissipation of energy in the walls and ends of the cylinder. With increase of resistivity of these surfaces the currents penetrate deeper and the resulting Q is lower.

A comparison of the relative Q's computed from the resistivity of several metals will show the importance of this factor:

Silver	1.03
Copper	1.00
Gold	0.84
Aluminum	0.78
Brass	0.48

Therefore, a brass cavity will have about one-half of the Q that a similar cavity would have if made of copper. Similarly, the silverplating of a copper cavity will gain about 3 per cent in Q.

Equation 5 may be made more convenient for calculations by combining

* This value reflects the higher Q required on ground radars.

terms which are a function of frequency and by assuming the conductivity of copper* for the cylinder walls. It then becomes

$$\frac{Q\sqrt{f}}{10^6} = 2.77 \frac{\left[1 + 0.168\left(\frac{D}{L}\right)^2 n^2\right]^{3/2}}{1 + 0.168\left(\frac{D}{L}\right)^3 n^2} \tag{6}$$

for $TE\ 01n$ modes; f is in megacycles.

Thus, it is seen that for the design parameters Q and f, selected values of n now define tentative useful points on the mode chart in terms of D/L and n.

Selection of Operating Area on Mode Chart

This will be more evident upon examination of Fig. 6, which is a basic design chart for cylindrical cavity resonators using $TE\ 01n$ modes. The coordinates of $(fD)^2$ and $(D/L)^2$ will be recognized from the previous discussion of the mode chart (Fig. 1) although in this case the range has been expanded by the use of logarithmic coordinates. Mode identification is obtained from equation 2; which, for $TE\ 01n$ modes becomes

$$(fD)^2 \times 10^{-8} = 2.0707 + 0.3480\, n^2\, (D/L)^2 \tag{7}$$

with D and L in inches and f in megacycles.

A family of $TE\ 01n$ modes has been drawn on the chart for selected values of n. To aid in designing for minimum volume a line labelled Max. $\frac{Q}{V}$ has been added (Equation 3). Lines of constant $Q\sqrt{f}$ are also shown as a series of dashed lines.

A tentative operating area on this chart may be selected on the basis of the required $Q\sqrt{f}$. Using mid-frequency for f, the intersection of the $Q\sqrt{f}$ line and the minimum volume line will define the operating mode, n, and also locate the center of the operating rectangle.

An enlarged plot of this area as in Fig. 7 will show all modes possible in the cavity. Some adjustment of the precise location of this area may then be desirable to eliminate certain types or numbers of unwanted modes. For example, it is extremely difficult to suppress $TE\ 02$ modes without affecting the $TE\ 01$ mode, because of the very close resemblance of the field configurations. On the other hand, most TM modes are easy to handle.

It may be possible to select an operating area such as the rectangle blocked out in Fig. 1 in which all extraneous responses (with the exception of the companion $TM\ 111$) are avoided. The largest rectangle which can be inscribed here is limited by the $TE\ 311$ and $TE\ 112$ modes. This will

* $\rho = 1.7241 \times 10^{-6}$ ohm-cm—the International Standard value for copper.

permit a ±3.6 per cent frequency range. Several designs of 3 *kmc* echo boxes have been based on this area.

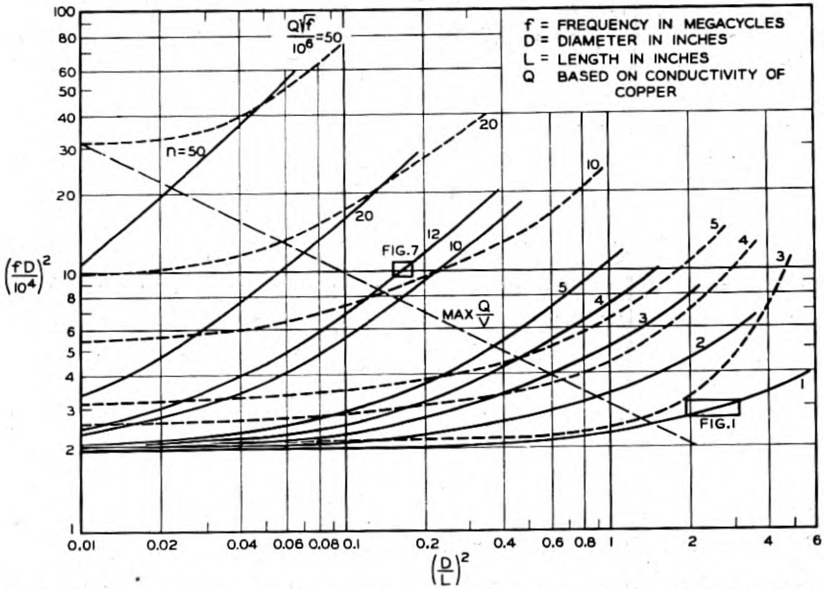


Fig. 6—Design chart for cylindrical cavity resonators operating in the *TE* 01*n* modes.

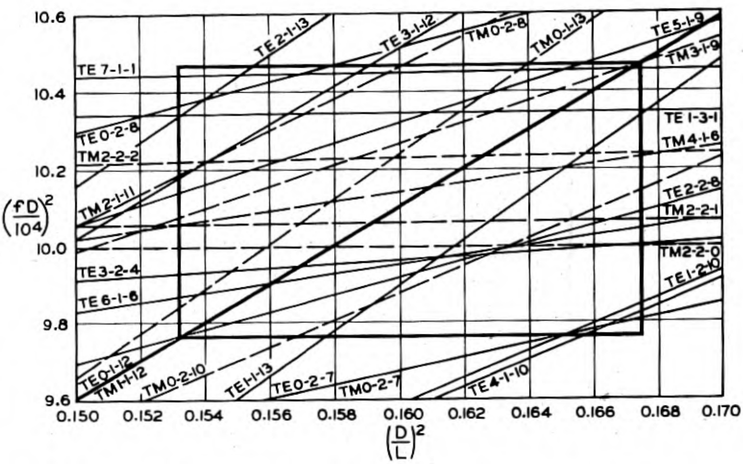


Fig. 7—Expanded mode chart covering the operating area of a 9 *kmc* echo box.

This type of solution is quite adequate for many of the simpler designs; but as the required Q becomes larger or the frequency coverage becomes greater it is generally not possible to locate such areas. An operating area for 9 *kmc* is shown on Fig. 7. Nine crossing modes and twelve interfering modes exist. For 25 *kmc* the crossing modes run into the forties with hundreds of interfering modes.

Suppression of the undesired modes requires a thorough knowledge of their field configurations and a number of effective techniques which may be applied on a practical engineering basis. Several examples are cited in later sections.

Since decrement is an important characteristic of these cavities, especially when applied to radar test sets, the uniformity of the decrement over the frequency range or "flatness of response" may be a significant design requirement. It will be seen from Fig. 3, that the MS factor of the wanted mode is not constant with varying D/L. In fact, if it were, the decrement would increase as the 3/2 power of frequency.

There are at least three attacks on this "flatness" problem: (1) to operate on the sloped portion of the MS curve in such a manner that its characteristic will tend to be complementary to the change with frequency; (2) to obtain compensation by varying the coupling with frequency—generally accomplished by selecting an appropriate coupling point along the side wall of the cavity; and (3) to overplate a portion of the cylinder's interior surface with a material of higher resistivity such as cadmium. For this third method, the formulas for $Q\frac{\delta}{\lambda}$ of Table I are no longer applicable since they assume a uniform resistivity of the cavity walls.

Thus, it will be seen that the final design of a cavity resonator is a compromise between a number of desired characteristics:

a) A cavity of minimum volume for a given Q.

b) A cavity having a minimum of extraneous responses of types difficult to suppress.

c) A cavity with compensation for flatness of decrement.

Engineering judgment is required to weigh the emphasis on each of these requirements which at times may be mutually exclusive.

Some Practical Considerations

Physically realizing the theoretical characteristics just described to obtain a satisfactory cavity brings forth a host of practical design problems. A number of these will be discussed in this section.

Description of Echo Box Test Sets

The schematics (Figs. 8 and 9) and photographs (Figs. 12 to 14) show the components and various construction methods of echo boxes in the 3 and 9 *kmc* bands. The cavity itself may be spun, drawn or turned of material such as brass or aluminum which will give mechanical stability without excessive weight.

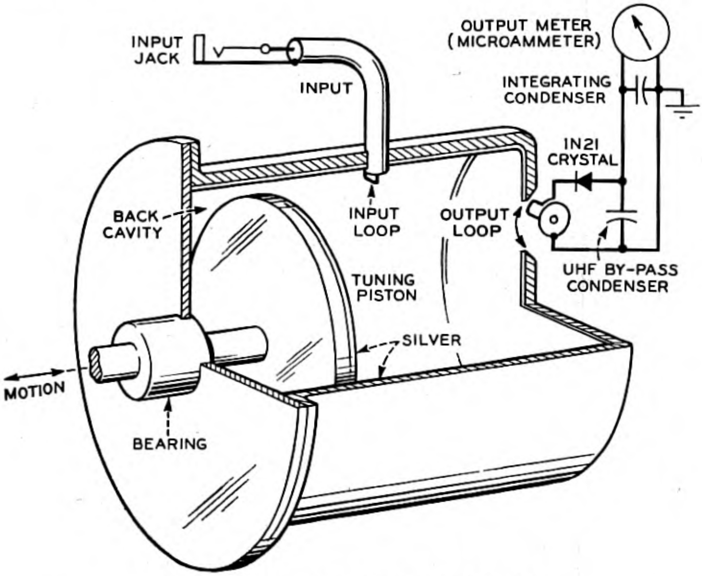


Fig. 8—Schematic of a 3 *kmc* echo box.

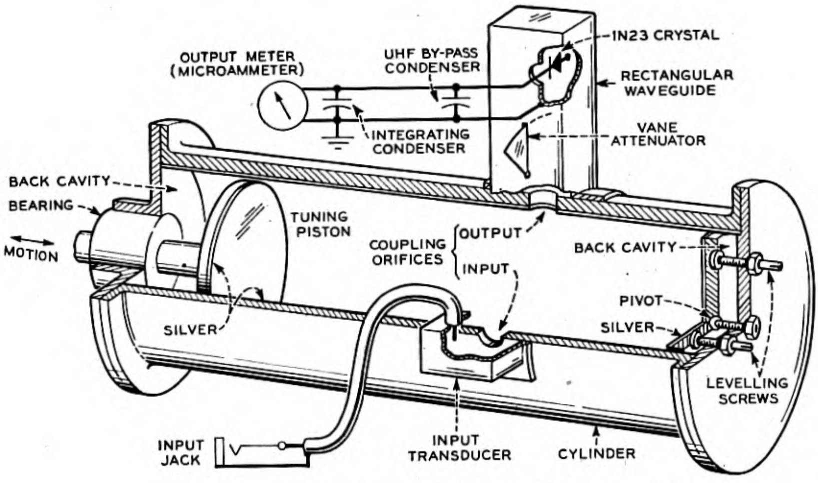


Fig. 9—Schematic of a 9 *kmc* echo box.

The movable plate is driven by the piston which operates in a close-fitting bearing. The drive mechanism translates the rotary motion of the tuning knob into linear motion of the piston through a crank and connecting rod assembly. Coupled to the drive shaft is an indicating dial to register frequency.

Silverplating is indicated on all interior surfaces of the cavity for minimum resistance to currents in the walls and ends of the cylinder.

For the 3 *kmc* bands, the input coupling is in the form of a loop protruding into the cavity and connected by coaxial microwave cable to the radar under

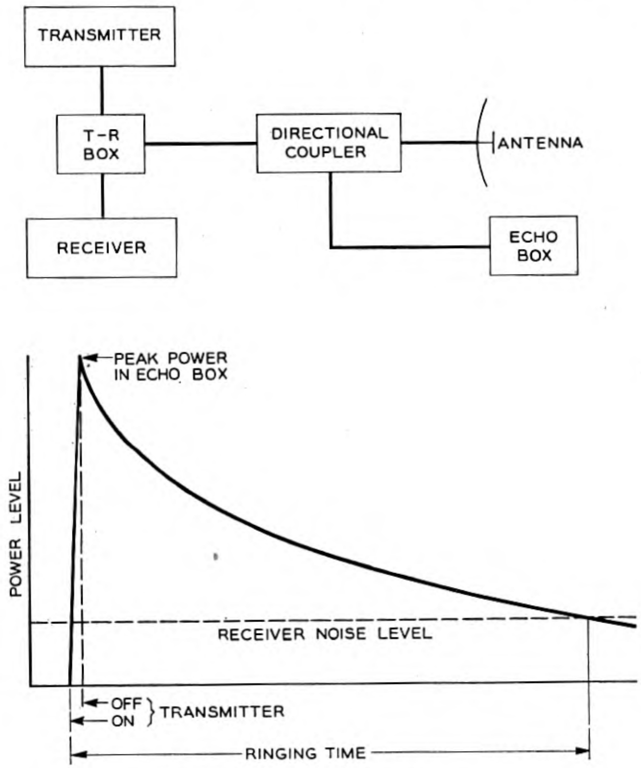


Fig. 10—Radar test with echo box.

test. For the 9 *kmc* bands, the input coupling is an orifice with which is associated a wave guide to coaxial cable transducer.

The output couplings for the two frequency bands also differ in construction. For the 3 *kmc* bands, the variable coupling is achieved by rotating the output loop before an aperture in the cavity. In this rotary mount is housed the rectifying crystal and by-pass condenser. An orifice is used for the 9 *kmc* band output coupling which feeds a crystal mounted in a wave guide. Amplitude control is by a vane attenuator in the wave guide. In all cases an integrating condenser is required to smooth out the *d-c* pulses delivered by the crystal to give output indication on the *d-c* microammeter.

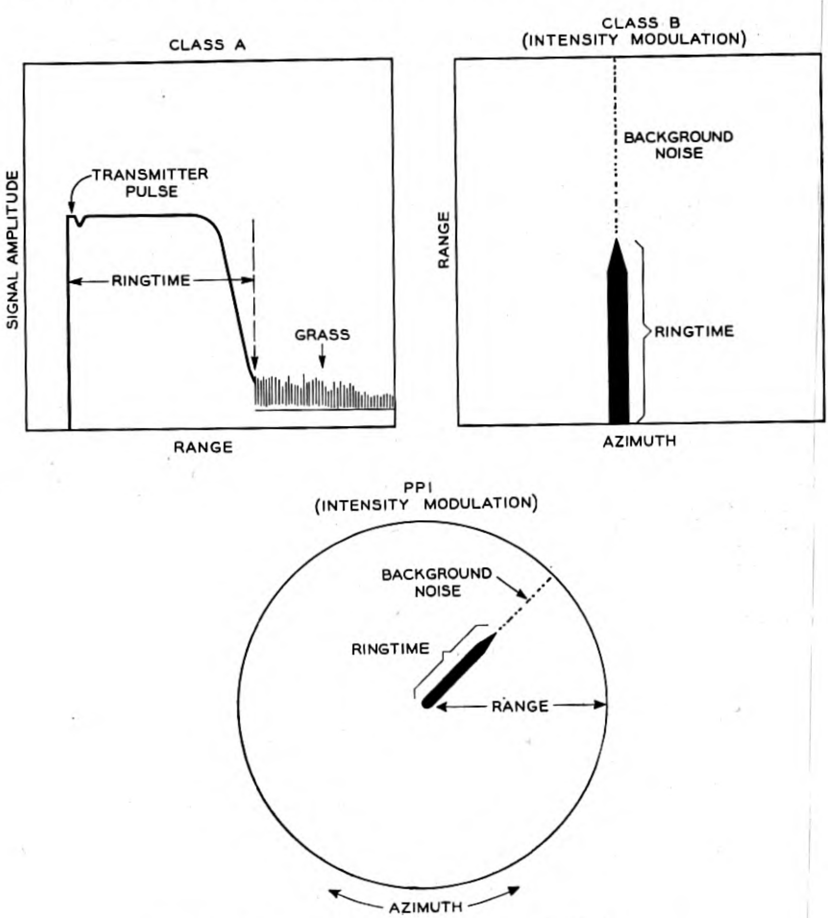


Fig. 11.—Typical ringtime patterns on radar indicators.

The cylinder in the 9 *kmc* band is an assembly of a tube and two end plates. One of these is driven by the piston as described and the other is adjustable for "levelling" i.e., parallelism of the two plates.

End Plate Gaps

The wanted *TE* mode is paired with a companion unwanted *TM* as described above. The fact that the *TM* mode has a Q substantially less than that of the *TE* makes the realization of the higher Q difficult. A method is needed of suppressing the TM in the presence of the TE. Since the resonant cavity is a cylinder, of variable length, the movable end plate (or reflector)

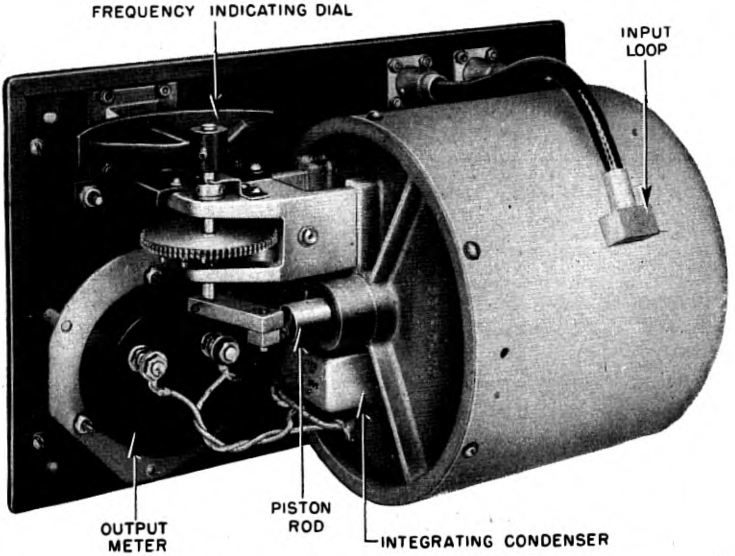


Fig. 12—Type of construction in a 3 *kmc* echo box.

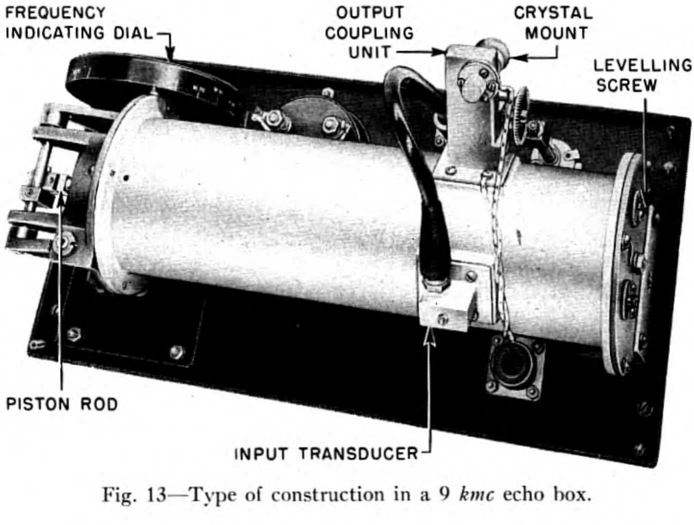


Fig. 13—Type of construction in a 9 *kmc* echo box.

can be modified to include a gap at its periphery. The gap perturbs the resonant frequencies of the two modes by different amounts so that they become separated. Secondly, the peripheral gap cuts through the surface currents at points of high density for *TM* modes and minimum density of *TE* 0*mn* modes and hence is a form of mode suppression. Thirdly, the gap greatly simplifies the mechanical design of a movable end plate by eliminating the need of physical contact with the side wall of the cylinder.

A similar gap may be used at the other end. This facilitates "levelling" of a false bottom in the cavity.

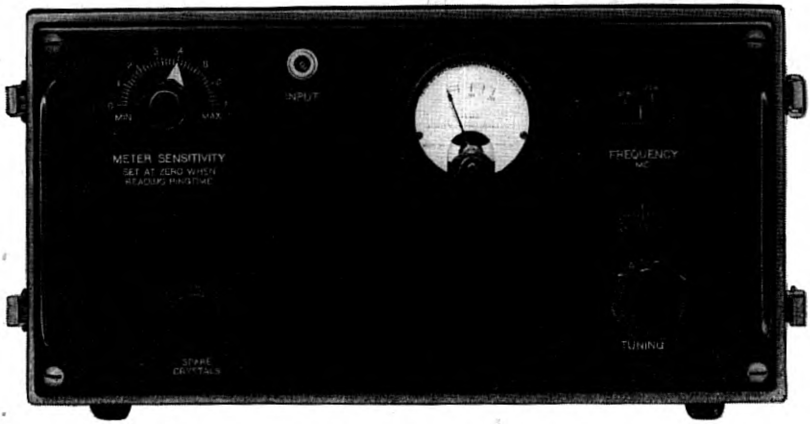


Fig. 14—Panel view of a 9 *kmc* echo box.

Back Cavity Effects

The cavity with a peripheral gap may give rise to further spurious resonances in the region behind the reflecting surface (known as the back cavity) if these responses are not damped. The addition of a lossy material such as bakelite or carbon loaded neoprene in the back cavity is a successful suppression method.

Cylinder Tolerances

The geometry of the structure is very important in realizing the potential Q of the cavity. The theoretical computations are based on a perfect right circular cylinder which in practice is seldom achieved. Distortions occur in various forms: e.g., the cylinder instead of being round may be elliptical; the ends may not be perpendicular to the axis of the cylinder or not parallel to each other; surface irregularities may be present causing field distortions within the cavity. Many of these effects have been minimized by requiring adherence to close dimensional tolerances.

For some designs, the requirement on the parallelism of the end plates is greater than can be commercially produced. An adjustable mechanism for levelling is a practical solution. Tilt adjustments in the order of 0.001 inch at the edge of the plate (about 3-inh diameter) are required in the 9 *kmc* band. Experience in the 25 *kmc* band shows evidence of the need for even finer control.

Plating

In addition, adequate control of the conductivity of the interior surfaces of the cavity is necessary to achieve a uniform manufactured product. This requires attention not only to the thickness and uniformity of the plating but also to the purity of the plating baths and the avoidance of introduction of foreign matter during buffing processes.

Couplings

The type and location of the coupling means can be used to discriminate between wanted and unwanted modes. Hence, this is a fertile field for mode suppression techniques. For example, since *TM* modes have $H_z = 0$, orifice coupling to the main mode at the side wall of the types shown in Table IV, cases 1A and 2A, will not couple to any *TM* modes. Again, if end coupling is used in a cavity which will support both the *TE* 01 and *TE* 02 modes, by locating this coupling at the point where $H_\rho = 0$ for the *TE* 02 mode (about 54% of the way out from the center), it will not be excited and coupling to the *TE* 01 will be only slightly below maximum.

For echo box test sets the magnitude of the input coupling to the wanted mode is a compromise between the incomplete buildup of the fields within the cavity during the charging interval and the loading of the cavity Q on discharge. This is carried on by varying the coupling and observing the "ringtime" (the echo box indication on the radar scope). Optimum coupling is achieved when ringtime is made a maximum.

Output couplings for echo boxes are made so that just enough energy is withdrawn from the cavity to give an adequate meter reading.

Drive Mechanism

The objective of the design of the tuning mechanism is to provide a smooth, fine control with a minimum of backlash. An illustration of the mechanical perfection required can be cited in a 9 *kmc* band design where $\frac{1}{4}$ inch of travel covered 200 megacycles in frequency. Hence, for frequency settings to be reproducible to within $\frac{1}{4}$ mc the mechanical backlash of the moving parts had to be held to about 0.0003 inches or 0.3 mil. To realize this in commercial manufacture and to maintain it after adverse operating conditions such as vibration and shock was a major mechanical design problem.

In the design of this drive mechanism it should be recognized that equal increments of cavity length will not produce uniform increments in frequency. The mode chart indicates graphically that a straight-line relationship exists between $(f)^2$ and $\left(\frac{1}{L}\right)^2$. Uniformly spaced markings on a dial reading directly in frequency can be realized by the use of such mechanisms as an eccentric operating on a limited arc. Adjustments are customarily provided to bring the cavity resonance and dial indication into agreement at some frequency of test. Frequency departures of the drive mechanism referred to this point are held commercially to about one part in 5000.

Application of Similitude

Echo box developments have often been undertaken at frequencies where adequate test equipment was not available. This has been especially true as the radar art progressed to higher and higher frequencies.

The principle of similitude has been utilized in the construction and test of models at the frequency of existing test facilities. The models have then been scaled to the assigned frequency band. This has been found to be a very practical expedient.

Use of Cavities For Radar Testing

The high Q resonant cavity when appropriately connected to a radar system returns to it a signal which may be used to judge the over-all performance of the radar. Its operation is as follows: During the transmitted pulse, microwave energy from the radar is stored in the cavity in the electromagnetic field. The charge of the cavity increases exponentially during this interval but fails to reach saturation for the cavity by a substantial margin because the pulse is too short. At the end of the pulse, the decay of this field supplies a signal of the same frequency as that of the radar transmitter (when the echo box is in tune) which is returned to the radar receiver as a continuous signal diminishing exponentially in amplitude.

The time interval between the end of the transmitted pulse and the point where the signal on the radar disappears into the background noise is the "ringtime." The term is used somewhat loosely since, in actual practice, the ringtime is measured on indicators whose range markings are generally in miles or yards referred to the beginning rather than the end of the pulse. The difference, of course, is small. It is customary to include the pulse length in all ringtime figures on operating radar systems. Typical ringtime patterns on radar indicators and a schematic of a radar test with an echo box are shown in Figs. 10 and 11.

As the output power of the radar either increases or decreases corresponding changes in the "charge" of the cavity will be reflected directly in ringtime changes. Similarly as the noise level of the radar receiver varies the merging point of the cavity signal and the noise will show proportional changes in ringtime. Hence, the ringtime indication measures these two factors on which the radar's ability to discern real targets so largely depends.

The exponential buildup and decay of the charge in the cavity occur at a rate determined by the working decrement of the cavity. As mentioned previously, a decrement of about 3 *db* per microsecond is a satisfactory value for the 3 *kmc* and 9 *kmc* bands. A one microsecond change in ringtime (roughly one-tenth mile) would, therefore, represent a change in system performance of 3 *db*.

Uniformity Control and Expected Ringtime

By introducing an adjustment for the working Q of the cavities it is now possible to control the uniformity of the manufactured product to very close limits. Other improvements have also been incorporated which insure that boxes which have been made alike as to Q will similarly give uniform ringtime indications on a test radar. If the test sets are all alike as to ringtime, it is then possible to quote an "expected ringtime" for each of the various radars to be serviced by the echo box. Initially a measuring tool indicating relative changes in day to day operation of the radars, the uniformity provision with its "expected ringtime" has made the echo box test set an absolute measuring instrument of moderate precision.

Other Uses

In addition to its use as a measure of over-all performance of a radar, a significant number of diagnostic tests may be performed when trouble develops, which aid in rapidly locating the source. One such test is spectrum analysis. The extreme selectivity of the high Q cavity permits examination of the spectrum of the pulsed wave and from this may be deduced characteristics of the pulse, including pulse length. Multiple-moding of the magnetron circuit is easily shown by this analysis.

The meter of the test set gives a relative indication of the output power of the radar and this in itself assists greatly in segregating transmitter troubles from receiver troubles.

Also of importance is the use of the echo box as a frequency meter. The high Q of the cavity plus the fine control of the drive mechanism and the direct reading dial give excellent results (comparable to that of a wavemeter).

Acknowledgement

The design of resonant cavities is a difficult and complex art. In bringing it to the present state, the number of individuals who have made significant mathematical, theoretical, engineering and mechanical contributions is so large that it is impossible to give them due credit by name. To them, however, the authors wish to acknowledge their indebtedness and to express their appreciation.

Bibliography

1. E. I. Green, H. J. Fisher and J. G. Ferguson, "Techniques and Facilities for Microwave Radar Testing", A. I. E. E. Transactions, 65, pp. 274–290 (1946), and this issue of the *B. S. T. J.*
2. G. C. Southworth, "Hyper-Frequency Wave Guides—General Considerations and Experimental Results," B. S. T. J., 15, pp. 284–309 (1936).
3. J. J. Thompson, "Notes on Recent Researches in Electricity and Magnetism," Oxford, 1893.
4. Lord Rayleigh, "On the passage of electric waves thru tubes or the vibration of dielectric cylinders," *Phil. Mag.*, 43, pp. 125–132 (1897).
5. A. Becker, "Interferenzrohren fur electrische Wellen," *Ann. d. Phys.*, 8, pp. 22–62 (1902).
6. F. K. Richtmyer and E. H. Kennard, "Introduction to Modern Physics," pp. 184–189, McGraw Hill, 1942.
7. R. I. Sarbacher and W. A. Edson, "Hyper and Ultra-high Frequency Engineering," p. 377, Wiley, 1943.
8. H. A. Wheeler, "Formulas for the Skin Effect," *Proc. I. R. E.*, 30, pp. 412–424 (1942).
9. W. L. Barrow and W. W. Mieher, "Natural Oscillations of Electrical Cavity Resonators," *Proc. I. R. E.*, 28, pp. 184–191 (1940).
10. L. J. Chu, "Electromagnetic Waves in Elliptical Hollow Pipes of Metal," *Jour. App. Phys.*, 9, pp. 583–591 (1938).
11. H. A. Bethe, M.I.T. Radiation Laboratory Reports V-15-S, 43–22, 43–26, 43–27, and 43–30.
12. H. Konig, "The Laws of Similitude of the Electromagnetic Field, and Their Application to Cavity Resonators," *Hochf:tech u. Elek:akus*, 58, pp. 174–180 (1941). Also *Wireless Engr.*, 19, pp. 216–217, No. 1304 (1942).

Techniques and Facilities for Microwave Radar Testing*

By E. I. GREEN, H. J. FISHER, J. G. FERGUSON

Methods and devices are described for testing microwave radars in the radio frequency range from about 500 mc to 25,000 mc, and at associated video frequencies. In general, the same instruments and techniques are applicable also in testing microwave communication systems.

Introduction

THAT radars are marvels of ingenuity has long since become common knowledge. This ingenuity is reflected, however, in complexity of circuits. A rough index of this is found in the number of vacuum tubes, which for a single radar may range from 50 to 250. Notwithstanding the most careful design, it is easy for the radar performance to become impaired under operating conditions.

Not only is radar complex, but its performance criteria are less tangible than those of conventional communication systems. Ordinary radio is to some extent self-testing in that reception of intelligible speech or signals frequently constitutes a sufficient check of satisfactory performance. With radar, the greater the range coverage and the more accurate the data, the more valuable the information is likely to be. However, the working range may fall to a fraction of the possible maximum or some other degradation or malfunctioning may occur, with nothing in the operation of the radar to tell that this has happened. Since lack of maximum performance may have serious military results, measurement of performance assumes the utmost importance in radar work.

The new techniques and new frequency ranges employed for radar necessitated the wartime development of a wide variety of new types of test equipment. A large part of this development work was concentrated at Bell Laboratories and at the N.D.R.C.'s Radiation Laboratory at M.I.T., working in close coordination with one another and with the technical services of the Army and Navy. In the manufacture of radar test equipment, Western Electric took a major part. This article discusses the techniques of radar testing and describes the types of test gear developed by Bell Laboratories and manufactured by Western Electric. These cover the radio frequency range from about 500 megacycles to 25,000 megacycles, together with associated video frequencies.

Because of its importance during the war, emphasis has been placed on

* Presented at A.I.E.E. Winter Convention, New York, N. Y., January 21–25, 1946. Published by *Elec'l. Engg., Trans. Sec.*, May 1946.

the testing of microwave radars. However, similar methods and instruments have also been employed in the testing of microwave communication systems and such applications can be expected to increase. In this situation the developers and users of microwave communication systems are fortunate in that almost all of the techniques and devices developed for radar testing are equally applicable to communication systems. This is even true of the video units, which are useful in connection with pulse modulated telephone systems and AM or FM television systems.

So many persons, both within and outside Bell Laboratories, have contributed to the developments described that the authors have reluctantly reached the conclusion that the assignment of individual credit should not be attempted.

Requirements

Operation of Typical Radar

Subsequent discussion may be simplified by first reviewing briefly the operation of a somewhat typical radar, as shown in Fig. 1. Under the control of d.c. or so-called video pulses from the modulator, short pulses of radio frequency energy are delivered by the magnetron transmitter to a highly directive antenna, ordinarily arranged to scan a section of space. Energy reflected from an object or "target" in the path of the beam is intercepted by the same antenna. The received pulses or echoes are converted to an intermediate frequency by heterodyning against a local oscillator, the frequency of which may be automatically controlled.

To enable the same antenna to serve for both transmitting and receiving, a TR tube or transmit-receive switch is usually provided. This consists of a partially evacuated resonant cavity containing a spark gap which breaks down during the transmitted pulse, thus preventing the transmitted power from injuring the sensitive receiver. An RT tube, consisting of a similar resonant cavity and spark gap, may be provided to prevent absorption of the received signal by the transmitter. After amplification and detection of the received signal, the resultant video pulses are applied to an indicator which may present information in any of several different ways. Customarily the direction of the target (determined by antenna orientation) and its range (determined by reflection interval, 10.7 microseconds per mile) are shown. The system may be used merely for searching, or for fire control, bomb direction, or other functions, with additional equipment as required.

Types of Tests and Test Sets

Figure 2 shows an early assemblage of radar test equipment for the 10 cm range, initially produced in 1942, which has seen wide usage.

The more important types of tests required in radar work, either at radar operating locations or at centralized service points, are: (1) Over-all Performance (Range Capability), (2) Transmitter Power, (3) Receiver Sensitivity, (4) Transmitter Frequency, (5) Transmitter Spectrum, (6) Standing Wave Ratio, (7) R. F. Envelope, (8) Receiver Recovery, (9) AFC Tracking, (10) I.F. Alignment, (11) Video Wave Shapes, (12) Range Calibration and (13) Computer Calibration.

The principal types of test sets required for carrying out the above are: (1) Signal Generators, (2) Echo Boxes, (3) Frequency and Power Meters (separate or combined), (4) Standing Wave Meters, (5) R.F. Loads, (6)

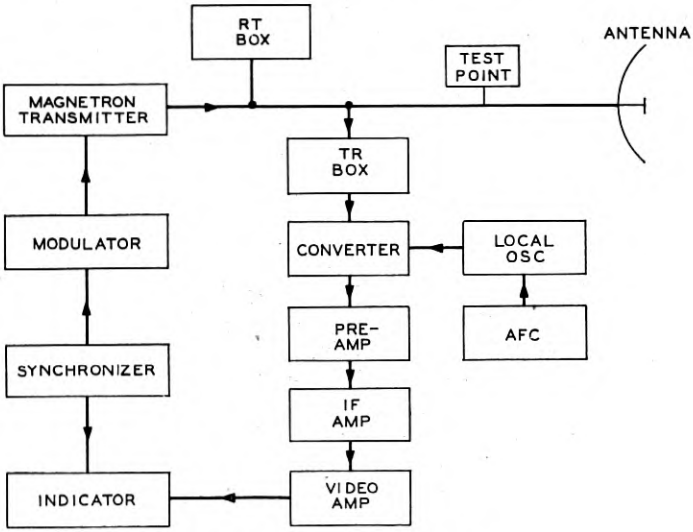


Fig. 1—Diagram of typical radar.

Oscilloscopes, (7) Video Dividers and Loads, (8) Range Calibrators, (9) Computer Test Sets and (10) Spectrum Analyzers. Various auxiliary testing instruments are also employed, including vacuum tube testers, I.F. signal generators, audio oscillators, flux meters, etc. Before discussing the above tests and devices individually, some of the requirements for radar test equipment, especially those resulting from military usage, will be summarized.

Generality of Application

Radars perform a great variety of functions, including search and surveillance, gun laying and fire control, bomb direction, and navigation. They are used in the air, on shipboard, and on the ground, in attack and in defense, in combat zones and in rear areas. To realize the advantages of different parts of the frequency spectrum, avoid interference, and keep ahead of enemy countermeasures, it has been necessary for radar to exploit many different frequency bands and sub-bands. These diversities have

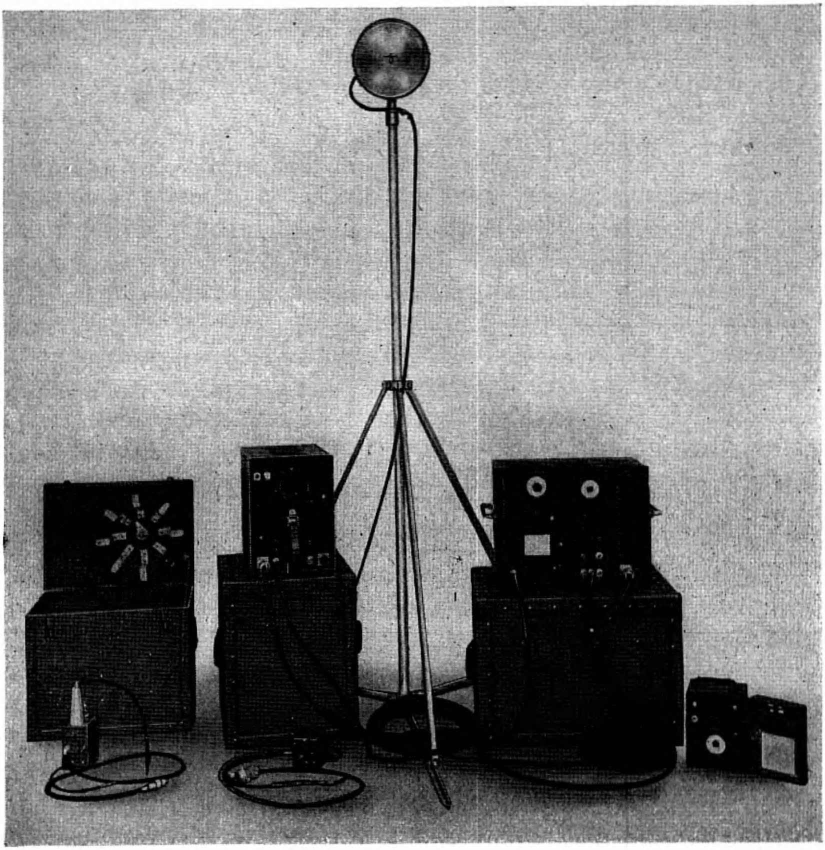


Fig. 2—Early radar test set for 3000 mc range—includes signal generator, oscilloscope, power meter, test antenna and auxiliary units.

bred a multiplicity of types of radar, with a corresponding variety of testing problems and requirements.

Maintenance and testing of radars must be performed at many different locations. In the Army these locations, known as echelons, include (a) the operating unit, (b) central service points, either fixed or mobile, for a number of operating units, and (c) large depots either in the military theatres or in the United States. Navy maintenance and testing are carried out on board the fighting ships or auxiliaries, and at advance and overhaul bases.

In developing test equipment, an offhand approach might have been to provide specialized equipment for testing each radar under specific conditions. Since this would have made the total burden of development, manufacture and field maintenance well nigh intolerable, a coordinated plan was followed whereby (with minor exceptions) each test set was made capable of widespread application in testing as many radars under as varied conditions as possible.

Broadbanding

Generality of application required the designing of test equipment for broad frequency bands, bands as a rule much wider than those of the radars themselves. It was necessary, therefore, not only to develop new microwave testing techniques, but to advance the art still further to render the testing components as far as possible insensitive to frequency.

Precision

A radar itself is an instrument of considerable precision. The test equipment used for checking the radar performance in the field has to have still higher accuracy. It is noteworthy that the measuring accuracy realized throughout the microwave range is comparable with that obtainable at lower frequencies where many years of background exist.

Packaging—Size and Weight

Light weight and compactness are of paramount importance where a test set has to be carried any distance by the maintenance man, where it is used in cramped quarters in a plane, truck or submarine, or where it has to be taken up ship ladders or through small hatchways. To permit portable use under such conditions, the design objective was established of a weight not exceeding about 30 lbs. (exclusive of transit case), combined with a ruggedness adequate for all conditions of use. Through rigorous attention to both mechanical and electrical design, this objective has been realized (in many cases with considerable margin) except for a few sets intended primarily for bench use. Figure 3 illustrates the use of lightweight test equipment in maintaining airborne radars.

Environmental Influences

Military usage requires that the test equipment be capable of efficient operation at any ambient temperature between a minimum of the order of $-40°$ to $-55°$ C and a maximum of the order of $+65°$ to $+70°$ C, as well as at any relative humidity up to 95%. In addition the set must withstand continued exposure to driving rain, dust storms and all other conditions encountered in tropical, desert or arctic climates. Often the test set in its transit case must be capable of submergence under water without ill effects. Fungus-proofing with a fungicidal lacquer is a standard requirement.

Simplicity, Reliability, Accessibility

Not only must the functioning of the test set be reliable, stable and trouble-free, but the set must make minimum demands for special skill or technique on the part of the maintenance man. Access for maintenance purposes, while important in radar test sets, is not as controlling as in the radars themselves and sometimes has to be sacrificed in part for compactness.

Fig. 3—Portable units used in checking airborne radars at Boca Raton, Fla.

Ruggedness

For general application, the test equipment must be capable of withstanding airplane vibration, the shock of heavy guns, depth charges and near misses, and the combinations of shock and vibration connoted by the requirement of "transportation over all types of terrain in any Army vehicle." Test and experience have made it possible to translate these general requirements into two specific requirements, namely the ability to withstand (1) vibration at frequencies from 10 to 33 cycles per second with $\frac{1}{16}''$ excursion for 30 minutes in each of three axes and (2) the shock produced by a 400 lb. hammer falling through distances of 1, 2 and 3 ft. in each of three axes, and striking an anvil to which the set is attached. These requirements have been met without using shock and vibration mounts, which are undesirable in test sets because they increase size and weight.

Range Capability

The range capability of a radar, like that of any radio system, depends upon three things; the transmitted power, the loss in the medium, and the minimum perceptible received signal. Two of these can be combined by taking the ratio of the radiated signal to the minimum perceptible received signal. This ratio, ordinarily expressed as a level difference in db, is variously termed the "system performance," "over-all performance" or merely the "level difference." It may be determined by separate measurement of transmitter power and receiver sensitivity, or by a single overall measurement. With the powers and sensitivities commonly employed in radar, the level difference is of the order of 150 to 180 db.

The actual range that can be spanned for a given performance ratio varies considerably. For a given transmitted power, the echo power received by a radar theoretically varies inversely as the fourth power of the range. The reason for this is simple. In free space the power intercepted by a target which is small in comparison with the area of the radar beam in its vicinity will vary according to the inverse square of the distance from the radar. Similarly, that fraction of the energy reflected from the target which is intercepted by the receiving antenna will vary as the inverse square law. Since the received power involves the product of these two factors, the relation becomes:

$$P_r = K\frac{P_t}{R^4}, \quad \text{or}, \quad R = \left(K\frac{P_t}{P_r}\right)^{\frac{1}{4}} \tag{1}$$

where P_t and P_r represent, respectively, the transmitted and received power, R the range and K a constant determined by antenna design, character of target, etc.

Under operating conditions considerable departure from the above relationship may be experienced, due to such factors as (1) the curvature of the earth, (2) interference between the direct beam and single or multiple reflections, and (3) attenuation due to atmospheric absorption. Except under conditions of severe attenuation such as may occur at the very short wavelengths, the received power commonly varies somewhere between the inverse fourth power and the inverse 16th power of distance. To state it another way, the change in effective range is somewhere between the fourth and the sixteenth root of the change in system performance. The former condi-

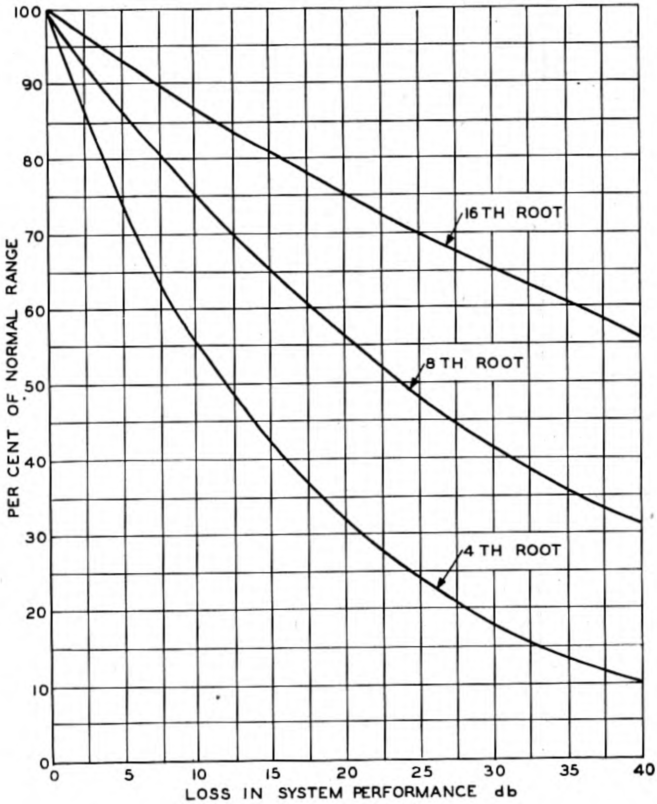


Fig. 4—Effect of reduction in system performance on radar range.

tion might hold for high angle plane-to-ship search in clear weather, the latter for ship-to-ship search in fog. The loss of range resulting from a given degradation of system performance is shown for 4th, 8th and 16th root laws by the curves of Fig. 4. Whatever the propagation law, reduction in performance always means loss of range.

Field surveys have shown that when test equipment is not available or not used, radars in the field are likely to give no more than $\frac{1}{2}$ to $\frac{1}{4}$ of the maximum range of which they are capable. Hence it is necessary to know with good accuracy the over-all performance. Known, or so-called standard, targets have often been used in the field for checking performance. Because of wide variations in transmission due to many factors, results obtained from such targets are frequently misleading.

Signal Generators

Signal generators for radar work deliver one or more different types of test signals, which may serve a variety of functions. More important among these functions are tuning or alignment of the radar components (TR and RT boxes, converter, beat oscillator, AFC, etc.), measurement of receiver sensitivity, checking TR and receiver recovery, measurement of loss, detection of frequency pulling, check of AFC following, measurement of IF bandwidth, check of automatic range tracking, measurement of standing wave ratio and check of video "gating" circuits. Many signal generators include means for measuring the power and frequency of the test signal, and also of an incoming signal.

Types of Signals

The test signals delivered by a signal generator may be CW, pulsed, or frequency-modulated. Occasionally square wave or sine wave modulation is provided.

Pulsed signal generators deliver a succession of single RF pulses or pulse trains, either of these generally synchronous with the pulses of the radar under test. Multivibrator or trigger techniques[1] are used to generate the pulses for modulating the microwave generator. The trigger pulse for synchronizing the pulsing circuits is commonly produced by rectifying RF pulses from the radar transmitter, thus avoiding a separate video connection to the radar. To avoid possible difficulties in video response, the RF test pulses should be of comparable width to those of the radar under test. For observing the test signals, either the radar indicator or an auxiliary oscilloscope may be used. With the single pulse method, provision is usually made for varying the delay of the test pulse with respect to the radar pulse. The width of the test pulse may also be adjustable or variable.

If the frequency of the signal generator is swept over a sufficiently wide frequency band, the IF output of the radar traces the curve of IF selectivity, thus producing a kind of pulse. With a suitable rate of frequency sweep, this pulse becomes comparable in width to the transmitter pulse, and when synchronized with the radar it can be used for test purposes. Since the pulse is produced in the radar, comparison of the shapes of receiver input and output pulses is not possible. The nominal duration of the pulse in the IF output is

$$T = B/\gamma \qquad (2)$$

where B is the width of IF band in cycles per second (for this purpose conveniently measured between 6 db points) and γ is the speed of frequency sweep in cycles per second per second. For best results this nominal width should be similar to the width of transmitter pulse for which the IF and video circuits are designed. This means that the scanning speed should be in the neighborhood of $B^2/2$.

A CW input of the same frequency as the transmitter produces in the output of the IF detector merely a direct current to which the video amplifier and radar indicator do not respond. However, CW test signals may be utilized by observing on a d.c. meter (built into the radar or separate) the change produced in detector current or converter current.

Receiver Sensitivity

Just as in radio, the sensitivity of a radar receiver is defined as the minimum received signal that is perceptible in the presence of set noise.* At microwave frequencies atmospheric disturbances are usually negligible, so that unless accidental or deliberate interfering signals are present, the operating sensitivity is the same as the intrinsic sensitivity of the receiver.

Receiver sensitivity is commonly stated as the minimum perceptible signal power in db referred to a milliwatt, (abbreviated as dbm). In practice, the receiver sensitivity depends upon the noise figure of the converter, the conversion loss, and the noise figure of the IF amplifier. If an RF amplifier is used, as is the practice at lower microwave frequencies, its noise figure is likely to be controlling. By noise figure in each case is meant the noise power in comparison with the thermal noise. The thermal noise in watts delivered to a load is kTB, where k is Boltzmann's constant, T is absolute temperature in degrees K, and B is the frequency bandwidth. Thus for a 4 mc band at 25° C the thermal noise is -108 dbm. With good design the over-all receiver noise is of the order of 10 to 15 db higher than thermal noise. The minimum detectable signal is usually not equal to the receiver noise but depends on the type of indicator, particularly on whether the presence of an echo is indicated by spot deflection or spot modulation.

With a CW signal generator, receiver sensitivity is measured by determining the minimum input power necessary to produce a perceptible change in meter reading. This affords a satisfactory relative measure of receiver performance, but since the radar indicator usually permits better visual discrimination against noise, the minimum input as read with the meter ordinarily differs from the minimum pulse input for barely discernible indicator response.

* The term noise is commonly used even though the disturbances are observed on a cathode-ray screen.

RF Oscillators

Beat oscillator tubes for radars deliver (with sufficient decoupling or isolation to prevent undue frequency pulling) a power of the order of milliwatts. This power being adequate for most test purposes, such tubes are well adapted for use as signal generators.

Throughout the greater part of the microwave range, reflex velocity-modulated tubes,[2] both the type with built-in cavity (Pierce-Shepherd) tuned mechanically or thermally, and that with external cavity (McNally) tuned by plugs, vanes or adjustment of dimensions, have been used. The former is more convenient for general use but the latter usually permits wider frequency coverage. Oscillation occurs when the repeller voltage is adjusted so that the round trip transit time corresponds to an odd number of quarter wavelengths. Ordinarily there are several different ranges of repeller voltage, corresponding to different numbers of quarter waves, each of which supports oscillation over a range of frequencies, called a mode of oscillation. Pulsing or frequency modulation is accomplished by applying a pulse or sawtooth wave to the repeller.

At the longer microwaves, a triode with closely spaced electrodes, or so-called "lighthouse" tube[3], has been employed in a tuned-plate tuned-grid oscillator of the positive grid type. Two coaxial lines, conveniently placed one inside the other, provide the tuning, with the feedback through interelectrode capacitances supplemented by loop coupling. The inner cavity (between plate and grid) controls the frequency of oscillation, while the outer cavity (between grid and cathode) provides a suitably high grid impedance. Mechanical arrangements are provided for tracking the tuning of the two cavities over a wide frequency range.

Some Design Principles

Standard signal generators which have been employed in the past for measuring the sensitivity of radio receivers usually deliver a known voltage across a low impedance. This voltage is applied in series with a dummy antenna to the receiver under test. In the microwave range this technique is inconvenient, and signal generators are designed to deliver test power on a matched impedance basis. Receiver sensitivity is stated in terms of power (dbm) instead of volts.

The components of a signal generator or other test unit are commonly arranged along a microwave transmission line. The wave guide type of line possesses certain advantages over a coaxial line in affording a lower loss, facilitating attenuator design as discussed in a subsequent section, etc. Hence the wave guide type of line is used in test equipment for those wavelengths where its size is not excessive, i.e. from about 4,000 mc upwards, and coaxial line for lower frequencies.

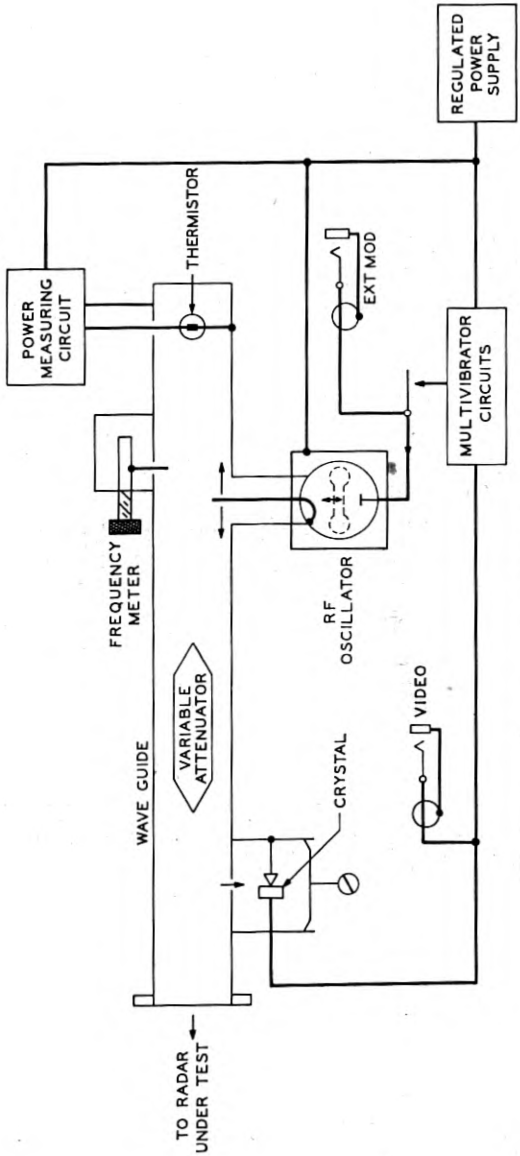


Fig. 5—Block diagram of TS-35A/AP signal generator.

A major problem in the design of microwave signal generators is the provision of shielding adequate to reduce leakage signals due to unwanted couplings or stray fields well below the minimum signal required for receiver testing. This minimum level may be as low as -70 or -80 dbm, depending on the coupling loss in the test connection.

A Pulsed and FM Signal Generator

To illustrate the functioning of a signal generator, there is shown in Fig. 5 a block schematic of a design (Army-Navy type TS-35A/AP) which covers a 12% frequency band in the vicinity of 9000 mc. An RF connection to the radar is established with a wave guide flange coupling. The frequency and power of the radar transmitter are measured by means of a coaxial-type frequency meter and thermistor power measuring circuit as described in subsequent sections. The attenuator and pad are adjusted to reduce the incoming average power to about 1 milliwatt, which gives a suitable deflection on the indicating meter. The thermistor is mounted across the wave guide.

An RF pulse train is employed in many tests. To produce this the RF oscillator output is modulated by a multi-vibrator which pulses continuously except when being synchronized. Synchronizing pulses are derived by crystal rectification of the RF pulses from the radar transmitter. The result is an initial RF pulse of 7 microseconds followed by an off period of about 10 microseconds followed by a train of RF pulses each 2 microseconds wide and recurring every 8 microseconds until resynchronization occurs at the next radar pulse.

Using the pulse train, the radar system components can be tuned for maximum sensitivity by maximizing the signal on the indicator. To check receiver sensitivity the CW power is first adjusted so that a power of 1 milliwatt is delivered to the pad and attenuator. Then with the set in the pulsed condition the amplitude of the test signal is adjusted by means of the attenuator and pad until the signal on the radar indicator is barely discernible. It is necessary for this test that the frequency of the test signal be equal to the magnetron frequency. The frequency meter is provided as part of the signal generator for this purpose.

The receiver recovery, i.e. the time required by the receiver to recover after disablement by the transmitter pulse, determines the minimum range at which a radar can be used. With this test set the receiver recovery characteristic is indicated by the amplitude of the test pulses in the interval immediately following the transmitter pulse.

The set is also adapted to serve as an FM signal generator. A sawtooth wave applied to the repeller gives a succession of frequency sweeps, each

Fig. 6—1942 model signal generator compared with one produced in 1945.

about 20 mc wide, and lasting about 6 microseconds. With this frequency modulated signal the width of receiver response may be observed on a Class A oscilloscope (i.e. one showing signal amplitude vs. time). However, with non-adjustable IF strips such measurement is seldom required. Failure of the radar AFC to follow frequency changes due to antenna scanning or other causes is indicated by a change in the indicator presentation. Pulling of the magnetron frequency due to changes in load impedance can be detected by turning off the AFC.

Signal Generator Designs

Designs of signal generators developed for the military arms during the war are interesting as landmarks of progress. The signal generator of the IE30 test set, deliveries of which began in May 1942, delivered pulsed RF signals in the 10 cm range, using sine wave synchronization. Following only three months later was the signal generator of the Army IE57A and Navy LZ test sets (Fig. 6), which covered a then very broad frequency band of 20% in the vicinity of 10 cm, and was designed to be triggered by the incoming RF pulse from the radar instead of by a separate synchronizing connection. This set and a redesigned version of it have seen wide usage in testing Army, Navy and Marine Corps radars.

Delivery of a test set for the 3 cm range, designated TS-35/AP, started in the fall of 1943. This set furnished both a train of pulses and a train of FM signals, both of which features have proved valuable. It covered a 9% frequency band with no tuning adjustment except for the oscillator. An improved design known as TS-35A (see Fig. 6) covered a 12% band.

Progress in reducing the size and weight of the test units is indicated by the fact that the IE30 signal generator weighed 121 lbs., IE57 74 lbs., whereas TS-35 and TS-35A weigh approximately 30 lbs.

Frequency Measurement

Usually a radar need not operate at a precise frequency. Accurate measurements are required in the field, however, to keep the operating frequency within limits, to set the local oscillator, to check the measuring frequency, etc. In the laboratory, accurate frequency measurement is fundamental.

Frequency measurement in the microwave range is ordinarily accomplished by (1) a resonant coaxial line or (2) a resonant cavity, generally cylindrical. These types are illustrated in Fig. 7. Sometimes a combination of the two, referred to as a hybrid or transition type resonator, is employed. The measurement is actually one of wavelength, with the scale calibrated in frequency or a conversion chart provided. Some specific designs of frequency meters are shown in Fig. 8.

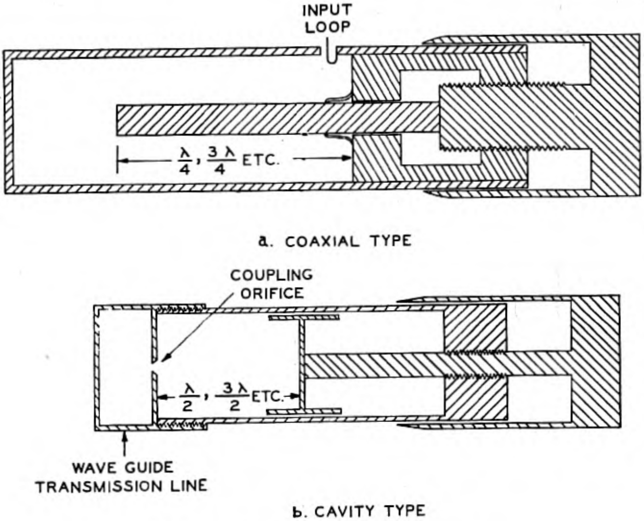


Fig. 7—Types of frequency meters.

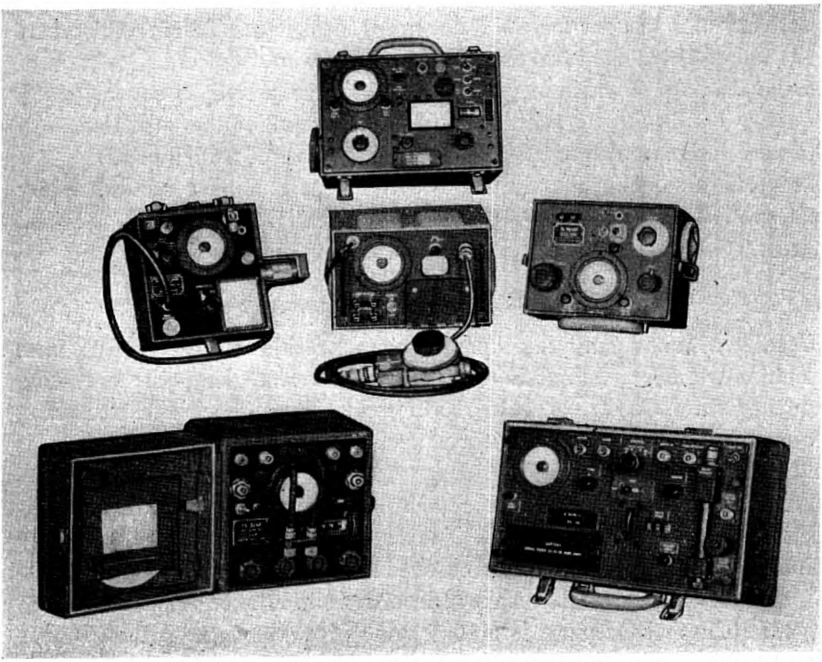


Fig. 8—A group of frequency and power meters designed for various bands in the frequency range 500 to 25,000 mc.

Coaxial Wavemeters

A coaxial wavemeter is formed of a section of coaxial transmission line of small enough diameter so that only the coaxial mode (in wave guide notation $TM_{0,0,n}$) can exist. Usually the line is short-circuited at one end and open at the other, in which case resonance occurs at the odd quarter wavelengths $\lambda/4$, $3\lambda/4$ etc.). The open circuit is obtained merely by terminating the inner coaxial conductor, the continuing outer conductor acting as a wave guide below cutoff. Sometimes the line is short-circuited at both ends, giving resonance at the even quarter-wavelengths ($\lambda/2$, λ, etc.).

Cylindrical Cavity Wavemeters

A cylindrical cavity wavemeter is merely a section of cylindrical wave guide transmission line[4] whose length is varied. In order to avoid confusion with other modes, it is preferable to use the dominant mode (described in wave guide notation, $TE_{1,1,n}$*) i.e. the mode with the lowest cutoff frequency. The cutoff wavelength (λ_c) of this mode is 1.706D, where D is the diameter in meters. For a higher Q it may be necessary to use the circular electric mode $TE_{0,1,n}$. The cutoff wavelength for this mode is .82D. No useful purpose is served by using modes with l and m subscripts above unity. *TM* modes are often used for fixed frequency cavities, but for variable cavities *TE* modes are preferable since these have zero current at the inner wall of the cylinder and thus obviate moving contact difficulty. If any mode higher than the dominant one is used, suppression of unwanted modes may be required.

The accuracy of a wavemeter is dependent on its resolving power. This in turn depends upon Q, which is an index of the decrement of the resonant circuit, and is equal to $f/\Delta f$, where Δf is the distance between 3 *db* points on the resonance curve.

In a coaxial wavemeter, maximum Q for a given inner diameter is obtained with a diameter ratio of about 3.6[5] The basic Q of a coaxial wavemeter, assuming copper of standard conductivity, is roughly[6]

$$Q_0 = 0.042D\sqrt{f} \qquad (3)$$

This expression neglects end effects and hence gives somewhat too high a value of Q.

The basic Q's for $TE_{1,1,n}$ and $TE_{0,1,n}$ cylindrical cavity resonators employing copper of standard conductivity are, respectively,

* *TE* and *TM* represent, respectively, transverse electric and transverse magnetic. The subscripts l, m, n denote, respectively, the number of wavelengths around any concentric circle in the cross section, the number of wavelengths across a diameter, and the number of half wavelengths along the length of the cylinder.

$$Q_0 = 0.0937 \times 10^{10} \frac{A^3}{\sqrt{f}} \frac{1}{1 + B^2\left(0.826 \frac{B}{n} + 0.295\right)} \tag{4}$$

$$Q_0 = 0.2762 \times 10^{10} \frac{A^3}{\sqrt{f}} \frac{1}{1 + B^2\left(2.439 \frac{B}{n}\right)} \tag{5}$$

with $A = \frac{\lambda_c}{\lambda}$, $B^2 = A^2 - 1$, and f is in cycles per second.

The value of Q which determines accuracy is not the basic Q, but the loaded or working Q, herein designated Q_L.

The resolving power of a wavemeter used for measuring a single frequency can be made considerably better than f/Q_L. With a sensitive meter it is readily possible to detect differences less than 1 db, which corresponds to a frequency interval of $f/2Q_L$.

The required accuracy in a wavemeter is generally absolute rather than a percentage. Hence increasingly large values of Q_L are required at the higher frequencies. Thus for a resolution of 1 mc, assuming 1 db discrimination, the values of Q_L required for different frequencies are:

Frequency	Q_L	Frequency	Q_L
1,000 mc	500	10,000 mc	5,000
3,000 mc	1,500	25,000 mc	12,500

An unnecessarily high value of Q_L has the disadvantage of making it more difficult to find the desired frequency.

Linearity

The displacement of the coaxial plunger of a coaxial type wavemeter for resonance is substantially a direct linear function of free space wavelength and if an ordinary centimeter micrometer drive is used it is possible to read wavelength differentials directly. Over bandwidths less than 20 per cent, displacement vs. frequency is also quite linear which is of considerable advantage for some uses.

For the cavity type wavemeter the displacement is a variable function of free space wavelength and becomes very non-linear as the cutoff frequency of the guide or cavity is approached. This is evident from the relation between wavelength in the guide, λ_g, wavelength in free space, λ, and cutoff wavelength, λ_c:

$$\lambda_g = \frac{\lambda}{\sqrt{1 - \frac{\lambda^2}{\lambda_c^2}}} \tag{6}$$

A cam or mechanical linkage may be employed to obtain a linear scale.

Frequency Coverage

Increasing the length of a wavemeter is desirable because this gives a larger mechanical displacement for a given frequency interval. The permissible increase is limited, however, by ambiguity with the next lower mode in going toward the upper end of the frequency scale and with the next higher mode in going toward the lower end. This means that, if ambiguity is to be avoided, the ratio of top to bottom frequency cannot exceed $(n + 2)/n$ for a coaxial line of n quarterwaves. For a cylindrical cavity resonator the ratio for the $0,1,n$ mode must be less than $n + 1)/n$, the exact limit depending on proximity to cutoff.

Guideposts

The following guideposts are suggested for choosing the type of wavemeter in the microwave range. Where limitations of size and Q permit, the coaxial quarter-wave type should be used because of its greater linearity. If this type is inapplicable, the cavity type with $TE_{1,1,n}$ should be used unless its Q is inadequate, in which case $TE_{0,1,n}$ should be employed.

Couplings

A loop, orifice or probe may be used for coupling to a wavemeter. Coupling to a coaxial wavemeter is generally effected by a loop placed near a short-circuited end so as to be in the maximum magnetic field. For coupling to a cylindrical cavity wavemeter, an orifice in or near the base of the cavity is usually employed. The coupling to the wavemeter is kept small enough to avoid serious reduction of loaded Q.

Types of Detectors

Various types of detectors may be associated with a wavemeter, the most commonly used being (1) a crystal rectifier and microammeter, or (2) a thermistor bolometer. When a crystal rectifier is employed with a cavity or coaxial wavemeter a circuit similar to that shown in Fig. 9 is used. Important items in such a circuit are the "RF by-pass" condenser, and the "video" condenser. The latter, by providing a low-impedance path to the video signals, improves rectification efficiency when the input signal is pulsed. The quarter-wave stub shown in the figure is used when the input or coupling circuit does not provide DC and video paths. When the signal is pulsed at a low duty cycle, high peak currents through the crystal are obtained even though the average current through the meter is small, and it is possible to impair or burn out the crystal unless extreme care is taken. The use of a thermistor, which is self-protecting for large overloads, avoids this danger. Another expedient is to limit the crystal current to a small value and employ a video amplifier and oscilloscope.

Methods of Use

A wavemeter may be used as either a transmission or a reaction instrument. In the former case (Fig. 10a) it is inserted directly in the transmission path, so that substantially no through transmission occurs except at resonance. In the latter case (Fig. 10b) the wavemeter is coupled to the transmission path or a branch circuit. When the meter is tuned off resonance it presents such a high impedance to the main path that its effect is negligible. At resonance, however, it offers a lower impedance which reflects energy in the main line so that less power reaches the detector and

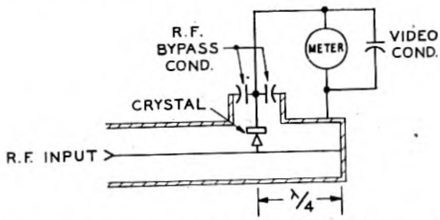


Fig. 9—Crystal detector for pulsed RF signals.

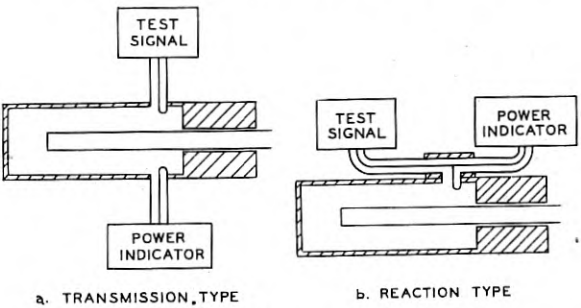


Fig. 10—Types of frequency meter circuits.

a dip in the reading occurs. For most applications the reaction arrangement is preferable since the power transmitted when the wavemeter is off tune may serve various purposes. For analysis of frequency spectrum the transmission method is necessary. This method requires two couplings to the wavemeter, which lowers the Q as compared with the reaction type.

Drive and Scale

A direct drive with a precision lead screw of the micrometer type is frequently used. Accuracy of reading is insured by spring loading to minimize backlash and by specifying close tolerances for threads and for concentricity of plunger and cavity. The scale on a wavemeter must be fine enough to permit utilization of the resolving power. The conventional micrometer type of scale sometimes makes accurate reading difficult in a small compact meter. Scale mechanisms which have been used include counter types, clock face or expanded drum types with gearing between vernier and coarse scales, and a single direct reading scale with divisions arranged in a spiral for compactness.

Effect of Temperature and Humidity

To minimize the effects of temperature on the accuracy of readings, invar has been employed for elements whose dimensions affect the wavelength. For accurate work, the scale reading must be corrected for temperature. A rough approximation is that the scale reading varies in accordance with the coefficient of expansion of the metal.

Water vapor included in the air dielectric of a wavemeter has an appreciable effect on the dielectric constant and hence on the resonant frequency.[7] Thus, for example, in going at sea level from 25° C, and 60% humidity to 50° C, and 90% humidity, the scale reading should be reduced by .03%. Correction can be made by means of a chart, a convenient form of which has been prepared by Radiation Laboratory.

Calibration

Frequency meters are calibrated against sub-standards which in turn are calibrated against a multiplier from lower frequencies for which a high order of accuracy can be obtained. Such multipliers have been made available at Radiation Laboratory and the National Bureau of Standards. The accuracy obtained at interpolation frequencies is of course less than at exact multiples of the base frequency. In the microwave range the accuracy is believed to be of the order of one part in 100,000.

POWER MEASUREMENT

There are two needs for power measurement in radar maintenance, namely (1) in evaluating transmitter performance and (2) in standardizing test signals. Power output is, of course, only one factor in transmitter performance, others being (a) frequency and (b) spectral distribution or shape of RF envelope. Ability to measure absolute power is desirable to permit interchangeable use of test sets in the field.

Measurement of Pulse Power

The transmitter power as used in Formula (1) is the average power during the pulse. The relationship of pulse to (long) average power is

$$\frac{P_t \text{ av.}}{P_t \text{ pulse}} = Tf_r \tag{7}$$

T is the pulse duration in seconds and f_r is the pulse recurrence frequency (P. R. F.) in cycles per second. The product Tf_r is the duty cycle. (Sometimes the reciprocal of this number is referred to as the duty cycle. The magnitude is usually such that no ambiguity arises.)

During the early days of radar it was the practice to measure pulse power. The test equipment was coupled to the radar by a path of known loss. The RF envelope was derived by means of a crystal rectifier and applied to an oscilloscope. With the aid of an RF attenuator the level applied to the crystal rectifier and oscilloscope was held constant. Calibration was obtained by using a signal generator whose output was standardized, prior to pulsing, with an averaging type of power meter. The procedure was rather involved, with several sources of possible error. Since it is much simpler to measure (long) average power, field measurement of pulse power was soon abandoned. Though the pulse power can be computed from average power if the pulse width, pulse shape and repetition rate are known, it soon became the practice to specify field performance requirements in terms of average power.

Thermistor Power Meters

A number of devices have been used for measuring average power in the microwave range. Those suitable for handling the small amounts of power normally involved in field tests include (1) thermistors, (2) platinum wires and (3) thermocouples. In each case the RF power to be measured is absorbed in the measuring element. The measurement consists in observing the resistance change in the thermistor or platinum wire, or the thermoelectric voltage from the thermocouple. By analogy with devices used for measuring minute quantities of radiant heat, either a thermistor or a platinum wire instrument is sometimes referred to as a bolometer. The platinum wire device has also been termed a barretter.

A thermistor for microwave power measurement is a tiny bead (about 5 mils in diameter) composed of a mixture of oxides of manganese, cobalt, nickel and copper, constituting a resistor with a very high negative temperature coefficient.[8]

The thermistor has a number of advantages for microwave work, namely: (1) resistance is highly sensitive to change of heating power, which obviates any need for amplification or a super-sensitive meter, and makes it possible to use a rugged d.c. meter; (2) reactance is low compared with RF resistance, which makes it possible to incorporate the thermistor in a power absorbing termination which matches the impedance of a microwave transmission line over a wide band; (3) resistance change is the same function of electrical heating power at any frequency, which permits direct comparison of the unknown microwave power with easily measurable d.c. power; (4) sensitivity to damage and burn-out is inherently low, and added protection results from impedance mismatch during overload. Because of these characteristics thermistors have been far more widely used than other detectors for microwave power measurement. Broad-band thermistor mounts have been designed to match both wave guide and coaxial transmission lines, the latter not only in the microwave range but also down to low frequencies. Some of the test sets specifically intended for power measurement or for combined power and frequency measurement are shown in Fig. 8.

The change in the thermistor resistance due to RF heating current is determined by placing the thermistor in one arm of a d.c. bridge. By noting the d.c. power necessary to balance the bridge with and without RF power in the thermistor, the magnitude of the RF power may be determined. For most purposes, however, a direct reading power meter is preferable. This can be obtained over a moderate range of power levels by employing an unbalanced bridge. The bridge is balanced for d.c. only and the measurement consists in noting the meter deflection when RF power is added.

The resistance of a thermistor is a highly sensitive function not only of electrical heating power but also of ambient temperature. For convenient field measurement, the effect of ambient temperature must be cancelled out in the indicator circuit so that the indication depends only on RF power.

Water Loads

A method which has been used in the laboratory and factory for measuring high-level microwave power consists in terminating the RF transmission line in a water load arranged as a continuous flow calorimeter. This method can be made quite accurate but is cumbersome. More recent practice is to terminate the RF line in a solid load of a type described later in this article, and to couple a thermistor power meter to the line by means of a directional coupler (described below) of known loss. Very close correlations have been obtained between the two methods over the entire microwave band.

Echo Boxes

A device unique to radar testing is a high Q resonant cavity, known as an "echo box" or "ring box." The cavity is coupled to the radar transmission line or antenna as indicated in Fig. 11. During the transmitted pulse, microwave energy is stored in the cavity. In the period immediately thereafter, energy is returned to the radar over the same path, producing a signal on the radar indicator. The energy in the cavity builds up exponentially to an amplitude dependent on the radar power. At the end of the pulse the returned energy decays exponentially, disappearing into the noise at a point determined by receiver sensitivity. The time interval between the end of the transmitted pulse and the point where the signal on the radar indicator disappears into the background noise, called the "ring time," therefore measures the over-all performance of the radar.

An echo box is a particularly useful instrument for radar testing because it measures over-all performance directly, because it permits a rapid tune-up, and because it utilizes the radar transmitter as its only source of power and therefore can be made extremely portable. Figure 12 shows typical ring-time patterns on different types of indicators. In actual practice the ring-time is read in miles on the radar range scale and hence is measured from

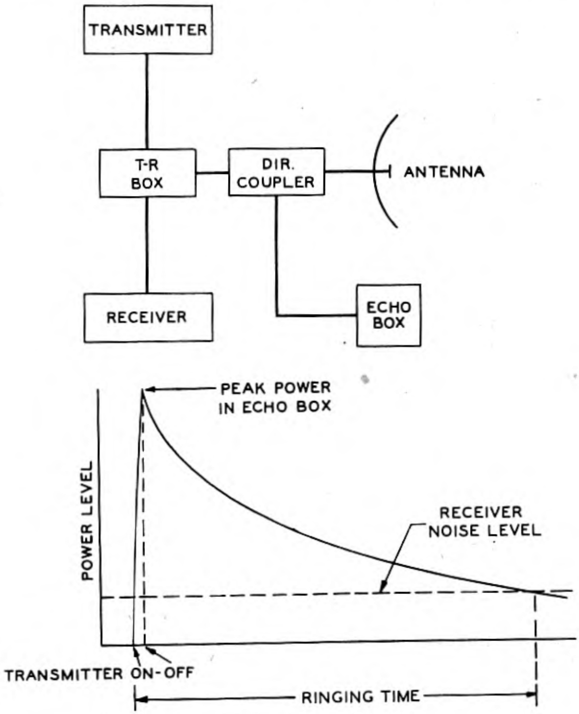


Fig. 11—Radar test with echo box.

the beginning of the transmitted pulse rather than the end. The difference is unimportant, however, since standard and limiting values of ringtime are established for operating conditions. It will be noted that the echo box does not return a true echo to the radar, so that the name "echo box" is not entirely appropriate.

Types and Uses

Echo boxes are of two general types, tuned and untuned. A tuned echo box is designed to resonate in a single mode adjustable over the operating frequency range. An untuned echo box is a fixed cavity of a size sufficient to support a very large number of modes within the working range. Tuned echo boxes are more versatile and more widely used than untuned boxes.

The most common type of tuned echo box is designed for hand tuning. While other shapes are possible, the most convenient one is a right cylinder whose length is adjusted by a movable piston. The $TE_{0,1,n}$ mode gives

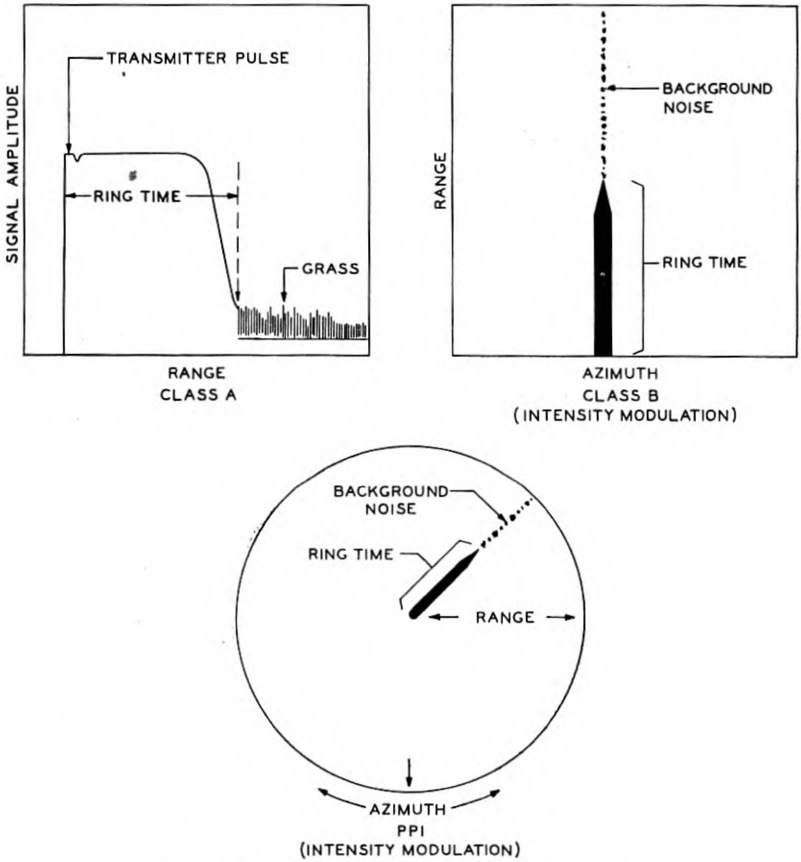


Fig. 12—Typical ringtime patterns on radar indicators.

maximum Q for a given volume and minimizes the number of unwanted modes present within the desired band. The value of n is determined by the desired value of Q (see formula 5). Unwanted modes can be partially avoided by choice of design parameters. However, for high values of Q, and especially for broad frequency bands, the suppression of unwanted modes involves design problems of the highest order.

As indicated in Fig. 13, a tuned echo box cavity is usually provided with two couplings. One of these is to the radar pick-up; the other to an attenuating device, crystal rectifier, and meter, which serve for tuning the cavity and for other purposes. With such an instrument, not only can the radar be tuned up and its over-all performance determined, but many other tests can be made, to wit: (1) the setting of the plunger at resonance indicates the transmitter frequency or wavelength; (2) calibration of the crystal affords a rough measure of output power; (3) since the Q required for adequate ringtime is so high that the cavity selects only a narrow segment of the transmitter spectrum, a spectrum analysis can be made by plotting frequency versus crystal current reading; (4) slow recovery of TR box and receiver after the transmitted pulse can be detected by noting the behavior of the ringtime pattern at short ranges as the echo box is detuned; (5)

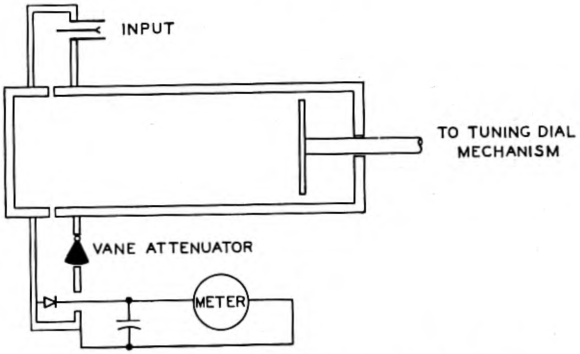


Fig. 13—Functional schematic of tuned echo box.

inability of the receiver to recover promptly after a strong signal (the result of imperfect d.c. reinsertion in the video amplifier or of overloading of the I.F. amplifier) is indicated by a blank following the end of the ring; (6) improper pulsing (e.g. double moding or misfiring) can be determined with a class A oscilloscope; (7) the frequency and power of the local oscillator can be measured. In tuned echo boxes, requirements for extreme fineness of tuning control and precise resettability have given rise to interesting problems in the design of the mechanical drive and indicating mechanism.

In another type of echo box, hand tuning is supplemented by motor-driven tuning or so-called "wobbling" over a frequency range wide enough to embrace expected variations in transmitter frequency. Operation is controlled by a single push-button which energizes the motor and actuates the cavity coupling. Such an instrument may be permanently installed in a plane and used to check the radar during flight.

For an untuned or multi-resonant echo box, rectangular shape is convenient. The box should be large enough to make it highly probable that over the operating band one or more modes will be present within any frequency interval of width equal to the main concentration of the transmitter spectrum. For a given rectangular volume a cube gives the largest number of modes. The total number of modes up to a frequency of wave length λ is

$$N_M = 8.38\ V/\lambda_0^3 \tag{8}$$

where V is the volume. However, because of the cubical shape many different modes tend to coincide in wavelength, a condition referred to as

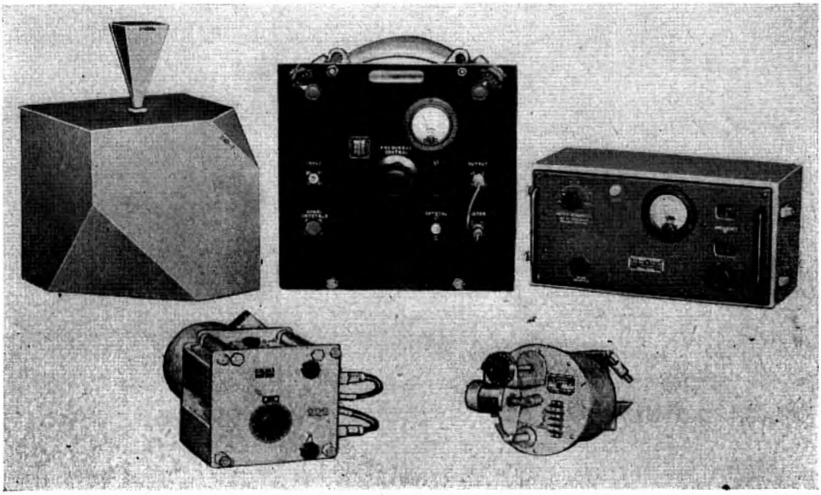


Fig. 14—A group of echo boxes of various types.

degeneracy. To spread out the modes, the box is made slightly off cube and one or more corners sliced off. At the longer microwaves the size of box is determined by the number of modes, and the size becomes quite awkward. For the shorter microwaves the size is determined by the required value of Q. Hence the use of untuned echo boxes has been limited to the frequency range from about 9000 mc upwards with sizes of the order of 12 to 24 inches on a side. Even with an extraordinarily high probability of finding modes within the radar band, substantial differences in response are found for relatively small changes in frequency. Accordingly untuned echo boxes are more useful for rough tune-up than for precise measurement.

A number of specific designs of echo boxes for different microwave bands are shown in Fig. 14.

Q and Ring Time

For satisfactory measurement the ringtime must extend beyond nearby echoes which would obscure the test signal. For most radars a ringtime of 20 to 30 microseconds (about 2 to 3 miles) has been found satisfactory although considerably higher values have sometimes been provided. Even apart from echoes, a long ringtime is desirable since this gives a lower decay rate and a more sensitive measurement.

Computation will show that an extremely high value of Q is necessary to obtain the desired ringtime. For maximum ringtime the cavity coupling should be such as to make the working Q (Q_L) about 90 per cent of the non-loaded Q. Values of working Q which have been provided in different frequency ranges are approximately as follows:

Frequency	Q_L	Frequency	Q_L
1,000 mc	70,000*	10,000 mc	100,000
3,000	40,000	24,000	200,000

* In this case a higher Q was needed for a long range ground search system.

The difference in performance corresponding to a given change in ringtime can be determined from the decay rate which is

$$d = 27.3\, f/Q_L \ db/\text{microsecond} \tag{9}$$

For a given frequency the ringtime is directly proportional, and the decay rate inversely proportional, to Q. For a given ringtime, the required Q is directly proportional to frequency.

Accurate measurement of extremely high Q's is essential in echo box work. A decrement method, in which a pulsed RF oscillator and oscilloscope are used to determine the loss corresponding to a known time interval, has proved most satisfactory.

Spectrum Analysis

The frequency components of a non-repetitive rectangular d-c. pulse may be determined by well known methods using Fourier integral analysis. The envelope of amplitudes is of the form $(\sin x)/x$ where $x = \pi f T$. This envelope is shown by the right-hand side of the curve of Fig. 15a, f_0 being assumed to represent zero frequency. The first zero occurs at the frequency $f = 1/T$.

Similarly the envelope of the spectrum of a rectangular a-c. pulse is given by the complete curve of Fig. 15a, f_0 in this case being the carrier frequency. For a non-repetitive pulse all frequencies are present in amplitude as shown by the envelope. When a stable carrier frequency is pulsed at uniform intervals and in precise phase relation, only harmonics of the repetition frequency are present under the envelope. In radar practice conditions are, as a rule, not sufficiently stable for this to occur.

Because of its bandwidth, an echo box cannot reproduce the ideal spectrum envelope of Fig. 15a. Instead the curve for a good spectrum may resemble that of Fig. 15b, while spectrum irregularities detrimental to radar performance may be revealed by curves such as those of Fig. 15c and 15d. Broadening of the spectrum is undesirable because less energy falls within the receiver band. Energy removed from the main concentration may result from double moding or from the occurrence of a different frequency during the rise or fall of the pulse. Frequency modulation due to a sloping

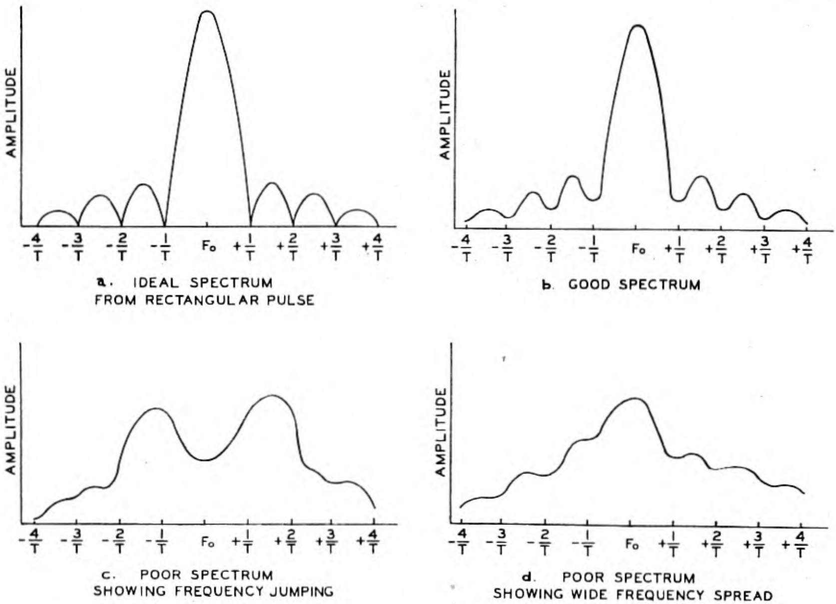


Fig. 15—Radar spectrum analysis with echo box.

or spiked input pulse produces a non-symmetrical spectrum, not infrequently characterized by a high side lobe. Frequency jump in the middle of the pulse, due to line reflection, may produce two distinct maxima.

Another device widely used for this purpose is the "Spectrum Analyzer" as developed by Radiation Laboratory, which provides an oscilloscope presentation of spectrum component amplitudes versus frequency.

Standing Wave Measurements

Theory

The expression for the distribution of current or voltage along a misterminated line of appreciable electrical length yields two terms which may be considered as representing two waves transmitted in opposite directions, one (the incident wave) from the generator toward the load, the other (the reflected wave) from the load toward the generator. The summation is a standing wave pattern. The standing wave ratio (SWR) is defined as the ratio of the wave amplitude at a maximum point (anti-node) to that at a minimum point (node). If the standing wave ratio is stated as a numeric, it is necessary to specify whether it applies to voltage (VSWR) or power (PSWR). Possibility of ambiguity is avoided by stating the ratio in db.

The ratio of the reflected current to the incident current is the *reflection coefficient*, here designated as ρ. The value of the reflection coefficient is given both in magnitude and phase by

$$\rho = \frac{Z_0 - Z}{Z_0 + Z} \tag{10}$$

where Z_0 is the characteristic impedance of the line and Z is the load impedance. The reflection coefficient is related to the standing wave ratio as follows:

$$VSWR = \sigma = \frac{1 + \rho}{1 - \rho}, \quad \text{or,} \quad \rho = \frac{\sigma - 1}{\sigma + 1} \tag{11}$$

Plots of the relationships are shown in Fig. 16.

The reduction of radiated power due to reflection losses in a radar transmission line, while important, is usually less serious than other effects of impedance irregularities. Since the load impedance reacts on the oscillator circuit, the frequency and output of most transmitter tubes are quite sensitive to load impedance. If the line is electrically long, so that its impedance varies rapidly with frequency, marked instability of oscillator frequency may occur, a condition referred to as "long line effect."

Since radar transmission lines contain many potential sources of impedance discontinuity, including not only the antenna but a variety of couplings, bends, wobble joints, rotating joints, switches, etc., measurements of standing wave ratio are frequently required. The need for such measurements depends in part on whether the line is "preplumbed" or is provided with field adjustments.

Devices

Standing waves may be detected and measured by several different types of devices, including (1) a slotted line, (2) a squeeze section, (3) a directional coupler and (4) a hybrid T. All of these furnish information on the magnitude of the standing wave ratio. In some cases phase information may be obtained also, which permits determination of impedance,[9] but this knowledge, while useful in the laboratory, is seldom required in field work.

A block diagram of an arrangement employing a slotted line for measuring standing waves is shown in Fig. 17. The oscillator source is commonly followed by a pad or attenuator to prevent frequency pulling. The slotted section may be either a coaxial or a wave guide line employing a mode which is not disturbed by the presence of the slot (e.g. normal coaxial mode; $TE_{1,0}$ in rectangular wave guide; $TM_{0,1}$ in round wave guide). A traveling pick-up probe or loop projects through the slot and couples energy from the line into a detector which delivers d-c. or audio-frequency to the indicator. The

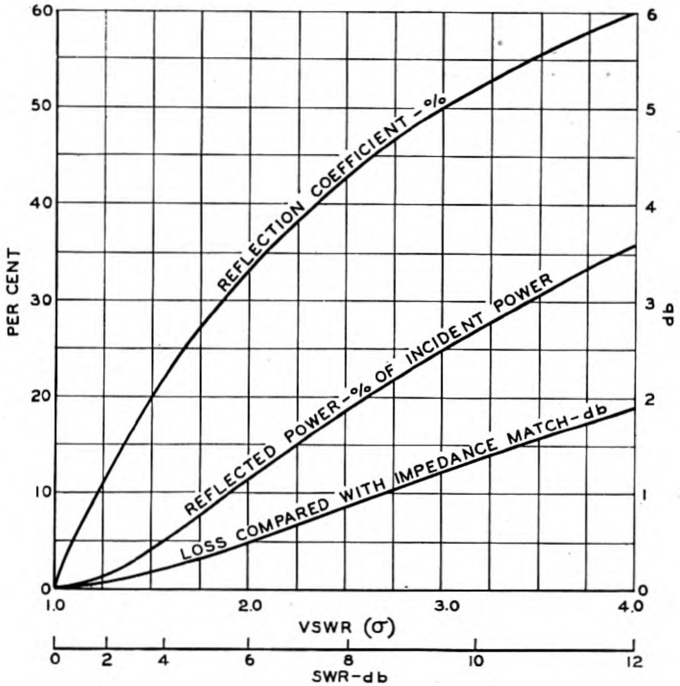


Fig. 16—Relations in mismatched transmission lines.

probe is moved longitudinally to find points of maximum and minimum field strength. To avoid distortion of the field within the line, the probe should be small and should project only a short distance inside the slot. For accurate results extreme care must be exercised in design and construction to avoid variation in depth of immersion as the probe is moved. Several slotted lines employed for standing wave measurements are shown in Fig. 18.

A squeeze section consists of a section of rectangular guide with slots milled in the center of both broad faces so that the width of the guide can be varied by external deforming means. This changes the wave length in

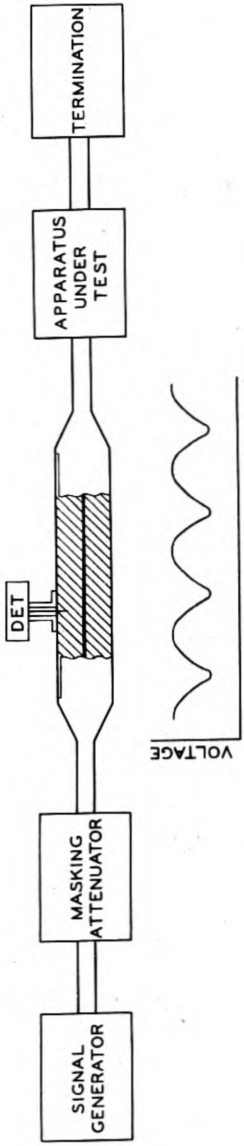


Fig. 17—Standing wave measurement.

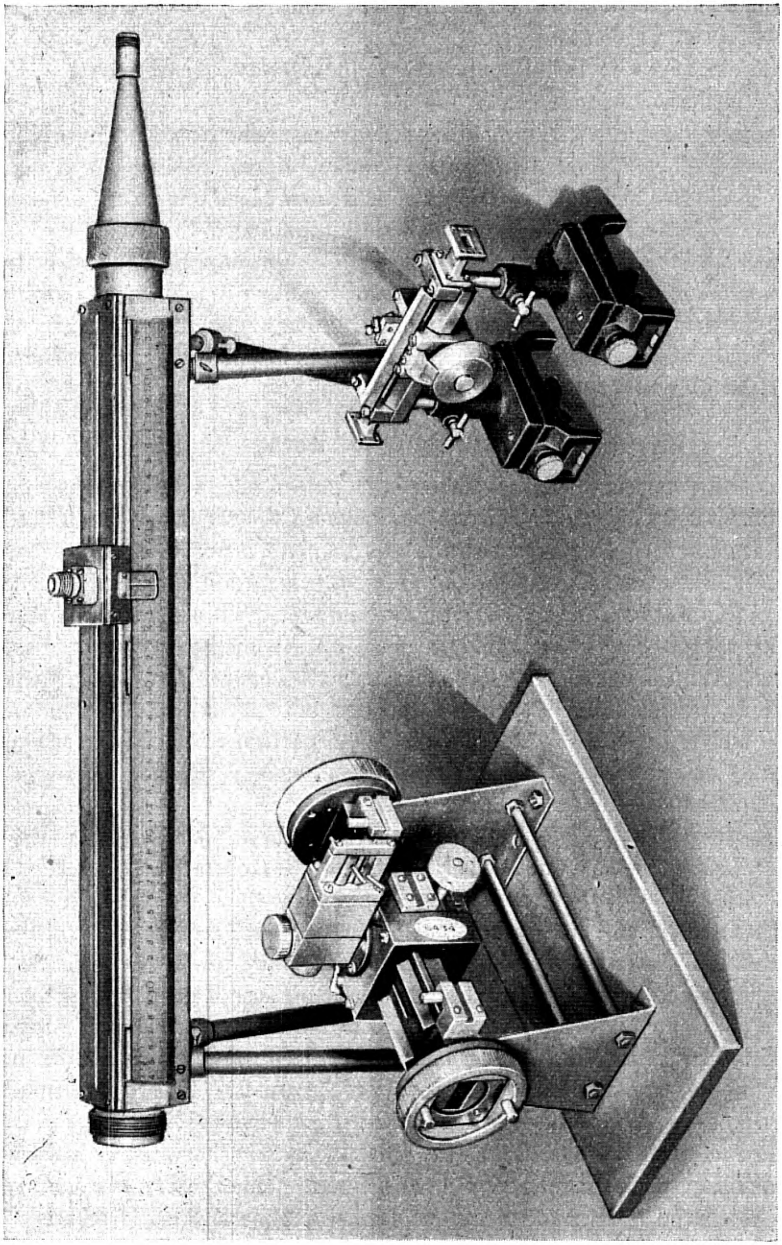


Fig. 18—Standing wave detectors used on coaxial and wave guide transmission lines.

the guide, so that maximum and minimum values may be determined by a fixed probe and indicator.

The use of a directional coupler for standing wave measurement is discussed in the following section.

Another useful device for standing wave measurement is the "hybrid T" or "magic T". This is a sort of microwave bridge, consisting of a main wave guide to which an E plane branch and an H plane branch are joined in the same physical plane. With matched terminations of the two ends of the main guide, the two branches are conjugate. Terminating one end of the main guide in the unknown impedance and the other in a matched termination, the degree of impedance mismatch of the unknown is indicated by the magnitude of the reflected wave which appears in one branch when energy is fed into the other.

Directional Couplers

Accurate measurement of transmitter power and receiver sensitivity requires a coupling path of known loss between the radar and the test set. The first method employed for this was to place a portable test antenna (see Fig. 2) in the field of the radar antenna. Depending on the frequency range, this test antenna took the form of a dipole,[10] with or without a small reflector, or an electromagnetic horn.[11] With this method it was necessary to calibrate the loss of the space coupling path between the two antennas. Since it proved difficult to locate the test antenna at exactly the same point and to be sure that the main antenna pattern remained the same, a separate calibration of the coupling loss was usually required whenever a measurement was made.

An alternative method was to place a single probe in the radar transmission line. This introduced another sort of difficulty. Accuracy of measurement was vitiated by the presence of standing waves which rendered the probe pick-up a function of frequency and of location with respect to the irregularities. A highly satisfactory answer to the entire problem was found in a device which is called a directional coupler because it couples only to the wave propagated in one direction. In its simplest form a directional coupler consists of two couplings to the main transmission line, which add for one direction of transmission and cancel for the other. Thus, for example, Fig. 10a shows a form of directional coupler for wave guide which is placed in the radar transmission line at the point indicated schematically in Fig. 1. An auxiliary wave guide is coupled to the main guide through two identical orifices spaced $\lambda_g/4$ between centers (or more generally $n\lambda_g/4$ where n is an odd integer). Assuming the incident and reflected waves in the main guide to be directed as shown, and the auxiliary guide to be terminated on one end, a test circuit connected to the other end will be coupled to the incident wave, while theoretically the two couplings to the reflected wave will differ by $\lambda/2$ and therefore cancel one another.

With such a device measurements may be made of the characteristics of the incident wave independently of reflections. If the coupling to the main line is not too close there is no appreciable effect on the incident wave, and continuous monitoring can be had. Conversely, test signals applied through the directional coupler will travel in the main guide in the proper direction for testing the radar receiver.

If the locations of the termination and the test connection point in Fig. 19a are reversed, the couplings to the main transmission line are also reversed. Such an arrangement therefore permits measurement of the reflected power which in turn makes it possible to adjust for minimum reflected power and hence for minimum SWR. Comparison of the reflected power with the direct power determines the SWR. For convenience in measurement, two directional couplers pointed in opposite directions are frequently used, the combination being referred to as a bi-directional coupler (Fig. 19b). One advantage of this arrangement is that the ability to measure the reflected power from the antenna and that part of transmission line beyond the coupler provides means for detecting trouble in that part of the system. Directional couplers may be applied to any type of transmission line. Figure 19c shows a simple form of directional coupler for a coaxial line.

One characteristic of importance in a directional coupler is the coupling loss. A small value of coupling loss affords increased sensitivity of measurement, while a sizable value is desirable to minimize reaction on the main transmission line as well as for other reasons. A loss of around 20 db has usually been found a good compromise. It is now the practice to incorporate a directional coupler in every radar to obtain a test connection point.

Due to unavoidable imperfections, a directional coupler never gives complete cancellation for the undesired direction of transmission. The departure from ideality is indicated by the directivity (also referred to as front-to-back ratio) which is defined as the scalar ratio of the two powers measured at the test connection point when the same amount of power is applied to the main guide, first in one direction and then in the other. For measurements of the direct wave and of receiver characteristics, a moderate directivity, of the order of 15 db or better, is sufficient. In measuring reflected power, however, the directivity determines the amount of direct power which appears at the point of measurement and therefore controls accuracy. The chart of Fig. 20 will facilitate determination of the maximum

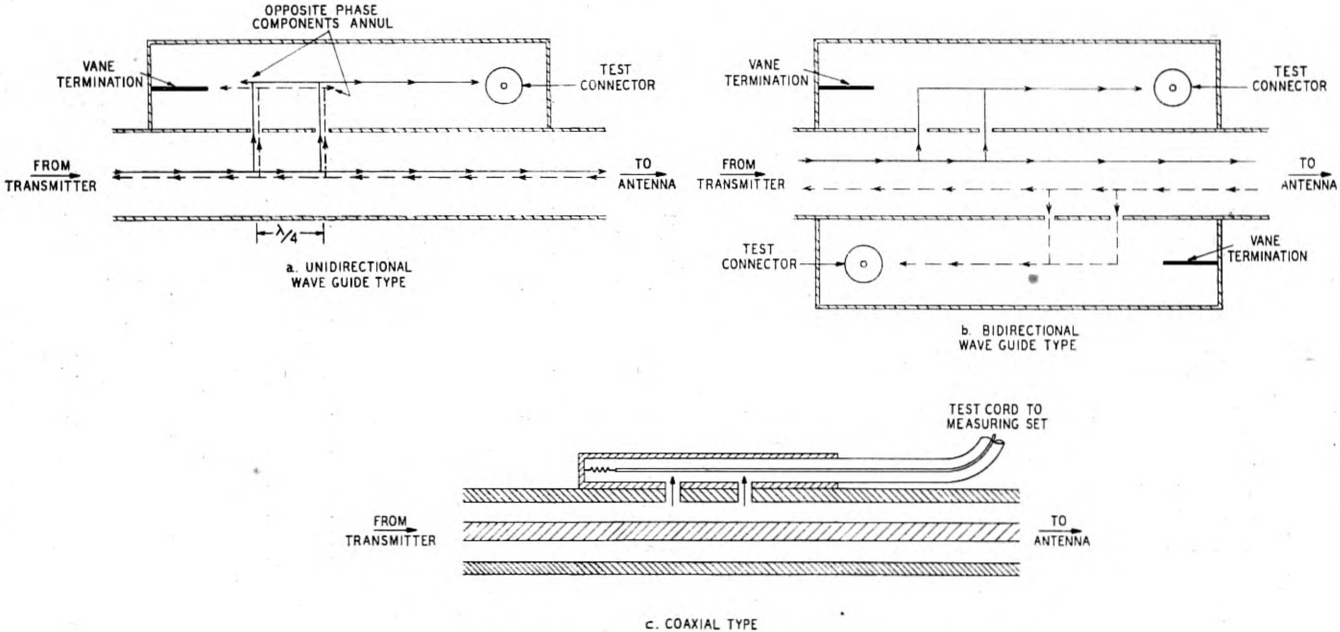


Fig. 19—Directional coupler arrangements.

error that may occur in measuring different values of SWR with various assumed directivities.

With a simple two-hole coupler, the directivity deteriorates rapidly as the frequency departs from that corresponding to quarter-wave spacing.

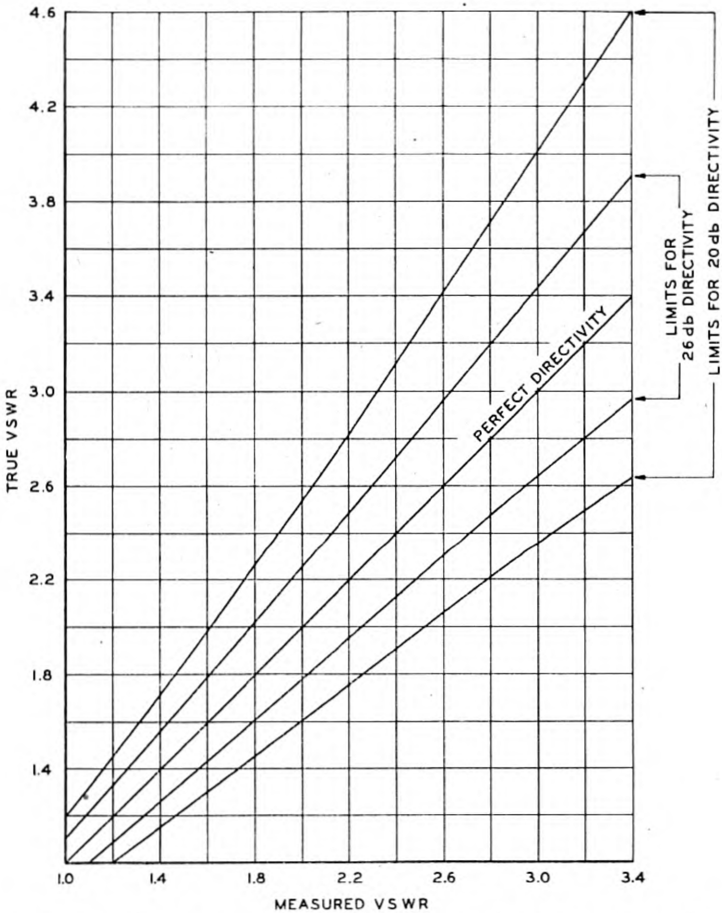


Fig. 20—Error in standing wave measurements caused by directivity of directional coupler.

By providing additional couplings suitably spaced, the residuals from different sets of couplings can also be cancelled against one another and the directivity versus frequency characteristic can be materially broadened. With multiple hole couplings a minimum directivity of 26 to 30 db over a frequency band of 10 to 20% is readily practicable in quantity production,

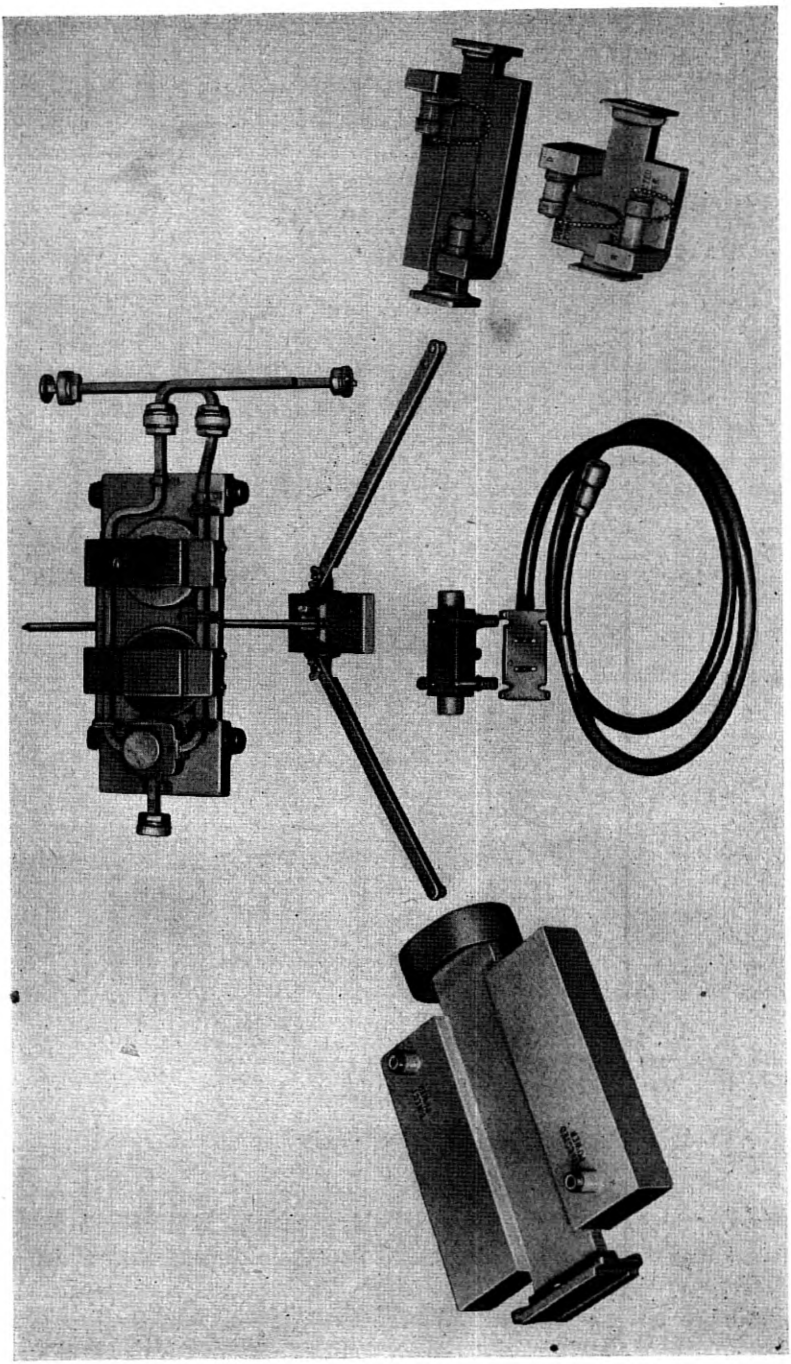


Fig. 21—A group of microwave directional couplers together with a directional coupler test set.

and much better values are obtained in the laboratory. Some of the numerous designs of directional couplers developed for association with operating radars are shown in Fig. 21.

Because they are more convenient than slotted lines and can be made more accurate, directional couplers have been extensively used for SWR measurement in the laboratory. A number of special arrangements have been devised to improve both accuracy and convenience. A directional coupler arrangement which has been provided for field measurement of SWR in the vicinity of 25,000 mc is also illustrated in Fig. 21. In this the direct power is brought to equality with the reflected power by an attenuator whose dial is calibrated directly in SWR. A wave guide switch facilitates the power comparison.

Auxiliaries and Components

RF Loads

An RF load (or dummy antenna) which will absorb the radar power in an impedance which matches the transmission line is very useful in radar work. Such a device permits testing the radar in operating condition without actual radiation which might give information to the enemy or interfere with other radars. It also makes it possible to test the radar in locations where reflections from the ground or nearby objects would otherwise hamper or prevent a test. RF loads for microwave work usually consist of a section of transmission line (either coaxial or wave guide, depending on wavelength) containing a high-loss dielectric. The impedance of such a load is necessarily low and must be matched to the radar line by tapering the dielectric over a distance of several wave lengths.[10] Moreover, if the line is to handle high power, tapering over a considerable length is necessary to distribute the heat.

A coaxial load is preferably tapered from outer conductor to inner conductor, since this both reduces the voltage gradient and facilitates heat dissipation. A dielectric consisting of a mixture of bakelite, silica and graphite, molded in place, has been found satisfactory. For wave guides a ceramic containing carbon may be preformed, with taper in one or two dimensions, and cemented in place.

Figure 22 shows a number of RF loads developed for different frequency bands. One of these, TS-235/UP, provides an excellent impedance match over the frequency range from 500 mc to above 3,000 mc. When equipped with a blower designed for uniform transverse ventilation, it will handle a peak power of the order of 750 kw with a duty cycle of about .001.

Microwave Attenuators and Pads

RF attenuators and pads are cornerstones of microwave testing. Attenuators are used to adjust unknown signals to levels suitable for measure-

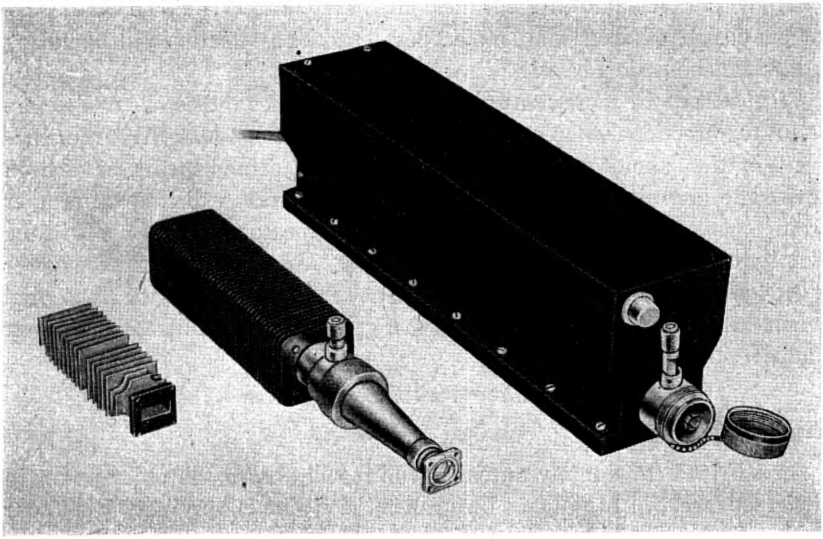


Fig. 22—RF loads for different bands in the microwave frequency range.

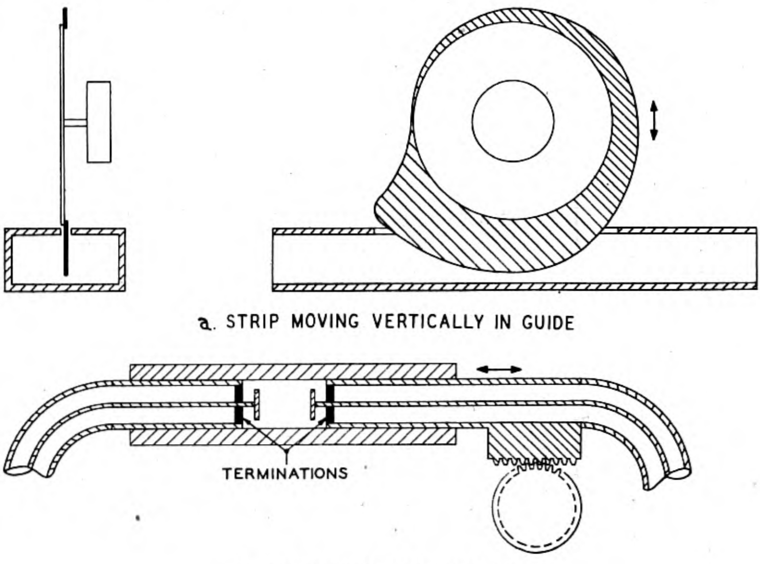


Fig. 23—Microwave attenuators.

ment and to obtain the minute test signals required for measuring receiver characteristics. Pads serve to change levels and to prevent interaction between testing components. Microwave attenuators and pads are of two general types (a) those which employ dissipative elements to absorb power and (b) non-dissipative devices which introduce propagation or coupling loss.

For the shorter microwaves the most convenient form of attenuator is of the dissipative type, employing a strip or vane of dielectric coated with a resistance material, as for example, carbon-coated bakelite. This is placed in rectangular wave guide with its plane paralleling the side of the guide. The attenuation is varied by varying the depth to which the vane is inserted in the guide (Fig. 23a) or by changing its position in the guide. A valuable feature of such attenuators is that the minimum loss can be made substantially zero. For good impedance match the strip must be tapered. By using two strips the over-all length of the attenuator can be reduced. Extremely satisfactory attenuators of this type covering a frequency range of 8 to 12%, with loss variable from 0 to 35 or 40 db, have been obtained in the frequency range 4,000 to 24,000 mc.

For the longer microwaves, where wave guides are inconveniently large, attenuators of the wave guide-below-cutoff type are very useful. These consist of a section of round wave guide whose diameter is small compared with wavelength and whose length is adjusted by telescoping (see Fig. 23b). The $TM_{0,1}$ mode has been found very satisfactory, and $TE_{1,1}$ has also been used. Connection is made to the attenuator by a coaxial circuit at each end, with disk excitation for the $TM_{0,1}$ mode and loop coupling for $TE_{1,1}$. The attenuation formulas are:[12]

$$TM_{0,1}. \qquad A = \frac{41.8}{D}\sqrt{1 - \left(\frac{1.31D}{\lambda}\right)^2}\ db/\text{meter} \tag{12}$$

$$TE_{1,1}. \qquad A = \frac{32.0}{D}\sqrt{1 - \left(\frac{1.71D}{\lambda}\right)^2}\ db/\text{meter} \tag{13}$$

where D = diameter of wave guide in meters. Because of the effect of other modes when the coupling is close, a minimum loss of 20 to 30 db is required before the attenuation becomes linear with displacement. The attenuation differentials are substantially independent of frequency. Attenuators of this type present a large impedance mismatch at either end, the effect of which may be alleviated by padding or by a termination.

Types of pads employed in microwave work include the following:

(1) Flexible coaxial cable, usually with high resistance inner conductor.
(2) Coaxial π with carbon coated rod and discs.
(3) Coaxial with carbon coated rod as inner conductor.
(4) Resistance strip in wave guide.
(5) Directional coupler.

In calibrating microwave attenuators and pads, comparison with an accurately calibrated IF attenuator, using a heterodyne test set, has been found to give excellent results.

RF Cables and Connectors

Flexible RF cables for connecting test equipment to equipment under test are an important adjunct of field testing. At frequencies of 10,000 mc and below, flexible coaxial cables of about .4″ over-all diameter with solid or stranded inner conductor, solid low-loss dielectric (polyethylene) and braided outer conductor have been used satisfactorily, although in the upper part of this range special measures have been necessary to prevent attenuation change due to flexure and aging. Over most of this range coaxial jack and plug connections have been found satisfactory but wave guide connectors are preferable at the upper end. In the range above 10,000 mc, coaxial

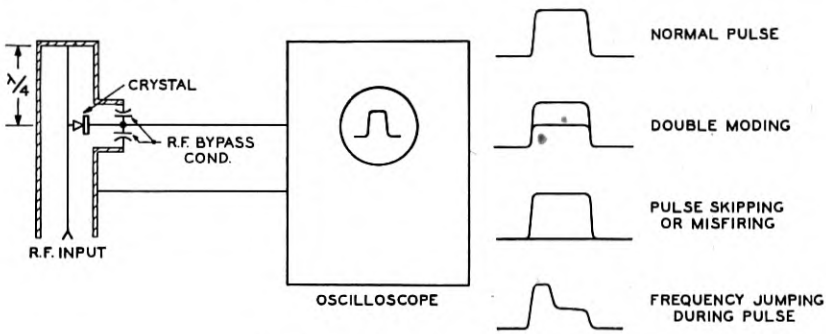


Fig. 24—Radar pulse envelopes.

cables of requisite stability have not yet been obtained and rubber covered wave guide with soldered articulated joints is the best form of flexible cable now available.

Oscilloscopes

Oscilloscopes are used extensively in radar maintenance (a) for examination of video waves and (b) for viewing RF envelopes. Satisfactory radar performance depends on a variety of video wave shapes which may include trapezoidal or triangular pulses, sawtooth waves, square waves or combinations of these. Observation of these wave shapes, supplemented if necessary by measurements of amplitude and duration, helps in diagnosing many troubles.

Examination of the envelope of the RF pulse is a convenient but less informative alternative to spectrum analysis. The envelope should be a clean, single trace of good shape. Figure 24 shows traces sometimes experienced. Double moding, i.e., oscillating at different frequencies on different pulses, is shown by a double trace. Frequency jumping during a pulse is shown by a break in the envelope. Misfiring gives a base line under the envelope. Other abnormalities in the RF envelope may result from incorrect video wave shape. Observation of the RF envelope requires a rectifier, usually a crystal, together with a suitable video amplifier. Since limitation of the scope to video functions permits general application to radars of all frequencies, the rectifier is generally provided externally.

The oscilloscopes available before the war did not meet the requirements of radar. Fast sweeps were necessary to permit viewing of pulses ranging from several microseconds to a fraction of a microsecond. Amplifiers were required for such pulses with low phase and amplitude distortion over a broad frequency band. Existing methods of synchronizing and phasing sweeps were also inadequate. The progress of the oscilloscope art during the war is illustrated in the successive designs of field test oscilloscopes shown in Fig. 25.

The BC910A oscilloscope, gotten out as a "stop gap" not long after the attack on Pearl Harbor, incorporates fast sweeps and broad-band amplification. Following close upon this was the BC1087A (Navy code CW60AAY) which replaced sine wave synchronization by a start-stop sweep triggered by the incoming pulses. This feature made it possible to superpose the erratic pulses produced by spark wheel and similar pulsers and at the same time avoided external synchronizing connections. A valuable feature conjoined with the start-stop sweep was a delay network in the main transmission path which gave the sweep time to start before the pulse reached the cathode-ray tube. This oscilloscope in original and modified form has seen wide service in all theaters. However, its weight of more than 60 pounds was a handicap for many uses.

Further advances in oscilloscope circuitry and in weight limitation resulted in TS-34/AP, weighing only 25 pounds. This combined the short pulse features of the previous design with those of the conventional oscilloscope for viewing slower waves. A schematic diagram is shown in Fig. 26. A redesign, coded as TS-34A/AP, incorporated variable start-stop sweeps and improved mechanical design. These two oscilloscopes, TS-34 and TS-34A, were produced to a total of some 12,000 and universally used by all branches of the service for both radar and radio testing. Toward the end of the war the trend toward shorter pulses, coupled with the need for precise measurement of wave amplitude and duration, led to a new design, TS-239/UP, which embodied wide advances over TS-34A in performance and versatility but with an increase in weight.

In association with different oscilloscopes, other video devices have been

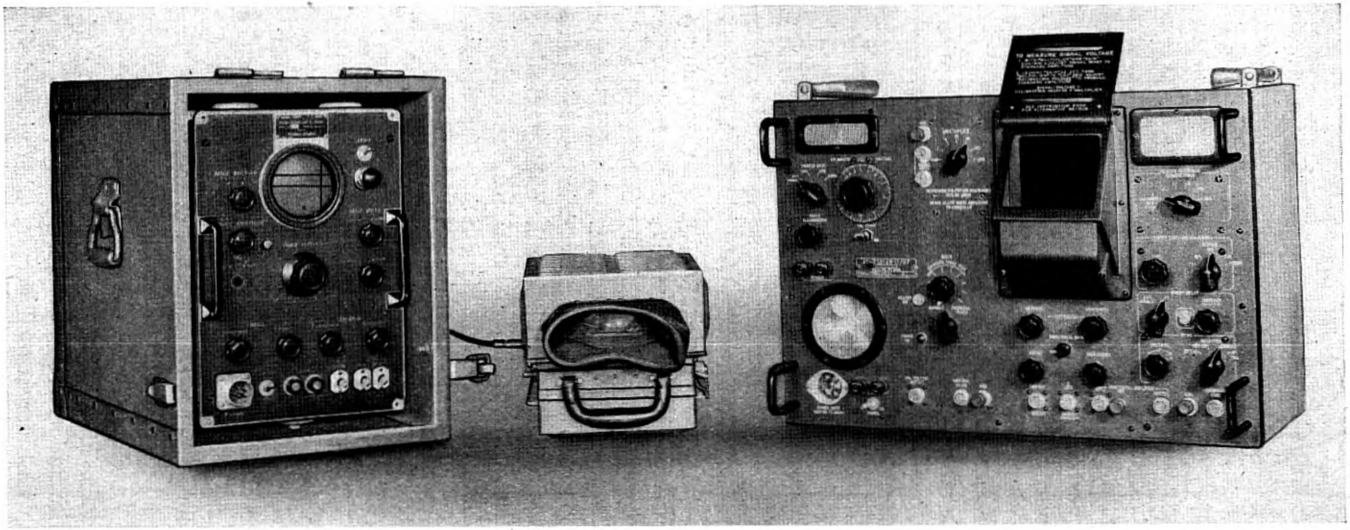


Fig. 25—Test oscilloscopes for viewing wave forms in radars. Left to right: BC-910-A (1942), TS-34A/AP (1944) and TS-239/Up (1945).

employed. The amplitude of pulse applied to the magnetron is thousands of volts. To derive a voltage suitable for application to the oscilloscope, a voltage divider of the condenser type is used (TS-89/AP). Suitable video terminations, dividers, and loads, sometimes of high voltage and power capacity, are required to obtain proper test conditions and provide convenient test points (TS-98/AP, TS-390/TPM-4, TS-90/AP, TS-234/UP). Originally a high-impedance connection to the oscilloscope was effected by a single-stage amplifier unit (BC1167A), but a simple divider type of probe was later found more satisfactory for this purpose.

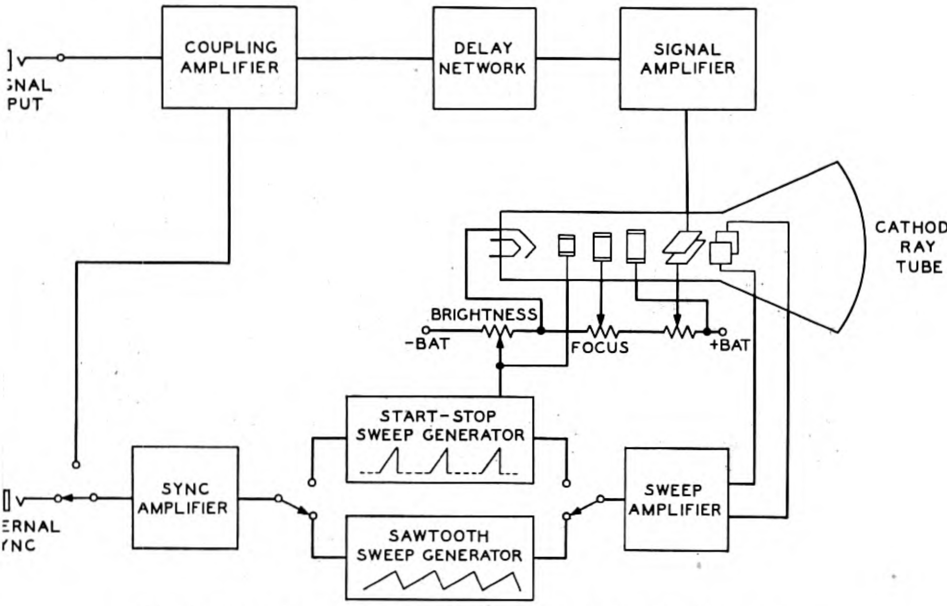


Fig. 26—Block diagram of TS-34/AP and TS-34A/AP oscilloscopes.

Range Calibration

Types of timing circuits used for radar range determination include (a) multi-vibrators, (b) coil and condenser oscillators (generally without but sometimes with temperature control) and occasionally (c) quartz crystal oscillators. The first two depend for their accuracy on condensers, resistances, coils and other elements which are subject to error due to aging, temperature, humidity, mechanical damage and the like. Nor is the quartz crystal oscillator wholly immune to error. Consequently, portable range calibrators are required for field maintenance.

TS-102A/AP (Fig. 27) and its predecessors TS-102 and TS-19 are precision calibrators which have been extensively used for checking a large

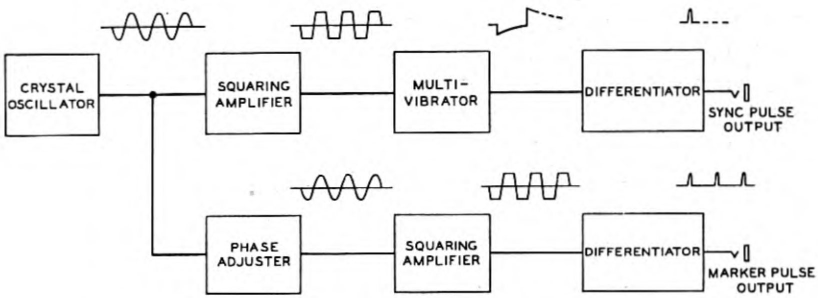


Fig. 27—Block diagram of TS-102/AP and TS-102A/AP range calibrators.

Fig. 28—Two test sets for checking the computers used in bombsight and fire control radars.

number of different airborne bombing and gunlaying radars and shipborne fire control and search radars. These sets deliver accurately spaced marker pulses, derived from a quartz crystal oscillator, for checking the radar range pulses. A trigger pulse derived from a multi-vibrator synchronized with the quartz oscillator is also provided for actuating the radar timing circuits. With certain radars the calibration procedure requires an oscilloscope as well. Extreme stability of marker pulses, better than $\pm.02$ microsecond, is obtained. A stop watch is included in these sets for checking rate of change in range.

Less precision is required in range calibration of search radars. For this purpose the TS-5/AP calibrator provides marker pulses of $\frac{1}{4}$, 1, 5 or 10 nautical or statute miles, derived from a coil and condenser oscillator with closely controlled temperature coefficients. This calibrator is designed to be triggered by the radar or some other external source.

Computer Test Sets, Etc.

A number of radars are equipped with computers which receive the data on location of target and its direction and rate of change, together with essential related information on such factors as wind velocity, ground speed, altitude, etc., and deliver the solution of the ballistic problem in the form of a voltage which releases bombs, points the guns or serves other purposes. Means for checking the accuracy of these computing devices are generally required. The type of test set needed depends upon the computer design, which has taken different forms according to the nature of the problem and the state of the art.

Two types of computer test set are shown in Fig. 28. TS-158/AP, designed for use with certain airborne bombing radars, furnishes to the computer a signal representing a target approaching at known speed and checks the accuracy of bomb release. TS-434/UP, designed for several airborne and ground radars, is an accurate instrument for determining the voltage ratios at various points in a computer and thus checking its performance.

Conclusion

More than 200 different designs of test sets were developed during the war by Bell Laboratories to meet the exacting requirements of radar field maintenance. These differed radically from previous art. Outstanding features were portability, precision and generality of application. The large number of designs is due partly to the varied functions of radar and to the varied conditions of use. Largely, however, it results from the fact that the frequency band that can be handled in any one set is limited, whereas many frequency ranges and subranges had to be covered in all.

Altogether more than 75,000 radar test sets were manufactured by Western Electric Company and these were used in all theatres of war by the

United Nations forces. The production rate at the end of the war exceeded 5,000 test sets a month. In numerous cases, moreover, small preproduction quantities of test equipment were built on a "crash" basis for special missions and for training purposes. The test equipment produced for the field had to be more precise than the radars, and the equipment used in the factory and laboratory to test the field test equipment had to be still more precise.

Trends of development at war's end were toward (a) further broad-banding, simplification and precising, and (b) coverage of new frequency ranges.

References

(1) "Ultra-High Frequency Techniques," by J. G. Brainard, G. Koehler, H. J. Reich, and L. F. Woodruff, D. Van Nostrand, 1942.
(2) "Reflex Oscillators," by J. R. Pierce, *Proc. I.R.E.*, Vol. 33, Feb. 1945, p. 112.
(3) "The Lighthouse Tube," by E. D. McArthur and E. F. Peterson, *Proceedings of National Electronics Conference*, Vol. I, 1944, p. 38.
(4) "Electromagnetic Waves," by S. A. Schelkunoff, D. Van Nostrand, 1943.
(5) "The Proportioning of Shielded Circuits," by E. I. Green, F. A. Leibe and H. E. Curtis, *Bell System Technical Journal*, Vol. 15, April 1936, p. 248.
(6) "Resonant Lines in Radio Circuits," by F. E. Terman, *Electrical Engineering*, Vol. 53, July 1934, p. 1046.
(7) "Ultra-Short-Wave Transmission Phenomena," by C. R. Englund, H. T. Crawford, and W. W. Mumford, *Bell System Technical Journal*, Vol. 14, July 1935, p. 369.
(8) "Thermistors in Electronic Circuits," by R. R. Batcher, *Electronic Industries*, Vol. 4, Jan. 1945, p. 76.
(9) "An Impedance Transmission Line Calculator," by P. H. Smith, *Electronics*, Jan. 1944.
(10) "Microwave Transmission," by J. C. Slater, McGraw-Hill, 1942.
(11) "Theory of the Electromagnetic Horn," by W. L. Barrow and L. J. Chu, *Proc. I.R.E.*, Vol. 27, Jan. 1939, p. 51.
(12) "Attenuation of Electromagnetic Fields in Pipes Smaller Than Critical Size," by E. G. Lindner, *Proc. I.R.E.*, Vol. 30, Dec. 1942, p. 554.

Performance Characteristics of Various Carrier Telegraph Methods

By T. A. JONES and K. W. PFLEGER

This paper describes laboratory tests of certain carrier telegraph methods, to determine their relative advantages from the standpoints of signal speed, and sensitivity to level change, carrier frequency drift, interchannel interference, and line noise.

INTRODUCTION

MOST of the carrier telegraph methods mentioned below are well known,[1,2,3,4,5] but the selection of the best method for a particular application is difficult without comparative tests on specific designs. It is the purpose of this paper to record data taken during such tests and to explain the results so that they may be helpful to those concerned with the selection of the optimum method for a given set of requirements.

The conclusions here reached regarding methods of telegraph transmission do not necessarily apply to transmission of sound, pictures, or television, because their requirements differ. In telegraph transmission it is important that signal transitions be received at approximately the correct times, and wave rounding is permissible.

Computations for a square cut-off band-pass filter with zero phase distortion[6] show that the shape and duration of the transient in the received wave are about the same for a sudden transition in both on-off and frequency-shift arrangements (explained in the next section), when the total frequency shift is not more than half the channel width. As telegraph distortion depends largely upon the transient, one might therefore infer that, if the transients are about alike, there is no particular advantage in frequency-shift over the on-off method as far as signal speed is concerned. However, the computation for the idealized filter gives no assurance that a physical filter will per-

[1] H. Nyquist: "Certain Topics in Telegraph Transmission Theory", *A. I. E. E. Trans.*, Vol. 47, pp. 617–644, April 1928.
[2] H. Nyquist and K. W. Pfleger: "Effect of Quadrature Component in Single Sideband Transmission", *Bell System Technical Journal*, Vol. XIX, pp. 63–73, Jan. 1940.
[3] E. H. Armstrong: "Methods of Reducing the Effect of Atmospheric Disturbances", *Proc. I. R. E.*, Jan. 1928, pp. 15–26.
[4] J. R. Carson: "Reduction of Atmospheric Disturbances", *Proc. I. R. E.*, July 1928, pp. 966–975.
[5] F. B. Bramhall & J. E. Boughtwood: "Frequency Modulated Carrier Telegraph System", *Electrical Engineering*, Vol. 61, No. 1, Jan. 1942, *Transactions Section*, pp. 36–39.
[6] Fig. 3 of H. Salinger: "Transients in Frequency Modulation", *Proc. I. R. E.*, August 1942, pp. 378–383.

form thus. In order to investigate this experimentally, as well as other factors that concern the choice of method, the effects on telegraph transmission of interchannel interference and of varying the signaling speed, transmission level, mean carrier frequency, and line noise, were determined for several different methods, using the same channel filters. In order to test the two-band methods* using the same frequency range occupied by the one-band arrangements, narrow-band filters would be required to divide the frequency range into two parts. Since such filters were not available, it was necessary to use two adjacent frequency bands each similar to that used with the on-off method. However, some tests were made of a two-band arrangement using somewhat narrower filter pass bands.

A special wide-band frequency-shift arrangement using filters of about twice the band width of the other frequency-shift arrangement, was tested mainly in order to observe the effect of band width on sensitivity to noise and interference.

In all of the noise tests, thermal or resistance noise was used. With noise of the impulse type it is possible that somewhat different results would have been obtained, but it is believed that the difference would not have been great.

Conclusions

A study of the test results leads to the following conclusions which apply for the conditions assumed, and which are thought to be of general application, except for modifications which may be made necessary by future technical advances:

1. There is no important advantage in frequency-shift carrier telegraph over the on-off method as used in the Bell System for stable, quiet circuits, either wire or radio. However, the frequency-shift method shows some improvement in operating through noise. The frequency-shift method has disadvantages as regards complication and cost. Furthermore, it may be seriously affected by carrier frequency drift and interchannel interference, although the effects of these can be mitigated to some extent by special devices.
2. For high-frequency radio transmission over long distances, which is subject to comparatively severe non-selective fading, a great advantage is realized from the use of frequency-shift telegraphy with a fast receiving limiter instead of the conventional "continuous wave" or on-off method. For satisfactory operation it is still necessary that the signal level be kept sufficiently higher than the noise level in the transmission band.

* See the section entitled "Explanation of Terms".

3. Single-sideband telegraphy has an advantage of providing somewhat higher speeds without increasing the band width. Whether it holds much promise for any general application in multi-channel systems utilizing narrow bands and moderate signal speeds is questionable in view of certain difficulties. For a single-channel high-speed circuit, single-sideband telegraphy might be found worth while from the standpoint of economical use of the frequency spectrum.
4. Certain other arrangements tested possessed some characteristics which have advantage under particular conditions. For example, two-band arrangements may sometimes be conveniently obtained by combining existing on-off arrangements. These two-band arrangements are capable of furnishing high-grade service over radio circuits subject to severe fading. The use of a single source of carrier instead of two sources on a two-band arrangement results in a substantial transmission improvement. The performance then is comparable to that of a single-band frequency-shift channel occupying the same frequency space.

A more complete discussion of the results is given under the heading "Summary of Results", at the end of this paper.

Explanation of Terms

The following is intended to explain what is meant by certain terms used in this paper. They apply specifically to carrier telegraph operation in the voice range but, in general, they could also apply to radio telegraphy. (It will be appreciated that various other combinations of the instrumentalities involved in the present discussion could be used.)

Channel

A telegraph channel is a path which is suitable for the transmission of telegraph signals between two telegraph stations. In the present discussion the term "channel" is restricted to mean one of a number of paths for simultaneous transmission in different frequency ranges as in carrier telegraphy, each channel consisting of an arrangement of carrier telegraph equipment designed for the transmission of one message at a time, in only one direction.

On-Off Method

This, the most common form of amplitude modulation, is the same as "continuous wave" in radio telegraphy. It is a method of signaling over a channel utilizing a single carrier frequency, normally located at the center of the transmission band of the channel filters. The presence of carrier current on the line corresponds to the marking condition of the channel, and its absence, to the spacing condition. A Fourier analysis of the line current during signaling would show a steady carrier frequency component and substantially symmetrical upper and lower sideband frequency components.

Single-Sideband Method

This is similar to the method just described except that: (1) the carrier frequency is located near one boundary of the channel filters, so that during signal transmission one of the sidebands is attenuated much more than the other before the signals reach the line, and (2) during the spacing condition, carrier current may be either absent from the line, or present with amplitude less than that of marking current. (The latter condition tends to reduce that part of the distortion which is due to the quadrature component.[2])

Frequency-Shift Method

This is a method of signaling over a channel utilizing a carrier current of substantially constant amplitude from a frequency-modulated oscillator. The carrier current has no phase discontinuity and its instantaneous frequency varies between two limits within the transmission band. In the present discussion the two limits, symmetrically located in the transmission band, correspond respectively to the marking and spacing conditions of the channel. Variation of the instantaneous frequency may be abrupt or gradual, for example, sinusoidal. Throughout this paper the reader should assume that the variation is substantially abrupt except where otherwise indicated. At the receiving terminal the variable frequency signals are converted to amplitude-modulated signals by means of a frequency detector.

Two-Source Method

This is also a frequency-shift method, but it differs from that described above in that it is made up by combining two on-off channels, each supplied with carrier current of a different but substantially constant frequency from a separate source. In the present discussion each of the two on-off channels has a separate oscillator and occupies a different frequency band on the same line, and the sending relays of the channels have their operating windings differentially inter-connected so that one oscillator delivers current to the line during marks, and the other during spaces. Thus the marking and spacing signals are confined to separate frequency bands on the line. This is sometimes referred to as two-band operation. The switching takes place abruptly. No attempt is made to control the phases of the two sources. Therefore phase discontinuities are likely to occur at the instants of switching, causing brief transients of varying shapes in the line current. The output circuits of the two receiving detectors are differentially interconnected to obtain polar signals for the operation of a common receiving relay.

One-Source Two-Band Method

This is also a frequency-shift method, and is substantially the same as the two-source method except that a single frequency-modulated oscillator is used at the sending end instead of two oscillators, and thus there is no phase discontinuity in the sent signals.

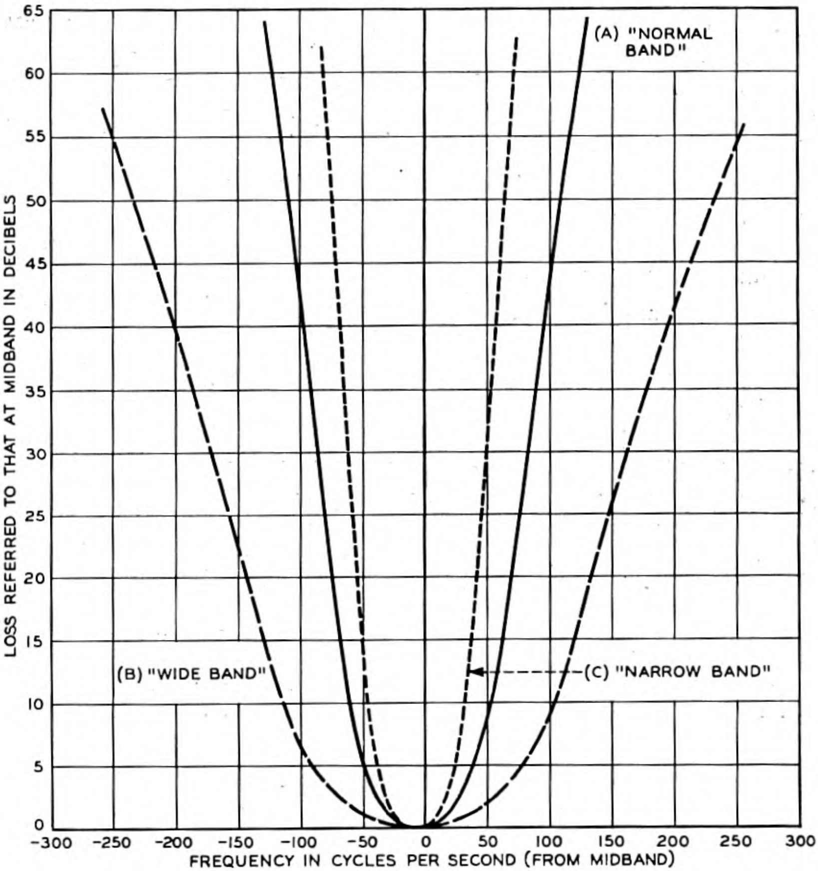


Fig. 1—Loss characteristics of channels tested, including sending and receiving filters and two repeating coils.

These methods will be more clearly understood from the following description of apparatus.

Apparatus

Channel Filter Characteristics

In Fig. 1, curve A represents the loss vs. frequency characteristic of the majority of channels used in the tests, including both sending and receiving filters and associated repeating coils. The two-source arrangements each occupied two such bands with midband frequencies spaced 170 cycles apart, except in the case of the narrow-band two-source arrangement which occupied two bands having characteristics similar to curve C with midband frequencies spaced 120 cycles apart. Curve B represents the loss vs. frequency characteristic of the wide-band frequency-shift arrangement, having approximately twice the band width of curve A. These characteristics were all measured between 600-ohm terminations, without adjacent channel filters present.

On-Off Terminal Apparatus

Figure 2 shows in block form the different circuit arrangements tested, together with the location of the carrier frequencies in the transmitted bands. In the on-off arrangement, the oscillator transmitted 1955-cycle current to a modulator which contained a polar telegraph relay controlled by signals from the local sending loop. During spacing signals this relay short-circuited the carrier supply, and during marking signals it allowed the carrier current to flow through the sending band-pass filter to the adjustable resistance line.

When it was desired to measure the effect of interference from other channels working in adjacent pass bands, their terminal equipment was added by connecting the line sides of their sending or receiving filters to the common sending or receiving bus (See dashed lines designated "bus" in Fig. 2), so that all the channels would transmit over the same line.

After the carrier signals passed through the receiving filter connected to the output of the line, they were converted by a detector-amplifier into direct current for operating a polar receiving relay, which, in the absence of incoming signals, was held on its spacing contact by local biasing current. The receiving relay contacts transmitted into the local receiving loop.

Single-Sideband Terminal Apparatus

In testing the single-sideband arrangement shown in Fig. 2, the terminal equipment of the on-off arrangement was used and the carrier frequency was placed 43 cycles above midband, at 1998 cycles. In some of the tests the modulator was modified to transmit during spacing intervals a carrier current 6 db below its marking value. Since the loss in the receiving filter was greater at the edge of the band than at the center, it was necessary to increase the gain of the linear detector-amplifier in order to have the same change of current in the line winding of the receiving relay of the single sideband arrangement as in the on-off arrangement. A further increase in the detector-amplifier gain was necessary when the single-sideband arrangement was operated with spacing carrier 6 db below the marking carrier.

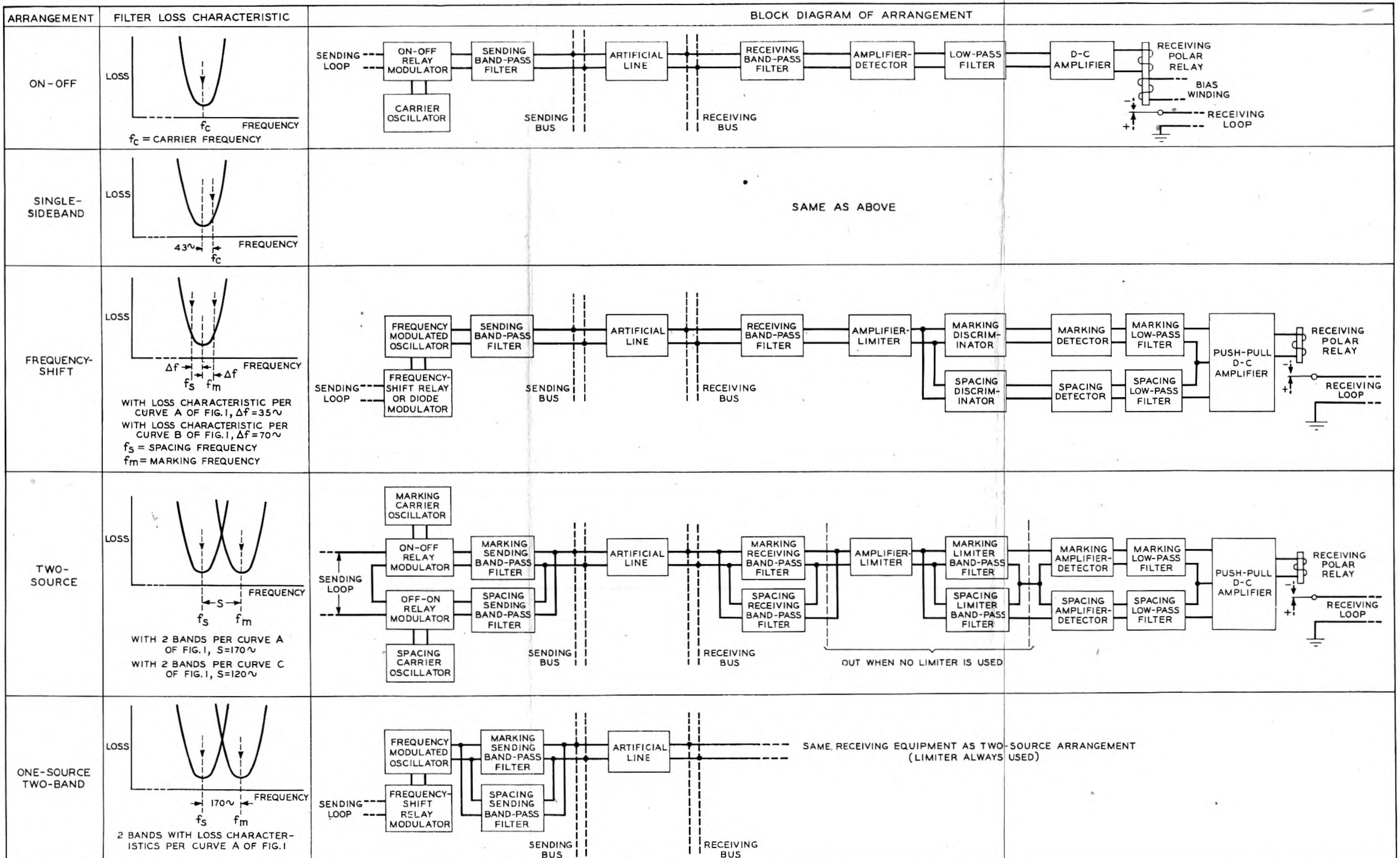


Fig. 2—Telegraph arrangements tested.

Frequency-Shift Terminal Apparatus

In most of the tests on the frequency-shift arrangement shown in Fig. 2, the oscillator frequency was caused to vary abruptly by a relay modulator. The sending relay was of the same type as used in the on-off arrangement and varied the tuning capacity of the oscillator. For some other tests the frequency variation was made more gradual by converting the sent signals into polar signals and passing these through a low-pass filter in order to round the wave so that the pulses had an approximately sinusoidal shape during reversals and attained steady-state value only at the center of each pulse. These polar signals were used to control the plate resistance of either of two diodes, thereby connecting a positive or negative reactance across the tuned circuit of the oscillator.[5] This caused the oscillator frequency to be either increased or decreased in proportion to the amplitude change of the control current. During most of the tests the spacing frequency was 1920 cycles and the marking frequency was 1990 cycles, both equally spaced from the midband frequency, 1955 cycles. The oscillator, when on marking frequency, was set to produce one milliwatt into the 600-ohm resistance artificial line. After passing through the receiving filter connected to the output of the artificial line, the signals entered a limiter (unless otherwise stated) delivering an output current which was practically constant for input levels between -55 and $+25$ dbm. From the limiter the signal passed into a frequency discriminator circuit having two output branches, each of which was connected to a diode detector tube followed by a low-pass filter. The two discriminator branch circuits in combination with their detectors and low-pass filters had output amplitude vs. input frequency characteristics of opposite slopes. After differential recombination of the two low-pass filter outputs, the resultant characteristic was linear over the range of fundamental frequencies transmitted by the limiter. The differentially recombined wave in the final d-c amplifier had an amplitude substantially proportional to the instantaneous deviation from the average value of the received carrier frequency over a range of $\pm$ 70 cycles. (Some calculations by one of the writers indicate that the use of discriminators of this type is helpful in reducing characteristic telegraph distortion.) In most of the tests the low-pass filters associated with the detector output had a cut-off frequency (about 503 cycles) low enough to suppress the carrier but high enough not to affect the telegraph transmission. The final d-c amplifier was substantially linear and increased the d-c wave to a suitable value for operating the polar receiving relay.

The wide-band frequency-shift arrangement was similar to that just described, except for the change in filters and tuning of the sending oscillator and the discriminator. The spacing frequency was 2055 cycles and the marking frequency was 2195 cycles which were equally spaced from the midband frequency, 2125 cycles. In this arrangement a limiter was always used at the receiving terminal. The discriminator was adjusted to be linear over twice the frequency range of the previously described discriminator, and the slope of the new discriminator characteristic was adjusted to give the same marking or spacing output as before. The wide-band frequency-shift arrangement was also tested with low-pass filters associated with the detector outputs which had a cut-off frequency (about 58 cycles) low enough to give a distortion vs. speed characteristic close to that obtained on the normal-band frequency-shift arrangement. The loss of each of these low-pass filters was about 10 db at 58 cycles. Since each filter consisted of only one section the cut-off was gradual.

Two-Source Terminal Apparatus

The sending circuits of the two-source arrangement shown in Fig. 2 included separate oscillators of different frequency for marking and spacing signals. Two sending relays were operated in synchronism by the signals in the sending loop. The relay contacts were so connected that marking carrier was transmitted to a marking band-pass filter and spacing carrier was cut off from a spacing band-pass filter, or vice versa. Thus either the marking or the spacing carrier frequency was transmitted to the line at any instant. At the receiving end of the line the incoming signals flowed through two receiving filters, one passing the marking current and the other passing the spacing current. When a limiter was not used, the outputs of these filters were connected directly to separate amplifiers, detectors and low-pass filters. The rectified marking and spacing signals were combined differentially, passed through a push-pull d-c amplifier, and operated the receiving relay just as in the frequency-shift arrangement. When a limiter was used the outputs of the receiving filters were recombined and passed through the limiter, after which they were again separated by means of additional band-pass filters whose losses were about half those of the receiving band-pass filters. The two-source arrangement was tested both with and without limiter when the loss characteristics of the spacing and marking paths were each similar to curve A of Fig. 1 and had midband frequencies of 1785 and 1955 cycles, respectively.

The two-source arrangement with limiter also was tested with filters having loss characteristics for the spacing and marking paths each similar to curve C of Fig. 1, and having midband frequencies of 1980 and 2100 cycles, respectively. This arrangement occupied a frequency band approximately 12/17 that used for the two-source arrangement with filters having characteristics similar to curve A.

One-Source Two-Band Terminal Apparatus

The one-source two-band sending circuit was similar to that of the frequency-shift arrangement with relay modulator, except that the frequency shift was 170 cycles, and the marking and spacing frequencies were adjusted to be at the centers of the pass bands of the marking and spacing sending filters, as shown in Fig. 2. The receiving circuit was the same as that used for the two-source arrangement. A limiter was always used.

Telegraph Transmission Measuring Apparatus

In all of the tests, the d-c open-and-close signals in the local sending loop were substantially rectangular and consisted either of reversals (a succession of alternate marks and spaces of equal duration) or of the test sentence: THE QUICK BROWN FOX JUMPED OVER A LAZY DOG'S BACK 1234567890 BTL SENDING. The customary 7.42 unit teletypewriter code was used, consisting of a stop pulse, a start pulse, and five code pulses per character[7]. The distributors supplying these signals were driven by synchronous motors controlled by an adjustable frequency oscillator. The speeds utilized experimentally ranged from 60 to 180 words per minute (about 23 to 68 dots per second).

In order to measure the telegraph distortion of the signals obtained in the receiving loop, a cathode-ray tube distortion measuring set was used which measured maximum total distortion in per cent of a unit pulse in much the same manner as a start-stop distortion measuring set previously described[8], except that electronic circuits were used to replace the distributor and all but one relay, which made possible precise measurements over a wide range of speeds. The bias of received signals was measured on reversals by means of a highly damped zero-center d-c milliammeter inserted in the receiving loop.

Source of Resistance Noise

In order to measure the effect of line noise, resistance noise was reproduced from a phonograph record, amplified, and combined with the carrier signals by means of a symmetrical three-way pad (part of the artificial line). A variable attenuator was used to regulate the amount of noise entering the line. The r.m.s. noise power or marking carrier power was measured with a thermocouple.

Measuring Precision

The signals generated by the dot and test sentence distributors were distorted less than 3 per cent of a dot length. As these distributors were of a

[7] E. F. Watson: "Fundamentals of Teletypewriters Used in the Bell System", *Bell Sys. Tech. Jour.*, Vol. XVII, Oct. 1938, pp. 620–639.

[8] R. B. Shanck, F. A. Cowan, S. I. Cory: "Recent Developments in the Measurement of Telegraph Transmission", *Bell Sys. Tech. Jour.*, Vol. XVIII, Jan. 1939, p. 149.

commercial type, they are believed to be representative. This distortion was erratic, depending upon the speed and wear of brushes and commutators. The distortion measured at the receiving relay could not be corrected by subtracting the distortion of the sent signals because during miscellaneous signals maximum distortion in the received signals might occur on a different transition from that in the sent signals. The small errors which existed in the sent signals are therefore believed to have been neither serious in their effect on the measured distortion, nor the sole cause of irregularity in the data.

The accuracy of the distortion measuring set itself was in the order of 1 per cent, as determined by measuring known amounts of bias in signals sent from a special distributor. Usually two observers took independent readings which were required to check closely or the observations were repeated. The average of the two observations was taken as the final measurement.

Although the line loss was constant in these tests, amplifiers, oscillators, power packs, and telegraph batteries were subject to slight voltage variations. Precautions were taken to reduce all variables as far as was practicable, yet it seems likely that the telegraph transmission measurements may be slightly in error due to such variations.

The individual sources of error mentioned in the last three paragraphs seem reasonable and sufficient to account for most of the irregularities in the following curves of telegraph transmission vs. speed. Yet it did not seem fair to draw smooth curves and neglect the irregularities, because these can also be due to the telegraph system itself, as was found by careful and repeated measurements. For example, it is known that relay performance is erratic and depends upon the speed. Chattering of relay contacts and periodic vibration of the armature have appreciable effect upon the distortion and can cause irregularities in distortion vs. speed characteristics, particularly at the higher speeds. Furthermore, such irregularities may not be wholly reproducible in repeated measurements due to changes in the relay temperature or contact surfaces and due to the occasional readjustments of relays. Another cause of irregularities in a distortion vs. speed characteristic may be the loss and phase characteristics of the channel filters. For example, consider an ideal transducer[1] which is distortionless at a speed s near the cut-off. It is also distortionless at speeds such as s/2, s/3, s/4, s/5, etc. At intermediate speeds distortion may exist, so that a curve of distortion vs. speed would show irregularities with minima at these optimum speeds. (Some computations by one of the writers for a frequency-shift arrangement using idealized filters show irregularities in the distortion vs. duration curve for a single dot.) No attempt was made to shape the channel filter characteristics to be perfectly distortionless at a particular speed; and it seems reasonable that there could likewise be a number of speeds where maxima and minima occur in the distortion.

It is practically impossible to sift out and measure all causes of irregularities in a reasonable time. Therefore the irregularities have been shown exactly as measured in all the following curves wherein noise and interchannel interference were absent.

Transmission fluctuations due to the causes just mentioned were small compared to those encountered when strong interference or noise was present, because the latter varied greatly with time. In resistance noise, for example, peaks of great amplitude occur occasionally, although most of the time the fluctuations are relatively minor. It was necessary to observe for several minutes the distortion measured in the presence of resistance noise before one could be sure of finding anything approaching the maximum distortion; and the longer the period of observation, the greater was the peak distortion. In order to complete the testing in a reasonable time, watching periods were restricted to five minutes per observer and his maximum distortion reading was recorded. The results of two such observation periods for the same noise condition were averaged to determine a point for an experimental curve of distortion vs. noise-to-carrier ratio. Such points when plotted failed to lie in a perfectly smooth curve, but a smooth curve disregarding irregularities was drawn through the available points in what was estimated to be the correct location. The curves are described under the heading "Noise Tests" and may be used for comparison purposes, but are not an exact measure of the worst distortion to be expected over a long period of time. A similar procedure was followed for distortion measurements with interchannel interference.

Distortion vs. Speed Tests

The distortion mentioned throughout this paper is the absolute value of the maximum total distortion measured with the test sentence, and for brevity is merely called distortion. Except where otherwise specified, the arrangements were previously adjusted to have zero bias on reversals at the same speed. In these tests line noise and interchannel interference were absent.

Frequency-Shift Arrangements

Limiter

Some preliminary measurements on a frequency-shift arrangement having the loss characteristic of curve A of Fig. 1, with carrier varied abruptly from 1920 to 1990 cycles, showed that the limiter has little effect on the speed of the channel whether or not channel filters are used. The fact that low distortion was measured without channel filters indicated that the channel and measuring devices were in good condition as far as could be reasonably expected over the range of speeds. The distortion measured without channel filters ranged from 2.5 per cent at 60 w.p.m. (23 d.p.s.) up to 10 per cent at 170 w.p.m. (65 d.p.s.).

Swing

Some measurements were made on the complete frequency-shift arrangement over the same range of speeds, using several values of abrupt frequency swing from $\pm$ 15 cycles to $\pm$ 55 cycles, keeping the marking and spacing frequencies equidistant from 1955 cycles. With a swing of $\pm$ 55 cycles the distortion was slightly worse than when the swing was $\pm$ 35 cycles. The measurements showed least distortion for a swing of $\pm$ 15 cycles. It has been previously shown[6] that the less the swing the smaller the amplitude of the oscillations in the transient for a given channel frequency band width. Accordingly one might expect distortion to be least when the swing is least. If only a small swing is used the signal bias change with carrier frequency drift is worse, unless automatic bias compensation is provided. Greater amplification is also required in the detector-amplifier in order to maintain the same relay operating current. A swing of $\pm$ 35 cycles was used in most of the frequency-shift tests as a good compromise between the distortion caused by the greater swings and the severe apparatus requirements and greater susceptibility to noise when using the lesser swings.

Types of Modulator

Figure 3 shows distortion characteristics of a frequency-shift arrangement using different types of modulator. Curve A was measured with sinusoidal frequency variation obtained by the use of a diode modulator and low-pass filter at the modulator input[5]. When the low-pass filter was omitted, the frequency variation was substantially abrupt, and curve B was obtained. When the diode modulator was replaced by a relay modulator, which also produced a substantially abrupt frequency change, curve C resulted. There is not much difference between these characteristics at low speeds. At high speeds the abrupt frequency variation appears to give somewhat lower distortion than sinuosidal variation. The distortion shown by curve A depends not only upon the speed and channel filter characteristic but also upon the characteristic of the low-pass filter used in rounding the sent wave in order to produce sinusoidal frequency variation. A considerable amount of care was necessary to prevent this low-pass filter from introducing too much distortion and at the same time to produce sufficient rounding. The cut-off frequency of this low-pass filter was adjusted at each signaling speed to be about three times the dot frequency. It was apparently low enough to

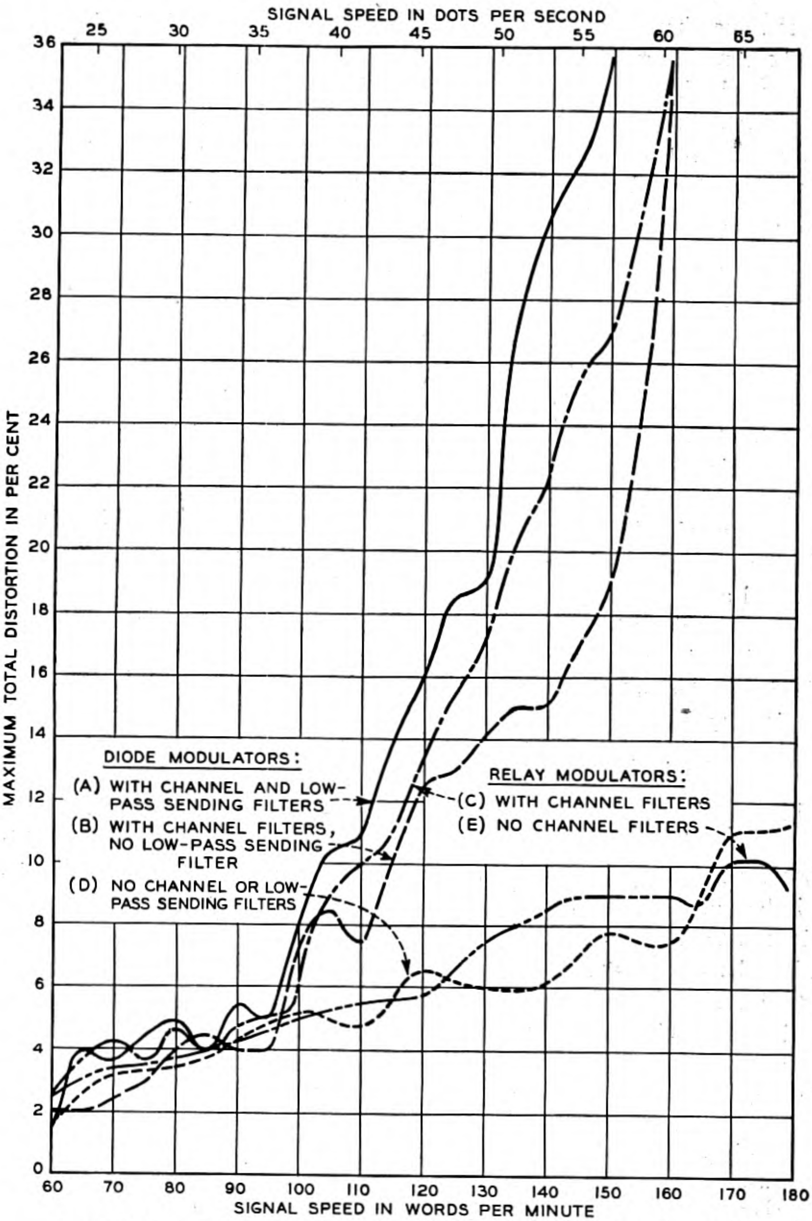


Fig. 3—Distortion vs. speed characteristics of normal-band frequency-shift arrangements, using two different types of modulator.

cause some characteristic distortion. No attempt was made to obtain an ideal distortionless filter characteristic because it would have been necessary also to take into account the characteristics of other parts of the circuit, which would have required considerably more time than was then available. Curves D and E of Fig. 3 show that the modulators caused very little distortion when channel and low-pass sending filters were absent.

Band Width

Figure 4 shows a comparison of the distortion vs. speed characteristics of the normal-band and wide-band frequency-shift arrangements when a relay modulator was used. Curve A of Fig. 4 is the same as curve C of Fig. 3 and shows the characteristic of the normal band arrangement. Curve C of Fig. 4 shows the characteristic of the wide-band arrangement when the low-pass filters at the detector output were the same as for the normal-band arrangement, their cut-off frequency being about 503 cycles. The distortion shown in curve C is much lower than that of curve A since the width of the side-bands transmitted was doubled. Curve B of Fig. 4 shows the characteristic of the wide-band arrangement when the low-pass filters at the detector output had a cut-off at about 58 cycles. As previously indicated, the latter cut-off was selected in order to give the wide-band arrangement about the same distortion vs. speed characteristic as the normal-band arrangement. The reason for the use of the lower cut-off is explained under the heading "Noise Tests".

Comparison of Frequency-Shift and On-Off Arrangements

Figure 5 is a comparison of the distortion vs. speed characteristic of the frequency-shift arrangement having a relay modulator (curve B), with characteristics of two on-off arrangements having commercial receiving circuits[9, 10, 11]. The 40B1 detector had no level compensator and included a triode detector having an output vs. input characteristic which roughly followed a square law. The other detector had a slow acting level compensator[11] designed to eliminate receiving bias due to slow changes in line equivalent. The output vs. input characteristic of this detector was much steeper than that of the 40B1 arrangement at the transition points of the signals. The level compensated arrangement was adjusted at each speed to

[9] B. P. Hamilton, H. Nyquist, M. B. Long and W. A. Phelps: "Voice Frequency Carrier Telegraph System for Cables", *Jour. A. I. E. E.*, Vol. XLIV, No. 3, Mar. 1925.

[10] A. L. Matte: "Advances in Carrier Telegraph Transmission", *B. S. T. J.*, Vol. XIX, pp. 161–208, Apr. 1940.

[11] A separate paper describing a commercial system using an improved level compensator is now in preparation by other Bell System authors. The function of the level compensator is to vary the gain of the receiving amplifier-detector so as to automatically compensate for relatively slow level changes. See also: V. P. Thorp: "A Level Compensator for Carrier Telegraph Systems", *Bell Laboratories Record*, Vol. XVIII, No. 2, October 1939, pp. 46–48.

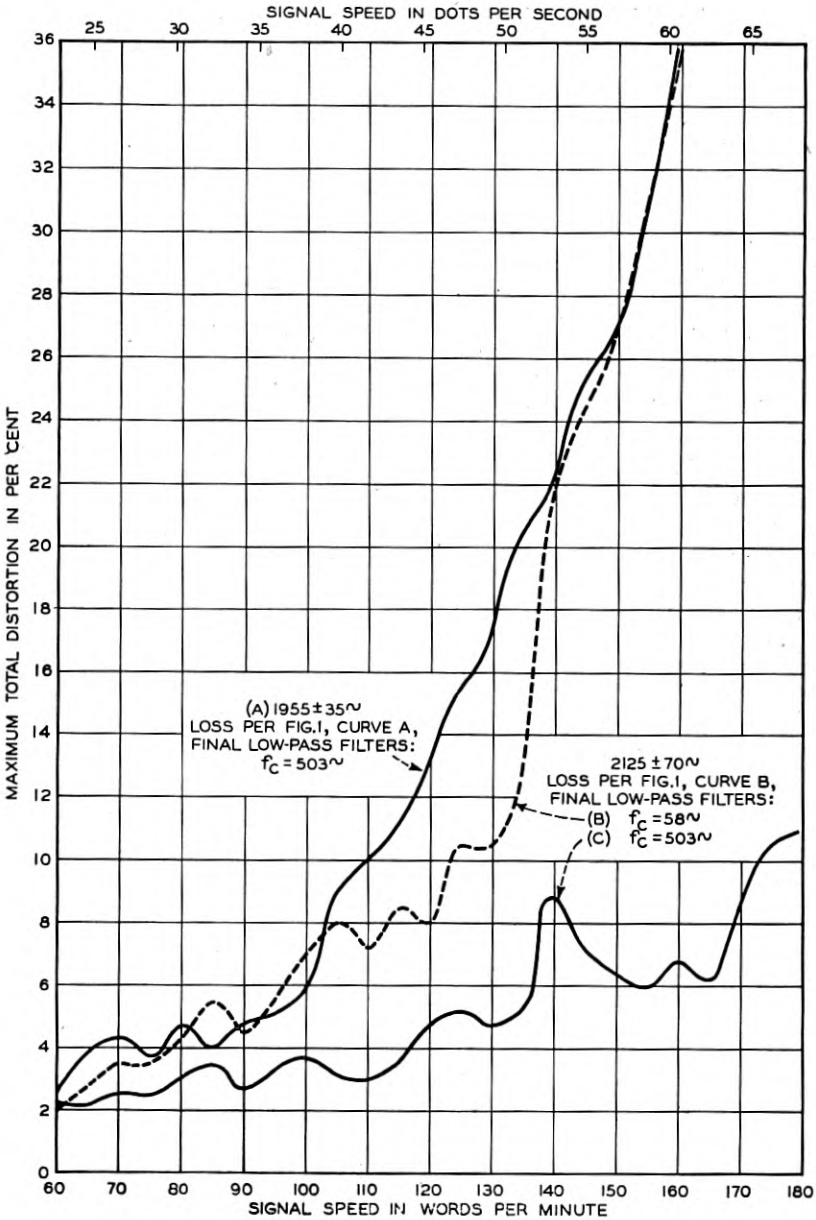


Fig. 4—Distortion vs. speed characteristics of normal- and wide-band frequency-shift arrangements.

have zero bias while transmitting the test sentence because the characteristic of the level compensator was such that reversals could not be used to adjust for zero bias of the received signals. The distortion vs. speed characteristics of the level compensated on-off and 40B1 arrangements are given by curves A and C, respectively, in Fig. 5. This figure indicates that at high speeds the frequency-shift arrangement is subject to somewhat greater distortion than the on-off method.

It may be argued that Fig. 5 is not a fair comparison between frequency-shift and on-off methods because of the difference in detector characteristics. In order to overcome this objection an experimental on-off arrangement was set up utilizing a linear detector and the same receiving relay as in the frequency-shift arrangement. The effective operating ampere-turns in the 255A receiving relay were kept the same for both frequency-shift and linear on-off methods of transmission. The line winding current varied from about 10 mils during spacing signals to 50 mils during marking signals, and the biasing winding current tending to move the armature toward spacing was about 30 mils for the on-off arrangement. The effective relay operating current in the frequency-shift arrangement was $+20$ mils in the marking condition and -20 mils in the spacing condition.

The distortion measurements are shown in Fig. 6. In order to compare frequency-shift with the linear on-off arrangement, consider curves A and B of Fig. 6. There is not much difference between them, but the frequency-shift characteristic shows slightly higher distortion over part of the speed range, as in Fig. 5.

In these tests the channel loss characteristic used was that of Fig. 1, curve A.

Two-Source and One-Source Two-Band Arrangements

The same linear detector, receiving relay, and effective relay operating current were used for these two-band arrangements as for the frequency-shift arrangement. Curves C, D, and E of Fig. 6 show the speed characteristics of various two-source and one-source two-band arrangements in which the marking and spacing paths had loss characteristics similar to curve A of Fig. 1. Curve C of Fig. 6 applies to the arrangement using two oscillators and no limiter and does not differ greatly from the characteristic for the on-off method, curve B. Curve D applies to the arrangement using two oscillators and limiter, and shows greater distortion than curve C because of modulation products arising in the limiter between the sidebands of the marking and spacing carriers. This type of interference was due to discontinuities in phase of the carrier wave at the signal transitions, and was eliminated by the use of a frequency modulated oscillator in place of the two independent oscillators, as indicated by curve E. The latter is somewhat similar to curve C on Fig. 4 measured on the frequency-shift arrangement with the wide

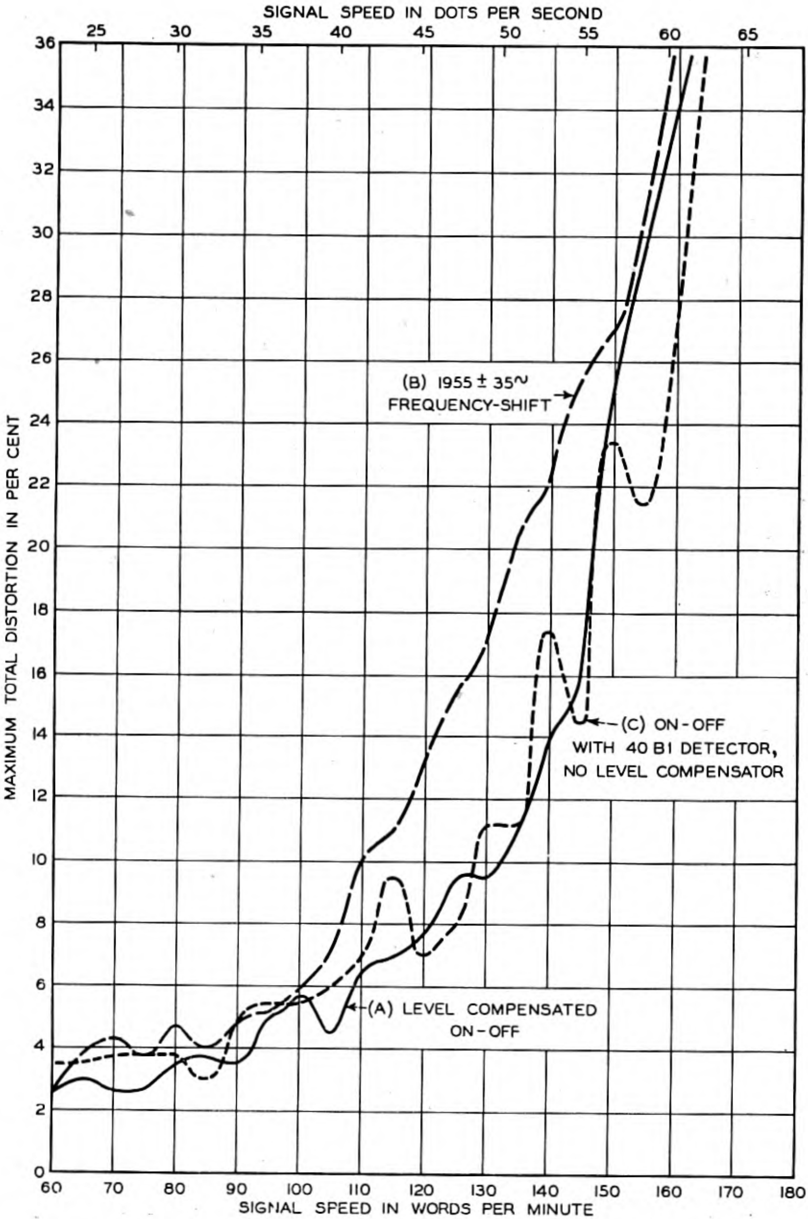


Fig. 5—Distortion vs. speed characteristics of frequency-shift and non-linear on-off arrangements.

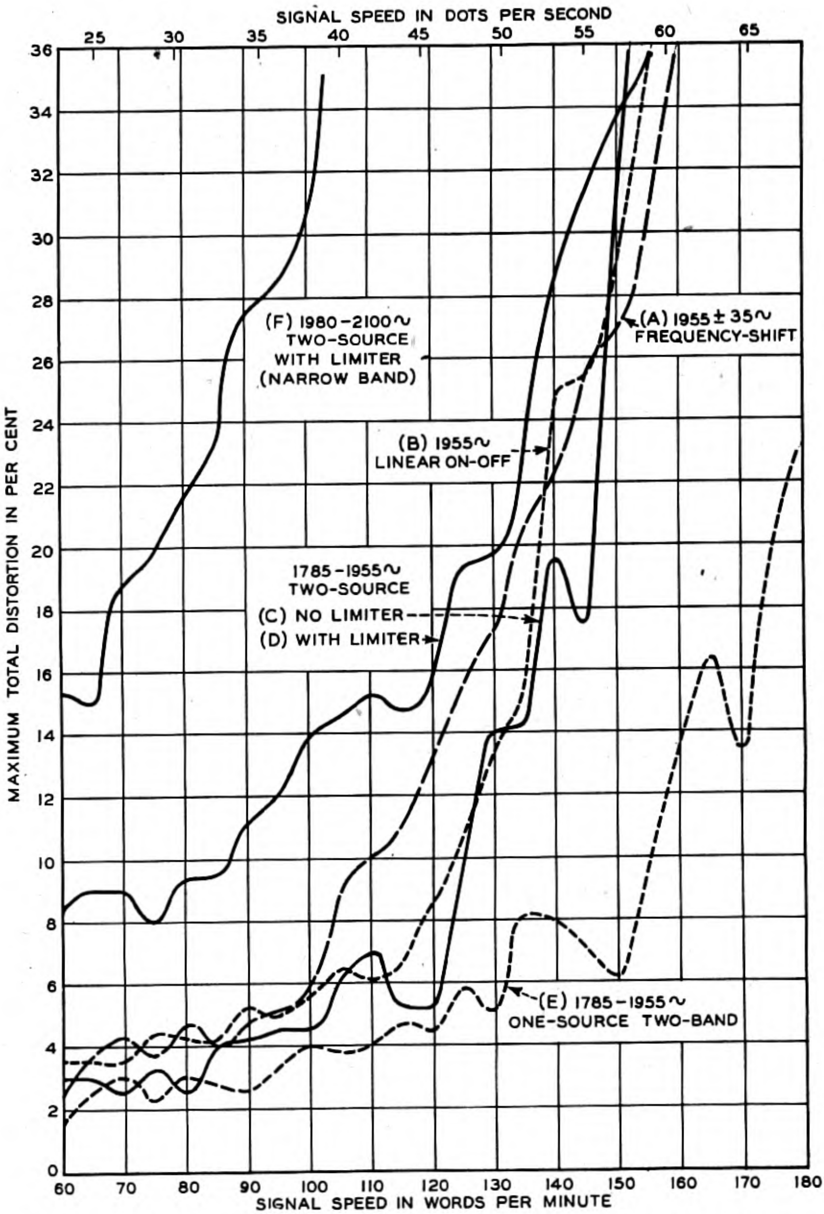


Fig. 6—Distortion vs. speed characteristics of frequency-shift, linear on-off, two-source and one-source two-band arrangements.

band, except that the distortion at the higher speeds is greater with curve E of Fig. 6 because sideband components in the middle portion of the total frequency band were considerably attenuated by the channel filters of the one-source two-band arrangement, but not by those of the wide-band frequency-shift arrangement. The distortion vs. speed characteristic of the two-source arrangement having paths with loss characteristic similar to curve C of Fig. 1, is shown by curve F of Fig. 6. The large distortion is due to the narrower sidebands transmitted and to the combined use of two independent oscillators and a limiter, the effect of which is discussed above.

Single-Sideband Arrangement

The same linear detector, receiving relay, and effective relay operating current were used for the single-sideband tests as for the linear on-off tests.

An ideal single-sideband arrangement should operate at twice the speed of the on-off arrangement for the same pass band, if the quadrature component[2] is eliminated. The cost of a phase discrimination method of reception[1] for this purpose would probably be prohibitive in practice. If the quadrature component is allowed to remain, it is a principal cause of distortion, so that the single-sideband method gives only a slight increase in speed. The effect of the quadrature component on telegraph distortion can be reduced[2] by the transmission of a certain amount of spacing carrier current. Curve B of Fig. 7, measured with a spacing current 6 db below the marking current, shows less distortion than curve A of Fig. 7, measured with no spacing current; and, in the range of speeds investigated, does not differ greatly from curve C, taken on the linear on-off arrangement without channel filters. Thus, it is apparent that the single-sideband arrangement is capable of higher speeds, for a given distortion and band width, than the other arrangements here considered.

Tests of Carrier Frequency Variations

When a carrier telegraph circuit contains a radio or carrier telephone link, some instability may occur in the average received carrier frequency. In order to investigate the effect of varying the mean carrier frequency, the carrier supply frequency was varied as a matter of convenience. Since the signals were transmitted through both sending and receiving channel filters the effects observed were doubtless about twice as bad as if only the received carrier frequency had been varied, except perhaps in the frequency-shift arrangements where the discriminator produced a large effect.

Distortion at 60 Words per Minute

In Figs. 8 and 9, the distortion obtained over the various arrangements is shown as a function of carrier frequency variation from the nominal value, when the speed was 60 w.p.m. The marking and spacing frequencies of the two-source, one-source two-band, and frequency-shift arrangements were

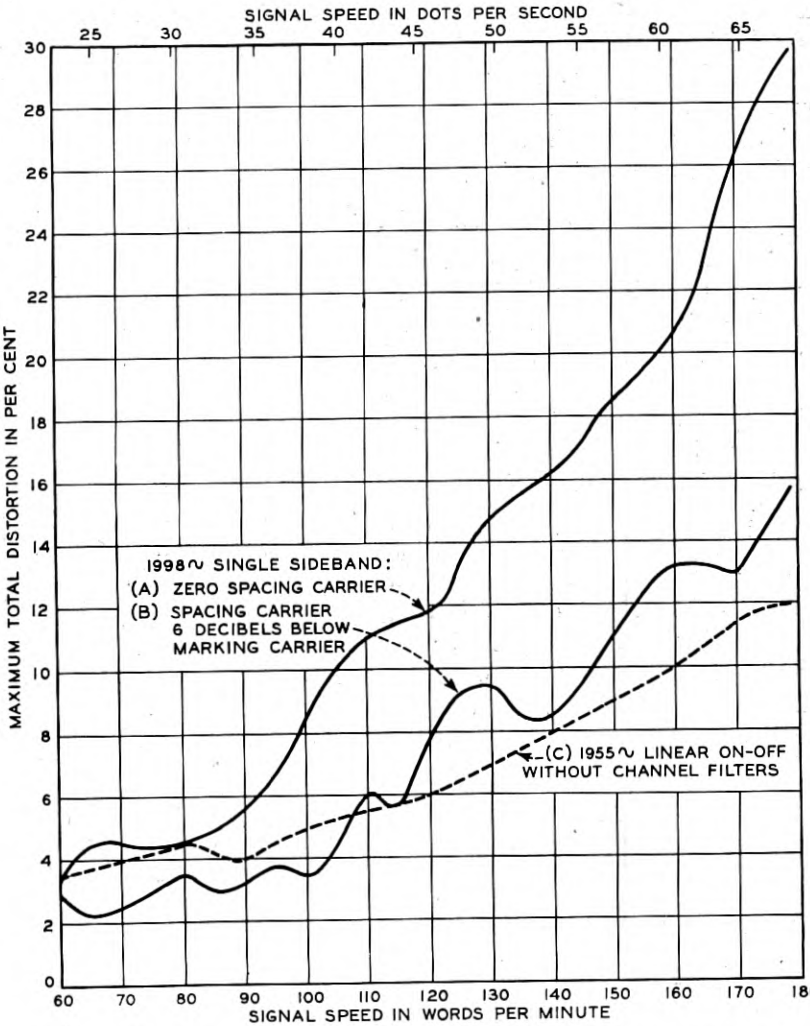


Fig. 7—Distortion vs. speed characteristics of single-sideband arrangements and on-off arrangement without channel filters.

both varied by the same amount from their normal values without changing their relative separation. It is evident from Fig. 8 that the arrangements ranked in the following order as regards the permissible carrier frequency

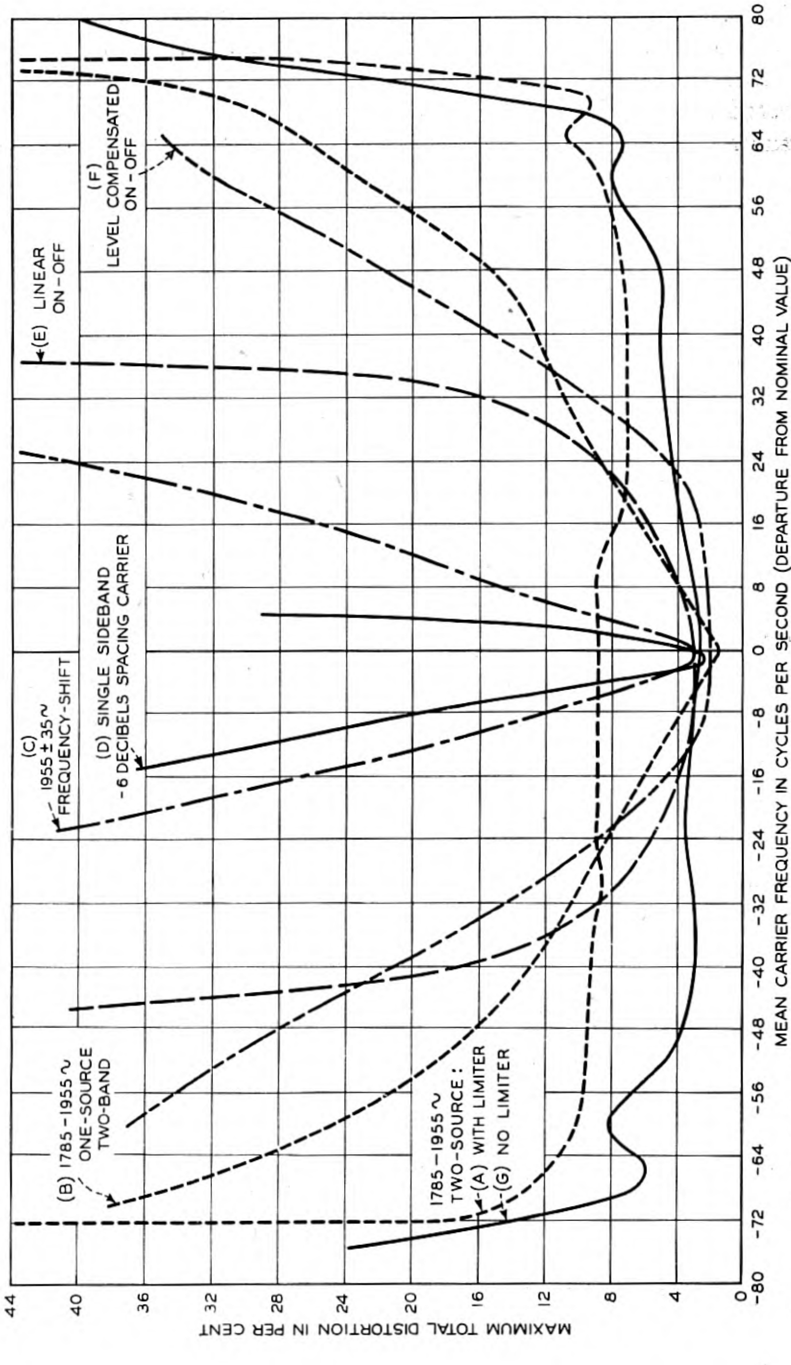


Fig. 8—Distortion vs. mean carrier frequency characteristics of normal-band arrangements at 60 w.p.m. (23 d.p.s.).

variations when the distortion was about 20 per cent: two-source without limiter on a par with two-source with limiter, one-source two-band with limiter, level compensated on-off, linear on-off, frequency-shift, and single-sideband with −6 db spacing carrier. This comparison includes only those arrangements using the same type of filter (with loss per curve A, Fig. 1). Figure 9 shows the results obtained by making similar tests on the wide-band frequency-shift (with loss per curve B, Fig. 1) and on the narrow-band two-

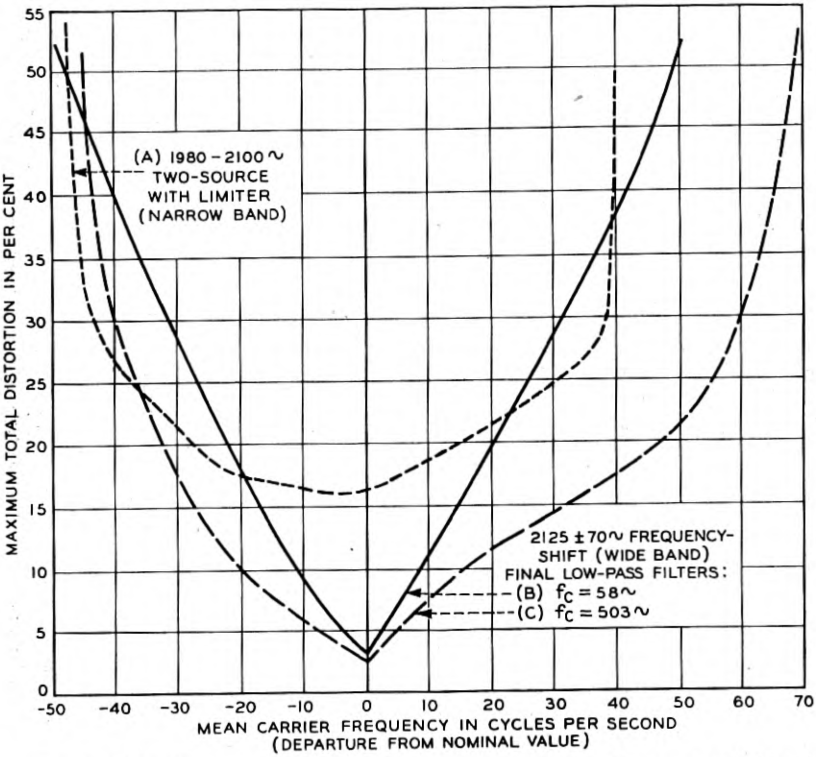


Fig. 9—Distortion vs. mean carrier frequency characteristics of narrow- and wide-band arrangements at 60 w.p.m. (23 d.p.s.).

source arrangements (with loss per curve C, Fig. 1). The distortion curves for the wide-band frequency-shift arrangement in the latter figure are more favorable than the curve for the normal-band frequency-shift arrangement shown in Fig. 8 because the wide-band arrangement had a discriminator with only half the slope of that used in the normal-band arrangement. The wide-band arrangement with the higher cut-off low-pass filter could tolerate a greater change in carrier frequency than the arrangement with the lower cut-off low-pass filter because of the steeper wave front of the signals delivered by the former low-pass filter, resulting in a lower signal bias when the detected amplitudes of the marking and spacing signals differed.

When the average carrier frequency of a two-source arrangement was varied, the marking and spacing frequencies moved toward one side of their respective pass bands. As they approached the cut-off frequencies of the band-pass filters, telegraph distortion occurred due to suppression of the carrier and adjacent components. The narrow-band two-source arrangement was more sensitive to variation in the average carrier frequency than the normal-band two-source arrangement, and the reason is obvious.

Bias at 60 Words per Minute

When tested with reversals at 23 d.p.s. (corresponding to 60 w.p.m.), the arrangements also ranked in the same order from the bias standpoint as from the distortion standpoint, as the mean carrier frequency was varied. In Figs. 8 and 9 the distortion due to carrier frequency variation consisted mainly of bias, except in the two-source arrangements, where the received marking and spacing pulses were substantially equal on reversals so that there was little bias. (Bias measurements referred to here and below have not been shown graphically in order to save space.)

The on-off, two-source, and one-source two-band arrangements were fairly insensitive to carrier frequency variations since the loss vs. frequency characteristics of the channel filters changed but slowly near the middle of the transmission band. Since the frequency-shift arrangement had a discriminator which was sensitive to frequency changes, drifting of the average carrier frequency resulted in a rise in detected current in one half of the push-pull detector and a reduction thereof in the other half, thus causing serious bias in the differentially combined rectified waves, since no frequency compensator was provided. It is outside the scope of this paper to describe such a compensator, but it is no more complicated than the level compensator used with an on-off arrangement. In the single-sideband arrangement the carrier was located at a point on the filter loss vs. frequency characteristic where the slope was steep. Consequently small frequency changes produced large amplitude variations in the operating current of the receiving relay. Since there was no compensating change in the biasing current of the relay, large bias variations resulted from small changes in carrier frequency.

Distortion at 120 Words per Minute

Figures 10 and 11 give distortion for the arrangements at 46 d.p.s. or 120 w.p.m. when the mean carrier frequency was varied. The arrangements ranked in the following order when the distortion was 20 per cent: two-source without limiter, two-source with limiter, linear on-off, one-source two-band,

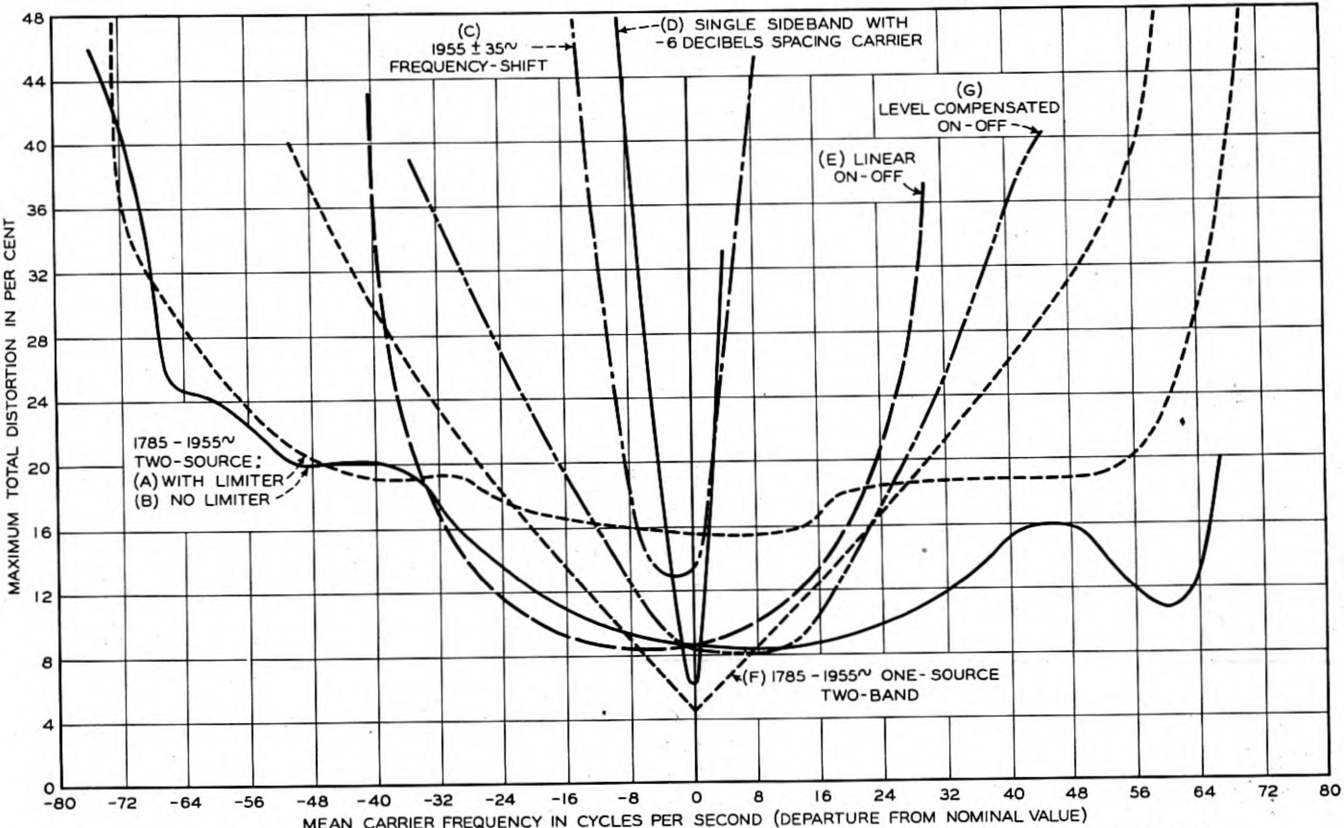


Fig. 10—Distortion vs. mean carrier frequency characteristics of normal-band arrangements at 120 w.p.m. (46 d.p.s.).

level compensated on-off, frequency-shift, single-sideband with -6 db spacing carrier.

In both Figs. 8 and 10 there appears to be considerable difference between the distortion vs. carrier frequency characteristics of the two-source and one-source two-band arrangements with limiter. The two-source arrangement was better for large carrier frequency variations, and the one-source two-band arrangement was better for small carrier frequency variations.

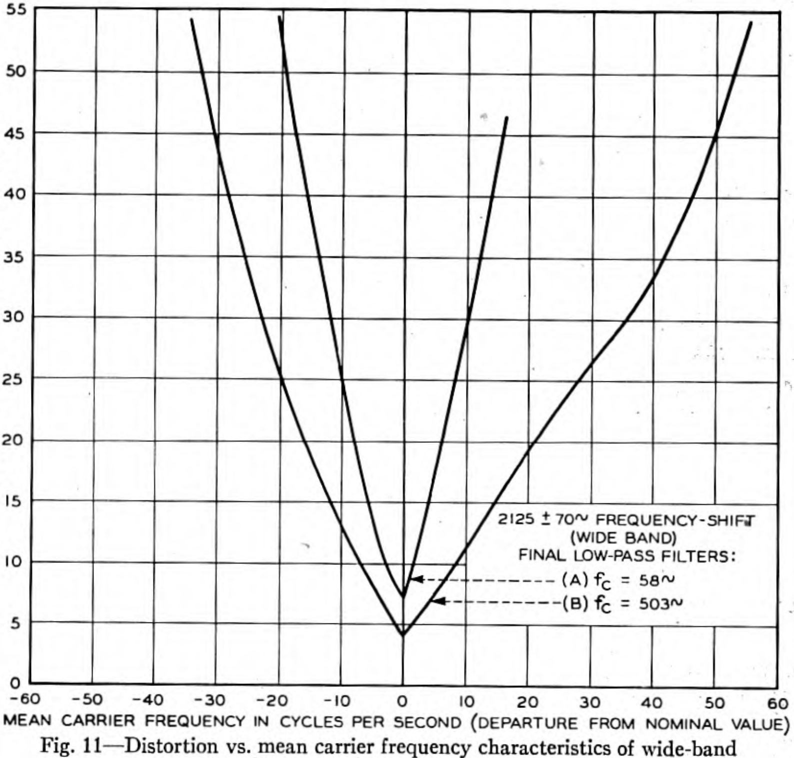


Fig. 11—Distortion vs. mean carrier frequency characteristics of wide-band frequency-shift arrangements.

The difference between the performance of these two arrangements was partly due to interference between sidebands in the two-source arrangement as previously mentioned, and partly due to the difference in the sending arrangements used. In the case of the two-source sending arrangement shown in Fig. 2, the sidebands of the marking and spacing paths were separated by the sending filters, and a small shift in carrier frequency affected each branch similarly, and relatively little bias or distortion resulted. In the case of the one-source two-band method also shown in Fig. 2, the sidebands of the marking and spacing frequencies were not completely separated by the sending filters and a shift in carrier frequency affected the two sets of sideband components dissymmetrically, thus causing bias and distortion.

Bias at 120 Words per Minute

The bias resulting from carrier frequency change in the various arrangements at 46 d.p.s. or 120 w.p.m. was in general worse than at half this speed because the received wave shape was more rounded. The distortion due to carrier frequency variation, as shown in Figs. 10 and 11, consisted mainly of bias, except in the two-source arrangements, as explained above.

Other Considerations Relating to Single-Sideband Arrangement

The single-sideband arrangement with −6 db spacing carrier was found to have the lowest distortion at high signal speeds. Consequently it was thought desirable to study this arrangement further in order to see what might be done to improve its stability during carrier frequency variations, and to select an optimum location for the average carrier frequency. First, curve A of Fig. 12 was plotted showing the distortion at 160 w.p.m. resulting from a change in carrier frequency. Then it was assumed that a level compensator might be provided. In order to simulate the effect of such a device without actually constructing one, the line loss was adjusted manually to keep the r.m.s. marking carrier power constant at the receiving filter output. Curve B of Fig. 12 shows an improvement in the change in distortion vs. carrier frequency under this condition. Large variations in bias still persisted in spite of level compensation, due to variations in shape of the envelope of the received carrier signals caused by various amounts of quadrature component and sluggish in-phase component[1] depending on the location of the carrier frequency. In order to simulate the effect of a level compensator providing automatic bias adjustment, the relay bias current was adjusted to give zero receiving bias at each setting of the carrier frequency, and also the receiving filter output was maintained at a constant value as before. The results are given by curve C in Fig. 12 which shows a considerably increased tolerance to carrier frequency changes when the arrangement was stabilized in this manner. This curve indicates that the carrier frequency could be increased about 17 cycles before the distortion started to increase rapidly, and could be reduced about 5 cycles before a gradual increase in distortion began to appear.

It has been shown that for a certain ideal filter[1], a suitable location of the carrier frequency for the single-sideband method is at a point either at the upper or lower side of the band where the loss is 6 db with respect to that in the middle of the band. It can be demonstrated that the envelope of the received wave is the same when the carrier frequency is set at either of these locations, if the filter characteristic is symmetrical about the midband frequency, if the pass band is narrow, and if the carrier frequency is high compared to the dot speed, as in the arrangement tested. Consequently there was no point in duplicating measurements for carrier locations at the lower edge of the band except perhaps to discover the second order effect of slight asymmetry in the channel filters. 1998 cycles is at the right-hand 6 db point of the filter characteristic given by curve A of Fig. 1; and as there is not much choice in the region from 5 cycles below to 17 cycles above this

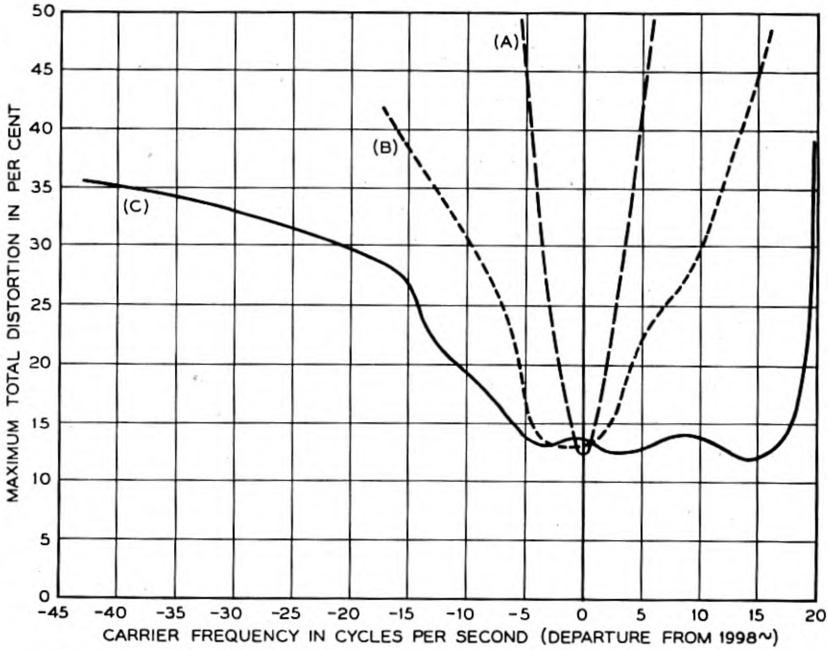


Fig. 12—Distortion vs. carrier frequency characteristics of 1998-cycle single-sideband arrangement using spacing carrier current 6 db below marking current at 160 w.p.m. (61 d.p.s.).

A: Fixed adjustments. B: Constant marking output level from receiving filter, and fixed relay bias current. C: Same as B, except receiving signal bias adjusted to zero by varying receiving relay bias current while transmitting reversals.

frequency, according to curve C of Fig. 12, it was considered satisfactory to locate the carrier at 1998 cycles for certain other single-sideband tests covered in this paper, although a slightly higher or lower value would probably have done about as well. As previously indicated, the reason for assuming a spacing carrier current 6 db below the marking value for single-sideband tests was to reduce the distortion due to the quadrature component[2]. Still less effect from the latter would have existed if the spacing carrier current had been further increased, but this would have reduced the difference between the marking and spacing current amplitudes on the line, causing a corresponding reduction in sideband power. If the quadrature component had been completely eliminated it is possible that 1998 cycles would have been found to be a more favorable carrier location than some of the other closely adjacent frequencies. However, no attempt was made to design the filters for the theoretical single-sideband requirement that the transfer admittance of the vestigial sideband should be complementary[1] to that of the other sideband near the carrier frequency.

Tests of Received Level Variations

Any transmission path is likely to have level variations caused by temperature and weather changes in case of wire lines and by fading in case of radio links. Each arrangement was therefore tested for susceptibility to level changes by varying the artificial line over a wide range. As the artificial line was made of resistances, the effect of equal fading over the entire frequency range was thereby simulated.

Distortion at 60 Words per Minute

In Fig. 13 is shown the total distortion at 60 w.p.m. vs. level change for different arrangements using the same type of filter. As may be seen from Fig. 13, the arrangements which had the same loss characteristic ranked in the following order at 20 per cent distortion and 60 w.p.m. as regards stability when the line level was varied: two-source with limiter on a par with one-source two-band, frequency-shift, level compensated on-off, two-source without limiter, linear on-off, and single sideband with −6 db spacing carrier. The range of levels over which the first three arrangements mentioned above were stable, in the absence of interference, was largely a function of the range of the limiter, which was over 80 db. The on-off arrangement including a level compensator had a range of approximately 40 db, but it should be remembered that a level compensator is effective only for level changes slow compared to the signal speed. The range of the two-source arrangement without limiter depended on the accuracy with which the marking and spacing halves of the receiving circuit were balanced. In the arrangement tested this range was about 30 db. The linear on-off arrangement without level compensator was very sensitive to level changes because the operating current in the receiving relay varied without a compensating variation in the bias current. The single-sideband arrangement had greater sensitivity than the linear on-off arrangement because a variation of 6 db in the line current of the single-sideband arrangement was accompanied by a 15 db variation in relay operating current due to the use of increased gain and a large grid bias in the receiving d-c amplifier.

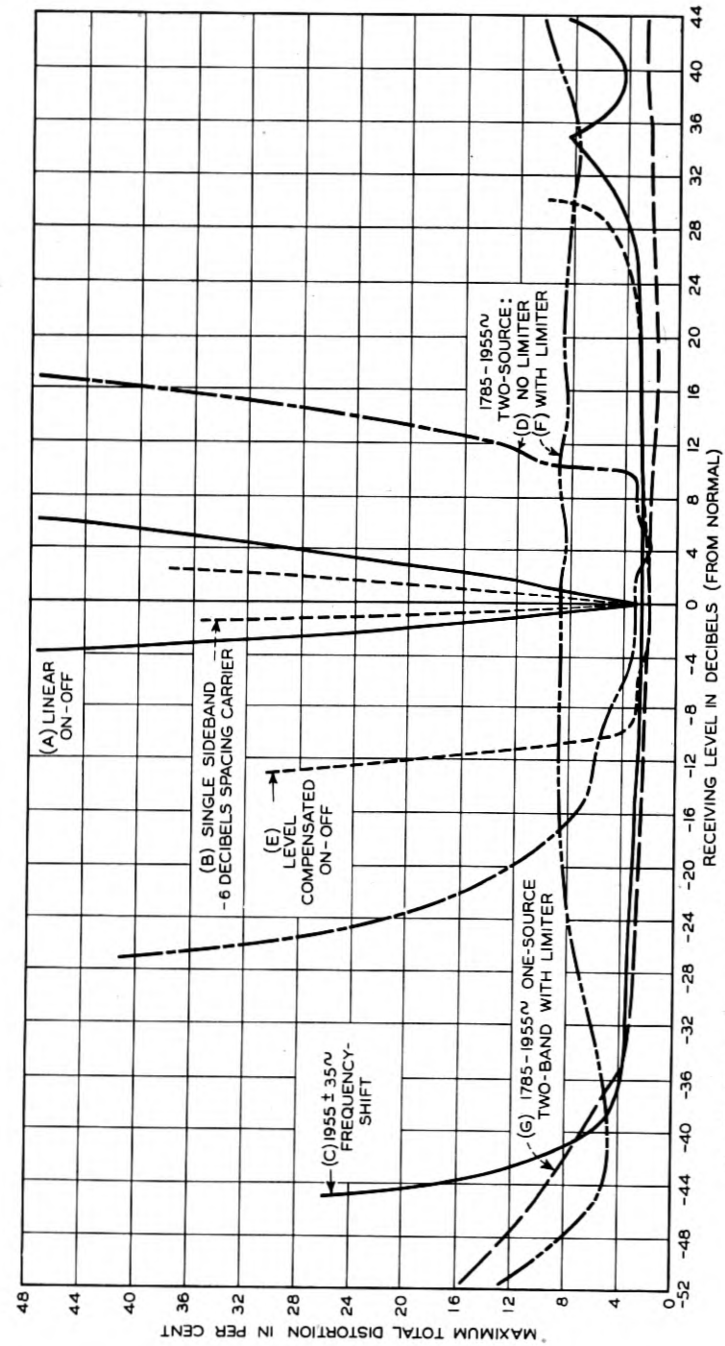


Fig. 13—Distortion vs. receiving level characteristics of normal-band arrangements at 60 w.p.m. (23 d.p.s.).

Just as the line level range of an on-off arrangement may be extended by the use of a level compensator, so it seems reasonable to expect that a similar improvement could be obtained by using level compensators on the linear on-off and single-sideband arrangements. However, when a level compensator is applied to such arrangements, it must be a slow acting device in order not to cause characteristic distortion. When large and rapid level changes are frequent, as may occur on a radio circuit due to fading, an arrangement including a fast acting device like a limiter is preferred, such as a two-source, one-source two-band, or frequency-shift arrangement.

The effect of level variations on the distortion of the narrow-band two-source arrangement and of the wide-band frequency-shift arrangement with and without low cut-off low-pass filter is shown in Fig. 14. At 60 w.p.m. the wide-band frequency-shift arrangement with low cut-off low-pass filter could

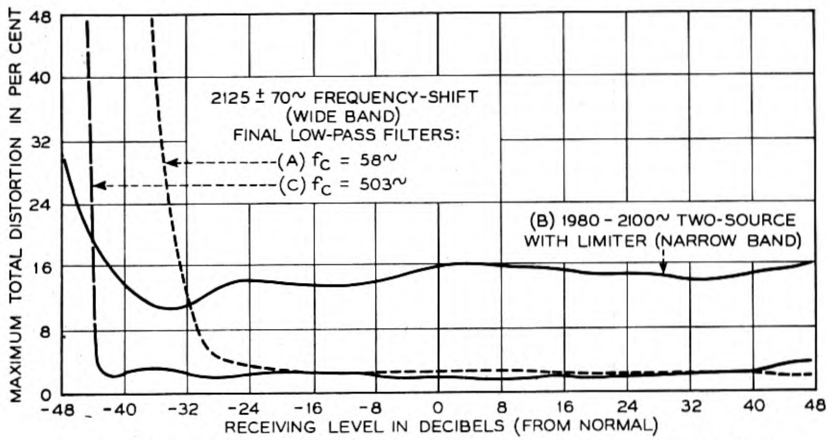


Fig. 14—Distortion vs. receiving level characteristics of narrow- and wide-band arrangements at 60 w.p.m. (23 d.p.s.).

tolerate the nearly same level change as the normal-band frequency-shift arrangement if 20 per cent distortion is the limit. When the low-pass filter of the wide-band frequency-shift arrangement had a high cut-off, the tolerance was somewhat greater due to the steeper wave front of the detected signals which rendered them less suceptible to bias caused by slight unbalance in the detector and d-c amplifier at levels below the cut-off of the limiter. The narrow-band two-source arrangement tolerated slightly less level change than the normal-band two-source arrangement with limiter, for the same reason.

Distortion at 120 Words per Minute

When the speed was doubled the effects of level change on distortion were found to be as shown by Figs. 15 and 16, except that tests on the narrow-band two-source arrangement were omitted because of high distortion at

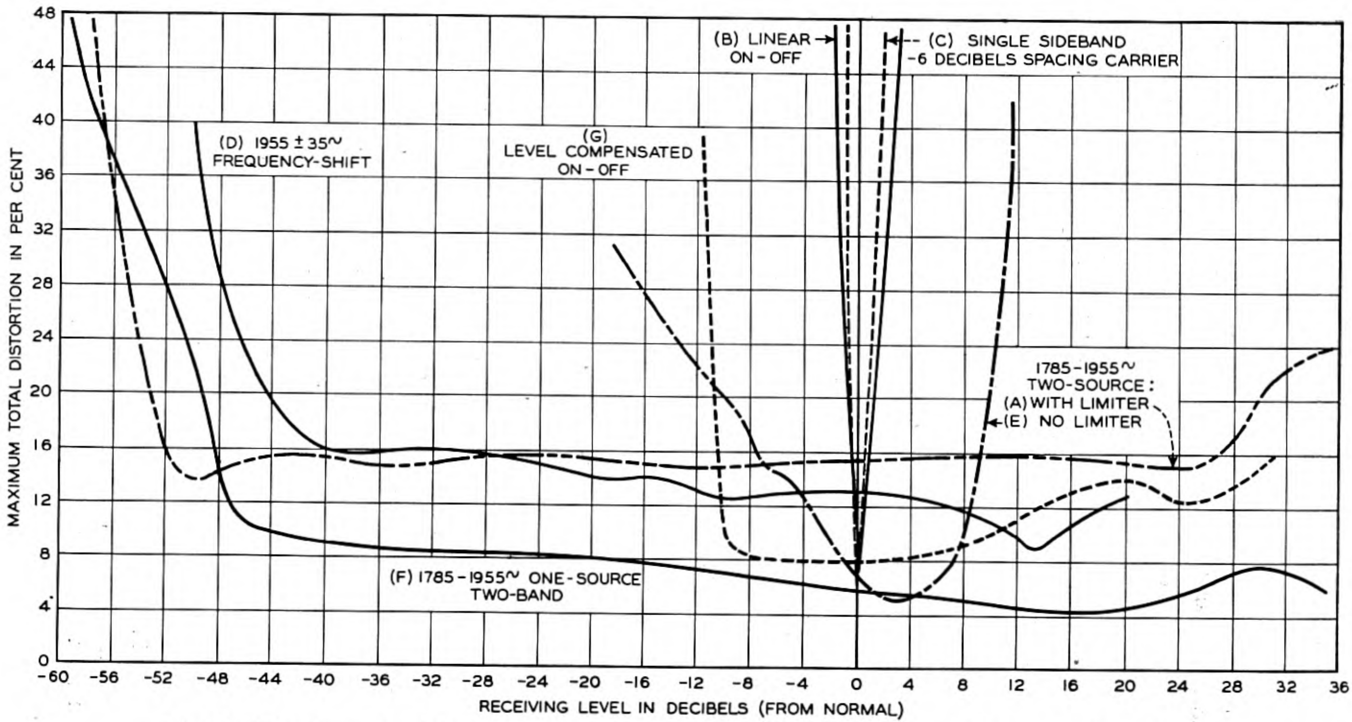


Fig. 15—Distortion vs. receiving level characteristics of normal-band arrangements at 120 w.p.m. (46 d.p.s.).

120 w.p.m. The arrangements tested were found to have about the same relative susceptibility to level changes at this speed as at 60 w.p.m.

The distortion shown in Figs. 13 and 15 consisted largely of bias for the arrangements having no limiter. The distortion shown in Figs. 13, 14, 15, and 16 for arrangements which had limiters rose faster than the absolute value of the bias as the level dropped below the cut-off of the limiter. This was partly due to extraneous noise in the laboratory apparatus.

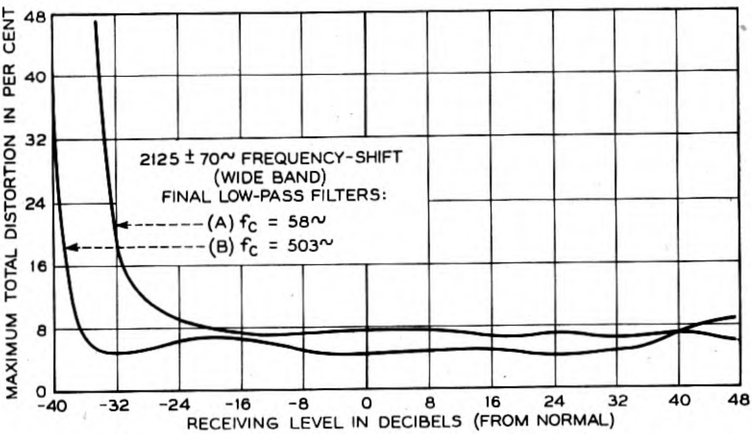


Fig. 16—Distortion vs. receiving level characteristics of wide-band frequency-shift arrangements at 120 w.p.m. (46 d.p.s.).

Selective Level Changes

When frequency-shift, two-source, or one-source two-band arrangements are operated over a radio link, selective fading may occur which affects the marking and spacing frequencies by different amounts. Tests were made only on the two-source arrangements to measure the effect on distortion and bias of differences between received marking and spacing levels. This effect was simulated by setting the level of one of the two carrier oscillators at different values with respect to the other, and measuring the distortion and bias for each level setting. The distortion measurements are shown in Fig. 17 and are summarized below in Table I.

It may be seen from Fig. 17 that the distortion rose rather rapidly as the marking level was changed with respect to the spacing level and thus the distortion due to selective fading was not greatly reduced by the use of a limiter. The increase in distortion was mainly bias.

Tests of Adjacent Channel Crossfire

In order to investigate the effects on distortion due to interference from adjacent channels, certain arrangements having the same loss characteristic were tested while upper and lower adjacent flanking channels were transmitting reversals produced by separate vibrating relays at roughly 23 d.p.s.

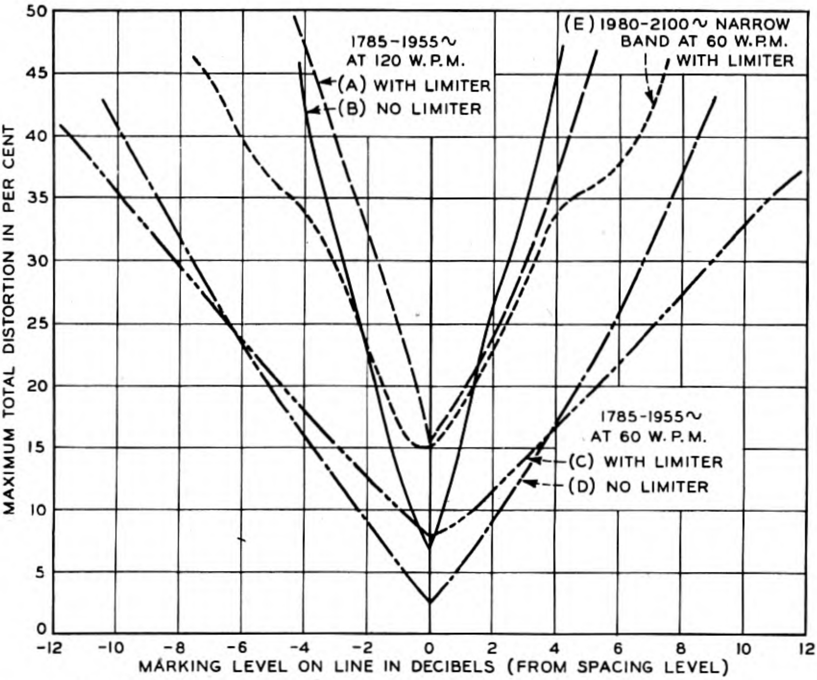


Fig. 17—Distortion vs. difference between marking and spacing level. Comparison of two-source arrangements.

TABLE I

COMPARISON OF TWO-SOURCE ARRANGEMENTS, EFFECT OF DIFFERENCE BETWEEN MARKING AND SPACING LEVELS

Mid-band Frequencies	Limiter	Marking Level on Line—DB from Spacing Level			
		For 10% Distortion		For 20% Distortion	
		At 60 W.P.M.	At 120 W.P.M.	At 60 W.P.M.	At 120 W.P.M.
1785–1955~	Out	−2.2 to +2.2	−0.6 to +0.4	−5.1 to +4.7	−1.7 to +1.4
1785–1955~	In	−1.0 to +1.4	*	−4.7 to +5.3	−0.4 to +1.0
1980–2100~ (narrow band)	In	*	failed	−1.5 to +1.3	failed

* Distortion exceeded 10%.

The distortion with and without the flanking channels operating and the difference in distortion are shown in the following tables, from which it is seen that the various arrangements tested ranked approximately in the following order as regards their insensitivity to flanking channel crossfire: linear on-off and two-source without limiter about the same, single sideband with no spacing carrier about the same as with -6 db spacing carrier, frequency-shift.

It may be observed from Tables II and III that the interchannel crossfire obtained for the linear on-off and two-source arrangements was small, due to the location of the carrier frequencies at the centers of the channel filter bands. The crossfire obtained for the single-sideband and frequency-shift methods given in Tables IV to VI was greater, due to the location of the marking or spacing frequencies near the edges of the channel filter bands.

TABLE II

1955-CYCLE LINEAR ON-OFF ARRANGEMENT, EFFECT OF OPERATING ADJACENT 1785-CYCLE AND 2125-CYCLE ON-OFF CHANNELS OVER LINE

Signal Speed, W.P.M.	Maximum Total Distortion, Per Cent		Increase in Distortion, Per Cent
	Flanking Channels Off	Flanking Channels On	
100	6.2	6.2	0
120	7.8	8.6	0.8
140	18.0	20.0	2.0

TABLE III

1785 AND 1955-CYCLE TWO-SOURCE ARRANGEMENT WITHOUT LIMITER, EFFECT OF OPERATING ADJACENT 1615-CYCLE AND 2125-CYCLE CHANNELS OVER LINE

Signal Speed, W.P.M.	Maximum Total Distortion, Per Cent		Increase in Distortion, Per Cent
	Flanking Channels Off	Flanking Channels On	
80	3.2	4.2	1.0
100	3.5	4.5	1.0
120	7.7	8.5	0.8
140	18.5	19.7	1.2

The other arrangements previously mentioned were not tested for flanking channel crossfire. The two-source arrangement with limiter and normal band width is thought to be no worse from this standpoint than that without limiter, since a limiter usually helps in discriminating against small spurious currents. From theoretical considerations the one-source two-band arrangement with limiter is thought to be better than the two-source arrangement without limiter, using the same channel filters and carrier frequencies. The narrow-band two-source arrangement had considerably sharper cut-off filters than the normal-band arrangement, thus causing greater attenuation of frequencies outside the desired band. Consequently the narrow-band two-source arrangement is thought to be no worse than the normal-band arrangement as far as interchannel interference is concerned. The wide-band frequency-shift arrangements are thought to be no worse than the normal-band frequency-shift arrangement from this standpoint, because of the greater separation between the sidebands of the adjacent channels. According to a study of the frequency spectra produced by abrupt or sinusoi-

TABLE IV

1998-CYCLE SINGLE-SIDEBAND ARRANGEMENT (NO SPACING CARRIER), EFFECT OF OPERATING ADJACENT 1828-CYCLE AND 2168-CYCLE SINGLE-SIDEBAND CHANNELS OVER LINE

Signal Speed, W.P.M.	Maximum Total Distortion, Per Cent		Increase in Distortion, Per Cent
	Flanking Channels Off	Flanking Channels On	
100	7.1	7.9	0.8
120	12.0	14.9	2.9
140	18.5	22.7	4.2
160	21.0	25.5	4.5

TABLE V

1998-CYCLE SINGLE-SIDEBAND ARRANGEMENT (WITH SPACING CARRIER 6 DB BELOW MARKING CARRIER), EFFECT OF OPERATING ADJACENT 1828-CYCLE AND 2168-CYCLE SINGLE-SIDEBAND CHANNELS OVER LINE

Signal Speed, W.P.M.	Maximum Total Distortion, Per Cent		Increase in Distortion, Per Cent
	Flanking Channels Off	Flanking Channels On	
100	5.7	9.2	3.5
120	8.5	12.0	3.5
140	14.0	17.7	3.7
160	14.0	18.2	4.2

TABLE VI

1955±35-CYCLE FREQUENCY-SHIFT ARRANGEMENT (WITH RELAY MODULATOR), EFFECT OF OPERATING ADJACENT 1785 ± 35-CYCLE AND 2125 ± 35-CYCLE FREQUENCY-SHIFT CHANNELS OVER LINE

Signal Speed, W.P.M.	Maximum Total Distortion, Per Cent		Increase in Distortion, Per Cent
	Flanking Channels Off	Flanking Channels On	
60	2.5	6.0	3.5
100	5.5	11.0	5.5
120	11.0	19.0	8.0
140	20.0	28.0	8.0

dal frequency-shift arrangements during transmission of reversals[12], it appears that the sinusoidal shift is better than abrupt shift from the standpoint of interference between adjacent channels, when the sending channel filter does not sufficiently attenuate undesired sideband components. But

[12] Balth. van der Pol: "Frequency Modulation", *Proc. I. R. E.*, Vol. 18, No. 7, July 1930, pp. 1194–1205.

there is probably little use in complicating the modulating arrangement to produce sinusoidal shift, merely for the purpose of simplifying the sending filter.

In the multi-channel (on-off type) voice frequency carrier telegraph used in the Bell System plant, the carrier currents of the different channels have frequencies which are odd multiples of an 85-cycle base frequency, and the channel filters have corresponding midband frequencies. Even order modulation of these carrier currents, which occurs to a certain extent in the line repeaters of the system, results in the production of interfering frequencies which are even multiples of 85 cycles. These products fall midway between the pass bands of the receiving channel filters and the loss which they encounter in these filters greatly reduces their effect. In a telegraph system having this channel frequency arrangement, but designed to operate on a frequency-shift basis, with the carrier frequencies shifted over a large portion of the channel frequency bands, even order modulation products originating in the line repeaters would, to a much greater extent, lie in frequency ranges freely passed by the receiving filters; and the effect of such interference would be correspondingly greater than in the on-off system.

Noise Tests

One way to judge the relative noise sensitivity of carrier telegraph arrangements is to subject each to measured amounts of noise on the line and then to compare the resulting signal distortions. Resistance or thermal noise was used in these tests because it is the most general kind of noise. It consists of a superposition of rapidly recurring random impulses, some of which may overlap. No tests were made using impulse noise such as caused by lightning, ignition, or sharp static, because it was thought that resistance noise tests would suffice. Impulse noise, when considered in a strict mathematical sense, consists of isolated pulses of very short duration and the component frequencies are so phased with respect to each other that their amplitudes add arithmetically at the instants of occurrence of the pulses. Atmospheric disturbances range all the way from isolated pulses to grinding static caused by dust storms which has characteristics approaching those of resistance noise. It is difficult to choose a representative type of impulse noise for testing. Another reason for not testing with impulse noise was that theoretical considerations[13] indicate there is not much difference in the advantage of frequency-shift over on-off methods whether the disturbance is of the impulse or resistance type.

In order to compare the sensitivities of the different arrangements to re-

[13] M. G. Crosby: "Frequency Modulation Noise Characteristics", Proc. I. R. E., Vol. XXV, No. 4, April 1937, pp. 472–514.

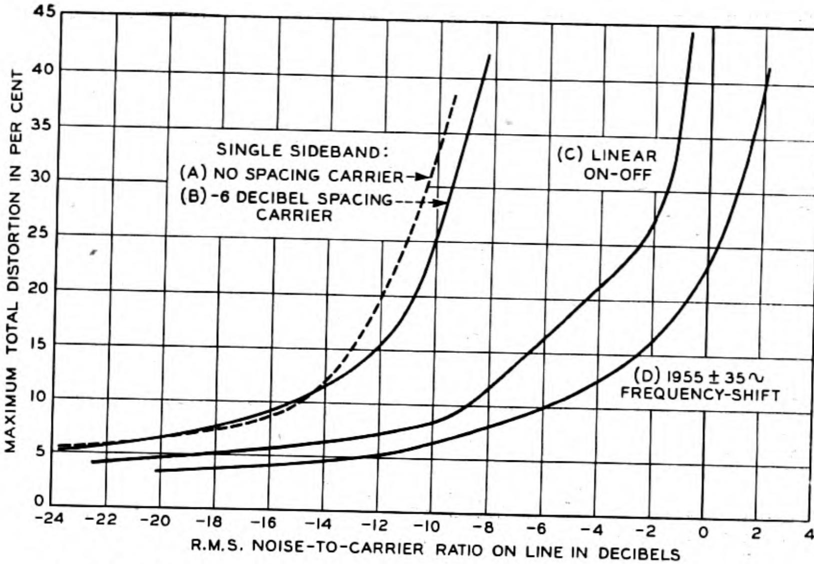


Fig. 18—Distortion vs. noise characteristics of single-sideband, linear on-off and frequency-shift arrangements at 60 w.p.m. (23 d.p.s.).

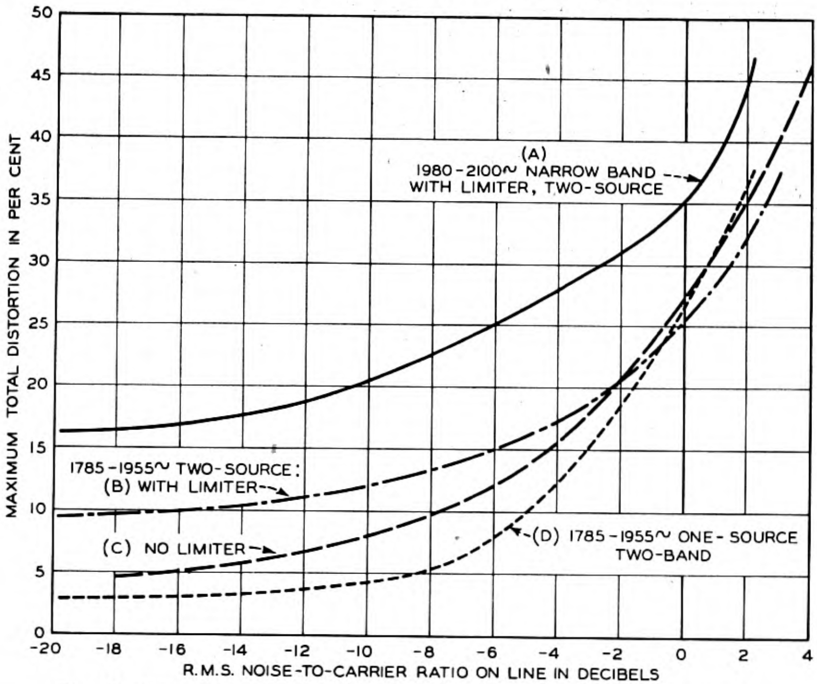


Fig. 19—Distortion vs. noise characteristics of two-band arrangements at 60 w.p.m. (23 d.p.s.).

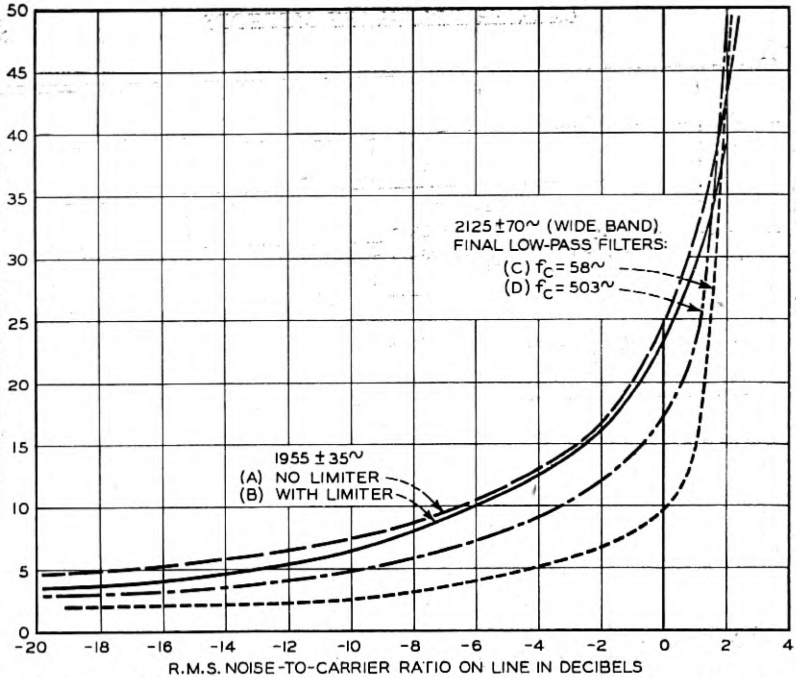


Fig. 20—Distortion vs. noise characteristics of frequency-shift arrangements at 60 w.p.m. (23 d.p.s.).

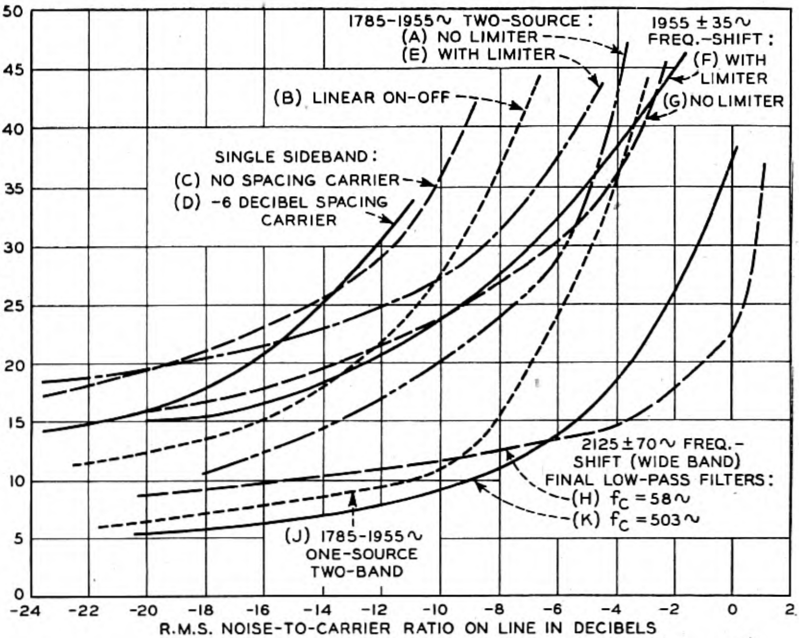


Fig. 21—Distortion vs. noise characteristics of various arrangements at 120 w.p.m. (46 d.p.s.).

sistance noise, the noise was introduced into the line through a symmetrical three-way pad, as previously mentioned. The marking carrier level at one input of the noise pad was kept at a constant value. The level of the noise current entering the other input of the noise pad was adjustable, and its r.m.s. value was measured with a thermocouple, permitting the computation of the r.m.s. noise-to-carrier ratio at the output of the three-way pad, since the loss through the pad was the same for both signal and noise. This ratio was used for abscissae in Figs. 18 to 21, inclusive, and the ordinates represent distortion for various arrangements. These curves are useful in comparing the relative noise sensitivities of the different arrangements, because the same noise source was used in all tests. The absolute noise sensitivities of the various arrangements were not known accurately because the band width of the noise source was not known exactly. The band width must have been about 3 kc. because the measured signal-to-noise power ratio of the on-off arrangement was 15 db better at the receiving filter output than on the line. Fifteen db corresponds to a power ratio of 31.6, which also should be the ratio of the band width of the noise source to the 95-cycle band width of the receiving filter. In comparing the various distortion vs. noise curves, one should note that certain arrangements have different amounts of distortion when noise is absent, which affects the comparison when noise is present.

Distortion at 60 Words per Minute

According to Figs. 18, 19, and 20 the arrangements having the same loss characteristic rank in the following order as regards their insensitivity to resistance noise at 60 w.p.m. and 20 per cent distortion: frequency-shift with limiter, closely followed by frequency-shift without limiter, one-source two-band with limiter, two-source without limiter on a par with two-source with limiter, linear on-off, single-sideband with −6 db spacing carrier, and single-sideband with no spacing carrier. All the tests recorded in Figs. 18, 19, and 20 were made on arrangements having linear receiving detectors. No noise data were taken by the writers on the level compensated on-off arrangement. Measurements made by other Bell System engineers on a similar level compensated arrangement follow the general shape of curve C of Fig. 18 for the linear on-off arrangement, except for an average displacement of about one db to the left. It is not known whether the difference in performance of the level compensated on-off arrangement was due to the difference in detector characteristics or to a difference in measuring technique.

The frequency-shift arrangements tested were all less sensitive to noise than the on-off arrangement. When a limiter was not used, this difference was due mainly to differential recombination of the rectified output currents of the two branches of the discriminator[14]. When a limiter was used and the noise was small compared to the carrier current, the limiter theoretically should have reduced the noise-to-carrier ratio at least 3 db. Therefore, one might expect curves A and B of Fig. 20 to be separated horizontally at least 3 db in the region of low noise and to come together as the noise approaches zero. However, since curves A and B are fairly flat in the region of low noise, small errors in distortion measurement may have caused appreciable errors in the horizontal separation between the curves. When noise is not small compared to signal current, it is expected from theory that a limiter would give only a small reduction in distortion caused by noise on a frequency-shift arrangement. This appears to be verified experimentally by curves A and B of Fig. 20.

The main purpose of testing the wide-band frequency-shift arrangements was to compare their sensitivity to noise with that of the normal-band frequency-shift arrangement with limiter. According to Fig. 20, curves B and D, at 60 w.p.m. the tolerance to noise interference in the wide-band arrangement with low-pass filters having a cut-off frequency of about 503 cycles was 2.7 db greater than that in the normal-band arrangement at 10 per cent distortion and 1.3 db greater at 20 per cent distortion. This improvement was unexpected because the wider band admitted more noise. The improvement must have been due to the smaller distortion in the wide-band arrangement when no noise was present. The theoretical difference in noise tolerance between these two frequency-shift arrangements, which had a two-to-one ratio of frequency band width and of frequency swing, would have been 6 db at low noise levels if low-pass filters had been used at the detector outputs of the wide-band channel to cut off components in the detected current which are higher in frequency than those transmitted by the normal-band arrangement[13]. However, the low-pass filters used in the tests had a cut-off well above this value in order to permit signaling at very high speeds. Consequently, there was some unnecessary noise passed by these filters which accounts for an increase in noise tolerance of less than 6 db at 60 w.p.m.

In order to verify the fact that low cut-off low-pass filters improve the noise tolerance[13] of the wide-band frequency-shift arrangement, the 503-cycle cut-off filters were replaced by filters having a cut-off at about 58 cycles. According to Fig. 20, curves B and C, at 60 w.p.m. the tolerance to noise interference in the wide-band arrangements was then 6.1 db greater than in the normal-band arrangement at 10 per cent distortion, and 2 db greater at 20 per cent distortion.

[14] J. R. Carson and T. C. Fry: "Variable Frequency Electric Circuit Theory with Application to the Theory of Frequency Modulation", *Bell Sys. Tech. Jour.*, Vol. XVI, No. 4, October 1937, pp. 513–540.

The superiority of the two-source arrangements over the on-off arrangement may be accounted for by differential recombination of the detected waves, and by the greater amount of sideband power transmitted with two separate carriers.

As shown in Fig. 19, the tolerance to noise of the narrow-band two-source arrangement at 60 w.p.m. was less than that of the normal-band two-source arrangement with limiter because of the greater distortion obtained when operating at this speed without noise. Also, the two-source arrangement illustrated by curves B and C in Fig. 19 appears to have been made worse at low noise values by use of a limiter. Actually the limiter probably did reduce the noise power, but the improvement in distortion obtained thereby was greatly outweighed by the distortion increase which occurred without noise when the limiter was added, as explained above under the heading "Distortion vs. Speed Tests", and illustrated in curves C and D of Fig. 6.

The single-sideband arrangements were both more sensitive to noise than the on-off arrangement. In the case of the single-sideband arrangement with no spacing carrier the presence of quadrature[2] component increased the marking bias; and in order to have zero bias on reversals it was necessary to increase the d-c bias current in the receiving relay, which reduced the effective marking current and therefore made the arrangement more susceptible to interference occurring during the marking intervals. In the case of the single-sideband arrangement with −6 db spacing carrier the amount of sideband power transmitted was less than in the on-off arrangement and consequently noise was more troublesome.

Distortion at 120 Words per Minute

Further tests were conducted at 120 w.p.m. on the arrangements having linear receiving detectors, and results are shown in Fig. 21. The increase in speed caused an increase in distortion on all arrangements at a given noise level, but the increase in speed has less effect on the single-sideband arrangements than on the others having the same loss characteristics, at medium and high noise levels. On account of the narrower sidebands transmitted on the two-source, normal-band frequency-shift, and on-off arrangements, an increase in speed was accompanied by a greater decrease in amplitude of received signal and therefore by a greater increase in noise-to-signal ratio than in the case of the single-sideband arrangements, each of which used a wider sideband. Consequently, at a given noise-to-signal ratio, the speed increase was accompanied by a greater distortion increase on the former group of arrangements than on the latter.

As was previously found at 60 w.p.m. the frequency-shift arrangement with limiter was less sensitive to noise than the arrangement without, when the noise was small. But the reduction in distortion was negligible, because the distortion was small in either case.

In Fig. 21, a comparison of curve F, applying to the normal-band frequency-shift arrangement with limiter, and curve K, applying to the wide-band frequency-shift arrangement with 503-cycle low-pass filter, shows that at 20 per cent distortion the increase in band width and frequency swing resulted in a 9 db improvement in noise-to-carrier ratio. This improvement may be explained on the same basis as the 1.3 db improvement obtained at 60 w.p.m. and was larger due to the higher distortion of the normal-band arrangement when operating at 120 w.p.m. When the cut-off frequency of the low-pass filter of the wide-band frequency-shift arrangement was changed from 503 cycles to 58 cycles, this improvement in noise-to-signal ratio was increased to 11.6 db. It is apparent that tolerance to noise was appreciably increased by this change of filters but, of course, this was accompanied by a reduction in maximum operating speed.

Noise-to-Signal Ratio at Receiving Relay

Another method that was used to measure noise-to-signal ratio on the on-off and single-sideband arrangements was as follows: With the carrier turned off, the receiving gain was increased until the receiving relay just operated on occasional noise peaks. Then with the carrier on the line, and with the noise absent, the receiving gain was readjusted until the receiving relay again just operated. The difference in receiving gain in these two tests was called the noise-to-signal ratio at the receiving relay. When this method of measuring the noise-to-signal ratio was used it was found that the on-off and single-sideband arrangements had distortion vs. noise-to-carrier ratio characteristics similar to those given in Figs. 18 and 21. However, in order to express these characteristics correctly when substituting the words "at receiving relay" for "on line" in the scale of abscissae, it was found experimentally that it was necessary to shift the curves for the on-off arrangement 7 db to the left and to shift those for the single-sideband arrangement 4 db to the left, relative to their present positions in these figures. Such a change of scale was necessary because the noise-to-signal ratio was smaller at the output of the receiving filter than it was on the line.

Summary Of Results

Explanation of Tables VII and VIII

Tables VII and VIII have been prepared from the test data in order to compare the various arrangements for a given amount of distortion on each. Table VII describes certain of their properties for 10 per cent maximum total distortion. In column A (see bottom line of the table) are listed the different arrangements. In columns B and C are given the speeds in dots per second and words per minute respectively at which these arrangements operated with 10 per cent distortion. In taking these data all other variables such as noise and level or frequency changes were absent. (Similarly in the other columns only the variable mentioned was allowed to change.) Column D gives the order of preference for arrangements having the same channel filter loss characteristic (but not necessarily the same total frequency band) on the basis that the highest speed is the most desirable. These data apply when no flanking channels were present. The one-source two-band arrangement and the single-sideband arrangement with -6 db spacing carrier were the fastest of those having the same channel filter loss characteristic, as might be expected from the widths of the transmitted sidebands. However, for a given band width, the single-sideband arrangement with -6 db spacing carrier was the fastest. If frequency band width is of paramount importance, consideration should be limited to arrangements with the same band width, which would change the order of preference.

Columns E, F, and G apply, respectively, in place of columns B, C, and D, when flanking channels were in operation. According to column G, the single-sideband arrangement lost its place in preferential rating due to interference from the adjacent channels; and the on-off and two-source arrangements were the best, as might be expected, because their carriers were located at midband.

In column H are listed the ranges of levels over which the received signal power could vary without causing more than 10 per cent distortion. The order of preference listed in column I indicates that arrangements with limiters were the most stable.

In column L are listed the ranges of carrier frequency variations which could be tolerated without causing the distortion to exceed 10 per cent. In column M it is seen that the two-source arrangements performed better than the others when the mean frequency changed.

In column P the arrangements are rated on the basis of the resistance noise which could be tolerated on the line compared to the linear on-off arrangement. For example, the frequency-shift arrangement in item 1.11 could tolerate 2.7 db more noise than the on-off arrangement for 10 per cent distortion at 60 w.p.m. According to column Q the one-source two-band arrangement performed the best in this respect.

Table VIII is arranged similarly to Table VII except that the maximum total distortion is 20 per cent throughout, and additional data are given in columns J, K, N, O, R, and S to cover speed at 120 w.p.m.

Discussion

It is believed that the circuit arrangements tested were reasonably representative of those commonly used in carrier telegraph practice, so that

TABLE VII

COMPARISON OF DIFFERENT ARRANGEMENTS WHEN MAXIMUM DISTORTION IS 10% ON EACH. UNLESS OTHERWISE STATED: CHANNEL SPACING 170 CYCLES & LOSS PER FIG. 1, CURVE A

Note: d.p.s. = dots per sec.; w.p.m. = words per min.; o.p. = order of preference based on performance. (In practise, consideration should also be given to economic advantages, and in this respect the on-off arrangements rank high.)

Arrangement	Basis of Comparison											
	Speed						Flat Level Change		Mean Carrier Freq. Change		Signal-to-Noise Advantage on Line Compared with On-Off at 60 w.p.m.	
	No Flanking Channel			With Flanking Channels			Range at 60 w.p.m.		Range at 60 w.p.m.			
	d.p.s.	*w.p.m.*	*o.p.*	*d.p.s.*	*w.p.m.*	*o.p.*	*db*	*o.p.*	*c.p.s.*	*o.p.*	*db*	*o.p.*
1. *Arrangements with Same Band Width*												
1.1. Frequency-Shift, 1955 ± 35 Cycles, with Relay Modulator:												
1.11. With Current Limiter	42	110	8	36	94	5	>82	3	11.6	5	+2.7	2
1.12. Without Current Limiter	41	107	9								+2.2	3
1.2. Frequency-Shift, 1955 ± 35 Cycles, with Limiter, Diode Modulator:												
1.21. Abrupt Frequency Change	44	115	7									
1.22. Sinusoidal Frequency Change	39	103	11									
1.3. Linear On-Off, 1955 Cycles	47	123	6	48	126	1	2.1	6	60	3	0	5
1.4. On-Off with 40B1 Detector, 1955 Cycles (No Level Compensator)	48	127	4									
1.5. Level Compensated On-Off, 1955 Cycles	50	132	3				>41	4	59	4		
1.6. Single-Sideband, 1998 Cycles:												
1.61. No Spacing Carrier	40	105	10	41	108	3					−6.3	6
1.62. Spacing Carrier −6 db	56	147	2	41	107	4	1.0	7	7	6	−6.7	7

2. *Arrangements with Different Band Widths*												
2.1. Two-Source, 1785 and 1955 Cycles:												
2.11. Without Current Limiter	47	125	5	47	124	2	29	5	138	1	+1.0	4
2.12. With Current Limiter	33	87	12				93	1	126	2	−7.3	8
2.2. One-Source Two-Band, 1785 and 1955 Cycles, with Current Limiter	59	155	1				>90	2	59	4	+3.7	1
2.3. Two-Source, 1980 and 2100 Cycles, 120-Cycle Spacing (Narrow-Band), 2 Bands per Fig. 1, Curve C, with Current Limiter	(Distortion always greater than 10% at speeds greater than 60 W.P.M.)											
2.4. Frequency-Shift, 2125 ± 70 Cycles, 340-Cycle Spacing (Wide-Band), with Limiter, Loss per Fig. 1, Curve B:												
2.41. With Final L.P. Filters, f_c = 503 cycles	66	173	*				>90	*	36	*	+5.4	*
2.42. With Final L.P. Filters, f_c = 58 cycles	47	123	*				>81	*	20	*	+8.8	*
Column No.: A	B	C	D	E	F	G	H	I	L	M	P	Q

* No order of preference given because the loss characteristics of these channels differ from those of the other channels.

TABLE VIII

COMPARISON OF DIFFERENT ARRANGEMENTS WHEN MAXIMUM TOTAL DISTORTION IS 20% ON EACH.
UNLESS OTHERWISE STATED: CHANNEL SPACING 170 CYCLES & LOSS PER FIG. 1, CURVE A

Note: d.p.s. = dots per sec.; w.p.m. = words per min.; o.p. = order of preference based on performance. (In practise, consideration should also be given to economic advantages, and in this respect the on-off arrangements rank high.)

Arrangement (same as Table VII)	Basis of Comparison																	
	Speed						Flat Level Change				Mean Carrier Frequency Change				Signal-to-Noise Advantage on Line Compared with On-Off			
	No Flanking Channels			With Flanking Channels			At 60 w.p.m.		At 120 w.p.m.		At 60 w.p.m.		At 120 w.p.m.		At 60 w.p.m.		At 120 w.p.m.	
							Range		Range		Range		Range					
	d.p.s.	*w.p.m.*	*o.p.*	*d.p.s.*	*w.p.m.*	*o.p.*	*db*	*o.p.*	*db*	*o.p.*	*c.p.s.*	*o.p.*	*c.p.s.*	*o.p.*	*db*	*o.p.*	*db*	*o.p.*
1. *Arrangements with Same Band Width*																		
1.1. Freq.-Shift, 1955 ± 35 Cycles, Relay Modulator:																		
1.11. With Current Limiter.........	51	134	9	46	121	5	>84	2	>64	3	25	5	10.6	6	+3.5	1	+0.3	3
1.12. Without Current Limiter......	55	144	7												+3.3	2	−0.5	5
1.2. Freq.-Shift, 1955 ± 35 Cycles, Diode Modulator:																		
1.21. Abrupt Frequency Change....	57	151	4															
1.22. Sinusoidal Frequency Change.	50	131	10															
1.3. Linear On-Off, 1955 Cycles.........	52	137	8	53	140	3	4.8	5	1.9	6	75	4	59	3	0	5	0	4
1.4. On-Off 40B1 Detector, 1955 Cycles (No Level Compensator).........	56	148	5															
1.5. Level Compensated On-Off, 1955 Cycles.........................	56	147	5				>42	3	>41	4	85	3	44	5				
1.6. Single-Sideband, 1998 Cycles:																		
1.61. No Spacing Carrier...........	60	157	3	50	131	4									−7.7	7	−6.4	7
1.62. Spacing Carrier − 6 db........	72*	190*	1	>68	>180	1	2.5	6	1.3	7	12.3	6	6.7	7	−6.4	6	−3.7	6

2. Arrangements with Different Band Widths																		
2.1. Two-Source, 1785 & 1955 Cycles:																		
2.11. Without Current Limiter......	55	146	6	54	141	2	37	4	19.7	5	145	1	117	1	+2.0	4	+2.8	2
2.12. With Current Limiter.........	50	132	10				>100	1	83	2	145	1	102	2	+2.0	4	−6.2	8
2.2. One-Source, Two-Band, 1785 & 1955 Cycles, with Current Limiter.....	66	175	2				>100	1	>84	1	110	2	57	4	+2.6	3	+6.0	1
2.3. Two-Source, 1980 & 2100 Cycles, 120-Cycle Spacing (Narrow-Band), 2 Bands per Fig. 1, Curve C, with Current Limiter..................	29	76	**				>94	**	fails		44		fails		−6.2		fails	
2.4. Frequency Shift, 2125 ± 70 Cycles, 340-Cycle Spacing (Wide-Band), with Current Limiter, Loss per Fig. 1, Curve B:																		
2.41. Final L.P. Filt., $f_c = 503\sim$....	>68	>180	**				>90	**	>89	**	80	**	37	**	+4.9	**	+9.3	**
2.42. Final L.P. Filt., $f_c = 58\sim$....	52	138	**				>84	**	>82	**	42	**	14	**	+5.6	**	+11.9	**
Column No.: A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S

* Extrapolated.

** No order of preference given because the loss characteristics of these channels differ from those of the other channels.

fairly general conclusions are warranted. With the use of other types of filters, relays, etc., somewhat different results would doubtless be obtained.

Relative advantages and disadvantages of frequency-shift, on-off, and single-sideband methods of carrier telegraphy were deduced from measurements utilizing the same channel filters and covering a range of signalling speeds, noise, interference, and other variables.

A frequency-shift arrangement with amplitude limiter is substantially unaffected by non-selective level changes and relatively insensitive to noise currents. However, interfering currents from similar flanking channels are greater than when using the other methods. It has bias instability when the mean carrier frequency drifts. None of the arrangements tested had any compensation for drifting of the mean carrier frequency. However, it is known that automatic compensation for this may be obtained by various special methods not here described.

A channel using the on-off method is less sensitive to frequency drift, slightly faster for a given band width, and cheaper in terminal equipment than that using the frequency-shift method. The interfering currents from similar flanking channels using the on-off method are quite small. Weaknesses of the on-off method are greater sensitivity to noise and level changes. However, on good wire lines, when a level compensator[11] is used, these weaknesses are unimportant, and the on-off method is satisfactory. Of course, these weaknesses become important on radio circuits when noise is strong and when fades are too rapid or too severe to be overcome by the level compensator.

The greatest speed for a given band width is attainable by the single-sideband method. Unfortunately this method is poor from the standpoint of interchannel interference, is the most susceptible to noise, and, unless special compensating devices are used, is the most sensitive to carrier frequency drift and level changes.

Several arrangements were also investigated which utilized approximately double the band width of those mentioned in the preceding paragraph. Among these, the two-source method is the best of all the methods herein mentioned from the standpoint of insensitivity to changes in carrier frequency, whether or not a limiter is used. If a limiter is used this arrangement ranks well in its ability to withstand non-selective level changes. Two-source arrangements are sensitive to differences between the marking and spacing levels, but some advantage is obtained by the use of a limiter. The distortion vs. speed characteristic of the two-source method without limiter is about the same as that of the linear on-off method utilizing half the band width of the two-source method. As previously explained, the use of a limiter with this method causes distortion and materially reduces the maximum operating speed. From a noise standpoint the two-source arrangement without limiter has a slight advantage over the linear on-off method. When a limiter is used, the two-source arrangement is inferior to the linear on-off method due to the distortion inherent in this arrangement.

In the one-source two-band arrangement the limiter does not cause distortion, and a wider range of frequency components is transmitted than by the two-source method. For these reasons, the one-source two-band arrangement has a maximum working speed considerably higher than the two-source arrangement. The one-source two-band arrangement ranks close to the two-source method with limiter in its ability to withstand non-selective level changes, but its susceptibility to carrier frequency changes is greater.

A frequency-shift arrangement utilizing approximately the same band width as the two-source arrangement is found to have appreciably higher speed and less sensitivity to noise, but unless compensation is provided this frequency-shift arrangement is considerably more susceptible to carrier frequency changes. A further reduction in noise sensitivity may be obtained by the use of a low-pass filter at the detector output with a cut-off frequency low enough to limit the speed to that attainable with the normal-band frequency-shift arrangement.

Economic considerations should also be given due weight in selecting an optimum arrangement for a specific application. For example, the on-off method has been widely used on certain wire lines of the Bell System, because appreciable level changes are gradual and the lines are relatively quiet. For this application, the level compensated on-off method therefore gives satisfactory service with a minimum amount of apparatus, and the terminal arrangements are fairly simple and easy to maintain. On radio links subject to noise and fading, the more expensive frequency-shift and two-source methods have frequently been selected because of their greater reliability under such adverse conditions. The on-off and two-source arrangements have the advantage that common carrier generators (or oscillators) may be used for a number of channels.

Abstracts of Technical Articles by Bell System Authors

Weathering of Soft Vulcanized Rubber.[1] JAMES CRABTREE and A. R. KEMP. Two separate and distinct processes are responsible for the breakdown of soft vulcanized rubber when exposed to outdoor weathering—light-energized oxidation and attack by atmospheric ozone. The former is independent of stress and controllable to any marked extent only by incorporation of opaquing fillers. The latter affects rubber only when under stress and is checked to a considerable degree by addition of certain hydrocarbon waxes as long as the stress is static. The conditions affecting these processes have been investigated, and suggestions for accelerated aging are made on the basis of the findings.

A Note on a Simple Transmission Formula.[2] HARALD T. FRIIS. A simple transmission formula for a radio circuit is derived. The utility of the formula is emphasized and its limitations are discussed.

Applications of Thin Permalloy Tape in Wide-Band Telephone and Pulse Transformers.[3] A. G. GANZ. The properties and uses of thin permalloy tapes ranging from two mils to as little as 1/8 mil thick in tape cores are described. Typical applications covered are in transformers and non-linear coils for radar and for telephone systems. Data are given on the steady a-c. properties of thin tapes up to one megacycle. Pulse magnetization of the tape is analyzed. The available flux density range with uni-directional pulses and the effects of appropriate air gaps and of reverse magnetization between pulses are illustrated. Equations are given for flux distribution, effective permeability and loss, assuming linear magnetic properties, and convenient graphs for these characteristics are included. Simple expressions are developed for effective permeability and loss, which are approximations for the high d-c. permeability and rapid transition to saturation which characterize the permalloys.

Derivation of the Lorentz Transformations.[4] HERBERT E. IVES. The Lorentz transformations were obtained by Lorentz as a succession of *ad hoc* inventions, to reconcile Maxwell's theory with the results of experiments on moving bodies. By Einstein they were derived after a discussion of the

[1] *Indus. & Engg. Chemistry*, March 1946.
[2] *Proc. I.R.E.*, May 1946.
[3] *Elec. Engg., Trans. Sec.*, April 1946.
[4] *Phil. Mag.*, June 1945.

nature of simultaneity, and the adoption of a *definition* of simultaneity which violates the intuitive and common-sense meaning of that term. It is the purpose of this paper to show that these transformations can be derived by imposing the laws of conservation of energy and of momentum on radiation processes as developed by Maxwell's methods.

The Effect of High Humidity and Fungi on the Insulation Resistance of Plastics.[5] JOHN LEUTRITZ, JR. *and* DAVID B. HERRMANN. The decrease in insulation resistance of methyl methacrylate, glass bonded mica, glass mat laminate phenolic, phenol fabric, phenol fiber, and wood flour filled phenol plastic is determined during prolonged exposure of the plastics to fungi and 97 per cent relative humidity at 25 C. The same plastics with fungi present also are exposed to 87, 76, and 52 per cent relative humidity to study their recovery, and then re-exposed to 97 per cent relative humidity. Samples with cleaned surfaces and with varnished surfaces are dried and then exposed to fungi and high humidity. The insulation resistance of a fungus network on methyl methacrylate is determined at 87, 76, and 52 per cent relative humidity.

Fungus growth occurs on all the test specimens except those with cleaned or varnished surfaces. The decrease in insulation resistance is retarded by the varnish. The degradation is due entirely to moisture. The rate of recovery is dependent on the composition and structure of the materials. None of the plastics is permanently affected by exposure to fungi and high humidity. Cleaning of surfaces and removal of moisture restore the insulation resistance to its original high value in every case. Water adsorption and absorption, not fungi, are the critical factors in the deterioration of the insulation resistance of these plastics.

The Elastic, Piezoelectric, and Dielectric Constants of Potassium Dihydrogen Phosphate and Ammonium Dihydrogen Phosphate.[6] W. P. MASON. Measurements have been made of all the elastic, piezoelectric, and dielectric constants of KDP and ADP crystals through temperature ranges down to the Curie temperatures. The piezoelectric properties agree well with Mueller's phenomenological theory of piezoelectricity provided the fundamental piezoelectric constant is taken as the ratio of the piezoelectric stress to that part of the polarization due to the hydrogen bonds. It is found that the dielectric properties of KDP agree well with the theory presented by Slater based on the interaction of the hydrogen bonds with the PO_4 ions. ADP undergoes a transition at $-125°C$ which results in fracturing the crystal. This transition cannot be connected with the H_2PO_4 hydrogen bond system

[5] *A.S.T.M. Bulletin,* January 1946.
[6] *Phys. Rev.,* March 1 and 15, 1946.

which controls the dielectric and piezoelectric properties, for these lie on smooth curves that do not change slope as the transition temperature is approached. It is suggested that two separate and independent hydrogen bond systems are involved in ADP. The transition temperature and specific heat anomaly appear to be connected with hydrogen bonds between the nitrogens and the oxygens of the PO_4 ions, while the dielectric and piezoelectric properties are controlled by the H_2PO_4 hydrogen bonds.

Nonlinearity in Frequency-Modulation Radio Systems due to Multipath Propagation.[7] S. T. MEYERS. A theoretical study is made to determine the effects of multipath propagation on over-all transmission characteristics in frequency-modulation radio circuits. The analysis covers a simplified case where the transmitted carrier is frequency-modulated by a single modulating frequency and is propagated over two paths having relative delay and amplitude differences. Equations are derived for the receiver output in terms of the transmitter input for fundamental and harmonics of the modulating frequency. Curves are plotted and discussed for various values of relative carrier- and signal-frequency phase shift and relative amplitude difference of the received waves.

The results show that a special kind of amplitude nonlinearity is produced in the input-output characteristics of an over-all frequency-modulation radio system. Under certain conditions, sudden changes in output-signal amplitude accompany the passage of the input-signal amplitude through certain critical values. Transmission irregularities of this type are proposed as a possible explanation of so-called "volume bursts" sometimes encountered in frequency-modulation radio circuits. In general, it appears that amplitude and frequency distortion are most severe where the relative delay between paths is large and the amplitude difference is small.

Propagation of 6-Millimeter Waves.[8] G. E. MUELLER. One step in the exploration of a new band of frequencies for communications purposes is a study of the transmission properties of the medium involved. This paper describes the methods and results of measurements of attenuation due to rainfall and atmospheric gases at a wavelength of 0.62 centimeter.

The one-way attenuation due to moderate rains at 0.62 centimeter is roughly 0.6 decibel-per-mile per millimeter-per-hour. The gas attenuation is probably less than 0.2 decibel per mile.

Vicalloy—A Workable Alloy for Permanent Magnets.[9] E. A. NESBITT. Alloys in the region of 30 to 52 per cent iron, 36 to 62 per cent cobalt, and 4

[7] *Proc. I.R.E.*, May 1946.
[8] *Proc. I.R.E.*, April 1946.
[9] *Metals Technology*, February 1946.

to 16 per cent vanadium were investigated with the result that permanent magnet materials of unusual mechanical as well as magnetic properties were discovered. The alloys differ from most age-hardening alloys in that the gamma phase, stable at high temperatures, is dispersed in the alpha phase, stable at low temperatures, instead of vice versa.

Distribution of Sample Arrangements for Runs Up and Down.[10] P. S. OLMSTEAD. Using the notation of Levene and Wolfowitz, a new recursion formula is used to give the exact distribution of arrangements of n numbers, no two alike, with runs up or down of length p or more. These are tabled for n and p through $n = 14$. An exact solution is given for $p > n/2$. The average and variance determined by Levene and Wolfowitz are presented in a simplified form. The fraction of arrangements of n numbers with runs of length p or more are presented for the exact distributions, for the limiting Poisson Exponential, and for an extrapolation from the exact distributions. Agreement among the tables is discussed.

Radar Systems Considerations.[11] D. A. QUARLES. In the broad field of radio technology, radar (object location) systems have come to occupy a relatively new but highly specialized area. Because radar is a seeing and measuring art, it has put a special premium on short wavelength and has thus tended to accelerate greatly the already rapid trend toward higher frequencies. Moreover, many radar systems are associated with computer and servo mechanisms for automatic control purpose such as gunfire, bomb release and the like. To meet these new needs, a dozen or more highly developed fields of specialization covering such components as antennas, pulse transmitters and display devices have been created. Planning a new radar system calls for an appraisal of these component arts and for selection and balancing of component characteristics to produce an integrated system. The present paper deals with such technical considerations involved in planning an overall radar system as a background for other more detailed technical expositions of the component arts.

The Effect of Rain upon the Propagation of Waves in the 1- and 3-Centimeter Regions.[12] SLOAN D. ROBERTSON and ARCHIE P. KING. This paper presents some experimental results which show the effect of rain upon the transmission of electromagnetic waves in the region between 1 and 4 centimeters.

At a wavelength of 1.09 centimeters, the waves are appreciably attenuated,

[10] *Annals of Mathematical Statistics*, March 1946.
[11] *Elec. Engg., Trans. Sec.*, April 1946.
[12] *Proc. I.R.E.*, April 1946.

even by a moderate rain. Attenuations in excess of 25 decibels per mile have been observed in rain of cloudburst proportions.

The attenuation of waves somewhat longer than 3 centimeters is slight for moderate and light rainfall. During a cloudburst, however, the attenuation may approach a value of 5 decibels per mile.

The Advancing Statistical Front.[13] W. A. SHEWHART. From the viewpoint of general education, statistics is not simply a tool as is so often stated, but a scientific way of looking at the universe; statistical method is not something apart from scientific method but *is* scientific method in which the three steps, hypothesis, experiment, and test of hypothesis, are adjusted to allow for the fact that scientific inference is only probable. Applications of statistics in this sense are rapidly extending to all fields of pure, background, and applied reserch.

A New Crystal Channel Filter for Broad Band Carrier Systems.[14] E. S. WILLIS. A new crystal channel filter for use in broad-band carrier telephone systems is described. It requires less than two-thirds as much mounting space as the earlier design and savings in materials and manufacturing effort are realized. The savings were made possible by assembling the four crystal units in one lattice-type filter section rather than two, resulting in a reduction in the number of component coils and capacitors.

[13] *Jour. Amer. Statis. Assoc.*, March 1946.
[14] *Elec. Engg., Trans. Sec.*, March 1946.

Contributors to This Issue

J. G. Ferguson, B.S. in Electrical Engineering, Queen's University, Canada. Northern Electric Company, 1923–1926; Bell Telephone Laboratories 1926–. During the war years, Mr. Ferguson developed various telephone, radio and radar equipments. Prior to the war he developed switching equipment for central offices and PBX's. He is currently interested in the design of No. 5 crossbar equipment.

H. J. Fisher, E.E., Cornell University, 1920. U. S. Signal Corps, 1917–1919. Western Electric Company, Engineering Department, 1920–1925. Bell Telephone Laboratories, 1925–. Mr. Fisher has been engaged in the development of toll transmission systems. Since 1940 and in his present capacity as Test Engineer his work has had to do principally with system testing equipment, and during the war with the development of radar testing equipment for the Armed Forces.

G. T. Ford, B.S. in Physics, Michigan State College, 1929; M.A. in Physics, Columbia University, 1936. Bell Telephone Laboratories, 1929–. Mr. Ford was concerned with some of the early work on thermistors, and, since 1934, has been engaged in development work on small high vacuum electron tubes.

Calvin S. Fuller, B.S., Chicago, 1926; Ph.D., 1929. Bell Telephone Laboratories, 1930–1942. Office of the Rubber Director, 1942–44. Bell Telephone Laboratories, 1944–. Dr. Fuller has been engaged in the development of organic insulations and the application of plastics to electrical apparatus.

E. I. Green, A.B., Westminster College, 1915; University of Chicago, 1915–1916; Professor of Greek, Westminster College, 1916–1917; U. S. Army, 1917–1919 (Captain, Infantry); B.S. in Electrical Engineering, Harvard University, 1921; Dept. of Development and Research, American Telephone and Telegraph Company, 1921–1934; Bell Telephone Laboratories, 1934–. Mr. Green's responsibilities have had to do principally with multiplex transmission systems, and during the war with radar and other projects for the Armed Forces. In his present capacity as Assistant Director of Transmission Development he is in charge of development work on carrier telephone systems and on test equipment for toll transmission systems.

Theodore A. Jones, University of Oregon, 1919–20; B.S., Oregon State College, 1924; M.A., Columbia University, 1928. Engineering Department, Western Electric Company, 1924–25; Bell Telephone Laboratories, 1925–. Since joining the technical staff Mr. Jones has been engaged in carrier telegraph development work.

J. P. Kinzer, M.E., Stevens Institute of Technology, 1925. B.C.E., Brooklyn Polytechnic Institute, 1933. Bell Telephone Laboratories, 1925–. Mr. Kinzer's work has been in the development of carrier telephone repeaters; during the war his attention was directed to investigation of the mathematical problems involved in cavity resonators.

K. W. Pfleger, A.B., Cornell University, 1921; E.E., 1923. American Telephone and Telegraph Company, Department of Development and Research, 1923–1934; Bell Telephone Laboratories, 1934–. Mr. Pfleger has been engaged in transmission development work, chiefly on problems pertaining to delay equalization, delay measuring, temperature effects in loaded-cable circuits, and telegraph theory.

C. W. Schramm, B.S. in Electrical Engineering, Armour Institute (now Illinois Institute) of Technology, 1927. Illinois Bell Telephone Company, 1927–29. Bell Telephone Laboratories, 1929–. Mr. Schramm has been concerned with the development of carrier telephone systems for both message and program use. During the war his attention was directed to the design of radar test equipment.

I. G. Wilson, B.S. in Electrical Engineering and M.E., University of Kentucky, 1921. Western Electric Company, Engineering Department, 1921–25. Bell Telephone Laboratories, 1925–. Mr. Wilson has been engaged in the development of amplifiers for broad-band systems. During the war he was project engineer in charge of the design of resonant cavities for radar testing.