THE BELL SYSTEM TECHNICAL JOURNAL

VOLUME XXXI MARCH 1952 NUMBER 2

Introduction to Formal Realizability Theory—I

By BROCKWAY McMILLAN

(Manuscript received October 15, 1951)

This paper offers a general approach to the realizability theory of networks with many accessible terminals. The methods developed are applied to give a complete characterization of all finite passive networks.

I. SUMMARY

1.0 A principal result of this paper is to characterize those matrices $Z(p)$, functions of the frequency parameter p, which can be realized as open-circuit impedance matrices of finite passive networks. This characterization is provided by the following theorem:

1.1 *Theorem*:* Let $Z(p)$ be an $n \times n$ matrix whose elements are $Z_{rs}(p)$, $1 \leq r, s \leq n$, where

(i) Each $Z_{rs}(p)$ is a rational function

(ii) $\overline{Z_{rs}(p)} = Z_{rs}(\bar{p})$ (the bar denotes complex conjugate)

(iii) $Z_{rs}(p) = Z_{sr}(p)$

(iv) For each set of real constants $k_1, \cdots, k_n$, the function

$$\varphi_Z(p) = \sum_{r,s=1}^{n} Z_{rs}(p) k_r k_s$$

has a non-negative real part whenever $\mathrm{Re}(p) > 0$.

Then there exists a finite passive network, a $2n$-pole, which has the impedance matrix $Z(p)$.

* Presented to the American Mathematical Society, April 17, 1948. Abstract 260, *Bulletin of the A.M.S. No. 54*, July, 1948.

Conversely, if a finite passive $2n$-pole has an impedance matrix $Z(p)$, that matrix has the properties (i), (ii), (iii), (iv).

A formally identical dual theorem holds for open-circuit admittance matrices $Y(p)$.

1.2 A general realizability theorem, applicable to and characterizing completely all finite passive networks, whether having impedance matrices or not, is also proved.

1.3 An effort is made to lay a foundation adequate for the realizability theory of both active and passive multi-terminal devices. To this end, a large part of the paper is devoted to the scrutiny of fundamental properties of networks.

II. INTRODUCTION AND FOREWORD

2.0 Network theory provides direct means for associating with an electrical network a mathematical description which characterizes the behavior of that network. Typically, this results in shifting engineering attention from a detailed, possibly quite intricate, electrical structure to a mathematical entity which succinctly describes the relevant behavior of that structure. An essential feature of this shift in focus is emphasized by the word "relevant": only those terminals of the network which are directly relevant to the problem at hand are considered in the mathematical description. Design work can then be done in terms of constructs relating explicitly to these accessible terminals, the effect of the internal structure being felt only by implication.

The physical origins of these mathematical constructs, and the implications of the internal structure upon them, cannot however be entirely forgotten, for they have mathematical consequences which are not always immediately evident. Until he knows these limitations—imposed upon him by the physical nature or the necessary structural form of the networks he is designing—a design engineer cannot make free use of the mathematical tools that network theory has provided.

We give the name "realizability theory" to that part of network theory which aims at the isolation and understanding of those broad limitations upon network performance, i.e., upon the mathematical constructs which describe that performance—which are imposed by limitations on the network structure. One may also include in the province of realizability theory some of the converse questions: the study of those structural features common to all networks whose performance is limited in some specified way.

Realizability theory would have little content were it not that "per-

formance" here must be construed to mean *performance as viewed from the accessible terminals only*. Were all branch currents and node potentials in a network available to observation, a mathematical statement of performance would be equivalent to stating the full system of differential equations governing these quantities, i.e., equivalent to giving the detailed network diagram.

2.1 With a few important exceptions, the converse kind of problem in realizability theory does not lead to a strict implication from functional limitations to structural features, because the field of equivalent structures for a specified performance is very broad. Typically, it is only by imposing some general *a priori* limitations on structure that further conclusions can be firmly drawn from a functional limitation. In studying this kind of problem one is rapidly led from those basic issues which are clearly part of realizability theory toward general, difficult, and usually unsolved problems of network synthesis. One cannot, and should not, draw a sharp boundary here, but Nature so far has provided a fairly definite one for us, in that most of these problems have proved too difficult of solution.

2.2 The direct realizability problems, the passage from structural properties to functional properties, have been somewhat more tractable. Here, again, there is no clear dividing line between general realizability theory and the sort of design theory in which, for example, one specifies a particular filter structure depending on a limited number of parameters and examines the performance of the structure as a function of these parameters. There is an extensive literature at or near this latter level of generality, most of it relating to filters or filter-like structures (e.g., interstage couplers in amplifiers).

At a more basic level, the limitations on a network's structure which are commonly met in practice are of the following kinds:

a. Limitations on the kind of elements appearing, e.g., to passive networks, networks without coupled coils, networks whose elements have specified parasitics, etc;

b. Limitations on the general form of the network diagram, e.g., to ladder or lattice structures, without limitation to a specified number of elements or parameters.

Here the problems are varied and difficult. We survey briefly the present status of some of them.

2.3 Networks with two accessible terminals, two-poles, are basic in network technology. Fortunately, also, two-poles are unique among networks in that there is always a simple way to describe their perform-

ance. Except for the trivial limiting case of an open circuit, every two-pole has a well-defined impedance, $Z(p)$, a function of the complex frequency parameter p, which describes its performance in a way which is by now well understood. Dually, except for the limiting case of a short circuit, every two-pole has a well-defined admittance function $Y(p)$. Even the limiting cases are tractable: every open circuit has the admittance function $Y(p) \equiv 0$ and every short circuit the impedance function $Z(p) \equiv 0$.

In other words, by exercising his option to speak in terms either of impedance or of admittance, one can always describe the performance of a two-pole by using a single function of frequency.

The descriptive simplicity and practical importance of two-poles led early to a fairly complete realizability theory for them. In 1924 R. M. Foster[7] gave a function-theoretic characterization of the impedance functions of finite passive two-poles containing only reactances. The corresponding problem for two-poles which are not at all limited as to structure, beyond being finite and passive, was solved by O. Brune[2] in 1931. The effects of various structural limitations have since been studied by several writers (cf. Darlington,[6] Bott and Duffin[13]).

2.4 Technology, and the promptings of conscience, have meanwhile urged the study of devices with more than two accessible terminals. Here, however, Nature has been less kind, in that no uniquely simple method is available for describing the performance of such devices as viewed from their terminals.

Indeed, basic network theory has been remiss here, in not even making available a mode of description which is generally applicable—whether simple or not.

W. Cauer[5] showed that, when one admits ideal transformers among his network components, it is sufficient to study networks which are natural and direct generalizations of two-poles, namely, $2n$-poles,* for arbitrary values of n. The corresponding natural generalization of the impedance function $Z(p)$ of a two-pole is the impedance *matrix* of a $2n$-pole: just as one multiplies a scalar current by a scalar impedance to get a scalar voltage, one multiplies a vector current by an impedance matrix to get a vector voltage.

2.41 Not all descriptive difficulties are resolved, however, by considering $2n$-poles and their impedance or admittance matrices. For the moment, a simple example will suffice to show this: the 2×2-pole which consists simply of one pair of short-circuited terminals and one pair of

* Defined in Cauer,[5] and also later here.

open-circuited terminals is a finite passive $2n$-pole ($n = 2$) which has neither an impedance matrix nor an admittance matrix.

2.42 When one eliminates this kind of descriptive difficulty by fixing his attention only upon $2n$-poles for which an impedance matrix (or, dually, an admittance matrix) is available, the general realizability problem for finite passive devices is solved. A partial solution, for the case $n = 2$, was published by C. M. Gewertz[8] in 1933. The solution (Theorem 1.1) of the problem for a general value of n has been announced recently by three authors, independently: Y. Oono,[10] in 1946,* the present author, in 1948,† and M. Bayard,[1] in 1949. The problem for reactive $2n$-poles is much simpler and was solved by Cauer,[3] in 1931.

2.5 Intermediate between the fairly specific problems of filter theory on the one hand and the general realizability theory of multi-terminal devices on the other, lies the study of four-poles as transducers. There is a considerable literature on the realization of transfer functions or transfer impedances under various structural limitations. The basic and extensive work of Bode[14] on active systems belongs also in this category since it is avowedly limited to single-loop structures.

2.6 Beyond the important result that, by sufficiently elaborate circumventions, one may avoid the use of transformers in the synthesis of any two-pole, (Bott and Duffin[13]) little in general is known about networks which do not have transformers.

2.7 The present paper is an essay in the realizability theory of devices with many accessible terminals. Ideal transformers are admitted as network elements; indeed, their use is essential. This fact is indicated by the adjective "formal" appearing in the title.

The availability of ideal transformers makes it possible to exploit the simplification noted by Cauer and to consider only networks which are $2n$-poles in his sense. The aim of the paper, therefore, is to set a foundation for realizability theory which is completely general within this framework.

2.71 The first problem is that of description. We observed above an example—entirely trivial—of a passive four-pole which had neither an impedance nor an admittance matrix. Unfortunately, opportunities

* Date of Japanese publication. This reference, and manuscript of Oono[10, 11], were sent by Professor Oono in a personal communication to R. L. Dietzold, who showed them to me in December, 1948, while an early draft of the present paper was in preparation. The recent (1950) American republication of Oono[10] unfortunately omits reference to the original.

† Cf. footnote to 1.1.

for this kind of degeneracy become manifold in multi-terminal devices, and some degree of degeneracy is the rule rather than the exception. Consider an entirely practical example: that of an amplifier chassis from which the tubes have been removed.* Here the degeneracy is essential and intrinsic; it would be highly artificial to regard it as the mere accident of a limiting case. True, given any *particular* degenerate network, there is usually evident a method for representing or describing its behavior. What is needed, however, is a mode of representation which is applicable generally to any $2n$-pole without *a priori* knowledge of its structure or peculiar degeneracies.

2.72 The mode of representation adopted in this paper, embodied in the notions of general $2n$-pole (Section 4) and linear correspondence (Section 6), is an obvious one, and so completely general that it solves no problems other than the elemental one for which it was introduced. It provides a definite mathematical construct whose properties one can discuss with mathematical precision. This is all that we ask of it.

Realizability theory begins and ends with the study of these properties. It would be more accurate to say that the notion of general $2n$-pole describes a particular, but still very large, class of mathematical entities; realizability theory consists in the study of certain subclasses of the whole class of these entities, the particular subclasses being distinguished by special, and to us interesting, properties.

2.73 Despite its avowed aim at generality, the paper is oriented toward the realizability theory of finite passive networks. It ultimately provides a proof of 1.1 and indeed a complete characterization of finite passive $2n$-poles, however degenerate. This characterization is accomplished in a sequence of postulates, each one delineating a property of general $2n$-poles, i.e., a subclass consisting of all $2n$-poles having this property. The class of $2n$-poles having all of these properties is then identified with the class of $2n$-poles obtained from finite passive networks.

2.74 If we have succeeded here in our hope to set an adequate foundation for the realizability theory of devices with many terminals, it will be because of the nature and organization of the postulates themselves. They describe what at present seem to be individually significant properties of $2n$-poles, of progressively greater specificity, which in the aggregate characterize finite passive devices. By eliminating them in various combinations one obtains larger classes of objects. Further re-

* It is exactly this example, and the practical need of an adequate theory for it, which led the author first to study the realizability theory of passive multi-terminal devices.

search alone will tell whether or not one obtains in this way the kinds of device which are significant. For example, one would like general realizability theorems for structures containing vacuum tubes with frequency-independent transconductances, vacuum tubes with non-vanishing transit times, unilateral devices with specified parasitics, etc.

2.75 Actually, the postulates as we have given them are certainly not adequate for such an ambitious program. Exigencies of the presentation have dictated a number of condensations and compromises. It is hoped that the basic ideas are still evident even if not isolated individually in separate and entirely independent postulates. In any event, it is the author's firm belief that the presentation as given is at least illustrative of the kind of approach, and the level of mathematical detail, which will be needed if one is ever to provide a truly adequate realizability theory: a theory which will cover, for example, the broad range of active linear systems which present-day technology allows us to consider.

2.8 Apart from the network theoretic concepts, which must be evaluated by their effectiveness in solving problems—an assessment which is by no means yet complete—this paper is strongly marked by an idiosyncracy of its author: a consistent and insistent use of geometric ideas and terminology. This is based on the personal experience that linear algebra achieves logical unity and a freedom from encumbering notation when viewed in this way. A general reference covering most of the linear algebra (geometry) required here is P. R. Halmos' elegant monograph[9].

2.9 For a proof solely of 1.1, which has already been three times proved in the literature,[1, 10, 11] this paper provides an apparatus which is too cumbersome. There is even a sense in which 1.1 alone provides a characterization of all finite passive devices, for it seems to be generally accepted that, by the use of ideal transformers, any finite passive network can be represented as a network which has an impedance matrix to which is adjoined suitable ideal transformers. Therefore we cannot claim that, in using this cumbrous apparatus to characterize all finite passive $2n$-poles (including the degenerate ones), we have offered anything not already provided by a simpler proof of 1.1.

Three things may be said in rebuttal. First, we have already emphasized that the apparatus here exhibited was designed for more problems than that to which it is here applied. It is presented in the belief that it will prove of further use.

Second, even in the study of passive networks, it has seemed to the author helpful to look at the manifold things which are *not* passive net-

works. One gets then a clearer view of the unique position occupied by passive devices among all linear systems.

Third, there is a kind of semantic issue here: the assertion that any finite passive *network* (sic) can be put in such a form that 1.1 applies seems to this author to give a kind of circular characterization of such devices. A characterization which did not itself involve the concept of a network seems more satisfying. Logically, there is no circle here, but this is a fact requiring proof. A careful reading of this paper will show that it provides a proof. This particular subtlety does not of itself justify the lengths to which we have gone. It is, however, no longer a subtlety if one wishes to consider devices which do not have a representation in terms of something non-degenerate to which ideal transformers have been added.

2.91 The present Part I of the paper is so organized that at the end of Section 8 the reader is in possession of all of its principal results and its basic ideas. The remaining Sections, 9 through 20, may then be regarded as an Appendix containing the details of proofs. Indeed, Part II will be largely devoted to further details of proof, though there will be there one important idea not mentioned, save casually, in Part I—the idea of degree for a matrix.

In Sections 4 through 11, technical paragraphs have been distinguished from explanatory or heuristic ones by starring the paragraph numeral.

Part II of the paper contains the bulk of the proof of 1.1. This proof is modelled after that of Brune[2] for the realizability of two-poles. One familiar with the Brune process will probably find Part II readable without extensive reference to Part I.

Let the reader be warned that the Brune process is not a practical one for realizing networks because of its critical dependence upon a difficult minimization and balancing operation. The same criticism applies to the generalized Brune process of Part II.

The Brune process is of theoretical importance because it does realize a network with the minimum number of reactive elements. These facts will be brought to light in Part II.

The proofs of Oono[10] and Bayard[1] are different from ours. That of Oono[11] again follows the Brune model.

III. INTRODUCTION TO PART I

3.0 We keep before us first the problem of finding a mathematical description applicable to and characterizing the behavior of all finite pas-

sive networks. Second, we seek to make mathematically precise those ideas which appear to form the basis of general realizability theory. Sections 4 through 7 introduce the immediate mathematical machinery for this. Section 8 states the fundamental realizability theorem and outlines its proof. At this point the reader has had an introduction to the results of the paper. The remainder of the paper is then devoted to the technical details of proof. Beginning with Section 12, the device of starring the technical passages will be dropped.

3.1 Cauer[5] distinguished precisely the class of networks called $2n$-poles from the class of all multi-terminal networks. He also showed that, by the use of ideal transformers, any multi-terminal network is equivalent to a network which is a $2n$-pole (for some n) in his sense. We shall in Section 4 define a class of objects to be called general $2n$-poles. This class includes all electrical networks which are $2n$-poles in Cauer's sense. Its definition abstracts the significant properties isolated by Cauer.

For the study, alone, of finite passive networks, this definition is unnecessary, since one can in fact so put the arguments as to deal only with $2n$-poles which are finite passive networks, and therefore to deal only with concepts already defined in Cauer[5]. The somewhat physical notion of a general $2n$-pole is a convenient backdrop against which to display the important physical properties of finite passive networks, and, indeed, of networks in general. Having it available, we use it throughout the realizability arguments.

IV. DEFINITION OF GENERAL $2n$-POLE

4.0* Network theory establishes a correspondence between oriented linear graphs and systems of differential equations. With each node of the graph is associated a potential $E_n = E_n(t)$ and with each oriented branch a current $I_b = I_b(t)$. These potentials and currents are constrained, first by Krichoff's laws, and second by differential equations which depend upon the nature of the branches but not upon the topology of the graph.

4.01* A finite passive network is one whose graph has the following properties:

(i) There are finitely many nodes, 1, 2, $\cdots$, N.

(ii) There are finitely many branches, 1, 2, $\cdots$, B.

* Technical paragraph as explained in Section 2.91.

(iii) Let the b-th branch begin at node n_b and end at n_b'. Let $v_b = E_{n_b} - E_{n_b'}$. Then for each b, one of

$$v_b = R_b I_b, \qquad R_b > 0 \tag{a}$$

$$I_b = C_b \frac{dv_b}{dt}, \qquad C_b > 0 \tag{b}$$

$$v_b = \sum_{b'} L_{bb'} \frac{dI_{b'}}{dt} \tag{c}$$

holds, where the matrix $L_{b,b'}$, is real, symmetric, and semi-definite.

Cauer has shown[5] how an ideal transformer can be defined as the limiting case of a finite passive network. It is indeed no more nor less ideal than an open circuit ($R_b = \infty$ or $C_b = 0$) or a short circuit ($R_b = 0$ or $C_b = \infty$).

4.02 We seldom deal with networks in the detail which is implicit in (iii) above. We are usually interested in the external characteristics, so to speak, of such networks as viewed from a relatively small number of terminals (nodes). These multi-terminal devices, however, we continue to incorporate into larger network diagrams. It is usually clear how Kirchoff's laws are to be applied in these cases, and what the differential equations of the resulting system are. We are obliged, however, to make these matters precise before we can deal intelligently with the most general physical properties of networks.

4.1 We have seen the two kinds of constaint that a multi-terminal device imposes on the branch currents and node voltages in a network in which it is incorporated: the topological ones contained in Kirchoff's laws and the dynamical ones described by differential equations. Correspondingly, there are two aspects to the concept of general $2n$-pole.

4.11* In its relation to Kirchoff's laws, a general $2n$-pole is indicated as an object with n pairs of terminals (T_r, T_r'), $1 \leq r \leq n$. Each terminal can be made a node in an arbitrary finite diagram constructed out of network elements and other general $2m$-poles, with arbitrary values of m. This diagram is not an oriented linear graph, so we have no basis for the use of Kirchoff's laws. From it, however, we construct an oriented linear graph, called the *ideal graph* of the diagram, by the following rule:

The nodes of the ideal graph are those of the original diagram. Every

* Technical paragraph as explained in Section 2.91.

oriented branch of the original diagram is repeated in the ideal graph, similarly situated and oriented. Between those nodes which, in the original diagram, were the (T_r , T'_r) of a $2n$-pole **N**, is drawn a branch β_r , called the r-th *ideal branch* of **N**, oriented from T_r to T'_r . This is done for each such terminal pair.

Kirchoff's laws now apply to this ideal graph.

4.12* Consider a particular $2n$-pole **N**. Let E_r be the potential of T_r , E'_r that of T'_r . Define

$$v_r(t) = E_r - E'_r .$$

Then $v_r(t)$ is the voltage across the ideal branch β_r so oriented that $v_r(t) \geq 0$ when T_r is positive relative to T'_r . Let $k_r(t)$ represent the current entering T_r . Then $k_r(t) = I_r(t)$, the current in β_r , so $k_r(t)$ is also the current leaving T'_r . This is the force of the notion of ideal branch and the fact which distinguishes a network which is a $2n$-pole from an arbitrary network with $2n$ terminals.

4.13 For example, the network at (a) of Fig. 1 is not a 2×2 pole because its currents are not constrained to meet the ideal branch requirement. The addition of ideal transformers in either of the ways shown in (b) or (c) of the figure converts it to a 2×2 pole. Of course in a circuit in which the currents are constrained externally—as they would be, for

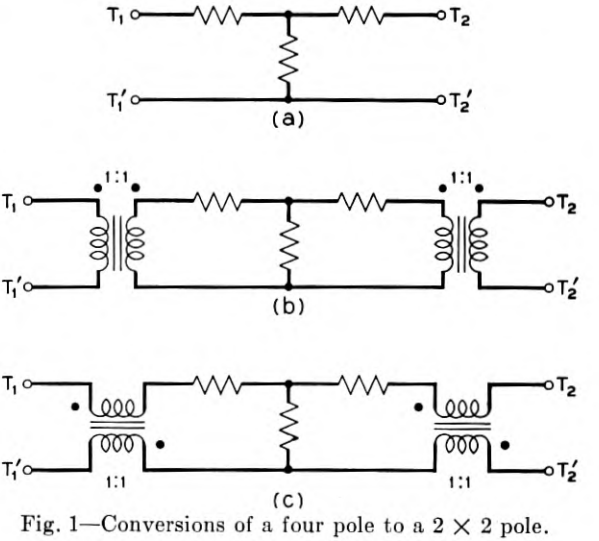


Fig. 1—Conversions of a four pole to a 2×2 pole.

* Technical paragraph as explained in Section 2.91.

example, when the 2×2 pole is driven by separate generators in the two external meshes—these transformers can be eliminated. The definition of $2n$-pole requires however that in every context the ideal branch concept is valid.

4.2* The second aspect of the concept of general $2n$-pole is that it imposes some kind of constraint—other than that implied by 4.11 and Kirchoff's laws—upon the voltages across and currents in its ideal branches. Define the symbols

$$\underline{v} = \underline{v}(t) = [v_1(t), v_2(t), \cdots, v_n(t)]$$

and

$$\underline{k} = \underline{k}(t) = [k_1(t), k_2(t), \cdots, k_n(t)]$$

as the n-tuples, respectively, of voltages across (T_r, T_r') and currents into T_r, $1 \leq r \leq n$. These are added and multiplied by scalars by the usual rules of vector algebra. If $\underline{v}$ and $\underline{k}$ represent *simultaneous* values of voltage and current in the $2n$-pole **N**—i.e., values satisfying all the constraints—then we say that **N** admits the pair $[\underline{v}, \underline{k}]$.

We say that **N** admits $\underline{v}$ if there is a $\underline{k}$ such that **N** admits the pair $[\underline{v}, \underline{k}]$. This $\underline{k}$ is said to correspond to $\underline{v}$. Dually, **N** admits $\underline{k}$ if there is a $\underline{v}$ (corresponding to $\underline{k}$) such that **N** admits $[\underline{v}, \underline{k}]$.

The constraints imposed by a general $2n$-pole **N** on voltages and currents are completely described by the totality of pairs $[\underline{v}, \underline{k}]$ which **N** admits. We shall *define* a general $2n$-pole, therefore, as

(i) a collection of n oriented ideal branches, as in 4.11, and

(ii) a list of pairs $[\underline{v}, \underline{k}]$ of voltages and currents admitted in these branches.

Hereafter we shall usually drop the adjective "general."

4.21 The definition we have just given is, in a way, a postulational form of an n-dimensional Thevenin's theorem. It postulates that a $2n$-pole is a thing† which, as far as the outside world is concerned, can be represented by a collection of two-poles β_r, $1 \leq r \leq n$, among which there exists a complicated agreement as to what currents and voltages will be admitted.

4.22 The passive networks (b) and (c) of Fig. 1 define 2×2 poles, because they satisfy 2.01 and clearly permit a complete specification of the admissible pairs $[\underline{v}, \underline{k}]$. Any equivalent network would also specify

* Technical paragraph as explained in Section 2.91.

† In fact, at this level of generality, *any* thing.

the same 2 × 2 pole, because—by its very equivalence—it would admit the same pairs. The closest association we can make between a $2n$-pole and a network, then, is to identify the $2n$-pole with an equivalence class of networks.

4.23 The completely symmetric role played by voltages and currents in this definition of general $2n$-pole will make it possible to take early advantage of the well-known duality principle of network theory. We shall do so freely.

4.3* We shall call a $2n$-pole physically realizable if its admissible pairs $[v, \underline{k}]$ are the solutions of a system of differential equations obtained from a finite passive network, admitting the limiting elements: ideal transformers, open circuits, and short circuits.

V. PHYSICAL PROPERTIES OF NETWORKS

5.0 There are clearly a great many properties of finite passive networks which are not yet possessed by the general $2n$-poles now introduced. It is instructive to examine these properties physically.

5.1 In the first place, the dynamical constraints (a), (b), and (c) of 4.01 are expressed by linear, time invariant, differential equations. Accordingly, the $2n$-poles of network theory are:

5.11 Linear, in that the class of admissible pairs $[v, \underline{k}]$ is a linear space;

5.12 Time invariant, admitting with each $[v(t), \underline{k}(t)]$ also all $[v(t + \tau), \underline{k}(t + \tau)]$ for aribtrary τ.

5.2 In the second place, a physical network **N** cannot predict the future, i.e., it cannot respond before it is excited. This can be formalized in terms of the pairs $[v, k]$ admitted by **N**, but to do so would require some digression. The reasons will be seen under 5.7 below.

5.3 We have already mentioned the constraints imposed on voltages and currents in a network by the topology of the network, through the medium of Kirchoff's laws. These constraints have three important properties:

5.31 They are workless, since they are imposed by resistanceless connections, leakless nodes, and, in the formal theory, by ideal transformers.

5.32 Though it seems scarcely necessary to say it, they are the only workless constraints. All other constraints are dynamical and have powers or energies associated with them.

* Technical paragraph as explained in Section 2.91.

5.33 They are frequency independent, that is, holonomic in the sense of dynamics.

5.4 The workless and the dynamical constraints in a physical network are all defined by relations with real coefficients. The space of admissible pairs is then a real linear space.

5.5 The positivities specified in 4.01 are characteristic of passive systems. They correspond to the fact that the power dissipation and the stored energies are all positive.

5.6 By definition, finite passive networks contain finitely many lumped elements. Correspondingly, their resonances and anti-resonances are finite in number.

5.7 We are accumstomed to dealing with networks which have, in addition to the properties listed above, a kind of non-degeneracy, in that the list of admissible pairs $[\underline{v}, \underline{k}]$ satisfies:

5.71 At least one of $\underline{v}$ or $\underline{k}$ can be specified arbitrarily—any real function is admitted;

5.72 When the free number of $[\underline{v}, \underline{k}]$ is specified, the other is uniquely determined.

For these non-degenerate networks, the property 5.2 above is easily formalized: if, say, $\underline{k}$ is determined by $\underline{v}$, then

$$\underline{v}^1(t) = \underline{v}^2(t) \quad \text{for} \quad t \leq t_0$$

implies

$$\underline{k}^1(t) = \underline{k}^2(t) \quad \text{for} \quad t \leq t_0 ,$$

where $[\underline{v}^i, \underline{k}^i]$ are admissible pairs, $i = 1, 2$. The general statement of 5.2 involves this condition and some discussion of the $\underline{v}$'s for which **N** admits $[\underline{v}, 0]$, and the dual notions.

5.8 The reason for speaking in terms of pairs $[\underline{v}, \underline{k}]$, instead of in terms of "cause" and "effect," or "impulse" and "response," is hinted at by 5.7 above. For the tacit implications of the cause and effect language completely obscure the fact that 5.71 and 5.72 are properties which are not automatically possessed by electrical networks. In fact, the simple four-pole of 2.41—a pair of unconnected terminals T_1 , T_1' , and a pair of shorted terminals T_2 , T_2'—has neither property, yet it is a perfectly good linear time invariant four pole. Its admissible pairs are

$$[(v_1 , 0), (0, k_2)],$$

where v_1 and k_2 are arbitrary real functions of the time

VI. LINEAR CORRESPONDENCES

6.0 In developing the formal properties of $2n$-poles which are equivalent to the physical ones just listed, it would be instructive to adjoin requirements piecemeal, much in the order given in Section 5. Space does not permit us full enjoyment of this luxury, but the reader will find a rough parallel between Section 5 and the developments of this Section and Section 7.

6.1 It is well known that linear time invariant systems are best studied by the tools of Fourier or Laplace analysis. We make this fact the basis of our first step in characterizing physically realizable $2n$-poles simply by phrasing our whole discussion in the frequency language. The content of the following paragraph will be obvious enough, but it does define terms to be used later.

6.11* Let v and k, without underscores, represent n-tuples of complex numbers:

$$v = [v_1 , v_2 , \cdots , v_n], \tag{1}$$

$$k = [k_1 , k_2 , \cdots , k_n]. \tag{2}$$

These are to be manipulated by the rules of vector algebra. Let p be a complex number. We shall say that a $2n$-pole $\mathbf{N}$ admits the pair $[v, k]$ at frequency p, if in the sense of 4.2 $\mathbf{N}$ admits the pair $[\underline{v}, \underline{k}]$ (*with* underscores) where $\underline{v}$ has components

$$v_r(t) = \mathrm{Re}(v_r e^{pt}), \qquad 1 \leq r \leq n, \tag{3}$$

and $\underline{k}$ has components

$$k_r(t) = \mathrm{Re}(k_r e^{pt}), \qquad 1 \leq r \leq n. \tag{4}$$

Also analogously to 4.2, we say that $\mathbf{N}$ admits v at frequency p if there is a k such that $\mathbf{N}$ admits $[v, k]$ at frequency p, and that this k corresponds to v (at frequency p). Similarly, $\mathbf{N}$ admits k at frequency p if there is a (corresponding) v such that $\mathbf{N}$ admits $[v, k]$ (at p).

6.12* Let $\mathbf{V}$ denote the aggregate of all n-tuples (1), and $\mathbf{K}$ the aggregate of all n-tuples (2). These are then complex linear spaces.

6.2* As our first step toward characterizing realizable $2n$-poles, let us consider a *linear correspondence* L between $\mathbf{V}$ and $\mathbf{K}$ described by the postulates:

P1. There is a set Γ_L of complex numbers and for each $p \epsilon \Gamma_L$ a list $L(p)$ of pairs $[v, k]$, $v \epsilon \mathbf{V}$, $k \epsilon \mathbf{K}$.

* Technical paragraph as explained in Section 2.91.

P2. If $[v^1, k^1] \epsilon L(p)$ and $[v^2, k^2] \epsilon L(p)$, then

$$[a_1 v^1 + a_2 v^2, a_1 k^1 + a_2 k^2] \epsilon L(p)$$

for any complex numbers a_1 , a_2 .

6.21* Given such a linear correspondence L, we can always describe a $2n$-pole $\mathbf{N}_L$ by:

$\mathbf{N}_L$ admits $[v, k]$ at frequency p if and only if $[v, k] \epsilon L(p)$.

That is, we can always interpret the pairs $[v, k]$ generated by (3) and (4) from the $[v, k] \epsilon L(p)$, for each $p \epsilon \Gamma_L$, as the voltages across and currents in a set of n ideal branches. We call $\mathbf{N}_L$ the $2n$-pole associated with L.

6.22* We call Γ_L the frequency domain of L (or of $\mathbf{N}_L$).

6.23 From here on, the words "$2n$-pole" can with some strain be regarded as suggestive but unnecessary. We in fact deal with linear correspondences—having properties as yet unspecified—and shall show how physical networks can be constructed which admit the pairs $[v, k] \epsilon L(p)$. Actually we use freely the concept of general $2n$-pole and thereby avoid some elaborate circumlocutions.

6.24* We identify two correspondences L_1 and L_2 as being the same if (i) their frequency domains differ only by a finite set, and (ii) for each p where both are defined the lists $L_1(p)$ and $L_2(p)$ are the same.

6.3 The simplest linear correspondences are those generated by matrices. For example, let $Z(p)$ be an $n \times n$ matrix with, say, elements $Z_{rs}(p)$ which are rational functions of p, $1 \leq r, s \leq n$. Let Γ_L consist of all the values of p at which $Z(p)$ is defined. For $p \epsilon \Gamma_L$, define $L(p)$ as the class of all pairs

$$[v, k] \tag{5}$$

obtained by letting k range over $\mathbf{K}$, where for each k, v is defined by the matrix equation

$$v = Z(p)k. \tag{6}$$

This kind of matrix equation will be used throughout to symbolize the n component equations

$$v_r = \sum_{s=1}^{n} Z_{rs}(p) k_s, \qquad 1 \leq r \leq n. \tag{7}$$

The list of pairs $L(p)$ defined by (5) clearly satisfies P1 and P2. It can therefore be used to define a $2n$-pole $\mathbf{N}_L$. It is easy to see that $\mathbf{N}_L$

* Technical paragraph as explained in Section 2.91.

in fact is non-degenerate in a sense similar to that of 5.7, for the current amplitudes k can be specified arbitrarily, and the resulting voltage amplitudes v are then fixed by k and p, by (6).

$Z(p)$ is called the impedance matrix of the $2n$-pole $\mathbf{N}_L$. It is also sometimes called the open-circuit impedance matrix, because each $Z_{rs}(p)$ is, by (7), the voltage amplitude across (T_r, T'_r) when the current amplitudes at all terminals save (T_s, T'_s) are zero—i.e., when all pairs save the s-th are on open circuit.

6.31 Dually, the pairs

$$[v, Y(p)v]$$

defined by an admittance matrix $Y(p)$ as v ranges over $\mathbf{V}$ define a linear time invariant $2n$-pole which is non-degenerate.

VII. WORK AND ENERGY

7.0* A linear correspondence satisfying P1 and P2 is something which abstracts the properties of linearity and time invariance. Most of the remaining properties of physical networks involve the mention of work or energy. These concepts enter our picture by way of the scalar product (v, k) between a voltage n-tuple (1) and a current n-tuple (2), of 6.11. This scalar product is defined by

$$(v, k) = \sum_{r=1}^{n} v_r \bar{k}_r. \tag{1}$$

7.01 If $p = i\omega$, one easily calculates from (3) and (4) of 6.11 that

$$2 \operatorname{Re}(v, k) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} \left[\sum_{r=1}^{n} v_r(t) k_r(t) \right] dt.$$

That is, when $p = i\omega$, the real part of $2(v, k)$ measures the average total power dissipated by the system of currents $k_r(t)$ against the driving voltages $v_r(t)$.

When p is not a pure imaginary, the interpretation of the scalar product (v, k) is not so clearly physical as this. The reader will ultimately observe that our significant statements about such products can all be reduced to statements applicable when $p = i\omega$, i.e., when the power interpretation is valid.

7.1* An important concept in what follows is that of the annihilator of a linear manifold (Halmos[9], par. 16). Let $\mathbf{V}_1 \subseteq \mathbf{V}$ be a linear manifold.

* Technical paragraph as explained in Section 2.91.

Then its annihilator $(\mathbf{V}_1)^0$ is the set of all k such that

$$v\epsilon\mathbf{V}_1 \quad \text{implies} \quad (v, k) = 0.$$

$(\mathbf{V}_1)^0$ is a linear manifold in $\mathbf{K}$.

Dually, given $\mathbf{K}_1 \subseteq \mathbf{K}$, $(\mathbf{K}_1)^0$ is the linear manifold of all $v\epsilon\mathbf{V}$ such that

$$k\epsilon\mathbf{K}_1 \quad \text{implies} \quad (v, k) = 0.$$

The annihilator concept is the analog in our general geometric framework of the idea of orthogonality. It clearly suggests a connection with workless constraints.

7.2* The complex conjugate of an n-tuple v (or k) is defined in the obvious way: if

$$v = [v_1, \cdots, v_n]$$

then

$$\bar{v} = [\bar{v}_1, \cdots, \bar{v}_n].$$

This conjugation operation clearly has the properties

$$\begin{gathered} \bar{\bar{\xi}} = \xi \\ \overline{a\xi + b\eta} = \bar{a}\bar{\xi} + \bar{b}\bar{\eta} \end{gathered} \tag{2}$$

where a and b are scalars and ξ and η are (consistently) elements of $\mathbf{V}$ or $\mathbf{K}$. Furthermore, at once from (1) of 7.0,

$$\overline{(v, k)} = (\bar{v}, \bar{k}). \tag{3}$$

7.21* A linear manifold will be called real if it contains, with any n-tuple also the conjugate of that n-tuple.

7.22* A real manifold is spanned by real n-tuples. This will be proved in the Appendix, Section 20.

7.23* The annihilator of a real manifold is real. For let $\mathbf{K}_1$ be real and $k^1, \cdots, k^r$ be real n-tuples which span $\mathbf{K}_1$. Then if $v\epsilon(\mathbf{K}_1)^0$ every

$$(v, k^s) = 0,$$

and conversely. But then also

$$(\bar{v}, k^s) = \overline{(v, k^s)} = \bar{0} = 0,$$

so $\bar{v}\epsilon(\mathbf{K}_1)^0$.

* Technical paragraph as explained in Section 2.91.

7.3* Given a linear correspondence L, we make several definitions:

$\mathbf{V}_L(p)$ is the set of all $v \epsilon \mathbf{V}$ such that there is a k with $[v, k] \epsilon L(p)$.

$\mathbf{K}_L(p)$ is the set of all $k \epsilon \mathbf{K}$ such that there is a v with $[v, k] \epsilon L(p)$.

$\mathbf{V}_{L0}(p)$ is the set of $v \epsilon \mathbf{V}_L(p)$ such that

$$[v, 0] \epsilon L(p).$$

$\mathbf{K}_{L0}(p)$ is the set of $k \epsilon \mathbf{K}_L(p)$ such that

$$[0, k] \epsilon L(p).$$

7.31* The postulate P2 implies that for each $p \epsilon \Gamma_L$, $\mathbf{V}_L(p)$, $\mathbf{K}_L(p)$, $\mathbf{V}_{L0}(p)$ and $\mathbf{K}_{L0}(p)$ are all linear manifolds.

7.32 $\mathbf{V}_L(p)$, for example, is the set of $v \epsilon \mathbf{V}$ such that $\mathbf{N}_L$ admits v at frequency p.

7.4* We now postulate

P3. There exist fixed linear manifolds $\mathbf{V}_L \subseteq \mathbf{V}$, $\mathbf{K}_L \subseteq \mathbf{K}$ such that

(A) For every $p \epsilon \Gamma_L$, $\mathbf{V}_L(p) = \mathbf{V}_L = (\mathbf{K}_{L0}(p))^0$

(I) For every $p \epsilon \Gamma_L$, $\mathbf{K}_L(p) = \mathbf{K}_L = (\mathbf{V}_{L0}(p))^0$.

7.41* We may henceforth write $\mathbf{V}_{L0}$, $\mathbf{K}_{L0}$, for $\mathbf{V}_{L0}(p)$, $\mathbf{K}_{L0}(p)$, knowing that, under P3

$$\mathbf{V}_{L0} = (\mathbf{K}_L)^0,$$

$$\mathbf{K}_{L0} = (\mathbf{V}_L)^0.$$

7.42 Linear correspondences satisfying P3 abstract the properties mentioned in 5.3. The equalities $\mathbf{V}_L(p) = \mathbf{V}_L$, $\mathbf{K}_L(p) = \mathbf{K}_L$ guarantee the frequency-independence of the workless constraints. The equalities $\mathbf{V}_L(p) = (\mathbf{K}_{L0}(p))^0$, $\mathbf{K}_L(p) = (\mathbf{V}_{L0}(p))^0$ in a sense guarantee that the only constraints imposed upon admissible currents and voltages (as opposed to constraints relating currents and voltages) are those which arise from open or short circuits, i.e., are workless.

7.43 An illustrative consequence of P3, for example, is that if L satisfies P3 and if $\mathbf{N}_L$ is such that all of the current amplitudes can be specified arbitrarily, then indeed the voltages are determined by the currents. This will appear as a consequence of 8.1. It is a very general theorem about networks of a kind that this author, at least, has not heretofore encountered.

* Technical paragraph as explained in Section 2.91.

7.5* Continuing toward realizability, we introduce

P4. If $p\epsilon\Gamma_L$, then $\bar{p}\epsilon\Gamma_L$. If $[v, k]\epsilon L(p)$, then $[\bar{v}, \bar{k}]\epsilon L(\bar{p})$.

This postulate embodies most of the reality properties of networks. It has as an immediate consequence the

7.51* *Lemma:* If L satisfies P1, P2, P3, and P4, then all of

$$\mathbf{V}_L\ ,\ \mathbf{V}_{L0}\ ,\ \mathbf{K}_L\ ,\ \mathbf{K}_{L0}$$

are real.

Proof: By P4, $v\epsilon\mathbf{V}_L(p) = \mathbf{V}_L$ implies $\bar{v}\epsilon\mathbf{V}_L(\bar{p}) = \mathbf{V}_L$. Hence $\mathbf{V}_L$ is real. Then $\mathbf{K}_{L0} = (\mathbf{V}_L)^0$ is real, and dually.

7.6* The three remaining postulates on L refer to scalar products. They are concerned with the energy questions related to passivity, rather than with the workless constraint questions.

P5. If $[u, j]\epsilon L(p)$ and $[v, k]\epsilon L(p)$, and if

(A) u and v are real, or if

(I) j and k are real,

then

$$(u, k) = (v, j).$$

7.61 This is the property which provides the reciprocity law. In its presence, the relations in P3 may be weakened to

$$\mathbf{V}_L(p) = \mathbf{V}_L \supseteq (\mathbf{K}_{L0}(p))^0,$$

$$\mathbf{K}_L(p) = \mathbf{K}_L \supseteq (\mathbf{V}_{L0}(p))^0.$$

This fact will appear as a consequence of the lemma of Section 12.

7.7* *Lemma:* A consequence of P2 and P3(A) is that if

$$[v, k_r]\epsilon L(p), \qquad r = 1, 2,$$

then for any $u\epsilon\mathbf{V}_L$,

$$(u, k_1) = (u, k_2).$$

For by P2 we have that

$$[v - v, k_1 - k_2] = [0, k_1 - k_2]\epsilon L(p),$$

hence $k_1 - k_2\epsilon\mathbf{K}_{L0}$. Then however, by P3(A), $u\epsilon\mathbf{V}_L$ implies $u\epsilon(\mathbf{K}_{L0})^0$, so

* Technical paragraph as explained in Section 2.91.

that

$$0 = (u, k_1 - k_2) = (u, k_1) - (u, k_2).$$

Q.E.D. A dual result follows from P3(I).

7.71* The result of 7.7 above means that the scalar product (v, k) is fixed by v alone when we know that $[v, k] \epsilon L(p)$. This means that, given $v \epsilon \mathbf{V}_L$, there is a unique function $F_v(p)$ defined for $p \epsilon \Gamma_L$ by

$$F_v(p) = \overline{(v, k)}$$

where $[v, k] \epsilon L(p)$. Dually,

$$J_k(p) = (v, k)$$

is defined for each fixed $k \epsilon \mathbf{K}_L$.

7.72* (P6.) The complement of Γ_L is finite and

(I) For each $v \epsilon \mathbf{V}_L$, $F_v(p)$ is rational

(A) For each $k \epsilon \mathbf{K}_L$, $J_k(p)$ is rational.

7.73* (P7.) (A) $\mathrm{Re}(p) \geq 0$ implies $\mathrm{Re}(F_v(p)) \geq 0$

(I) $\mathrm{Re}(p) \geq 0$ implies $\mathrm{Re}(J_k(p)) \geq 0$.

VIII. THE FUNDAMENTAL REALIZABILITY THEOREM

8.0* We can now state our fundamental realizability theorem: If a linear correspondence L satisfies P1, $\cdots$, P7, the associated $2n$-pole $\mathbf{N}_L$ is physically realizable. Conversely, given a physically realizable $2n$-pole $\mathbf{N}$, the associated linear correspondence satisfies P1, $\cdots$, P7.

8.01 Actually, the postulates P1, $\cdots$, P7 are not unique nor even entirely independent. Many changes may be rung on them. We indicated one above. At the expense of apparent asymmetry, the (A) or (I) portions, in various combinations, can be deleted or weakened. We shall not pursue this subject further at this point, but must come back to it in Section 12.

8.02 We close this Section by outlining the proof of 8.0. The details are then contained in the remainder of the paper.

8.03 The proof that P1 through P7 are necessary for physical realizability will be a direct one: it will be shown that, considered individually, each network branch and each ideal transformer satisfies the postulates.

* Technical paragraph as explained in Section 2.91.

By an application of Kron's method (described by Synge[12]), it will then be shown that the imposition of Kirchoff's laws preserves the postulates. This work is most efficiently performed after the full machinery of the sufficiency proofs is available, and will be done in Section 19.

8.04 The sufficiency of P1 through P7 can be deduced—and we will do so—from the lemmas to be quoted below. Apart from Section 19 on necessity, the remainder of the paper is devoted to the proofs of these lemmas.

8.1* *Lemma:* If L is a linear correspondence satisfying P1, P2, P3, and P4, then there exists a fixed real nonsingular matrix W such that

8.11 The list $L_W(p)$ of all pairs†

$$[W^{-1}v,\ W'k],$$

where $[v,\ k]\epsilon L(p)$, describes a linear correspondence L_W satisfying P1, P2, P3, and P4.

8.12 The $2n$-pole $\mathbf{N}_W$ $(= \mathbf{N}_{L_W})$ associated with L_W consists of

(i) Some number r of open-circuited terminal pairs $(T_1, T_1'), \cdots, (T_r, T_r')$,

(ii) Some number s of short-circuited terminal pairs $(T_{n-s+1}, T'_{n-s+1}), \cdots, (T_n, T_n')$,

(iii) A set of $m = n - r - s$ terminal pairs $(T_{r+1}, T'_{r+1}), \cdots, (T_{r+m}, T'_{r+m})$.

8.13 Either $m = 0$, or the terminal pairs in (iii) are those of a $2m$-pole $\mathbf{N}_1$ which has a nonsingular impedance matrix $Z_1(p)$.

This lemma, and the following, will be proved in 13.2.

8.2* *Lemma:* If L satisfies P5, P6, and P7, then $Z_1(p)$ is a positive real‡ matrix, that is, $Z_1(p)$ satisfies (i), $\cdots$, (iv) of 1.1.

8.3* *Lemma:* If a $2m$-pole $\mathbf{N}_1$ has a positive real impedance matrix, then $\mathbf{N}_1$ is physically realizable.

This is the sufficiency half of the matrix realizability theorem 1.1. Part II will be devoted to its proof.

8.4* *Lemma:* If $\mathbf{N}_W$ is physically realizable, then $\mathbf{N}$ can be constructed from it by the use of ideal transformers.

This is Cauer's Transformation Theorem[5] about which we shall say more in Section 9.

---

* Technical paragraph as explained in Section 2.91.

† W^{-1} and W' are respectively the reciprocal and the transpose of W.

‡ Gewertz's terminology[8], by now traditional.

8.5* The sufficiency half of 8.0 is now clear. By 8.2 and 8.3, $\mathbf{N}_1$ is physically realizable. Clearly then $\mathbf{N}_W$ is, simply by the adjunction of the necessary open and short circuits. Finally $\mathbf{N}$ is by Cauer's theorem, 8.4.

8.6* We can see now how to prove the necessity of positive reality for the realizability of a positive real matrix $Z(p)$. This is the necessity half of the matrix theorem 1.1. Let $Z(p)$ be the matrix of a realizable $\mathbf{N}$. Then $\mathbf{N}$ has an associated linear correspondence L satisfying P1, $\cdots$, P7, by the necessity half of 8.0. The pairs of L are the pairs

$$[Z(p)k,\ k]$$

generated as k ranges over all n-tuples. By definition, then, the pairs of L_W are

$$[W^{-1}Z(p)k,\ W'k].$$

As k ranges over all n-tuples, the nonsingularity of W implies that $W'k$ does also. Let $U = W^{-1}$. Then the pairs above are the same as

$$[UZ(p)U'k,\ k]$$

as k ranges over all n-tuples. Hence L_W has the impedance matrix $UZ(p)U'$, where $U = W^{-1}$ is real and nonsingular. Because L_W has an impedance matrix, $r = 0$ in 8.12.

Now by 8.1 and 8.2, $Z_1(p)$ is positive real and the matrix $UZ(p)U'$ of L_W is just $Z_1(p)$ bordered by s rows and columns of zeros. It is then easy to see that $UZ(p)U'$ is positive real, and finally also that $Z(p)$ is. These last two facts will be proved formally in Section 16.

IX. CAUER'S TRANSFORMATION THEOREM

9.0 Cauer's transformation theorem[5] is the cornerstone of formal realizability theory. In one form, the theorem reads:

9.1* Let $Z(p)$ be the impedance matrix of a physically realizable $2n$-pole $\mathbf{N}$. Let U be a real, constant, nonsingular matrix. Then

$$UZ(p)U' \tag{1}$$

is again the impedance matrix of a physically realizable $2n$-pole, $\mathbf{N}_U$. $\mathbf{N}_U$ can be constructed from $\mathbf{N}$ by the use of ideal transformers.

9.2* A superficial generalization of this theorem can be obtained at once from Cauer's proof. It asserts that if $\mathbf{N}$ is physically realizable and is described by the linear correspondence L, then there is a physically realizable $2n$-pole $\mathbf{N}_W$, obtainable from $\mathbf{N}$ by the use of ideal trans-

* Technical paragraph as explained in Section 2.91.

formers, which is described by the linear correspondence L_W whose pairs at each p are the pairs

$$[W^{-1}v,\ W'k], \tag{2}$$

where $[v,\ k]\epsilon L(p)$.

We refer to Cauer[5] for the proof. It is straightforward.

9.21 We shall use the second form (9.2) of Cauer's theorem in our realization process. Notice that it is in a sense a "physical" theorem, about the way one physical network is related to another. It is used in this way: we shall always solve a realizability problem by finding some network $\mathbf{N}$ which is easily realized, and then a W such that $\mathbf{N}_W$, which is now realizable, provides a solution to the given problem.

9.22* We shall call the $2n$-pole $\mathbf{N}_W$ a Cauer equivalent of $\mathbf{N}$.

9.3 Although Cauer's theorem will be applied, in a sense, only *a posteriori*, its effect is fundamental. For it implies that formal physical realizability is a property of matrices which is invariant under the operation (1) or a property of correspondences which is invariant under (2). There is an extensive classical literature on the properties of matrices invariant under operations like that of (1), and the effect of Cauer's theorem is to make these results all available to formal realizability theory.

9.31* It is worth observing here that we are already well set up to use Cauer's theorem:

Lemma: If L is a linear correspondence satisfying P1, $\cdots$, P7, then the correspondence L_W of 9.2 also satisfies P1, $\cdots$, P7.

Proof: Let $M = L_W$. P1 and P2 for M are obvious, with $\Gamma_M = \Gamma_L$. By definition of M,

$$\mathbf{V}_M(p) = W^{-1}\mathbf{V}_L(p) = W^{-1}\mathbf{V}_L$$

$$\mathbf{K}_M(p) = W'\mathbf{K}_L(p) = W'\mathbf{K}_L$$

$$\mathbf{V}_{M0}(p) = W^{-1}\mathbf{V}_{L0}(p) = W^{-1}\mathbf{V}_{L0}$$

$$\mathbf{K}_{M0}(p) = W'\mathbf{K}_{L0}(p) = W'\mathbf{K}_{L0}$$

where $W^{-1}\mathbf{S}$ for a manifold $\mathbf{S}$ consists of all n-tuples $W^{-1}v$, where $v\epsilon\mathbf{S}$. Hence in P3,

$$\mathbf{V}_M(p) = \mathbf{V}_M = W^{-1}\mathbf{V}_L$$

$$\mathbf{K}_M(p) = \mathbf{K}_M = W'\mathbf{K}_L$$

for fixed manifolds $\mathbf{V}_M$, $\mathbf{K}_M$ as defined.

* Technical paragraph as explained in Section 2.91.

Now if $v \epsilon \mathbf{V}_{L0}$, then

$$(v, k) = 0$$

for every $k \epsilon \mathbf{K}_L = (\mathbf{V}_{L0})^0$. Then, however, by direct calculation from Section 7.0,

$$(W^{-1}v, W^*k) = 0,$$

where W^* is the adjoint, i.e. transposed conjugate matrix of W. But because W is real, $W^* = W'$. Hence if $v \epsilon \mathbf{V}_{L0}$, then

$$(W^{-1}v, k) = 0$$

for every $k \epsilon W' \mathbf{K}_L = \mathbf{K}_M$. Hence

$$\mathbf{K}_M = (W^{-1}\mathbf{V}_{L0})^0 = (\mathbf{V}_{M0}(p))^0.$$

By this and its dual, P3 is completed for M.

The reamining postulates for M follow from those for L by the simple equality

$$(v, k) = (W^{-1}v, W'k)$$

already established, combined with $\Gamma_M = \Gamma_L$.

9.32 For fixed $Z(p)$, the matrices (1), as U ranges over a group, form an equivalence class. Classical matrix theory treats of such equivalence classes. This author's predilection is to regard this theory from a geometrical point of view. In part this prejudice may be justified by the ease with which that slightly more general object, a linear correspondence, can be treated by geometrical methods. In any event we shall begin our program of proofs with a brief introduction to the geometrical approach.

X. GEOMETRICAL PRELIMINARIES

10.0* We now wish to consider $\mathbf{V}$ and $\mathbf{K}$ as complex n-dimensional linear spaces† respectively of voltage vectors v and current vectors k. The distinction here is in point of view. A vector v is regarded as an absolute geometrical object; an n-tuple $[v] = [a_1, \cdots, a_n]$ is regarded as a set of coordinates for the vector v, relative to some coordinate basis. Given a fixed coordinate basis, there is a one-to-one correspondence between vectors v and the n-tuples $[v]$ which represent them in that basis, a correspondence which preserves the operations of vector algebra.

* Technical paragraph as explained in Section 2.91.

† For a reference concerning the ideas in this section, see Halmos[9], Chapters I and II.

10.01 The effect of attaching a geometric identity to vectors, rather than to n-tuples, is to make it possible to choose coordinate bases freely and as convenient, without elaborate constructions or even interpretations. We can then discuss properties of n-tuples (and other objects, e.g. matrices) which are invariant under the kind of operations exemplified by (1) and (2) of Section 9 as *properties of a single geometric object*, rather than as properties shared by an extensive class of concrete objects which are converted into each other by the group of operations.
10.1 This change in point of view need not change formally anything we have said to date; it simply erects a conceptual superstructure, or provides a conceptual foundation, depending on the reader's personal attitude.

We shall support this statement by going through the important ideas of Sections 4, 6, and 7 and examining their geometrical meanings or counterparts. It is convenient to consider first and at some length the notions of scalar product and complex conjugate. The geometric structure will then be complete enough to permit a rapid survey of the remaining ideas.

10.11* The geometrical counterpart of the scalar product introduced in 7.0 is a numerically valued function $\sigma = \sigma(v, k)$ of two vector variables. Its first argument v ranges over $\mathbf{V}$ and its second argument k ranges over $\mathbf{K}$. The function $\sigma(v, k)$ is linear in v and conjugate linear in k:

$$\begin{aligned} \sigma(au + bv, k) &= a\sigma(u, k) + b\sigma(v, k), \\ \sigma(v, ak + b\ell) &= \bar{a}\sigma(v, k) + \bar{b}\sigma(v, \ell). \end{aligned} \tag{1}$$

We denote this function $\sigma(v, k)$ by the simple bracket notation (v, k).

10.12 With this scalar product, the geometry of $\mathbf{V}$ and $\mathbf{K}$ is that of a space $\mathbf{K}$ and the space $\mathbf{K}^* = \mathbf{V}$ of conjugate linear functionals over $\mathbf{K}$. This is analogous to the real geometry of space and conjugate space discussed at length in Halmos[9]. In fact, in the introduction to Chapter III of Halmos[9], the modifications introduced by the conjugate linearity of (v, k) over $\mathbf{K}$ are treated in detail.

10.13* Because of its importance, we quote here a paraphrase of the results covered in Halmos[9], par. 12.

(i) If $f(v)$ is any numerically valued homogeneous linear function of $v \epsilon \mathbf{V}$, then there is a unique vector $k_f \epsilon \mathbf{K}$ such that

$$f(v) = (v, k_f)$$

for all $v \epsilon \mathbf{V}$.

* Technical paragraph as explained in Section 2.91.

(ii) If $g(k)$ is any numerically values homogeneous conjugate-linear function of $k\epsilon\mathbf{K}$ (i.e., if $\overline{g(k)}$ is linear in k) then there is a unique $v_g\epsilon\mathbf{V}$ such that

$$g(k) = (v_g , k)$$

for all $k\epsilon\mathbf{K}$.

10.2* The annihilator $(\mathbf{V}_1)^0$ of a manifold $\mathbf{V}_1 \subseteq \mathbf{V}$ is, as in 7.1, the set of all $k\epsilon\mathbf{K}$ such that

$$v\epsilon\mathbf{V}_1 \quad \text{implies} \quad (v, k) = 0.$$

10.21* It is shown in Halmos[9] that to each basis $v^1, \cdots, v^n$ in $\mathbf{V}$ there exists a unique dual basis $k^1, \cdots, k^n$ in $\mathbf{K}$ such that

$$(v^r, k^s) = \delta_{rs} , \tag{2}$$

where δ_{rs} is the Kronecker symbol: $\delta_{rs} = 0$ if $r \neq s$, $\delta_{rr} = 1$, $1 \leq r, s \leq n$.

10.22 If

$$\begin{aligned} [v] &= [a_1 , \cdots , a_n] \\ [k] &= [b_1, \cdots , b_n] \end{aligned} \tag{3}$$

are the n-tuples representing v and k relative to a pair of dual bases, then it is easily computed from (1) and (2) that

$$(v, k) = \sum_{r=1}^{n} a_r\bar{b}_r. \tag{4}$$

Therefore the concrete scalar product of 7.0 is indeed the geometric scalar product here considered, when we restrict our pairs of bases in $\mathbf{V}$ and $\mathbf{K}$ always to be dual in the sense of (2).

10.23* We shall use the words "coordinate frame" or simply "frame" to denote a pair of dual bases in $\mathbf{V}$ and $\mathbf{K}$. Any basis in $\mathbf{V}$ (or $\mathbf{K}$) specifies a frame by the uniqueness result quoted above.

10.24 We shall henceforth deal always with coordinate frames, in fact, ultimately, real coordinate frames, rather than arbitrary pairs of bases. This means in classical language that we are considering as "geometrical properties" all properties which are preserved under the group of linear transformations which leave the bilinear form (4) invariant. The properties related to physical realizability will turn out to be invariant only under the subgroup of real linear transformations preserving (4).

* Technical paragraph as explained in Section 2.91.

10.3* Conjugation is an operation which to each $v \epsilon \mathbf{V}$ associates a vector $\bar{v}$ uniquely determined by v with the properties

$$\begin{aligned} \bar{\bar{v}} &= v, \\ \overline{(au + bv)} &= \bar{a}\bar{u} + \bar{b}\bar{v}, \end{aligned} \tag{5}$$

where a and b are any complex numbers and $\bar{a}$, $\bar{b}$ their conjugates.

10.31* Given any such conjugation operation in $\mathbf{V}$, and given any $k \epsilon \mathbf{K}$, define a function $g_k(v)$ by

$$g_k(v) = \overline{(\bar{v}, k)} \tag{6}$$

for $v \epsilon \mathbf{V}$. Then $g_k(v)$ is linear in v, by (5) above and (1) of 10.11. Therefore, by 10.13, there is a unique vector $\bar{k} \epsilon \mathbf{K}$ such that

$$g_k(v) = (v, \bar{k}). \tag{7}$$

10.32* Directly from (1) of 10.11 and (6) above, if $j = ak + b\ell$, then

$$g_j(v) = a g_k(v) + b g_l(v).$$

From (7), therefore

$$(v, \bar{j}) = a(v, \bar{k}) + b(v, \bar{\ell})$$

for all $v \epsilon \mathbf{V}$. Comparing this with (1) of 10.11, we see that

$$\bar{j} = \bar{a}\bar{k} + \bar{b}\bar{\ell}. \tag{8}$$

The second item of (5) above then holds for vectors $k \epsilon \mathbf{K}$.

That $\bar{\bar{k}} = k$ follows easily: We have from (6) and (7), written for the vector $\bar{k}$, that

$$\overline{(\bar{v}, \bar{k})} = (v, \bar{k}). \tag{9}$$

We also have, by writing (6) and (7) for vectors $\bar{v}$ and k that

$$\overline{(\bar{\bar{v}}, k)} = (\bar{v}, \bar{k}).$$

Taking complex conjugates of these two numbers, and using $\bar{\bar{v}} = v$ from (5), we have

$$(v, k) = \overline{(\bar{v}, \bar{k})}. \tag{10}$$

Then (9) and (10), which hold for all $v \epsilon \mathbf{V}$, identify k and $\bar{\bar{k}}$ by 10.13.

10.34* We have now showed in (5), (8) and (10) that this complex conjugate satisfies the formal properties of the conjugate for n-tuples introduced in 7.2.

* Technical paragraph as explained in Section 2.91.

10.35. The abstract scalar product of 10.11 turned out in the end to be no more than the concrete one of 7.0 when we restrict our attention to n-tuples derived from vectors by the use of coordinate frames. In a similar way, it is not hard to show that there always exists a coordinate frame in which the abstract conjugation now introduced has the form of 7.2. This will be done in the Appendix (20.2).

10.36* Our need for writing out the components of vectors has now almost vanished. Henceforth we shall use subscripts to denote particular vectors, e.g. v_1, rather than components.

10.4* A vector will be called real if it is equal to its own conjugate. A manifold will be called real if it contains with each vector also the conjugate of that vector. **V** and **K** are then real. A basis will be called real if it is made up of real vectors, and a frame will be called real if its bases are real. Any frame in terms of which our conjugation operation takes the form of 7.2 is real by definition because its basis vectors *in that frame* have components which are 0 or 1. The vector 0 is real, similarly.

10.41* The basis dual to a real basis is real, for if

$$(v_r, k_s) = \delta_{rs},$$

then by (10) of 10.3 and the hypothesis that $v_r = \bar{v}_r$, we have

$$(v_r, \bar{k}_s) = \delta_{rs} = \delta_{rs}$$

so the $\bar{k}_s$ satisfy the same equations as the k. The uniqueness of the basis dual to $v_1, \cdots, v_r$ then proves that $\bar{k}_s = k_s$, $1 \leq s \leq n$.

10.42* Any vector v can be written

$$v = v_1 + iv_2$$

where v_1 and v_2 are real. Namely

$$v_1 = \frac{1}{2}(v + \bar{v}),$$

$$v_2 = \frac{1}{2i}(v - \bar{v}).$$

10.5* It is shown in Halmos[9], par. 34, that if $v \epsilon \mathbf{V}$, $k \epsilon \mathbf{K}$ are represented by $[v]$, $[k]$ in some coordinate frame, and by $[v]_1$, $[k]_1$ in some other frame, then there is a nonsingular matrix $[W]$, which (a) depends only upon the

* Technical paragraph as explained in Section 2.91.

two frames, and (*b*) relates these n-tuples as follows:

$$\begin{aligned} [v]_1 &= [W]^{-1}[v], \\ [k]_1 &= [W]^*[k]. \end{aligned} \tag{11}$$

It is easy to show that if $[W]$ has real elements, so that $[W]^* = [W]'$, then the two frames involved above are either both real, or else neither is real. Also, conversely, if both frames are real, then necessarily the $[W]$ of (11) has real elements and $[W]^* = [W]'$.

10.6* Some further important geometrical notions must be mentioned before we proceed.

If $\mathbf{V}_1$ and $\mathbf{V}_2$ are disjoint linear manifolds in $\mathbf{V}$—i.e. linear manifolds having in common only the single vector 0—we write

$$\mathbf{V}_1 \oplus \mathbf{V}_2$$

for the linear manifold consisting of all vectors $v = v_1 + v_2$, where $v_i \epsilon \mathbf{V}_i$, $i = 1, 2$. The circle around the plus sign is used to denote the disjointness of $\mathbf{V}_1$ and $\mathbf{V}_2$.

It is shown in Halmos[9], par. 19, that if

$$\mathbf{V} = \mathbf{V}_1 \oplus \mathbf{V}_2 \tag{12}$$

then

$$\mathbf{K} = \mathbf{K}_1 \oplus \mathbf{K}_2, \tag{13}$$

where $\mathbf{K}_1 = (\mathbf{V}_2)^0$, $\mathbf{K}_2 = (\mathbf{V}_1)^0$ and the dimension of $\mathbf{K}_i$ is equal to that of $\mathbf{V}_i$, $i = 1, 2$. We call (13) the decomposition dual to (12). We sometimes write $\mathbf{K}_i = \mathbf{V}_i^*$ to denote the $\mathbf{K}_i$ dual to $\mathbf{V}_i$ in the decomposition (13). It is shown in Halmos[9], loc. cit., that there exists a basis $v_1, \cdots, v_n$ in $\mathbf{V}$ and its dual $k_1, \cdots, k_n$ in $\mathbf{K}$ such that, if r is the dimension of $\mathbf{V}_1$,

$$\begin{aligned} v_1, \cdots, v_r &\quad \text{is a basis for } \mathbf{V}_1 \\ v_{r+1}, \cdots, v_n &\quad \text{is a basis for } \mathbf{V}_2 \\ k_1, \cdots, k_r &\quad \text{is a basis for } \mathbf{K}_1 \\ k_{r+1}, \cdots, k_n &\quad \text{is a basis for } \mathbf{K}_2. \end{aligned} \tag{14}$$

Furthermore, if $v_1, \cdots, v_n$ is any basis in $\mathbf{V}$ satisfying the first half of (14), its dual basis satisfies the second half, and dually.

We shall show in the Appendix that if any one of $\mathbf{V}_1$, $\mathbf{V}_2$, $\mathbf{K}_1$, or

* Technical paragraph as explained in Section 2.91.

$\mathbf{K}_2$ is real, then they all are, and that in this case the bases (14) can be chosen to be real.

Similar considerations apply to decompositions into more summands: if

$$\mathbf{V} = \mathbf{V}_1 \oplus \mathbf{V}_2 \oplus \cdots \oplus \mathbf{V}_m$$

then

$$\mathbf{K} = \mathbf{K}_1 \oplus \mathbf{K}_2 \oplus \cdots \oplus \mathbf{K}_m ,$$

where

$$\mathbf{V}_i^* = \mathbf{K}_i = \bigcap_{j \neq i} \mathbf{V}_j^0 = \left(\sum_{j \neq i} \mathbf{V}_j \right)^0 .$$

XI. GEOMETRICAL CORRESPONDENCES

11.0 With the geometry of $\mathbf{V}$ and $\mathbf{K}$ now in hand, we consider the geometric aspects of our network theoretic concepts.

The definition in Section 4 of general $2n$-pole describes a concrete thing and stands unaltered in our geometric view. The definitions in 6.11 of the terminology typified by "$\mathbf{N}$ admits $[v, k]$ at frequency p" are unchanged except that we should now explicitly indicate that we are discussing concrete n-tuples of complex numbers by placing brackets around the vector symbols, thus: $[v]$, $[k]$. In other words, a $2n$-pole is described by a concrete relation between n-tuples.

11.1* All of the postulates P1, $\cdots$, P7 are stated in a language which now has been given an absolute geometric meaning. In this meaning, P1 and P2 describe a *geometrical linear correspondence* between vectors $v \epsilon \mathbf{V}$ and $k \epsilon \mathbf{K}$. This is the geometric counterpart of the concrete notion of a linear correspondence between n-tuples.

11.11 An impedance matrix, as in 6.3, describes a particularly tightly knit linear correspondence, namely a linear function from $\mathbf{K}$ to $\mathbf{V}$. The geometrical counterpart is an *impedance operator* which for each p is by definition a linear homogeneous function which assigns to each vector $k \epsilon \mathbf{K}$ a unique $v = Z(p)k \epsilon \mathbf{V}$. That is: an operator is a functional relationship between vectors and as such has a geometric identity.

11.12 It is easy to prove† that, given an impedance operator $Z(p)$, and given any coordinate bases in $\mathbf{V}$ and $\mathbf{K}$ respectively, there is a matrix $[Z(p)]$, with elements $Z_{rs}(p)$, $1 \leq r, s \leq n$, such that relative to these bases the coordinates k_s of a vector k and the coordinates v_r of $v = Z(p)k$ are related by (7) of 6.3. We call $[Z(p)]$ the matrix of $Z(p)$

* Technical paragraph as explained in Section 2.91.

† Cf. Halmos[9], par. 26.

relative to the given pair of bases. A strong analog of this observation is contained in the following lemma.

11.13* *Lemma:* (i) Let L be a geometrical linear correspondence. Fix any real coordinate frame and let $[L]$ be the linear correspondence whose paired n-tuples are

$$[[v], [k]],$$

where

$$[v, k] \epsilon L(p).$$

(ii) Alternatively, let $[L]$ be a (concrete) linear correspondence between n-tuples. Interpret the n-tuples related by $[L]$ as representing vectors in some real coordinate frame. Let L be the geometrical correspondence whose pairs, expressed as n-tuples in this frame, are those of the concrete correspondence $[L]$.

In either case, (i) or (ii), the geometric correspondence L satisfies the geometric postulates P1 , $\cdots$, P7 if and only if the concrete correspondence $[L]$ satisfies the concrete forms of these postulates.

The proof of this lemma consists essentially in reading the postulates carefully. We shall not reproduce it.

11.2 Our position is now this: We have on the one hand geometrical objects, vectors v, k, operators $Z(p)$, $Y(p)$, and geometrical correspondences L. On the other hand, we have concrete n-tuples $[v]$, $[k]$, matrices $[Z(p)]$, $[Y(p)]$, and linear correspondences $[L]$. Given any pair of bases in **V** and **K**, in particular, given any coordinate frame, each geometric object generates a corresponding concrete object which represents it relative to those bases or that frame. Conversely, given a concrete object $[\xi]$, we can choose a frame in **V** and **K** and find that geometric object ξ whose coordinates in the chosen frame are given by $[\xi]$.

11.21* The concrete object, linear correspondence, defines a linear time-invariant $2n$-pole by 6.21. To complete the picture, we might say that a geometrical correspondence L defines a *Cauer class* of $2n$-poles.

11.22* This terminology is motivated by the following observation: if $[L]$ and $[L]_1$ are linear correspondences representing L in two distinct real frames, then there exists a real nonsingular matrix $[W]$ relating the

$$[[v], [k]] \epsilon [L](p)$$

and the

$$[[v]_1 , [k]_1] \epsilon [L]_1(p)$$

* Technical paragraph as explained in Section 2.91.

by the formulas of 10.5. This means that $[L]$ and $[L]_1$ are related like the $[L]$ and $[L_W]$ of 9.2. The $2n$-pole associated with $[L]_1$ therefore is a Cauer equivalent of that associated with $[L]$.

11.23 The observation of 11.22, combined with (ii) of 11.13, gives an alternative proof of 9.31. This proof is deceptively free of calculation, but of course the calculations are concealed in the extensive geometrical developments of Section 10, many of which are there offered on faith.

XII. THE FUNDAMENTAL LEMMA

12.0 This section is devoted to the statement, and the proof in part, of a lemma which, on the face of it, looks like an exercise in manipulating the postulates. In fact, the content of the lemma, and most of the details of its proof, are essential in what follows. To postpone them would force us into needless duplication of effort.

Lemma: Let L be a geometrical linear correspondence satisfying P1, P2, P4, P5(I), P6(I), P7(I) and the following weak form of P3(I):

P3′(I): If $p \epsilon \Gamma_L$, then $\mathbf{K}_L(p) = \mathbf{K}_L \supseteq (\mathbf{V}_{L0}(p))^0$.

Then there is a frequency domain $\Gamma'_L \subseteq \Gamma_L$, differing from Γ_L by a finite set, such that L satisfies all of the postulates for $p \epsilon \Gamma'_L$.

The statement of the dual result is evident and will be omitted.

The proof that L satisfies P3 will be given in this section. Verification of the remaining postulates will follow in paragraph 16.6.

We assume that the properties of positive real (PR) functions are known. They are summarized for later use in Section 15. We make occasional advance references thereto.

To the proof:

12.01 First, $\mathbf{K}_L$ is a real manifold and for $p \epsilon \Gamma_L$

$$\mathbf{K}_L \subseteq (\mathbf{V}_{L0}(p))^0. \tag{1}$$

This, with P3′(I), gives P3(I) for L.

Proof: $\mathbf{K}_L$ is real, as in 7.51. Consider now a $p \epsilon \Gamma_L$ and a $v \epsilon \mathbf{V}_{L0}(p)$; then $[v, 0] \epsilon L(p)$. Consider any real $j \epsilon \mathbf{K}_L$; then there is a $u \epsilon \mathbf{V}_L(p)$ such that $[u, j] \epsilon L(p)$. Now 0 and j are real. Hence by P5(I)

$$(v, j) = (u, 0) = 0.$$

Therefore any real $j \epsilon \mathbf{K}_L$ has a vanishing scalar product with every $v \epsilon \mathbf{V}_{L0}(p)$. Since $\mathbf{K}_L$ is real, it is spanned by real j and (1) follows.

12.1 By the dual of 7.7, if we know that

$$[v, k] \epsilon L(p),$$

then the value of (v, k) is determined by k. This makes it possible to state P6(I) and P7(I) for L (we take P6(I) to include the hypothesis that Γ_L has a finite complement).

12.11 If $k \epsilon \mathbf{K}_L$, then $J_k(p)$ is PR.

Proof: if k is real then

$$\overline{J_k(p)} = \overline{(v, k)} = (\bar{v}, k), \tag{2}$$

where, of course, $[v, k] \epsilon L(p)$. Then however $[\bar{v}, k] \epsilon L(\bar{p})$, by P4. Hence by 12.1, (2) gives us

$$\overline{J_k(p)} = J_k(\bar{p}).$$

From this and P6(I), P7(I) we conclude that $J_k(p)$ is PR for any real $k \epsilon \mathbf{K}_L$.

Now, given any $k \epsilon \mathbf{K}_L$, we have $\bar{k} \epsilon \mathbf{K}_L$ by 12.01. Then

$$k = k_1 + ik_2$$

where k_1 and k_2 are real and in $\mathbf{K}_L$, since $\mathbf{K}_L$ is a linear manifold (see 10.42). Let

$$[v_r, k_r] \epsilon L(p),$$

$r = 1, 2$. Then we have (P2)

$$[v_1 + iv_2, k] \epsilon L(p).$$

Then

$$J_k(p) = (v_1, k_1) + (v_2, k_2) + i(v_1, k_2) - i(v_2, k_1).$$

Now by P5(I), $(v_1, k_2) = (v_2, k_1)$. Hence

$$J_k(p) = (v_1, k_1) + (v_2, k_2) \tag{3}$$

for any $p \epsilon \Gamma_L$. Since each summand in (3) is a PR function, it follows that $J_k(p)$ is PR for any $k \epsilon \mathbf{K}$.

12.12 Let $\mathbf{K}_1$ be the set of all vectors $k \epsilon \mathbf{K}_L$ such that

$$J_k(p) = 0 \qquad \text{for every } p \epsilon \Gamma_L.$$

Notice that we do not assert that $\mathbf{K}_1$ is a linear manifold.

If $k \epsilon \mathbf{K}_1$ then $k \epsilon \mathbf{K}_L$ and (3) above applies. Then

$$(v_1, k_1) + (v_2, k_2) = 0$$

and, using this and the PR property of each summand, we conclude that k_1 and k_2 are in $\mathbf{K}_1$.

12.13 We wish now to show that $\mathbf{K}_1 \subseteq \mathbf{K}_{L0}(p)$. Consider a real $j \epsilon \mathbf{K}_L$ and a real $k \epsilon \mathbf{K}_1$. Let

$$\begin{aligned} &[u(p), j] \epsilon L(p), \\ &[v(p), k] \epsilon L(p). \end{aligned} \tag{4}$$

Then, given any real λ, by P2

$$[u(p) + \lambda v(p), j + \lambda k] \epsilon L(p).$$

Then, because $k \epsilon \mathbf{K}_1$,

$$(u + \lambda v, j + \lambda k) = (u, j) + \lambda(v, j) + \lambda(u, k).$$

Since j and k are real, by P5(I) this can be written

$$(u + \lambda v, j + \lambda k) = (u, j) + 2\lambda(v, j). \tag{5}$$

Choose any p_1 such that $\mathrm{Re}(p_1) \geq 0$. Then P7(I) implies that the left side of (5) has a non-negative real part at $p = p_1$. The right side, by suitable choice of λ, can have any chosen real part unless

$$\mathrm{Re}(v(p_1), j) = 0. \tag{6}$$

Hence P7(I) implies (6). Now $(v(p), j)$ is a rational function, by P6(I) applied to the other members of (5). By (6), this rational function has a vanishing real part throughout the right half p-plane. Hence it is an imaginary constant:

$$(v(p), j) \equiv ia. \tag{7}$$

Then

$$\overline{(v(p), j)} = (\overline{v(p)}, j) = -ia. \tag{8}$$

But $[v(p), k] \epsilon L(p)$, so $[\overline{v(p)}, k] \epsilon L(\bar{p})$ by P4. Since also $[v(\bar{p}), k] \epsilon L(\bar{p})$, by 12.1, we have from (8) that

$$(v(\bar{p}), j) = -ia.$$

Comparing this with (7) written for $\bar{p}$, we have $a = 0$ and

$$(v(p), j) = 0 \quad \text{for} \quad p \epsilon \Gamma_L. \tag{9}$$

Now $v(p)$ was determined by (4) wherein k is real. For any $k \epsilon \mathbf{K}_1$, $k = k_1 + ik_2$, where k_1 and k_2 are real and in $\mathbf{K}_1$ (12.11). A corresponding $v(p)$ satisfying (4) can be written

$$v(p) = v_1(p) + iv_2(p), \tag{10}$$

by P2, where $[v_r(p),\ k_r]\epsilon L(p)$, $r = 1, 2$. Then (9) holds for each of $v_1(p)$, $v_2(p)$ and therefore also for the $v(p)$ of (10).

We have showed now that for any $p\epsilon\Gamma_L$ and any $k\epsilon\mathbf{K}_1$, the $v(p)$ of (4) has a vanishing scalar product with every real $j\epsilon\mathbf{K}_L$. Since $\mathbf{K}_L$ is spanned by real j,

$$v(p)\epsilon(\mathbf{K}_L)^0 = \mathbf{V}_{L0}\,. \tag{11}$$

12.14 By (11),

$$[v(p),\ 0]\epsilon L(p).$$

Comparing this with (4), and applying P2,

$$[v(p) - v(p),\ k - 0] = [0,\ k]\epsilon L(p).$$

Since k is now any vector in $\mathbf{K}_1$, we have

$$\mathbf{K}_1 \subseteq \mathbf{K}_{L0}(p) \subseteq \mathbf{K}_L \tag{12}$$

for every $p\epsilon\Gamma_L$.

12.15 We can now also show that $\mathbf{V}_L(p) \subseteq (\mathbf{K}_1)^0$. We return to 12.13 and read (9) thereof as originally derived for real j and k. Applying P5(I), we have from (9) that

$$(u(p),\ k) = 0 \qquad \text{for } p\epsilon\Gamma_L\,. \tag{13}$$

By the argument immediately following (9), (13) also holds for any $k\epsilon\mathbf{K}_1$, provided j is real. As in 12.11 any $j\epsilon\mathbf{K}_L$ can be written $j = j_1 + ij_2$, where j_1 and j_2 are real, and the corresponding

$$u(p) = u_1(p) + iu_2(p)$$

where $[u_r(p),\ j_r]\epsilon L(p)$. Therefore, finally, (13) holds for any $u(p)$ satisfying (4)—i.e., any $u(p)\epsilon\mathbf{V}_L(p)$—and any $k\epsilon\mathbf{K}_1$. Therefore

$$\mathbf{V}_L(p) \subseteq (\mathbf{K}_1)^0 \tag{14}$$

for any $p\epsilon\Gamma_L$.

12.2 We now fix our attention on a specific *real* $p_0\epsilon\Gamma_L$

12.21 By P4, if

$$[v,\ k]\epsilon L(p_0)$$

we have also

$$[\bar{v},\ \bar{k}]\epsilon L(\bar{p}_0) = L(p_0).$$

In particular, $\mathbf{K}_{L0}(p_0)$ is real.

12.22 We can now show that $\mathbf{K}_1$ is a real linear manifold. Consider a real $k \epsilon \mathbf{K}_{L0}(p_0)$. Then $[0, k] \epsilon L(p_0)$ and by 12.1

$$J_k(p_0) = 0.$$

Then by 12.11 (and 15.12), $J_k(p) = 0$, so $k \epsilon \mathbf{K}_1$. Since $\mathbf{K}_{L0}(p_0)$ is spanned by real k (12.21), we have

$$\mathbf{K}_{L0}(p_0) \subseteq \mathbf{K}_1 .$$

Comparing this with (12) gives us

$$\mathbf{K}_{L0}(p_0) = \mathbf{K}_1 . \tag{15}$$

Since $\mathbf{K}_{L0}(p_0)$ is a real linear manifold by definition and 12.21, we see that $\mathbf{K}_1$ is.

12.3 Let us now write, by (12) and (15),

$$\mathbf{K}_L = \mathbf{K}_2 \oplus \mathbf{K}_1 \tag{16}$$

where $\mathbf{K}_2$ is an arbitrary fixed manifold disjoint from $\mathbf{K}_1$ and with it spanning $\mathbf{K}_L$. All three manifolds are real (12.21, (15), 10.6).

Choose a $\mathbf{K}_3$ disjoint from $\mathbf{K}_L$ such that

$$\mathbf{K} = \mathbf{K}_3 \oplus \mathbf{K}_2 \oplus \mathbf{K}_1 . \tag{17}$$

Let the decomposition of $\mathbf{V}$ dual to (17) be (10.6)

$$\mathbf{V} = \mathbf{V}_3 \oplus \mathbf{V}_2 \oplus \mathbf{V}_1 .$$

Then $\mathbf{V}_3 = (\mathbf{K}_2 \oplus \mathbf{K}_1)^0 = (\mathbf{K}_L)^0 = \mathbf{V}_{L0}$ by 12.01. Hence

$$\mathbf{V} = \mathbf{V}_{L0} \oplus \mathbf{V}_2 \oplus \mathbf{V}_1 . \tag{18}$$

By (14) and the definitions,

$$\mathbf{V}_{L0} \subseteq \mathbf{V}_L(p) \subseteq \mathbf{V}_{L0} \oplus \mathbf{V}_2 . \tag{19}$$

12.31 Consider a real p_0. Then by P3′(I), (15) and (16) we have

$$\mathbf{K}_{L0}(p_0) \subseteq \mathbf{K}_L(p_0) \subseteq \mathbf{K}_2 \oplus \mathbf{K}_{L0}(p_0). \tag{20}$$

This is now an expression dual to (19). We shall prove next that, given any $k \epsilon \mathbf{K}_L(p_0) \cap \mathbf{K}_2 (= \mathbf{K}_2)$, there is a unique $v_k \epsilon \mathbf{V}_L(p_0) \cap \mathbf{V}_2$ such that

$$[v_k , k] \epsilon L(p_0). \tag{21}$$

Dually, given any $v \epsilon \mathbf{V}_L(p_0) \cap \mathbf{V}_2$, there is a unique $k_v \epsilon \mathbf{K}_L(p_0) \cap \mathbf{K}_2$ such that

$$[v, k_v] \epsilon L(p_0).$$

The proof is a standard one in algebra and depends only upon P2, (19), and (20).

Proof: Given $k \epsilon \mathbf{K}_L(p_0) \cap \mathbf{K}_2$, there is some $v \epsilon \mathbf{V}_L(p_0)$ such that

$$[v, k] \epsilon L(p_0). \tag{22}$$

By (19), then,

$$v = v_0 + v_2$$

where $v_0 \epsilon \mathbf{V}_{L0}$, $v_2 \epsilon \mathbf{V}_2$. Then

$$[v_0, 0] \epsilon L(p_0)$$

so, applying P2 to this and (22),

$$[v - v_0, k - 0] = [v_2, k] \epsilon L(p_0). \tag{23}$$

Hence $v_2 \epsilon \mathbf{V}_L(p_0) \cap \mathbf{V}_2$ and $v_k = v_2$ satisfies (21). Suppose now $v_3 \epsilon \mathbf{V}_L(p_0) \cap \mathbf{V}_2$ and

$$[v_3, k] \epsilon L(p_0).$$

Then using this with (23) and P2

$$[v_2 - v_3, 0] \epsilon L(p_0).$$

Hence $(v_2 - v_3) \epsilon \mathbf{V}_{L0}$. Now $\mathbf{V}_L(p_0) \cap \mathbf{V}_2$ is a linear manifold and contains v_2, v_3. Hence

$$(v_2 - v_3) \epsilon \mathbf{V}_{L0} \cap \mathbf{V}_L(p_0) \cap \mathbf{V}_2 = 0.$$

Therefore $v_2 = v_3$.

The dual argument completes the proof.

12.32 The argument actually exhibited in 12.31 uses only P2 and (19), hence the v_k of (21) is unique whether or not p_0 is real. Indeed, this is true even when $k \epsilon \mathbf{K}_L$.

12.33 The result of 12.31 establishes a bi-unique linear mapping between $\mathbf{K}_2$ and $\mathbf{V}_L(p_0) \cap \mathbf{V}_2$. Hence these two manifolds are of the same dimension. Since $\mathbf{K}_2$ and $\mathbf{V}_2 = \mathbf{K}_2^*$ are of the same dimension by construction, it follows that

$$\mathbf{V}_L(p_0) \cap \mathbf{V}_2 = \mathbf{V}_2$$

and, by (19), that

$$\mathbf{V}_L(p_0) = \mathbf{V}_{L0} \oplus \mathbf{V}_2.$$

12.4 Let us now introduce a real frame in $\mathbf{V}$ and $\mathbf{K}$ which provides real bases in $\mathbf{K}_1$, $\mathbf{K}_2$, $\mathbf{K}_3$ and in $\mathbf{V}_1$, $\mathbf{V}_2$, $\mathbf{V}_{L0}$ of (17) and (18). Let $k_1, \cdots, k_m$

be the basis vectors spanning $\mathbf{K}_2$. By 12.32, there are unique vectors $u_1(p), \cdots, u_m(p)$ in $\mathbf{V}_2$ such that

$$[u_r(p), k_r] \epsilon L(p).$$

Let $v_1, \cdots, v_m$ be the (real) basis vectors in $\mathbf{V}_2$ dual to the $k_1, \cdots, k_m$:

$$(v_r, k_s) = \delta_{rs} \qquad 1 \leq r \leq s. \tag{24}$$

Since the $u_r(p)$ are all in $\mathbf{V}_2$ we have for each $p \epsilon \Gamma_L$

$$u_s(p) = \sum_{r=1}^{m} a_{rs}(p) v_r \tag{25}$$

where the coefficients $a_{rs}(p)$ are calculated by (24) to be

$$a_{rs}(p) = (u_s(p), k_r). \tag{26}$$

12.41 Because the k_r are real, P5(I) implies that

$$a_{sr}(p) = (u_r(p), k_s) = (u_s(p), k_r) = a_{rs}(p). \tag{27}$$

By the reasoning just following (8) and by the uniqueness of the $u_s(p) \epsilon \mathbf{V}_2$, since $\mathbf{V}_2$ is real, we have $\overline{u_s(\bar{p})} = u_s(\bar{p})$. Then

$$\overline{a_{rs}(p)} = \overline{(u_s(p), k_r)} = (u_s(\bar{p}), k_r) = a_{rs}(\bar{p}).$$

12.42 We have by P2 that

$$[u_r(p) + \lambda u_s(p), k_r + \lambda k_s] \epsilon L(p), \tag{28}$$

for any λ. The identity

$$\begin{aligned}(u_r + u_s, k_r + k_s) - (u_r - u_s, k_r - k_s) \\ = 2(u_r, k_s) + 2(u_s, k_r)\end{aligned} \tag{29}$$

holds in fact for any vectors u_r, u_s, k_r, k_s. Using (27), (28) and P6(I), it exhibits $a_{rs}(p)$ as a rational function.

12.5 Consider the $m \times m$ matrix $[Z_1(p)]$ whose elements are the $a_{rs}(p)$. the s-th column of this matrix consists of the components of $u_s(p)$. The rank of the matrix is by definition the dimension of the space spanned by these columns.

12.51 Now the rank of $[Z_1(p)]$ can be expressed in terms of the vanishing or not of its various minor determinants. There are finitely many such minors and each is a rational function. Each is either identically zero or else vanishes at only finitely many points. Hence the rank of $[Z_1(p)]$, except at these finitely many points, and at the p in the comple-

ment of Γ_L, is a constant. We call this constant the nominal rank of $[Z_1(p)]$.

12.52 Let Γ'_L consist of all $p \epsilon \Gamma_L$ where $[Z_1(p)]$ has its nominal rank. Then Γ'_L has a finite complement. By the reality result of 12.41, if $p \epsilon \Gamma'_L$ then $\bar{p} \epsilon \Gamma'_L$.

It is clear that at any $p \epsilon \Gamma_L$ the rank of $[Z_1(p)]$ does not exceed its nominal rank.

12.53 By construction, the vectors $u_1(p), \cdots, u_m(p)$ all lie in $\mathbf{V}_L(p) \cap \mathbf{V}_2$. By the reasoning of 12.33, at any real $p_0 \epsilon \Gamma_L$ they span $\mathbf{V}_2$. Hence the nominal rank of $[Z_1(p)]$ is m. Therefore, for any $p \epsilon \Gamma'_L$, $[Z_1(p)]$ has rank m and the $u_1(p), \cdots, u_m(p)$, lying in $\mathbf{V}_2$, still span $\mathbf{V}_2$. Therefore for all $p \epsilon \Gamma'_L$

$$\mathbf{V}_L(p) \cap \mathbf{V}_2 = \mathbf{V}_2 .$$

By (19), then,

$$\mathbf{V}_L(p) = \mathbf{V}_{L0} \oplus \mathbf{V}_2 = \mathbf{V}_L , \tag{30}$$

a fixed manifold, for all $p \epsilon \Gamma'_L$.

12.54 It is clear by its construction (cf. Halmos, par. 26) that $[Z_1(p)]$ describes the mapping of 12.32 from $\mathbf{K}_2$ to $\mathbf{V}_2 = \mathbf{V}_L(p) \cap \mathbf{V}_2$ by

$$[v_k] = [Z_1(p)][k].$$

Here the m-tuples $[v_k]$ and $[k]$ are the components of v_k and k relative to the bases now available in $\mathbf{V}_2$ and $\mathbf{K}_2$.

12.55 We repeat

$$\mathbf{K}_1 \subseteq \mathbf{K}_{L0}(p) \subseteq \mathbf{K}_L = \mathbf{K}_1 \oplus \mathbf{K}_2 . \tag{12}$$

Fix a $p \epsilon \Gamma'_L$ and a $k \epsilon \mathbf{K}_{L0}(p) \cap \mathbf{K}_2$. Then $[0, k] \epsilon L(p)$. Since $0 \epsilon \mathbf{V}_2$, it follows from 12.54 that $[Z_1(p)]$ annihilates k. Suppose $m \neq 0$. Since the rank of $[Z_1(p)]$ is m, it follows that $k = 0$. Hence for $p \epsilon \Gamma'_L$

$$\mathbf{K}_{L0}(p) \cap \mathbf{K}_2 = 0.$$

By (12), then, (31)

$$\mathbf{K}_{L0}(p) = \mathbf{K}_1 = \mathbf{K}_{L0} ,$$

a fixed manifold. This, with the result of 12.53, proves that L satisfies P3(A), when $m \neq 0$.

If $m = 0$ then $\mathbf{V}_2 = 0$, $\mathbf{K}_2 = 0$ and (31) follows from (12) and (16).

12.56 $[Z_1(p)]$ is of dimension m and rank m for any $p \epsilon \Gamma'_L$. Therefore

the correspondence of 12.32 and 12.54 between $\mathbf{V}_2$ and $\mathbf{K}_2$ is bi-unique for any $p\epsilon\Gamma'_L$. This extends 12.31 to any $p\epsilon\Gamma'_L$.

12.57 If $m = 0$, i.e., if $\mathbf{V}_2 = \mathbf{K}_2 = 0$, then $\mathbf{V}_{L0} = (\mathbf{K}_{L0})^\gamma$ and the fact that L satisfies all the postulates is trivial because all scalar products (v, k) for $v\epsilon\mathbf{V}_L = \mathbf{V}_{L0}$ and $k\epsilon\mathbf{K}_L = \mathbf{K}_{L0}$ are zero. If $m \neq 0$, we have yet to show that L satisfies P5(A), P6(A), P7(A).

12.6 Since now L satisfies P3, 7.7 as given is applicable and we find (with 12.1) that if $p\epsilon\Gamma'_L$ and

$$[v, k]\epsilon L(p),$$

then (v, k) is fixed by either v or k. Furthermore,

$$(v, k) = (v + v_0 , k + k_0)$$

for any $v_0\epsilon\mathbf{V}_{L0}$, $k_0\epsilon\mathbf{K}_{L0}$.

12.61 If $p\epsilon\Gamma'_L$ and $[v, k]\epsilon L(p)$, then $v\epsilon\mathbf{V}_L$, $k\epsilon\mathbf{K}_L$. By (30), (31), and (16), therefore, there exist $v_0\epsilon\mathbf{V}_{L0}$, $k_0\epsilon\mathbf{K}_{L0}$ such that $u = v - v_0\epsilon\mathbf{V}_2$, $j = (k - k_0)\epsilon\mathbf{K}_2$. Then by P2

$$[u, j]\epsilon L(p). \tag{32}$$

By 12.6, then, any value assumed by a scalar product (v, k) with $[v, k]\epsilon L(p)$ is also assumed by a product (u, j), where (32) holds and $u\epsilon\mathbf{V}_2$, $j\epsilon\mathbf{K}_2$.

XIII. SUFFICIENCY OF THE POSTULATES

13.0 We suppose that L satisfies the postulates of 12.0. Then the results of Section 12 are applicable. The ones of first importance are contained in the facts from (15), (30) and (31), that

$$\mathbf{V}_L = \mathbf{V}_{L0} \oplus \mathbf{V}_2 ,$$

$$\mathbf{K}_L = \mathbf{K}_2 \oplus \mathbf{K}_{L0} ,$$

where the choice of $\mathbf{K}_2$ was governed only by the requirement that the second of these formulae hold.

13.01 Considering $\mathbf{K}_2$ and $\mathbf{V}_2$ as separate spaces, $\mathbf{V}_2 = \mathbf{K}_2^*$ by 10.6. Let M be the geometrical linear correspondence between them with frequency domain Γ'_L and pairs described by 12.31 and 12.56 (or 12.54). That is, as vectors in $\mathbf{V}_2$ and $\mathbf{K}_2$

$$[v, k]\epsilon M(p)$$

if and only if, as vectors in $\mathbf{V}$ and $\mathbf{K}$,

$$[v, k] \epsilon L(p).$$

13.02 In the real frame of 12.4 let us renumber the basis vectors so that

$$\begin{aligned} v_1 , \cdots , v_r \quad & \text{span } \mathbf{V}_{L0} , \\ v_{r+1} , \cdots , v_{r+m} \quad & \text{span } \mathbf{V}_2 , \\ v_{r+m+1} , \cdots , v_n \quad & \text{span } \mathbf{V}_1 . \end{aligned}$$

Then

$$\begin{aligned} k_1 , \cdots , k_{r+m} \quad & \text{span } \mathbf{K}_2 , \\ k_{r+m+1} , \cdots , k_n \quad & \text{span } \mathbf{K}_{L0} . \end{aligned}$$

We say that such a frame *reduces* L.

13.1 Let us now interpret the s-th components of $[v]$ and $[k]$ in this frame respectively as the voltage across and the current in an ideal branch β_s of a $2n$-pole $\mathbf{N}$, $1 \leq s \leq n$.

By construction, the vectors $v \epsilon \mathbf{V}_L$ in this frame have components $a_{r+m+1} = \cdots = a_n = 0$, since $v_1 , \cdots , v_{r+m}$ span $\mathbf{V}_L$. At the same time, the components $b_{r+m+1} , \cdots , b_n$ of $[k]$ may be chosen arbitrarily without altering the fact that $[[v], [k]] \epsilon [L](p)$ because of 12.06. Therefore, the ideal branches $\beta_{r+m+1} , \cdots , \beta_n$ can each be realized physically by a short circuit.

In a dual way, since $k_{r+1} , \cdots , k_n$ span $\mathbf{K}_L$, any $k \epsilon \mathbf{K}_L$ has components $b_1 , \cdots , b_r$ all zero in our chosen frame. Furthermore, the components $a_1 , \cdots , a_r$ of $[v]$ can be chosen at will. Hence the ideal branches $\beta_1 , \cdots , \beta_r$ can each be realized physically by an open circuit.

Let $\mathbf{N}_1$ now be the $2m$-pole whose ideal branches are $\beta_{r+1} , \cdots , \beta_{r+m}$. Let the pairs $[[v], [k]]$ admitted by $\mathbf{N}_1$ at each $p \epsilon \Gamma'_L$ be the $[[v], [k]]$, where $[v, k] \epsilon M(p)$ (13.01). The representation just found for $\mathbf{N}$ shows that $\mathbf{N}$ is physically realizable if and only if $\mathbf{N}_1$ is.

13.11 The matrix $[Z_1(p)]$ of 12.54 is the impedance matrix of the $2m$-pole $\mathbf{N}_1$.

13.12 We now show that $[Z_1(p)]$ is a positive real matrix. The displayed formulae of 12.41 show (ii) and (iii) of 1.1, and 12.42 shows (i). Now suppose that $[v, k] \epsilon M(p)$. Then, as vectors in $\mathbf{V}$ and $\mathbf{K}$, $[v, k] \epsilon L(p)$ by definition of $M(p)$. Then, however, if k is fixed

$$J_k(p) = (v, k)$$

is a PR function (12.11). Regarding v and k in $\mathbf{V}_2$ and $\mathbf{K}_2$ let

$$[b_{r+1}, \cdots, b_{r+m}] = [k].$$

Then by (1) of 7.0

$$(v, k) = \sum_{t=1}^{m} \sum_{s=1}^{m} a_{st}(p) b_{t+r} \bar{b}_{s+r}$$

and this has a non-negative real part if $\mathrm{Re}(p) > 0$. This is (iv) of 1.1.

13.2 We can now prove the lemmas 8.1 and 8.2. Given a linear correspondence $[L]$ which satisfies P1, $\cdots$, P7 by 11.13 we can interpret $[L]$ as the concrete correspondence representing a geometrical correspondence L in some chosen real frame, and L satisfies P1, $\cdots$, P7. Then by the results in 13.01–13.12 there exists a real frame in which the representative $[L]_1$ of L has the properties claimed in 8.1 and 8.2 for L_W. But we saw in 11.22 that $[L]$ and $[L]_1$ are related by a real matrix W like the L and L_W of Section 8. Q.E.D.

13.21 With the proofs of 8.1 and 8.2 we have reduced the sufficiency claimed for P1, $\cdots$, P7 in 8.0 to the sufficiency of positive reality of $[Z(p)]$ claimed in 1.1, by the argument outlined in 8.5.

XIV. OPERATOR-VALUED FUNCTIONS OF p

The next three sections are directed principally toward the proof of the matrix theorem of 1.1. They do however, contribute to 12.10 and to the necessity proof.

14.0 We continue to use the geometric language. The reader who regards this as unduly pedantic is free to place a concrete interpretation upon every argument, for all of the arguments are either frankly based on matrix representations or upon the three identities:

14.01 $(Zj, k) = \overline{(Z^*k, j)}$ for all j, $k \epsilon \mathbf{K}$.

1402 $\overline{Z}\overline{k} = \overline{(Zk)}$ for all $k \epsilon \mathbf{K}$.

14.03 $Z' = (\overline{Z})^* = \overline{(Z^*)}$

14.04 These identities are obvious for matrices using 7.0 and 7.2. Geometrically, the first and second define Z^* and $\overline{Z}$, and the third defines Z' in two ways. The equivalence of these two ways is a theorem based on (10) of 10.33.

14.05 The symbol Z will always denote an impedance (operator, matrix, scalar), and Y will always denote an admittance. An impedance oper-

ates from **K** to **V**, an admittance dually. The operators in Halmos[9] are physically dimensionless, in that they operate, e.g., from **V** to **V**. This difference is scarcely noticeable.

We shall regularly omit the duals to concepts or proofs given in terms of impedances. In doing so, we adopt the rule that the dual to an expression

$$(Zk, k)$$

is

$$\overline{(v, Yv)}.$$

14.1 An operator is called symmetric if $Z = Z'$. Such operators have three useful special properties:

14.11 If Z is symmetric and j and k are real, then

$$(Zj, k) = \overline{(\bar{Z}j, k)} = ((\bar{Z})^*k, j) = (Z'k, j) = (Zk, j)$$

by (10) of 10.33, 14.02, 14.01, 14.03, and hypothesis.

14.12 Let $k = k_1 + ik_2$, where k_1 and k_2 are real (10.42). If Z is symmetric then

$$(Zk, k) = (Zk_1, k_1) + (Zk_2, k_2),$$

for, by 14.11,

$$\begin{aligned}(Zk_1, ik_2) = -i(Zk_1, k_2) &= -i(Zk_2, k_1)\\ &= -(Z(ik_2), k_1).\end{aligned}$$

(Cf. the similar identity in 12.11.)

14.13 The symmetric operator Z is completely defined by the quadratic form

$$(Zk, k) \tag{1}$$

as a function of *real* $k \epsilon \mathbf{K}$. For 14.11 permits the formula (29) of 12.42 in any real frame, where $u_s = Zk_s$. The matrix elements of $[Z(p)]$ in that frame are then defined by that formula in terms of values of (1) for real k.

The form (1) specifies any Z (symmetric or not) if k is allowed to range over all of **K** (Halmos[9], par. 53).

14.2 Let $Z(p)$ now be an impedance operator depending on p. We say that $p_0 \neq \infty$ is a pole of order m of $Z(p)$ if

$$\ell(k) = \lim_{p \to p_0} (p - p_0)^m (Z(p)k, k) \tag{2}$$

exists for every $k \epsilon \mathbf{K}$ and is not identically zero. By 15.13, this limit $\ell(k)$ defines an operator R_0, the residue* of $Z(p)$ at p_0, by

$$(R_0 k, k) = \ell(k) \quad \text{for} \quad k \epsilon \mathbf{K}.$$

The changes in (2) required to define a pole at $p = \infty$ are obvious.

14.21 A pole p_0 of order m of $Z(p)$ is a pole of some matrix element of $[Z(p)]$, of order m, in any frame, and no element of $[Z(p)]$ has a pole at p_0 of order exceeding m. For the elements of $[Z(p)]$ are defined by the values of $(Z(p)k, k)$, by 14.11 and Halmos[9] loc. cit.

XV. POSITIVE REAL FUNCTIONS

15.0 Let $f(p)$ be a scalar function of the complex variable p. Following Brune[2] we define $f(p)$ to be positive real if

(i) $f(p)$ is a rational function of p,

(ii) $\overline{f(p)} = f(\bar{p})$,

(iii) $\mathrm{Re}(p) > 0$ implies $\mathrm{Re}(f(p)) \geq 0$.

The property (i) of being rational is of course on a quite different level of ideas from the other properties, but it saves words later to include it specifically in the meaning of positive real.

We abbreviate the words positive real to PR.

15.01 The open region of the complex plane consisting of all finite p such that $\mathrm{Re}(p) > 0$—the right half plane—we denote by Γ_+.

15.1 Brune, loc. cit., established a number of properties of PR functions $f(p)$ which will be useful to us here:

15.11 $f(p)$ has no poles in Γ_+.

15.12 If $\mathrm{Re}(f(p)) = 0$ for some $p \epsilon \Gamma_+$, then $f(p) \equiv 0$ for all p.

15.13 If it exists, $\dfrac{1}{f(p)}$ is PR.

15.14 If $f(p)$ has a pole at $p = p_0$, it has one at $p = \overline{p_0}$.

15.15 If $f(p)$ has a pole at $p = i\omega_0$, that pole is simple and

$$f(p) = \frac{2p}{p^2 + \omega_0^2} r + f_1(p),$$

where $r > 0$, and $f_1(p)$ is PR.

* Properly, R_0 is a residue only when $m = 1$. There is no convenient name available for general m.

15.16 If $f(p)$ has a pole at $p = \infty$, that pole is simple and

$$f(p) = pr + f_1(p),$$

where $r > 0$, and $f_1(p)$ is PR.

15.17 We shall use all of these in the next section, save 15.13. Our aim is to prove properties analogous to 15.11, $\cdots$, 15.16 for PR matrices and operators.

The reader familiar with the Brune process[2] for realization of a 2-pole will remember the importance of the properties 15.11, $\cdots$, 15.16 for the success of that process. Correspondingly, we must establish the analogs of these properties to implement the general Brune process for $2n$-poles.

XVI. POSITIVE REAL OPERATORS

16.0 An operator $Z(p)$ from $\mathbf{K}$ to $\mathbf{V}$ will be called positive real (PR) if in some real coordinate frame the matrix $[Z(p)]$ is a PR matrix in the sense of 1.1—that is

(i) $[Z(p)]$ has rational elements $Z_{rs}(p)$

(ii) $\overline{Z_{rs}(p)} = Z_{rs}(\bar{p})$

(iii) $Z_{rs}(p) = Z_{sr}(p)$

(iv) For any real $k \epsilon \mathbf{K}$ and any $p \epsilon \Gamma_+$

$$\mathrm{Re}(Z(p)k, k) \geq 0.$$

We intend in this section to establish for PR operators the properties listed below. By subtracting 0.9 from the designation of each property one obtains the designation of the analogous property of a PR scalar function, stated earlier.

16.01 $Z(p)$ has no poles in Γ_+.

16.02 If $\mathrm{Re}(Z(p)k, k) = 0$ for some $p \epsilon \Gamma_+$, then $Z(p)k \equiv 0$ for all p.

16.03 If it exists, $Z^{-1}(p) = Y(p)$ is PR.

16.04 If $Z(p)$ has a pole at $p = p_0$, it has one at $p = \bar{p}_0$.

16.05 If $Z(p)$ has a pole at $p = i\omega_0$, that pole is simple* and

$$Z(p) = \frac{2p}{p^2 + \omega_0^2} R + Z_1(p),$$

where R is real, symmetric, and semi-definite, not zero, and $Z_1(p)$ is PR.

* i.e., of order one.

16.06 If $Z(p)$ has a pole at $p = \infty$, that pole is simple and

$$Z(p) = pR + Z_1(p)$$

where $R = R' = \bar{R}$, $R \geq 0$ and $Z_1(p)$ is PR.

16.07 There is property of rational scalar functions $f(p)$, whether PR or not, that is essential in the Brune theory: the existence of a finite integer, the degree of f. Each step in the Brune reduction of $f(p)$ leaves an unreduced portion which is of lower degree than the function upon which the step was performed. The finiteness of the original degree of f then guarantees the termination of the process in finitely many steps.

There exists also for rational matrices (and operators) a concept of degree. This degree plays the same role in the general Brune process for $2n$-poles as the degree of a scalar function does in the process for 2-poles. To define this degree and develop its properties requires an excursion into classical algebra. Since we shall not need these ideas until Part II we defer further discussion of them to that part.

16.1 If $Z(p)$ is PR it follows at once that the matrix $[Z(p)]$ is PR in any real frame.

Proof: Two such matrices are related by

$$[Z(p)]_1 = [U][Z(p)][U]'$$

where U is real, by 11.22 and the argument in 8.6. The PR properties of $[Z(p)]$ are obviously preserved by this operation.

16.11 If $Z(p)$ is PR, then

$$Z(p) = Z'(p) = \overline{Z^*(p)} = \overline{Z(\bar{p})}.$$

Proof: Use 16.0 and 14.03 in a real frame.

16.12 If $Z(p)$ is PR, then for any given $k \epsilon \mathbf{K}$ the function

$$J_k(p) = (Z(p)k, k)$$

is a PR scalar function. It follows that the limitation in (iv) of 16.0 to real k is a simplification, not a restriction.

Proof: $J_k(p)$ is independent of coordinate representation. By use of a real frame, (i) of 16.0 implies (i) of 15.0.

By 14.01 and 16.11

$$\overline{J_k(p)} = (Z^*(p)k, k) = (Z(\bar{p})k, k) = J_k(\bar{p}).$$

This is (ii) of 15.0. For any k, 14.12 and (iv) of 16.0 imply (iii) of 15.0.

16.13 Conversely to 16.12, if $Z(p)$ is symmetric and $J_k(p)$ is PR for every real k, then $Z(p)$ is PR, and $J_k(p)$ is PR for all k.

Proof: $J_k(p)$ is rational so (i) of 16.0 holds in any frame by 14.13. Clearly (iv) of 16.0 holds.

Now for real k, by (10) of 10.33 and 14.02

$$J_k(\bar{p}) = \overline{J_k(p)} = \overline{(Z(p)k, k)}.$$

Hence $Z(\bar{p}) = \overline{Z(p)}$ by 14.13. This is (ii) of 16.0, and (iii) there holds by hypothesis.

16.2 *Proof of* 16.01: By 15.11 and 16.12, $J_k(p)$ has no poles in Γ_+. This is 16.01 by the definition 14.3 of pole.

16.21 *Corollary:* Any PR $Z(p)$ can be considered as defined throughout Γ_+: for any k, $J_k(p)$ is defined throughout Γ_+ by 16.2. For each p, as a function of k, $J_k(p)$ defines $Z(p)$ (14.13).

16.3 *Proof of* 16.03: In any frame $[Z^{-1}(p)] = [Z(p)]^{-1} = [Y(p)]$ consists of rational elements, by direct calculation of the inverse matrix. In a real frame $[Y(p)] = [Z^{-1}(p)]$ is symmetric and real for real p by the same argument (both facts are also deducible geometrically). Hence we have the duals of (i), (ii) and (iii) of 16.0 for $Y(p)$. Clearly $Y(p)$ is defined throughout Γ_+.

Now suppose that for some $v \epsilon \mathbf{V}$ and some $p_0 \epsilon \Gamma_+$ we have

$$\mathrm{Re}\overline{(v, Y(p_0)v)} < 0.$$

Then there is a $k \epsilon \mathbf{K}$ such that $v = Z(p_0)k$. Therefore

$$\mathrm{Re}\overline{(Z(p_0)k, k)} = \mathrm{Re}(Z(p_0)k, k) < 0.$$

Since this is impossible, we have the dual of (iv) of 16.0 for $Y(p)$ and $Y(p)$ is PR.

16.4 *Proof of* 16.04: This is immediate from 15.14, 14.3, and 16.12.

16.5 *Proofs of* 16.05 *and* 16.06: Suppose $Z(p)$ has a pole at $p = i\omega_0$. Then $(Z(p)k, k)$ does and that pole is simple by 15.15 and 16.12. Then by 14.3 we can write

$$Z(p) = \frac{1}{p - i\omega_0} R_0 + Z_0(p)$$

where $Z_0(p)$ is regular at $p = i\omega_0$. Now $Z_0(p)$ has a pole at $p = -i\omega_0$ by 16.5, so a similar argument gives

$$Z(p) = \frac{1}{p - i\omega_0} R_0 + \frac{1}{p + i\omega_0} R_1 + Z_1(p), \tag{1}$$

where $Z_1(p)$ has no pole at $i\omega_0$ or $-i\omega_0$. The symmetry of Z and linear independence of the terms above then imply the symmetry of R_0, R_1 and $Z_1(p)$.

For any $k\epsilon\mathbf{K}$, now,

$$(Z(p)k, k) = \frac{1}{p - i\omega_0}(R_0k, k) + \frac{1}{p + i\omega_0}(R_1k, k) + (Z_1(p)k, k).$$

Applying 16.12 and 15.15,

$$(R_0k, k) = (R_1k, k) \geq 0$$

for all k. Hence $R_0 = R_1 = R$ (say) and R is semi-definite. Also, $(Z_1(p)k, k)$ appears as the residue $f_1(p)$ in 15.15 and is therefore PR. Then $Z_1(p)$ is PR by 16.13. With R_0 and R_1 identified, (1) above is the expansion given in 16.05. We have now proved all of 16.05 save the reality of R. But

$$\frac{2p}{p^2 + \omega_0^2} R$$

is PR, by 16.13, hence is real for real p. Therefore R is real.

The proof of 16.06 is similar.

16.6 To prove 16.02 we appear to digress somewhat, by first completing the proof of the fundamental lemma of 12.0. It was established in Section 13 that the matrix $[Z_1(p)]$ describing $M(p)$ in the chosen basis is PR. The case in which it is nonsingular (i.e., $m \neq 0$, cf. 12.56, 12.57) remains to be examined.

16.61 If $[Z_1(p)]$ is nonsingular then its inverse is PR (16.3). Then for any $v\epsilon\mathbf{V}_2$,

$$\overline{(v, k)} = \overline{(v, Y(p)v)} \tag{2}$$

is PR (16.12 dual). By 12.61, for any $u\epsilon\mathbf{V}_L$, the values of the function $F_u(p)$ are the values of (2) for some $v\epsilon\mathbf{V}_2$. Hence $F_u(p)$ is PR. This is P6(A) and P7(A) for L.

16.62 To settle P5 for L in 12.0, consider $p\epsilon\Gamma'_L$ and

$$[v, k]\epsilon L(p), \qquad [u, j]\epsilon L(p),$$

where u and v are real. Then, say,

$$v = v_0 + v_1,$$

where $v_0\epsilon V_{L0}$, $v_1\epsilon V_2$. But then

$$v = \bar{v} = \bar{v}_0 + \bar{v}_1$$

and, because $\mathbf{V}_{L0}$ and $\mathbf{V}_2$ are real, $\bar{v}_0 = v_0$, $\bar{v}_1 = v_1$, and these vectors are real. Using similar reasoning for u,

$$(v, j) = (v_1, Y(p)u_1), \qquad (u, k) = (u_1, Y(p)v_1), \tag{3}$$

by 12.61. The equality $(u, k) = (v, j)$ now follows from (3) and the duals of 16.11, 14.11. Hence we have P5(A) for L and 12.0 is proved.

16.7 We now prove an important

Lemma: Let $Z(p)$ be a PR operator from $\mathbf{K}$ to $\mathbf{V}$. Let Γ_L be the set of p where $Z(p)$ is defined and has a rank equal to its nominal rank. Let L be the correspondence with domain Γ_L and pairs

$$[Z(p)k, k], \qquad k\epsilon\mathbf{K}_L .$$

Then L satisfies P1, $\cdots$, P7.

Proof: L satisfies P1 and P2 (6.3). Γ_L satisfies P4 by the argument of 12.52. Then L satisfies P4, for by 16.11

$$\overline{Z(p)k} = Z(\bar{p})\bar{k}.$$

L satisfies P5(I) by 14.11 and 16.11. Γ_L satisfies P6 by 12.52. Then L satisfies P6(I) and P7(I) by 16.12. The fundamental lemma, 12.0, now proves that L satisfies all the postulates.

16.71 We call a correspondence satisfying all the postulates PR.

16.72 *Proof of* 16.02: Suppose $\mathrm{Re}(Z(p_0)k, k) = 0$ for some $p_0\epsilon\Gamma_+$. Because this function of p is PR (16.12) we have

$$J_k(p) = (Z(p)k, k) \equiv 0.$$

Hence $k\epsilon\mathbf{K}_1 = \mathbf{K}_{L0}$ (12.12, 12.55). Hence $[0, k]\epsilon L(p)$ for every $p\epsilon\Gamma_L$. That is

$$Z(p)k = 0 \quad \text{for} \quad p\epsilon\Gamma_L .$$

16.73 *Corollary:* If $Z(p_0)k = 0$ for some $p_0\epsilon\Gamma_+$, then $Z(p)k \equiv 0$. For the hypothesis here implies that of 16.72. This is the analog of 15.12; the result of 16.02 is stronger.

16.8 An important consequence of 16.7 is the

*Lemma**: If $Z(p)$ is PR and of rank m, then there exists a real coordinate frame in which the matrix $[Z(p)]$ is an $m \times m$ nonsingular PR matrix $[Z_1(p)]$ bordered by zeros.

* Proved by Cauer[5].

Proof: Consider the PR correspondence L defined by $Z(p)$. Then $\mathbf{V}_{L0} = 0$, because $Z(p)0 = 0$ for every $p \epsilon \Gamma_L$. Consider the real frame of 13.02. $[Z(p)]$ in this frame takes any of $k_{r+m+1}, \cdots, k_n$ into 0 because these span $\mathbf{K}_{L0}$. Within $\mathbf{K}_2$, $[Z(p)]$ must describe the same operation as the $[Z_1(p)]$ of 12.54. Because $[Z(p)]$ is symmetric the lemma follows.

XVII. THE JUXTAPOSITION OF CORRESPONDENCES

17.0 This section and the next will consider ways of constructing new correspondences from old. This will provide the basis of the necessity proof of Section 19.

17.01 It is obvious that if two physical networks are set side by side and their accessible terminals regarded as the terminals of a single larger network, that enlarged network is again a physical network. This is the gist of the present section.

17.1 Suppose that

$$\mathbf{V} = \mathbf{V}_1 \oplus \mathbf{V}_2, \qquad \mathbf{K} = \mathbf{K}_1 \oplus \mathbf{K}_2,$$

where $\mathbf{K}_i = \mathbf{V}_i^*$ and all spaces are real (10.6). Let E_1 project on $\mathbf{V}$ along $\mathbf{V}_2$ (Halmos[9], par. 33) and $E_2 = 1 - E_1$ project on $\mathbf{V}_2$ along $\mathbf{V}_1$. Then E_i^* projects on $\mathbf{K}_i$ along $\mathbf{K}_j$, $j \neq i$ (Halmos[9], loc. cit.). It is easily verified that $E_i = \bar{E}_i$, $E_i^* = \bar{E}_i^*$, from the analog of 14.02 for dimensionless operators.

Considering $\mathbf{V}_i$ and $\mathbf{K}_i$ as separate spaces, let L_i be a geometrical linear correspondence between them with frequency domain Γ_i, $i = 1, 2$.

Consider the correspondence L between $\mathbf{V}$ and $\mathbf{K}$ defined by

(i) The frequency domain $\Gamma_L = \Gamma_1 \cap \Gamma_2$

(ii) $[v, k] \epsilon L(p)$ if and only if $[E_i v, E_i^* k] \epsilon L_i(p)$, $\quad i = 1, 2.$

In (ii), of course, we regard $E_i v$ and $E_i^* k$ as elements of $\mathbf{V}_i$, $\mathbf{K}_i$.

17.11 L so defined is called the juxtaposition of L_1 and L_2.

17.2 *Lemma:* L is PR if and only if each of L_1 and L_2 is PR.

17.21 *Proof of* "if": It is clear that L satisfies P1 and P2. Further notation is now simplified if we put $L_1 = M$, $L_2 = N$. Consider the manifolds

$$\mathbf{V}_M \oplus \mathbf{V}_N, \qquad \mathbf{V}_{M0} \oplus \mathbf{V}_{N0}, \qquad \mathbf{K}_M \oplus \mathbf{K}_N, \qquad \mathbf{K}_{M0} \oplus \mathbf{K}_{N0},$$

where $\mathbf{V}_M \subseteq \mathbf{V}_1$ is the manifold of voltages admitted by $L_1 = M$ considered as a correspondence between $\mathbf{V}_1$ and $\mathbf{K}_1$, and $\mathbf{V}_{M0}$ the manifold

of voltages $v \epsilon \mathbf{V}_1$ such that $[v, 0] \epsilon L_1(p)$ for all $p \epsilon \Gamma_1$. Dual definitions for $\mathbf{K}_M$, $\mathbf{K}_{M0}$, and symmetrical ones for $\mathbf{V}_N$, $\cdots$, $\mathbf{K}_{N0}$ need not be repeated.

It is clear from these definitions that the four manifolds above are, in the order listed, the manifolds

$$\mathbf{V}_L, \mathbf{V}_{L0}, \mathbf{K}_L, \mathbf{K}_{L0}$$

for L. Now, for example,

$$(\mathbf{K}_{L0})^0 = (\mathbf{K}_{M0} \oplus \mathbf{K}_{N0})^0 = (\mathbf{K}_{M0})^0 \cap (\mathbf{K}_{N0})^0$$

by 10.6. This last manifold, *in* $\mathbf{V}$, is $(\mathbf{V}_M \oplus \mathbf{V}_2) \cap (\mathbf{V}_N \oplus \mathbf{V}_1)$, byP3 for M and N, and by 10.6. But by direct calculation

$$(\mathbf{V}_M \oplus \mathbf{V}_2) \cap (\mathbf{V}_N \oplus \mathbf{V}_1) = \mathbf{V}_M \oplus \mathbf{V}_N = \mathbf{V}_L .$$

The dual of this result then completes P3 for L.

P4 for L is immediate because the E_i and E_i^* are real.

The duality of the decompositions of $\mathbf{V}$ and $\mathbf{K}$ implies the identity

$$(v, k) = (E_1 v, E_1^* k) + (E_2 v, E_2^* k)$$

(that is $E_1 E_2 = E_2 E_1 = 0$, and dually. This is Halmos[9], par. 33). All of P5, P6, and P7 for L follow at once from this identity.

17.22 The "only if" of 17.2 is a special case of the result of Section 18. Its proof will be deferred to 18.4.

17.23 It is obvious that the notion of juxtaposition and the lemma of 17.2 extend to juxtapositions of more than two correspondences.

17.3 Even without the "only if" part of 17.2, we have enough for the following characterization of PR correspondences:

Theorem: A correspondence L is PR if and only if it is the juxtaposition of

(i) a correspondence defined by a nonsingular PR matrix between a $\mathbf{V}_1$ and a $\mathbf{K}_1 = \mathbf{V}_1^*$,

(ii) a correspondence consisting of short circuits: that is of pairs $[0, k]$ for all $k \epsilon \mathbf{K}_2$ and all p,

(iii) a correspondence consisting of open circuits: that is, of pairs $[v, 0]$ for all $v \epsilon \mathbf{V}_3$ and all p.

Proof: If L is PR, the decomposition indicated is that of 13.1, 13.11, 13.12. If L is the juxtaposition indicated, then it is PR by 16.6 and the "if" in 17.1, provided the short and open circuits are PR correspondences. The verification of the postulates for these latter is easy and will be omitted.

17.31 The labor of considering PR correspondences instead of matrices has yielded the disappointingly simple result of 17.3. We have already been warned of this, however, by our knowledge of the properties of physical networks (2.9).

XVIII. THE OPERATION OF RESTRICTION

18.0 In addition to juxtaposition, which is an operation on correspondences clearly motivated by physical considerations, there is an operation, here called restriction, which has important use in the next section. There the physical meaning of the operation will become clear.

18.1 Let $\mathbf{V}$ and $\mathbf{K} = \mathbf{V}^*$ be a pair of dual spaces. Let $\mathbf{U}$ and $\mathbf{J} = \mathbf{U}^*$ be another pair. Suppose that C is a given fixed linear operation from $\mathbf{J}$ to $\mathbf{K}$: given any $j \epsilon \mathbf{J}$, there is a unique $k(j) \epsilon \mathbf{K}$, written

$$k(j) = Cj,$$

such that if $k_r = Cj_r$, $r = 1, 2$, then

$$a_1 k_1 + a_2 k_2 = C(a_1 j_1 + a_2 j_2)$$

for any complex scalars a_1, a_2.

18.11 Let $(v, k)_1$ denote the scalar product between $\mathbf{V}$ and $\mathbf{K}$, and $(u, j)_2$ that between $\mathbf{U}$ and $\mathbf{J}$. Given C, and any $v \epsilon \mathbf{V}$, let us find that unique vector $u(v) \epsilon \mathbf{U}$ for which

$$(u(v), j)_2 = (v, Cj)_1 \tag{1}$$

for every $j \epsilon \mathbf{J}$. That such a vector $u(v)$ exists and is unique follows from 10.13 when we notice that the right-hand side of (1) defines a function conjugate linear in j. Now for fixed j, the right-hand side of (1) is linear in v, hence so also is the left side. That is, there is a linear operation C^* from $\mathbf{V}$ to $\mathbf{U}$ such that

$$u(v) = C^* v.$$

The following chart illustrates the situation:

$$\begin{array}{ccc} \mathbf{V} & & \mathbf{K} \\ {\scriptstyle C^*}\downarrow & & \uparrow{\scriptstyle C} \\ \mathbf{U} & & \mathbf{J} \end{array}$$

18.12 We suppose now that C takes real j into real k, i.e., that C is real. Then by (1)

$$\overline{(C^* v, j)_2} = (C^* v, j)_2 = \overline{(v, Cj)_1} = (\bar{v}, Cj)_1 .$$

By comparison with (1), we have

$$\overline{C^*v} = C^*\bar{v}.$$

Hence C^* also takes real vectors into real vectors and is real.

18.2 Now let L be a PR correspondence between **V** and **K**. We define one, say M, between **U** and **J**, as follows: For each $p\epsilon\Gamma_L$, let $M(p)$ consist of all pairs

$$[u, j]$$

such that $u = C^*v$ and

$$[v, Cj]\epsilon L(p).$$

This definition can be illustrated by enlarging the chart of 18.11:

$$\begin{array}{ccc} \mathbf{V} & \xleftarrow{L} & \mathbf{K} \\ C^* \downarrow & & \uparrow C \\ \mathbf{U} & \xleftarrow[M]{} & \mathbf{J} \end{array}$$

The u's corresponding to $j\epsilon\mathbf{J}$ can be constructed by going around through C, L and C^*. This then defines a direct mapping from **J** to **U**.

18.21 We call the M defined by 18.2 a restriction of L, since its pairs are images under C^* and C^{-1} (which is not defined over all of **K**) of a restricted set of pairs drawn from L.

18.22 Clearly there is a dual operation defined by an operator D from **U** to **V**. We might distinguish the operation of 18.2 by calling it a current restriction, its dual by calling it a voltage restriction.

18.23 The restriction M of L is defined by lists $M(p)$ which exist for any $p\epsilon\Gamma_L$. The frequency domain of M has not yet been specified, however.

18.3 *Theorem:* If L is PR, then there is a frequency domain Γ_M for M such that M is PR.

Proof: P1 and P2 for M are evident at once, for any $p\epsilon\Gamma_L$. The remainder of the proof is divided among 18.31, $\cdots$, 18.37 below.

18.31 For P3, let $\mathbf{J}_M$ be all $j\epsilon\mathbf{J}$ such that $Cj\epsilon\mathbf{K}_L$. Then, given $j\epsilon\mathbf{J}_M$, for each $p\epsilon\Gamma_L$ there is a v such that

$$[v, Cj]\epsilon L(p),$$

whence

$$[C^*v, j]\epsilon M(p).$$

Therefore $\mathbf{J}_M(p)$, the space of currents admitted by M at frequency p, coincides with the fixed $\mathbf{J}_M$ at each $p\epsilon\Gamma_L$.

Clearly $\mathbf{J}_M$ is a real linear manifold.

18.32 Consider now $\mathbf{U}_{M0}(p)$: if $[u, 0]\epsilon M(p)$, then there is a v such that $u = C^*v$ and

$$[v, C0] = [v, 0]\epsilon L(p).$$

Hence $v\epsilon\mathbf{V}_{L0}(p) = \mathbf{V}_{L0}$ for each $p\epsilon\Gamma_L$. Therefore, for each $p\epsilon\Gamma_L$,

$$\mathbf{U}_{M0}(p) \subseteq C^*\mathbf{V}_{L0}. \tag{2}$$

Now suppose, conversely, that $p\epsilon\Gamma_L$ and $v\epsilon\mathbf{V}_{L0} = \mathbf{V}_{L0}(p)$. Then $[v, 0]\epsilon L(p)$. Now $0 = C0$, so $[v, C0]\epsilon L(p)$. Hence $[C^*v, 0]\epsilon M(p)$, so $C^*v\epsilon\mathbf{U}_{M0}(p)$. This proves the inequality opposite to that of (2), so for $p\epsilon\Gamma_L$

$$\mathbf{U}_{M0}(p) = C^*\mathbf{V}_{L0} = \mathbf{U}_{M0}, \tag{3}$$

a fixed space.

18.33 Now consider $(\mathbf{U}_{M0})^0$. If $j\epsilon(\mathbf{U}_{M0})^0$, then

$$(u, j)_2 = 0$$

for every $u\epsilon\mathbf{U}_{M0}$. That is, by (3),

$$(C^*v, j)_2 = (v, Cj)_1 = 0$$

for every $v\epsilon\mathbf{V}_{L0}$. Therefore $Cj\epsilon(\mathbf{V}_{L0})^0 = \mathbf{K}_L$, and $j\epsilon\mathbf{J}_M$ by 18.31. That is, we have proved

$$\mathbf{J}_M \supseteq (\mathbf{U}_{M0})^0,$$

and, combining 18.31 with this and (3),

$$\mathbf{J}_M(p) = \mathbf{J}_M \supseteq (\mathbf{U}_{M0}(p))^0 = (\mathbf{U}_{M0})^0. \tag{4}$$

This is the weak form P3′(I) of 12.0 for M. It is as far as we can go with P3 at the moment.

18.34 Consider P4. If for $p\epsilon\Gamma_L$ we have

$$[u, j]\epsilon M(p)$$

then $[v, Cj]\epsilon L(p)$ and $u = C^*v$. But then $[\bar{v}, C\bar{j}]\epsilon L(\bar{p})$ and $\bar{u} = C^*\bar{v}$, by 18.12. Then however

$$[\bar{u}, \bar{j}]\epsilon M(\bar{p})$$

by definition of M. This is P4.

18.35 Consider P5(I): if

$$[u_r , j_r] \epsilon M(p),$$

where j_r is real, $r = 1, 2$, then

$$(u_r , j_s)_1 = (C^* v_r , j_s)_1 = (v_r , Cj_s)_1 , \tag{5}$$

where $[v_r, Cj_r] \epsilon L(p)$. Since Cj_r is real

$$(v_1 , Cj_2)_1 = (v_2 , Cj_1)_1$$

by P5(I) for L. This with (5) for $r \neq s$ proves P5(I) for M.

18.36 Fix a $j \epsilon \mathbf{J}_M$ and for each $p \epsilon \Gamma_L$ a $u(p)$ such that

$$[u(p), j] \epsilon M(p).$$

Then $u(p) = C^* v(p)$ and

$$[v(p), Cj] \epsilon L(p),$$

for some $v(p)$. Then as in (5) above

$$(u(p), j)_2 = (v(p), Cj)_1 .$$

P6(I) and P7(I) for L then imply that P6(I) and P7(I) hold for M, using Γ_L for Γ_M in P6.

18.37 We now have M satisfying the hypotheses of 12.0. Therefore there is a Γ_M such that M satisfies all the postulates. This is 18.3.

18.4 *Proof of* "only if" *in* 17.2: Suppose that L between $\mathbf{V}$ and $\mathbf{K}$ is the juxtaposition of L_1 between $\mathbf{V}_1$ and $\mathbf{K}_1$, L_2 between $\mathbf{V}_2$ and $\mathbf{K}_2$. Let, say, $\mathbf{U} = \mathbf{V}_1$ and $\mathbf{J} = \mathbf{K}_1$. Let C be the identity map from $\mathbf{K}_1$ to $\mathbf{K}$: if $j \epsilon \mathbf{J} = \mathbf{K}_1$, then Cj is just j considered as a vector in $\mathbf{K}$. Then C is real. It is easily computed that C^* is E_1.

Consider the restriction M of L based on this C. Its pairs for $p \epsilon \Gamma_M \subseteq \Gamma_L$ are all the pairs $[u, j]$ such that $j = E^* j \epsilon \mathbf{K}_L$ and $u = Ev$, where

$$[v, j] \epsilon L(p). \tag{6}$$

But then

$$[u, j] = [Ev, E^* j]$$

and this is in $L_1(p)$ by (6) and the definition of juxtaposition. Therefore the list $M(p)$ is contained in $L_1(p)$.

Suppose that $[u, j] \epsilon L_1(p)$. We have $[0, 0] \epsilon L_2(p)$ so by P2 and the defi-

nition of juxtaposition

$$[u, j]\epsilon L(p).$$

But then $j = E^*j$, $u = Eu$, and by definition of M

$$[u, j]\epsilon M(p).$$

Therefore for every $p\epsilon\Gamma_M$, $M(p) = L_1(p)$. Therefore there is a frequency domain (Γ_M) for L_1 such that L_1 is PR.

XIX THE NECESSITY PROOF

19.0 Fortunately for this section, those parts of network theory which we require have recently been very succinctly stated by J. L. Synge[12]. We shall paraphrase them here, referring the reader to the source[12] for details of definition.

19.01 First, we observe that in Cauer's definition[5], which we shall repeat in detail below, an ideal transformer with m windings is a $2m$-pole whose terminal pairs are the termini of the respective windings.

A system of m coupled coils is a $2m$-pole with similarly defined terminal pairs.

19.02 Given a $2n$-pole **N** which is a finite passive network, let us adjoin ideal transformers as in Figure 1(b). We then draw the ideal graph of this network. Adjoin to the graph ideal *generator branches* $\gamma_1, \cdots, \gamma_n, \gamma_r$ between T_r and T_r', $1 \leq r \leq n$. Let β_r be the ideal branch representing the transformer winding between T_r and T_r', $1 \leq r \leq n$. Enumerate the remaining branches of the graph $\beta_{n+1}, \cdots, \beta_b$.

19.03 The branch γ_r is in a mesh with β_r and no other branches. Let us call this the r-th external mesh. Any basic set of meshes must include each of these.

19.04 Let $\ell_1, \cdots, \ell_n$ be the currents in the generator branches, $k_1, \cdots, k_b$ the currents in the branches $\beta_1, \cdots, \beta_b$ and

$$[\ell] = [\ell_1, \cdots, \ell_n, k_1, \cdots, k_b], \qquad [k] = [k_1, \cdots, k_b].$$

Let $w_1, \cdots, w_n$ be the voltages across the generator branches, $v_1, \cdots, v_b$ the currents in the $\beta_1, \cdots, \beta_b$ and

$$[w] = [w_1, \cdots, w_n, v_1, \cdots, v_b], \qquad [v] = [v_1, \cdots, v_b].$$

19.05 Let us choose a basic set of meshes, let $j_1, \cdots, j_s$ be the respective mesh currents, and

$$[j] = [j_1, \cdots, j_s].$$

Let

$$[u] = [u_1 , \cdots , u_s]$$

be the s-tuple of mesh voltages. We suppose that $j_1 , \cdots , j_n , u_1 , \cdots , u_n$ refer respectively to the n external meshes. (Cf. 19.03.)

19.06 The results of Synge[12] can now be stated as follows:

There exists a real constant matrix $[C_1]$ of s columns and $b + n$ rows (having, in fact, elements which are $+1$, -1, or 0) such that for any $[j]$

$$[\ell] = [C_1][j] \tag{1}$$

is a set of branch currents satisfying Kirchoff's node law, and for any $[w]$

$$[u] = [C_1]'[w] \tag{2}$$

is a set of mesh voltages satisfying Kirchoff's mesh law. Furthermore, given any $[\ell]$ which satisfies the node law, there is a $[j]$ such that (1) holds.

19.07 If we interpret the $[\ell]$, $[j]$, etc., as representations in real bases then $[C_1]$ is real and $[C_1]' = [C_1]^*$.

19.08 The matrix $[C_1]$ has the form

$$[C_1] = \left[\begin{array}{c|c} C_2 & 0 \\ \hline 0 & C \end{array}\right]$$

where $[C_2]$ is an $n \times n$ diagonal matrix (having diagonal elements ± 1, in fact).

Proof: By construction, $j_1 , \cdots , j_n$ are mesh currents in the external meshes. These are then equal, save for sign, to the currents $\ell_1 , \cdots , \ell_n$ in the generator branches.

19.09 By 19.08, (1), and the definitions in 19.04,

$$[k] = [C][j], \qquad [u] = [C]'[v],$$

and by 19.07, $[C]' = [C]^*$.

19.1 Let us suppose that we have enumerated the branches $\beta_{n+1} , \cdots , \beta_b$ in 19.02 in such a way that $\beta_{n+1} , \cdots , \beta_c$ are all the two poles in the graph, $\beta_{c+1} , \cdots , \beta_d$ are all the branches containing coils which are magnetically coupled, and $\beta_{d+1} , \cdots , \beta_b$ the remaining ideal branches of ideal transformers.

Let $[Z_d(p)]$ be the $(d - n) \times (d - n)$ impedance matrix relating the voltages across the branches $\beta_{n+1} , \cdots , \beta_d$ to the currents in them when

we consider the individual two-poles and the system of coupled coils as separate unconnected networks. Then $[Z_d(p)]$ is composed of a $(c - n) \times (c - n)$ diagonal matrix in the upper left field and a $(d - c) \times (d - c)$ matrix in the lower right, with zeros elsewhere.

19.11 The diagonal part of $[Z_d(p)]$ has elements drawn from the following list:

(i) $f(p) = \rho$

(ii) $f(p) = \delta p$

(iii) $f(p) = \lambda p$

where ρ, δ, λ are non-negative constants, possibly different for each branch.

19.12 It is shown in texts on electromagnetic theory that the matrix representing a system of coupled coils is of the form

$$p[G],$$

where $[G]$ is a real, constant, symmetric, and semi-definite matrix. The lower right field of $[Z_d(p)]$ is then such a matrix.

19.13 It is obvious from this description that $[Z_d(p)]$ is PR. It therefore describes a PR correspondence between $(d - n)$-tuples of current and voltage.

19.2 We must at last consider ideal transformers in detail. Let $\mathbf{V}_1$ and $\mathbf{K}_1$ be m-dimensional spaces represented as aggregates of m-tuples.

Let $\rho_1, \rho_2, \cdots, \rho_m$ be m real numbers. Let $\mathbf{V}_T$ consist of all m-tuples $[a] = [a_1, \cdots, a_m] \epsilon \mathbf{V}_1$ such that

$$\frac{a_1}{\rho_1} = \frac{a_2}{\rho_2} = \cdots = \frac{a_m}{\rho_m}.$$

We interpret these relations as follows:

(a) If any $\rho_r = 0$, then $a_r = 0$

(b) If any two ρ_r, ρ_s are not zero, then

$$\frac{a_r}{\rho_r} = \frac{a_s}{\rho_s}$$

(c) If only one $\rho_r \neq 0$, then a_r is arbitrary.

Let $\mathbf{K}_T$ consist of all m-tuples $[b] = [b_1, \cdots, b_m] \epsilon \mathbf{K}_1$ such that

$$\rho_1 b_1 + \rho_2 b_2 + \cdots + \rho_m b_m = 0.$$

$\mathbf{V}_T$ and $\mathbf{K}_T$ are linear manifolds.

Let $[L_T]$ be the concrete linear correspondence defined by the list $[L_T](p)$ which consists for each complex p of all pairs $[[a], [b]]$ where $[a]\epsilon\mathbf{V}_T$, $[b]\epsilon\mathbf{K}_T$.

The correspondence described by $[L_T]$ is what Cauer[5] defines as an ideal transformer. He shows, loc. cit., how it can be defined as the limiting case of a physical transformer.

There is also a dual kind of device, described by a correspondence admitting all $[b]\epsilon\mathbf{K}_1$ for which

$$\frac{b_1}{\lambda_1} = \frac{b_2}{\lambda_2} = \cdots = \frac{b_m}{\lambda_m}$$

and all $[a]\epsilon\mathbf{V}_1$ for which

$$\lambda_1 a_1 + \cdots + \lambda_m a_m = 0.$$

This also is an ideal transformer obtainable as a limiting case of a physical one.

19.21 The correspondence L_T is PR.

Proof: We observe that $\mathbf{V}_T = (\mathbf{K}_T)^0$, for let $[a]\epsilon\mathbf{V}_T$, $[b]\epsilon\mathbf{K}_T$, and let t be the common value of the a_r/ρ_r. Then

$$(a, b) = \Sigma a_r \bar{b}_r = t\Sigma\rho_r\bar{b}_r = t\overline{(\Sigma\rho_r b_r)} = 0.$$

The postulates are now all easily proved. We omit the details.

19.3 Let $\mathbf{V}$ and $\mathbf{K}$ be b-dimensional spaces. We interpret the b-tuples $[v]$ and $[k]$ of 19.04 as representing vectors $v\epsilon\mathbf{V}$, $k\epsilon\mathbf{K}$ in a real frame.

Let L be the correspondence between $\mathbf{V}$ and $\mathbf{K}$ formed by juxtaposing

(i) the correspondence described by $[Z_d(p)]$ relating components with indices in the range $n + 1$ to d,

(ii) the several correspondences described by ideal transformers, relating components with indices in the ranges 1 to n and $d + 1$ to b.

L is PR because it is the juxtaposition of PR correspondences.

19.31 Let $\mathbf{U}$ and $\mathbf{J}$ be $s - n$-dimensional spaces. We interpret the $[u]$ and $[j]$ of 19.04 as representing $u\epsilon\mathbf{U}$, $j\epsilon\mathbf{J}$ in a real frame.

19.32 Let C be the operation from $\mathbf{J}$ to $\mathbf{K}$ whose matrix in our chosen frames is $[C]$. Then C^* operates from $\mathbf{V}$ to $\mathbf{U}$ with the matrix $[C]^* = [C]'$. By these definitions, C is real. Let M be the correspondence between $\mathbf{U}$ and $\mathbf{J}$ obtained by restricting L with C. Then there is a frequency domain Γ_M such that M is PR (18.3).

19.4 By 19.09, $[M]$ in our chosen frame is the correspondence established between mesh currents and mesh voltages by the network of the

$2n$-pole $\mathbf{N}$. When this network operates as a $2n$-pole, the only mesh voltages which are not zero are those relating to the external meshes, since there are no internal sources of voltage. We must now account for this.

19.41 Let $\mathbf{V}_2$, $\mathbf{K}_2$ be n-dimensional spaces. Choose a real frame and let D be the operation which takes

$$[a_1 , \cdots , a_n]\epsilon\mathbf{V}_2 \tag{3}$$

into

$$[a_1 , \cdots , a_n , 0, \cdots , 0]\epsilon\mathbf{U} \tag{4}$$

in the frame of 19.31. Then D is real and D^* in the chosen frames takes

$$[b_1 , \cdots , b_s]\epsilon\mathbf{J} \tag{5}$$

into

$$[b_1 , \cdots , b_n]\epsilon\mathbf{K}_2 . \tag{6}$$

19.42 We interpret the n-tuples (3) and (6) as voltages and currents in the external meshes of $\mathbf{N}$. Their relations to (4) and (5) are consistent with this interpretation.

Let us restrict M by D, to get a correspondence M_1 between $\mathbf{V}_2$ and $\mathbf{K}_2$. In our chosen frame, the passage to $[M_1]$ corresponds, by (3) and (4) of 19.41, to considering mesh voltages in $\mathbf{N}$ which vanish for every internal mesh, and, correspondingly letting the mesh currents adjust themselves to this situation. We of course observe only the external mesh currents (6).

19.43 M was PR. So, therefore is M_1 (18.3 dual). Since $[M_1]$ is the correspondence established by the physically realizable $2n$-pole $\mathbf{N}$, the necessity of P1, $\cdots$, P7 for formal realizability is established.

XX. APPENDIX TO PART I

20.0 We must prove 7.22 and those assertions of 10.6 which are not covered in Halmos[9]. These concern reality.

20.1 Let $\mathbf{V}_1$ be a real manifold and

$$\mathbf{V} = \mathbf{V}_1 \oplus \mathbf{V}_2 , \qquad \mathbf{K} = \mathbf{K}_1 \oplus \mathbf{K}_2$$

where $\mathbf{K}_1 = (\mathbf{V}_2)^0$, etc. The basis (14) of 10.6 exists by Halmos[9], par. 19. We show that it can be chosen to be real. We have linearly independent vectors

$$v_1 , \cdots , v_r , v_{r+1} , \cdots , v_n ,$$

where the first r span $\mathbf{V}_1$, the last $n - r$, $\mathbf{V}_2$. Let

$$v_s = u_s + iw_s, \qquad 1 \leq s \leq n,$$

where u_s, w_s are real (10.42). Since $\mathbf{V}_1$ is real and a linear manifold,

$$u_s = \tfrac{1}{2}(v_s + \bar{v}_s)\epsilon\mathbf{V}_1, \qquad 1 \leq s \leq r,$$

and, similarly, $w_s\epsilon\mathbf{V}_1$, $1 \leq s \leq r$. Among the $2n$ real vectors

$$u_1, u_2, \cdots, u_r, w_1, \cdots, w_r, u_{r+1}, \cdots, u_n, w_{r+1}, \cdots, w_n, \tag{1}$$

the first $2r$ are in $\mathbf{V}_1$, and they span $\mathbf{V}_1$ because the v_s, $1 \leq s \leq r$, can be constructed from them. The whole list (1) spans $\mathbf{V}$, because from it all the v_s, $1 \leq s \leq n$, can be constructed. Since the $v_s\epsilon\mathbf{V}_2$ do not use in their construction any of the first $2r$ vectors (1), it follows that the last $2(n - r)$ vectors in that list must contain a set spanning $\mathbf{V}_2$. The reality of the vectors (1) then establishes the existence of a real basis, say,

$$v_1', \cdots, v_r', v_{r+1}', \cdots, v_n' \tag{2}$$

which provides a basis in $\mathbf{V}_1$ and $\mathbf{V}_2$.

20.11 We now have 7.22. The unique dual basis

$$k_1', \cdots, k_n'$$

to (2) is real by 10.41. Hence all of $\mathbf{V}_1$, $\mathbf{V}_2$, $\mathbf{K}_1$, $\mathbf{K}_2$ are real. The proof of 10.6 is then complete.

20.2 If in a real basis (2) (dropping primes)

$$v = a_1v_1 + a_2v_2 + \cdots + a_nv_n,$$

that is, if

$$[v] = [a_1, \cdots, a_n],$$

then by (5) of 10.3

$$\bar{v} = \bar{a}_1v_1 + \cdots + \bar{a}_nv_n,$$

hence

$$[\bar{v}] = [\bar{a}_1, \cdots, \bar{a}_n].$$

The geometrical conjugation of 10.3 is therefore simply the concrete one of 7.2 in any real basis. This proves the remark of 10.35.

BIBLIOGRAPHY

1. M. Bayard, "Synthèse des Réseaux Passifs a un Nombre Quelconque de Paires de Bornes Connaissant Leurs Matrices d'Impedance ou d'Admittance," *Bulletin, Société Francaise des Electriciens*, **9**, 6 series, Sept. 1949.
2. O. Brune, *Jour. Math. and Phys., M.I.T.*, **10**, Oct. 1931, pp. 191–235.
3. W. Cauer, *Ein Reaktanztheorem, Sitzungsberichte Preuss. Akad. Wissenschaft*, Heft 30/32, 1931.
4. W. Cauer, "Die Verwirklichung von Wechselstromswiderständen vorgeschriebener Frequenzabhängigkeit," *Archiv für Elektrotechnik*, **17**, 1926.
5. W. Cauer, "Ideale Transformatoren und Lineare Transformationen," *Elektrische Nachrichten-Technik*, **9**, May, 1932.
6. S. Darlington, *Journal of Mathematics and Physics, M.I.T.* **18**, No. 4, Sept. pp. 257–353.
7. R. M. Foster, *Bell System Tech. J.*, April, 1924, pp. 259–267.
8. C. M. Gewertz, *Network Synthesis*, Baltimore, 1933.
9. P. R. Halmos, *Finite Dimensional Vector Spaces*, Princeton, 1942.
10. Y. Oono, "Synthesis of a Finite 2n-Terminal Network by a Group of Networks Each of Which Contains Only One Ohmic Resistance," *Jour. Inst. Elec. Comm. Eng. of Japan*, March, 1946. Reprinted in English in the *Jour. Math. and Phys., M.I.T.*, **29**, Apr., 1950.
11. Y. Oono, "Synthesis of a Finite 2n-Terminal Network as the Extension of Brune's Theory of Two-Terminal Network Synthesis," *Jour. Inst. Elec. Comm. Eng. of Japan*, Aug., 1948.
12. J. L. Synge, "The Fundamental Theorem of Electrical Networks," *Quarterly of Applied Mathematics*, **9**, No. 2, July, 1951.
13. R. Bott, and R. J. Duffin, "Impedance Synthesis Without the Use of Transformers," *Jour. Appl. Phys.*, **20**, Aug., 1949, p. 816.
14. H. W. Bode, *Network Analysis and Feedback Amplifier Design*, New York, 1945.

An Application of Boolean Algebra to Switching Circuit Design

BY ROBERT E. STAEHLER

(Manuscript received January 10, 1952)

This paper discusses the application of switching (Boolean) algebra to the development of an all-relay dial pulse counting and translating circuit employing the minimum number of relays. An attempt is made to outline what appears to be the most promising method of obtaining beneficial results from the use of the algebra in the design of practical switching circuits.

INTRODUCTION

The demands made upon telephone switching systems in regard to improvements in handling capacity, speed, flexibility and economy are continually increasing. In order to meet design objectives enabling the fulfillment of these demands, switching circuits have of necessity become more and more complex and intricate. As certain types of relay switching circuits increase in complexity, the problem of control and output contact network design becomes more and more laborious and time consuming. This is especially true in those circuits in which an attempt has been made to achieve the ultimate in efficiency and economy in that the number of relays used therein approaches the absolute minimum necessary to provide the required number of distinct output combinations. In this type of near-minimum combinational or sequential relay circuit there are numerous parallel control and output contact paths which thread through the same relays repeatedly, thereby causing the individual relay contact loads to become relatively large. Thus the designer's problem becomes that of first developing a workable control and output contact network and then manipulating and minimizing contacts within that network so that the maximum number of contacts used on any one relay is within that permissible on any commercially available relay having the necessary speed characteristics.

Even in those combinational and sequential relay circuits which are not near-minimum and therefore probably have fairly light individual relay contact loads, there are, of course, advantages to be gained by using the least number of contacts possible. Although the initial cost per additional contact (assuming that a few added contacts per relay will not impair the relay speed or space characteristics to an extent that the circuit requirements are not met) is almost negligible, there are other

economic savings possible. Since each contact must be connected to the remainder of the contact network, minimizing contacts and consequently soldered connections means a saving in wiring time and labor. Furthermore, if the designer will manipulate the contacts so that the relays can be chosen from a comparatively few standardized codes, which are in large demand, it is possible to avoid the expensive stockpiling of numerous special designs having only a limited demand. In addition, using the least number of contacts minimizes the focal points of most relay circuit failures which are the contacts themselves (i.e., dirty or worn contacts).

It might also be noted at this point that electronic combinational or sequential circuits usually require electronic gating networks to perform functions which are completely analogous to those of relay contact networks. Hence, the same problem of minimization exists. However, in electronic circuits, gate minimization is even more advantageous since the cost per additional electronic gate is much higher than the cost per additional relay contact.

It is rather obvious that the multiplicity of paths in most combinational and sequential circuits can cause their design to become an extremely difficult and time consuming problem if the contact paths are developed with the aforementioned considerations in mind.

The circuit designer's usual approach to the solution of such contact minimization and manipulation problems is that of inspection. The method of inspection presupposes a background of considerable experience in that the designer must recognize certain contact network arrangements that may allow further rearrangements and thereby he must mentally develop his own rules. In order to check on any of his manipulations he must repeatedly redraw the network during this inspection design process. It is evident that this is often a long and tedious method and, depending on the skill of the designer, may or may not result in an optimum or even adequate solution.

Suitable contact network arrangements often appear only after consideration of several alternative schemes and the rearrangements of the network interconnections of these schemes. Realization of this makes it quite evident that any means of obtaining and comparing these various schemes quickly and with a mathematical accuracy which does not require continuous checking of network paths permits a more rapid and complete exploration of the particular problem. Switching algebra, first codified by C. E. Shannon[1], is the systematic application of G. Boole's[2]

[1] C. E. Shannon, *A Symbolic Analysis of Relay and Switching Circuits*, Trans. AIEE, **57**, 1938.

[2] G. Boole, *The Mathematical Analysis of Logic* (Cambridge 1847) and *An Investigation of the Laws of Thought* (London 1854).

"Algebra of Logic" to switching circuits and is just such a means. It is a tool which can be used to investigate the complex combinational and sequential networks to determine satisfactory contact arrangements or reject unsatisfactory ones with a minimum of time and effort. It should be emphasized, however, that as with any tool, satisfactory results depend upon the judgment, ingenuity and logical reasoning of the user. Furthermore, as will be evident from the following development, switching (Boolean) algebra in its present state is not to be considered entirely selfsufficient but, for the most beneficial results, should be applied, when warranted, in conjunction with inspection techniques so that the latter may fill in any limitations in the algebra techniques which have not been completely systematized as yet due to the newness of this field.

The problem of solving the contact requirements of a minimum relay dial pulse counting and translating circuit recently developed as a component of the originating register of the No. 5 Crossbar System will be used as a means of illustrating the practical use now being made of switching algebra and of indicating exactly where the application of the algebra enters the design problem.

BASIC DIAL PULSE COUNTER REQUIREMENTS

The primary function of the originating register is to receive pulse signals representing digits from a telephone dial or similar calling device and to store a record of the digits in a form suitable for use by an external circuit. The dial pulse counting and translating circuit, an integral part of the originating register, is oriented with respect to other parts of the register by the block diagram of Fig. 1. The *L* relay is the pulse detecting relay. When the subscriber's switchhook contact is closed due to the lifting of the phone, the originating register is connected to the line and the *L* relay is operated. Thereafter it follows the breaks and makes of the subscriber's dial and feeds these repeated dial pulses into the counter. After the pulses are counted they are translated to a new code. In switching systems it is advantageous to translate from the basic dial ten pulse decimal code to a "two out of five" self-checking code. In this latter code any single error within the circuit will result in either one or three relays operated in the associated storage circuit rather than two and thus an error can readily be detected. The output of the translator is fed via a steering circuit to the register or storage circuit. The slow release *RA* relay is the pulse train detecting relay which holds between the individual pulses of a digit and releases only at the end of the pulse train. When it releases it activates the translating circuit and thereby transfers the translated code information to the storage circuit. The *RA1* relay

in operating terminates the output from the translator and simultaneously releases the relays in the counter to prepare it for the next digit.

Specific requirements imposed by the originating register circuit necessitate the counting of one to eleven pulses; the use of a driving source consisting of a single break-make (or transfer) contact with ground on the armature spring; and outputs as follows:

1. Count of 1 through 10: ground on two of the 0, 1, 2, 4, 7 output leads in the combination corresponding to the count.
2. Count of 10: ground on the *Z*0 lead.
3. Count of 11: ground on the 0 lead only (this is a trouble-detecting feature).

In addition, the design of the steering and register-storage circuit requires that no output leads be connected together until the second pulse is received. Furthermore, each relay is limited to a combination of simple make and break contacts not exceeding a total of twelve. This utilizes the maximum number of springs obtainable on presently available relays and also avoids the larger armature gaps imposed by transfers which would result in a reduction in the relay speed of operation. Speed requirements also do not permit the use of shunt release in the circuit operation.

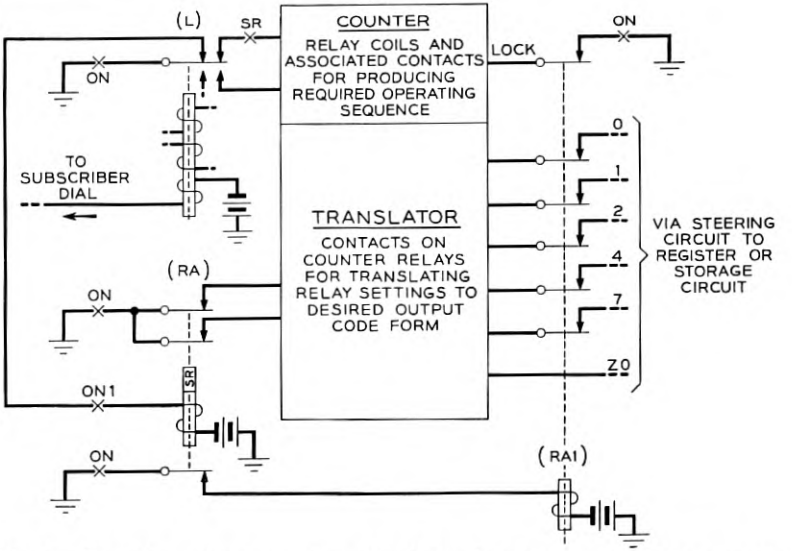


Fig. 1—The schematic of a portion of a dial pulse register circuit for counting decimal code pulses and translating them to "two out of five" signals. (In the symbolism used in the illustrations a cross indicates a "make" contact and a vertical bar indicates a "break" contact.)

THEORY OF A MINIMUM RELAY COUNTER

The counting circuit under consideration does not contemplate the use of any circuit elements other than relays that react to the beginning or end of a pulse. Therefore it must establish a distinct combination of relays operated or released during and between successive pulses. The minimum number of ordinary "two-position" relays, R, required to count P pulses can be obtained from the expressions (1) $2P \leq 2^R$ if the counter is to lock up during, or recycle after, the last pulse or (2) $2P \leq 2^R - 1$ if the counter is to lock up after the last pulse.

The usual counting circuit used for determining the number of pulses in a dial train is required to count ten pulses, however there are certain advantages in regard to trouble indications if the counter counts eleven pulses. In either case the minimum number of relays necessary, according to the preceding formulae, is five. It should be noted that the ease with which this minimum number can be attained depends upon whether the input is derived from a single, double or transfer contact source.

DETERMINATION OF OPERATING SEQUENCE

Having determined that the minimum number of relays necessary is five, the first step in design is to develop an operating sequence pattern from the resulting 2^5 or 32 possible relay combinations. These combinations may be utilized in any order deemed desirable to obtain the 23 distinct combinations needed to differentiate between eleven pulses (22 for the eleven makes and breaks plus an all-relays-normal combination). In this phase of the design switching algebra is not involved. The optimum sequence to meet a particular set of requirements can only be determined by repeated trials guided by an intimate knowledge of objectives.

Initial studies, made by Joseph Michal, of various possible sequence patterns for a five relay circuit, including those having a three relay "ring" followed by two auxiliary relays and those having a two relay pulse divider followed by three auxiliary relays, resulted in the conclution that the latter approach was the most fruitful. The sequence pattern adopted is shown in detail in Table I. The pattern is extended through 12 pulses, and it can be seen that the nature of the sequence is such that this employs all 32 combinations of the 5 relays. Several of these are transient and occur during part of a pulse or inter-pulse interval. Examination of the tail end of the sequence indicates that it will be simpler to design on the basis of a full 12 pulses than attempt to block at the end of the 11 pulses specified by the requirements. If trouble con-

TABLE I

SEQUENCE OF OPERATION

	Pulsing Relay	Counting Relays					Relay Combination	Two out of Five Code
	L	*A*	*B*	*C*	*D*	*E*		
Seizure	0	1	1	1	1	1	1	
1st pulse	1	0	1	1	1	1	2	
	0	0	0	1	1	1	3	0,1
2nd pulse	1	1	0	1	1	1	4	
	1	1	0	0	1	1	5	
	0	1	1	0	1	1	6	0,2
3rd pulse	1	0	1	0	1	1	7	
	0	0	0	0	1	1	8	
	0	0	0	0	0	1	9	1,2
4th pulse	1	1	0	0	0	1	10	
	0	1	1	0	0	1	11	0,4
5th pulse	1	0	1	0	0	1	12	
	1	0	1	1	0	1	13	
	0	0	0	1	0	1	14	1,4
6th pulse	1	1	0	1	0	1	15	
	0	1	1	1	0	1	16	
	0	1	1	1	0	0	17	2,4
7th pulse	1	0	1	1	0	0	18	
	0	0	0	1	0	0	19	0,7
8th pulse	1	1	0	1	0	0	20	
	1	1	0	0	0	0	21	
	0	1	1	0	0	0	22	1,7
9th pulse	1	0	1	0	0	0	23	
	0	0	0	0	0	0	24	
	0	0	0	0	1	0	25	2,7
10th pulse	1	1	0	0	1	0	26	
	0	1	1	0	1	0	27	4, 7-ZO
11th pulse	1	0	1	0	1	0	28	
	1	0	1	1	1	0	29	
	0	0	0	1	1	0	30	0
12th pulse	1	1	0	1	1	0	31	
	0	1	1	1	1	0	32	0

Total of $2^5 = 32$ combinations used.

Note: 0 is used to indicate that the relay listed at the head of the column is operated, and 1 is used to indicate that the relay is released.

ditions introduce pulses beyond 12, the circuit will without difficulty recycle through combinations corresponding to pulses 11 and 12.

Table I also indicates the leads which must be grounded in order to provide the translations to the "two out of five" and "single lead" codes.

The characteristics of this circuit may be summarized as follows: It contains only five relays which is the absolute minimum necessary. It

uses all 32 of its available combinations. Its control and translating job is complex enough to indicate the need for a considerable number of contacts and hence the need for extensive contact manipulation to minimize and distribute these contacts.

It is apparent that a great deal of time would be necessary to accomplish this manipulation by inspection methods, thereby indicating the need for an additional tool such as switching algebra to assist the designers in this task.

ALGEBRAIC METHODS APPLIED TO CONTROL CIRCUIT

The sequence of operations of Table I is used as the starting point in the application of the algebra. The exact calculations necessary to develop the control and translating circuit by this means are shown in detail later. However, the individual steps in the solution might well be outlined here. First, the design of the control and translating networks will be regarded as separate problems. In theory these can be integrated together, but the resultant network is likely to be so complex that understanding and maintenance of the circuit would suffer. Each of the two networks can be individually considered as a multi-terminal network of the single input type. That is, the control network is an associated set of contacts which connects a single ground input to the windings of five relays, and the translating network is an associated set of contacts which connects a single ground input to the six output leads. Since switching algebra is directly applicable to two-terminal networks rather than multi-terminal networks, the approach to this particular problem is of necessity somewhat indirect.

The most satisfactory method of attack is to develop first a two-terminal network for each of the output paths of the multi-terminal network under consideration. The two-terminal networks can be expressed algebraically and manipulated into their simplest form by means of the switching algebra theorems to be given later. The individual networks can then be inspected carefully, either in algebraic or circuit form, with the objective of combining them in the most advantageous fashion. It will be found, in general, that the simplest network configurations do not readily combine and that further manipulation is necessary to obtain an economical circuit. It is at this point that the algebra achieves its greatest utility, since its application permits the simple and rapid changing of a given two-terminal network into a large variety of different forms with mathematical assurance that circuit equivalence is maintained. Inspection of the networks in the several forms provides clues

as to the preferable combining forms and often indicates additional manipulations that might be desirable.

This network development is a combination of mathematics and integration by inspection. It is characterized by repeated trials of alternative forms and at no stage is there any definite assurance that the optimum circuit has been attained. However, the ease of manipulation provided by the algebra greatly enhances the probability of designing a better circuit than would be possible by inspection alone. In combining the two-terminal networks, care must be taken not to introduce "sneak" paths which improperly connect outputs together. The algebra usually offers means of introducing one or two additional contacts which permit combining networks and yet eliminate the adverse effects of the sneak paths.

The above procedure will now be carried out in detail with the switching algebra theorems that are used in all the following algebraic manipulations noted at the margin by the number which corresponds to the number of the theorem in the complete listing in Table II. This table is

TABLE II

SWITCHING (BOOLEAN) ALGEBRA*

Definitions	Postulates
Addition $(+) = AND$ = Series Multiplication $(\cdot) = OR$ = Parallel Circuit States 0 = Closed Circuit 1 = Open Circuit	(1) $X = 0$ or $X = 1$, where X is a contact or a network. (2a) $0 \cdot 0 = 0$ (2b) $1 + 1 = 1$ (3a) $1 \cdot 1 = 1$ (3b) $0 + 0 = 0$ (4a) $1 \cdot 0 = 0 \cdot 1 = 0$ (4b) $0 + 1 = 1 + 0 = 1$

Theorems	
(1a) $X + Y = Y + X$	(8a) $X' + X = 1$
(1b) $XY = YX$	(8b) $X'X = 0$
(2a) $X + Y + Z = (X + Y) + Z = X + (Y + Z)$	(9a) $0 + X = X$
(2b) $XYZ = (XY)Z = X(YZ)$	(9b) $1 \cdot X = X$
(3a) $XY + XZ = X(Y + Z)$	(10a) $1 + X = 1$
(3b) $(X + Y)(X + Z) = X + YZ$	(10b) $0 \cdot X = 0$
(4a) $X + X = X$	(11a) $(X + Y')Y = XY$
(4b) $XX = X$	(11b) $XY' + Y = X + Y$
(5a) $X + XY = X$	(12a) $(X + Y)(X' + Z)(Y + Z) = (X + Y)(X' + Z)$
(5b) $X(X + Y) = X$	(12b) $XZ + X'Y + YZ = XZ + X'Y$
(6a) $(X)' = X'$	(13) $(X + Y)(X' + Z) = XZ + X'Y$
(6b) $(X')' = X$	(14a) $f(X) = A \cdot f(X)_{A=1,\ A'=0} + A' \cdot f(X)_{A=0,\ A'=1}$
(7a) $(X + Y + Z + \cdots)' = X' \cdot Y' \cdot Z' \cdot \cdots$	(14b) $f(X) = [A + f(X)_{A=0,\ A'=1}][A' + f(X)_{A=1,\ A'=0}]$
(7b) $(X \cdot Y \cdot Z \cdot \cdots)' = X' + Y' + Z' + \cdots$	

* Reprinted from *The Design of Switching Circuits* by Keister, Ritchie and Washburn with the permission of D. Van Nostrand Co., Inc.

taken from *The Design of Switching Circuits* by Keister, Ritchie and Washburn*. The development of the algebraic expressions from the sequence of operations table will be in exact parallel to the methods suggested in the aforementioned text.

The symbolism adopted in the following development is basically that of using the notation A for all the make contacts on the A relay, and A' for all the break contacts on the A relay. Contacts or groups of contacts in series are related by the symbol of addition $(+)$ and contacts or groups of contacts in parallel are related by the symbol for multiplication $(\cdot)$ which may or may not be explicitly written, as in ordinary algebra. Therefore $(A + B')$ symbolizes a series contact path that is closed when the A relay is operated *and* the B relay is released, while (AB') symbolizes the parallel contact path that is closed when either A is operated *or* B is released. Switching algebra includes only two numerical values, 0 and 1, with the quantity 0 assigned to represent a closed path and 1 to represent an open path. For the tabular notation of Table I, 0 is used to indicate that the relay listed at the head of the column is operated and 1 is used to indicate that the relay is released.

As stated earlier, the present application of switching algebra utilizes the sequence of operation chart of Table I. The operate and release combinations for controlling the A, B, C, D and E relays can be selected from this table by observing where each relay to be controlled changes state. For example, the operate combination for relay D is relay combination 8 and the release combination for relay D is relay combination 24. It is not necessary to include the contacts of a relay in its own operate and release combinations. Note that the A and B relays which serve as a pulse divider can be controlled solely by the L relay and contacts on A and B without reference to C, D, E. However the C, D and E relays are internally controlled by all five counting relays. The development of all these control paths uses the following abbreviations:

$$g(X) = \text{operating combinations for the } X \text{ relay}$$

$$r(X) = \text{releasing combinations for the } X \text{ relay}$$

$$h(X) = \text{holding combinations for the } X \text{ relay}$$

$$X = \text{make contact on the } X \text{ relay}$$

Furthermore as expressed by theorem (6a and 6b) the negative of a contact network X is defined as a network which is a closed path under all conditions for which X is open, and is open under those conditions

* D. Van Nostrand, 1951. The Bell Telephone Laboratories Series.

for which X is closed. Hence $h(X)$ may be obtained from $r(X)$ by noting that $h(X)$ is the negative of $r(X)$. Therefore the entire control path of any relay can be expressed generally as

$$f(X) = g(X)[X + h(X)] = g(X)[X + (r(X))']$$

Thus for the A relay

$$\begin{aligned} g(A) &= L' + B' \\ r(A) &= L' + B \\ h(A) &= [L' + B]' = LB' \end{aligned} \tag{7a}$$

and

$$\begin{aligned} f(A) &= (L' + B')(A + LB') \\ &= (L' + B')(A + L)(A + B') && (3b) \\ &= (L + A)(L' + B') && (12a) \end{aligned}$$

Also for the B relay

$$\begin{aligned} g(B) &= L + A \\ r(B) &= L + A' \\ h(B) &= [L + A']' = L'A \end{aligned} \tag{7a}$$

and

$$\begin{aligned} f(B) &= (L + A)(B + L'A) \\ &= (L + A)(B + L')(B + A) && (3b) \\ &= (L + A)(L' + B) && (12a) \end{aligned}$$

The schematic forms of the A and B control circuits as represented by the above algebraic expressions are shown in Fig. 2a and 2b. Since the general requirements of the basic problem specify only one transfer on the L relay, only simple makes and breaks on the A and B relays and no shunt release paths (to avoid reduction in speed of operation), the combination of the above specific circuits is not possible without recourse to double windings. Another factor which affects the practical form of the circuit is the finite transit time of the L relay armature spring. Switching algebra presupposes instantaneous action of relay contacts and in certain cases, when the use of a break-make transfer is required, additional contacts are necessary to cover the open contact interval. The final circuit form, conceived by F. K. Low, is shown on Fig. 2c and uses an added A

contact. Algebraic equivalence of this circuit with the original is shown below.

$$\begin{aligned} f(A) &= (L'A + B')(L + A) \quad \text{(Fig. 2c)} \\ &= (L' + B')(A + B')(L + A) \quad (3b) \\ &= (L' + B')(L + A) \quad (12a) \\ f(B) &= (L'A + B)(L + A) \quad \text{(Fig. 2c)} \\ &= (L' + B)(A + B)(L + A) \quad (3b) \\ &= (L' + B)(L + A) \quad (12a) \end{aligned}$$

For the C relay which operates and releases twice in the entire 32 combination cycle

$$\begin{aligned} g(C) &= (A' + B + D' + E')(A' + B + D + E) \\ r(C) &= [(A + B' + D + E')(A + B' + D' + E)] \\ h(C) &= [(A + B' + D + E')(A + B' + D' + E)]' \\ &= (A'BD'E + A'BDE') \quad (7a, 7b) \end{aligned}$$

and

$$\begin{aligned} f(C) &= [(A' + B + D' + E')(A' + B + D + E)] \\ &\qquad [C + A'BD'E + A'BDE'] \\ &= [A' + B + (D' + E')(D + E)] \\ &\qquad [C + A'B(D' + E')(D + E)] \quad (3a, 3b, 13) \\ &= [A' + B + (D' + E')(D + E)][C + A'B] \\ &\qquad [C + (D' + E')(D + E)] \quad (3b) \\ &= [(A' + B)C + (D' + E')(D + E)][C + A'B] \quad (3b) \end{aligned}$$

The schematic circuit which the above represents is shown in Fig. 3a. Circuits of this type which use certain contacts more than once can some-

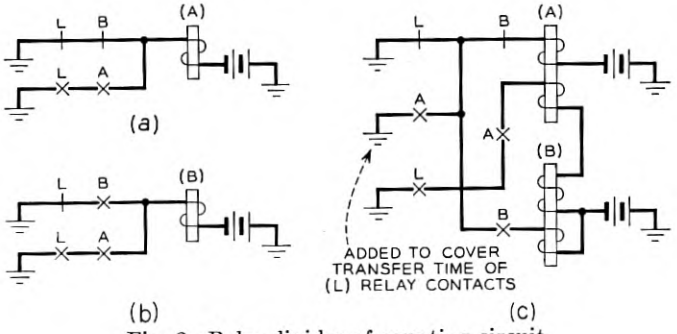


Fig. 2—Pulse divider of counting circuit.

times be drawn in bridge form with a consequent saving of contacts. One method is to manipulate the expression into a form which is known to be the series-parallel equivalent of a bridge. However, following usual algebraic procedures it is often difficult to recognize where this is possible. In the present case a method developed by G. R. Frost (not yet published) was used effectively. This resulted in the bridge circuit of Fig. 3b which has the series-parallel equivalent:

$$f(C) = [C + (D' + E')(D + E)][A' + B + (D' + E')(D + E)][C + A'][C + B]$$

By use of theorem (3b) this is seen to be equivalent to the previous expression for $f(C)$.

For the D relay which only operates and releases once in the entire cycle

$$\begin{aligned} g(D) &= (A + B + C + E') \\ r(D) &= (A + B + C + E) \\ h(D) &= (A + B + C + E)' \\ &= A'B'C'E' \end{aligned}$$

and $f(D) = (A + B + C + E')(D + A'B'C'E')$.

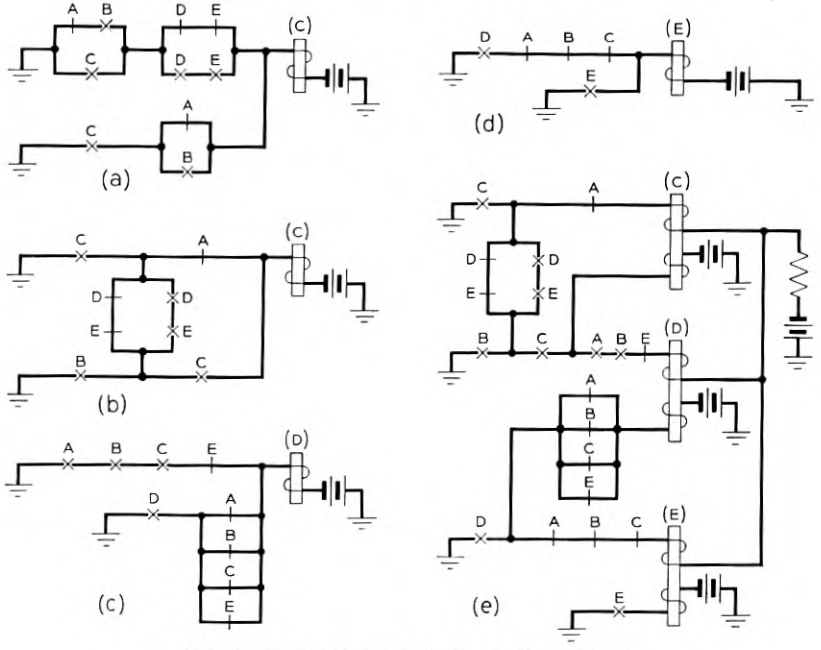


Fig. 3—Internal control of counting circuit.

By noting that $A + B + C$ is the negative of $A'B'C'$, this can be reduced to

$$f(D) = (A + B + C + E')(D + A'B'C') \qquad (3b, 12a)$$

However, this transformation introduces a hazard caused by the transit time of A relay contacts in passing from relay combination 9 to 10. Therefore the original expression will be used for relay D. The control path is shown in Fig. 3c.

For the E relay which operates and locks only once in the cycle

$$g(E) = (A' + B' + C' + D)$$
$$h(E) = E$$

and

$$f(E) = (A' + B' + C' + D)(E + E)$$
$$(A' + B' + C' + D)E \qquad (4a)$$

This control path is shown in Fig. 3d.

Apart from the problem of developing the required contact network, the practical problem of what operating power must be given to the relays in order to meet speed requirements must be dealt with. Since the use of low resistance windings in series with protective external resistors is called for to obtain the speed required, it appears that the use of two windings per relay might prove advantageous. By operating on the low resistance winding while locking on the high resistance winding, the current drain may be reduced (thereby saving a fuse) and furthermore some code reduction may be made possible as shown later. If double windings are used, two of the external series resistors may be eliminated by combining the control network so as to make certain that only one of the low resistance windings on the C, D, or E relays is energized at any one time. This would permit the use of one common external resistance with the aforementioned relays instead of three.

Keeping these practical considerations in mind, further savings may be made by combining the control circuits as shown in Fig. 3e. Although there is in this circuit a possibility of contact stagger on the A relay contacts causing the C and D low resistance windings to be energized at the same time, this will not be harmful since, when the stagger occurs, both relays are firmly locked operated by their high resistance holding windings.

TRANSLATING CIRCUIT

The translating circuit is particularly adaptable to switching algebra manipulation. Table I shows the combinations which prevail at the end

of each pulse and the necessary "code" leads that must be grounded at these times. Reference to the block diagram of Fig. 1 shows that the output of the translator is not activated until the slow release RA relay releases after the last pulse of a digit has been received. Therefore the A relay can be eliminated from these combinations since at the end of every pulse the A and B relays are either both operated or both released and hence only one is needed to indicate the condition of both. Table III lists the numerous combinations which must close a ground path through to each of the five code leads and the $Z0$ lead. At the conclusion of the algebraic manipulation, the A and B contacts may be redistributed evenly since they perform interchangeable functions in translation.

The objective in the design of the translating circuit is to obtain the most economical contact network subject to a spring distribution that

TABLE III
TRANSLATION

Output Lead Grounded	Counting Relays				Decimal Pulse
	B	C	D	E	
0	0	1	1	1	1
	1	0	1	1	2
	1	0	0	1	4
	0	1	0	0	7
	0	1	1	0	11
	1	1	1	0	12
1	0	1	1	1	1
	0	0	0	1	3
	0	1	0	1	5
	1	0	0	0	8
2	1	0	1	1	2
	0	0	0	1	3
	1	1	0	0	6
	0	0	1	0	9
4	1	0	0	1	4
	0	1	0	1	5
	1	1	0	0	6
	1	0	1	0	10
7	0	1	0	0	7
	1	0	0	0	8
	0	0	1	0	9
	1	0	1	0	10
$Z0$	1	0	1	0	10
Inconsequential combinations that may be used for simplification					
	0	0	1	1	
	1	1	0	1	
	0	0	0	0	
	1	1	1	1	

fits in with the control contact network. Again, this is a multi-terminal network problem and the procedure is to design two-terminal networks that combine most readily. Since it is impractical to illustrate all the repeated trials that led to the final design, each network will be designed separately with the understanding that some of the steps are imposed by the form of all networks viewed collectively.

The procedure adopted for developing the "0" lead network is as follows. First set up the miniature table repeating the portion of Table III that corresponds to the "0" lead. These parallel combinations should then be manipulated algebraically to obtain the greatest simplification possible. It is rather easy to apply some of the algebraic rules by observing the condition of the relay in the several combinations in the table. A simple "shorthand" rule to follow is: if in the table of combinations describing a particular two terminal network, all possible combinations of certain relays appear in conjunction with a single combination of other relays, the network contacts on the former relays may be neglected. In other words when 2^n different combinations of any number of variables m, are identical in all but n columns, contacts on the corresponding n relays are not required. This procedure is carried out below.

B	C	D	E	
0	1	1	1	(B + C′ + D′)
1	0	1	1	(B′ + C + E′)
1	0	0	1	
0	1	0	0	(B + C′ + E)
0	1	1	0	
1	1	1	0	(C′ + D′ + E)

Thus we have the following algebraic expression for the "0" lead, which can be simplified as shown.

$$(B + C' + D')(B' + C + E')(B + C' + E)(C' + D' + \mathrm{E})$$

$$[C' + (B + D')(B + E)(D' + E)](B' + C + E') \tag{3b}$$

$$[C' + (E + BD')(B + D')](B' + C + E') \tag{3b}$$

This is shown on Fig. 4a. A somewhat different manipulation of the equation permits placing the network in the bridge form of Fig. 4b. The algebraic equation, given below, can easily be shown to be the equivalent of the original.

$$[E + C' + B(B' + D')](B' + C + E')(B + C' + D')$$

In certain cases the use of theorem (14b), normally employed to reduce the contacts of a particular relay to a single make and break, can produce simplifications difficult to accomplish otherwise. This is shown below, with the theorem applied with respect to relay E since E tended otherwise to be heavily loaded.

$$(B + C' + D')(B' + C + E')(B + C' + E)(C' + D' + E)$$

$$[E + (B + C' + D')(B' + C + 1)(B + C' + 0)(C' + D' + 0)]$$
$$[E' + (B + C' + D')(B' + C + 0)(B + C' + 1)(C' + D' + 1)] \qquad (14b)$$

$$(E + C' + BD')[E' + (B' + C)(B + C' + D')] \qquad (9a, 10a, 3b, 9b, 5b)$$

By modifying the first factor of the final expression in accordance with theorem 11a, this equation can be put in bridge form as shown on Fig. 4c.

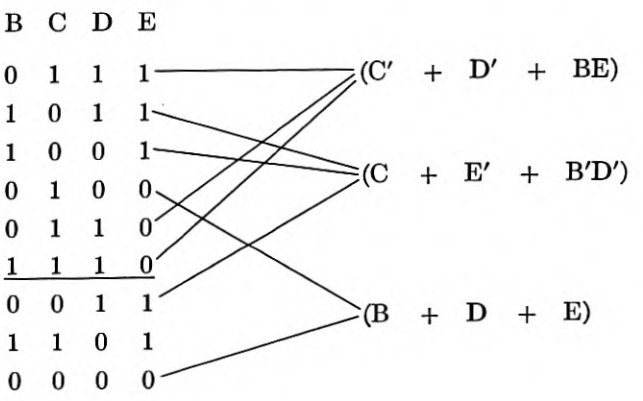
$$[E + C' + B(B' + D')][E' + (B' + C)(B + C' + D')]$$

The above equation uses the same contacts as the previous expression, and although the right hand member is in a slightly different form, the expression is equivalent to the one obtained earlier.

When it is known that output conditions are inconsequential for some relay combinations, these inconsequential relay combinations may be combined with valid combinations to eliminate contacts in the network. Inconsequential means that the output during these particular combinations does not affect the proper functioning of the circuit. Four such combinations are listed in Table III. Only those inconsequential combinations which will combine readily with the actual combinations, thereby resulting in a reduction in the number of contacts, are to be used. Although the use of all the all-relays-released condition may be helpful in certain cases, it will not be used in the circuit under consideration since its use makes the requirement that no tie shall exist between output leads until the second pulse is received hard to meet.

With this in mind the "0" lead network is again examined. Note the use of another "shorthand" rule which states that if a part of the 2^n possible combinations is used in closing a path, the negative of the unused part of the 2^n possible combinations is equivalent to the original combinations. Thus if in the case at hand three of the possible four combinations of the B and C relays occur in series with the same combination of the D and E relays, the expression used is that for the series path of the D and E relays plus the negative of the missing combination of the B

and C relays. In the following tabulation the combinations below the horizontal line are inconsequential.



The expression becomes:

$$(C + E' + B'D')(C' + D' + BE)(B + D + E)$$

$$[E + (C + 1 + B'D')(C' + D' + B0)(B + D + 0)]$$
$$[E' + (C + 0 + B'D')(C' + D' + B1)(B + D + 1)]$$
(14b on E)

$$[E + (C' + D')(B + D)][E' + (C + B'D')(C' + D' + B)]$$
(9a, 9b, 10a, 10b)

$$[E + (C' + D')(B + D)][E' + CD' + CB + B'D'] \quad (8b, 9a, 4b, 5a)$$

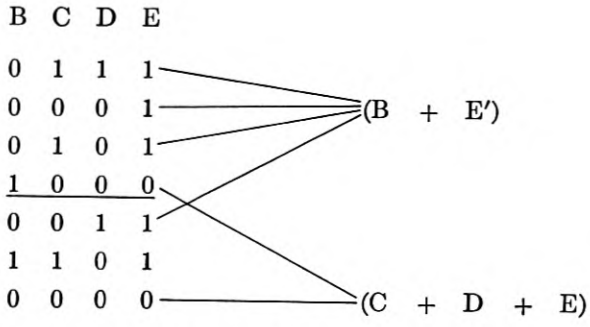
$$[E + (C' + D')(B + D)][E' + BC + B'D'] \quad (12b)$$

$$[E + (C' + D')(B + D)][E' + (B + D')(B' + C)] \quad (13)$$

Fig. 4d shows the schematic of the above expression. It is possible to put this in a bridge form without other changes because of the manner in which the front and back contacts of D are related to the other contacts. Comparison of all the circuits of Fig. 4 indicates that they all use the same number of contacts although final decision should be postponed until all the output circuits are obtained and the ease of combination of the different circuits can be compared.

The procedure for determining the remaining code leads is carried out on the following pages.

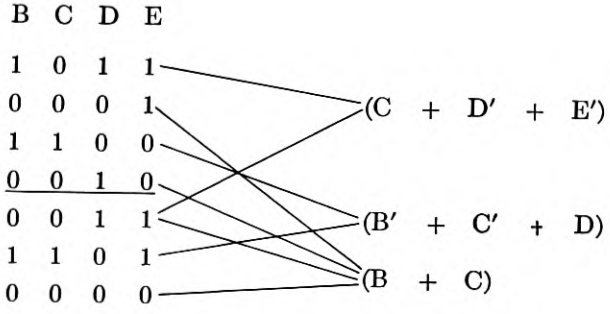
"1" lead—



resulting in $[E + C + D][E' + B]$ which is shown on Fig. 5a. It will later be found advantageous, in combining, to include the B' term in the first factor, giving the expression:

$$(E + B' + C + D)(E' + B) \quad \text{shown on Fig. 5b}$$

"2" lead—



hence

$$(C + D' + E')(B' + C' + D)(B + C)$$

or

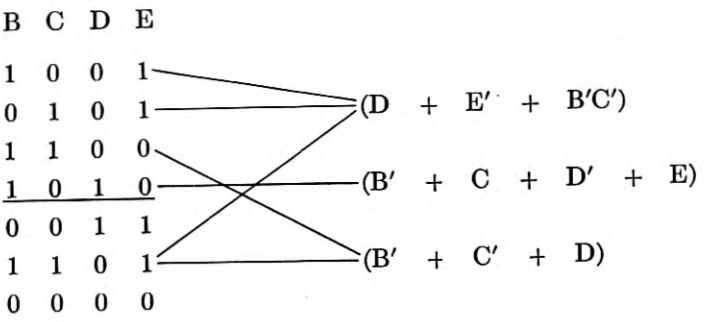
$$[C + B(D' + E')][C' + B' + D] \tag{3b}$$

For later ease in combining, this is changed to:

$$[C + B(B' + D' + E')][C' + B' + D] \tag{11a}$$

shown on Fig. 5c.

"4" lead—

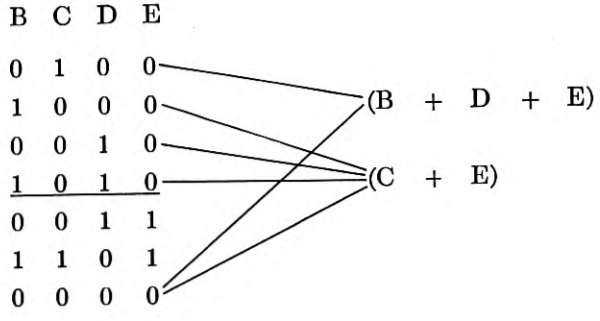


hence

$$(D + E' + B'C')(B' + C' + D)(B' + C + D' + E)$$
$$[D + (B' + C')(B'C' + E')][D' + B' + C + E] \quad (3b)$$

which is shown on Fig. 5d.

"7" lead—



hence

$$(B + D + E)(C + E)$$

or

$$E + C(B + D) \quad (3b)$$

which is shown on Fig. 5e.

"Z0" lead—

B	C	D	E
1	0	1	0

hence one has $(B' + C + D' + E)$ which is shown in Fig. 5f.

The final contact savings are achieved by combining the various output paths. The combined translation circuit that appears to be as reduced as possible is shown in Fig. 6. Note that certain forms of the individual output paths combine more readily than others. For example Fig. 4c and 5b combine more readily with the remaining paths than Fig. 4b and 5a. Note also that sometimes it is not the most reduced form of the individual output paths that permits efficient combining. This is

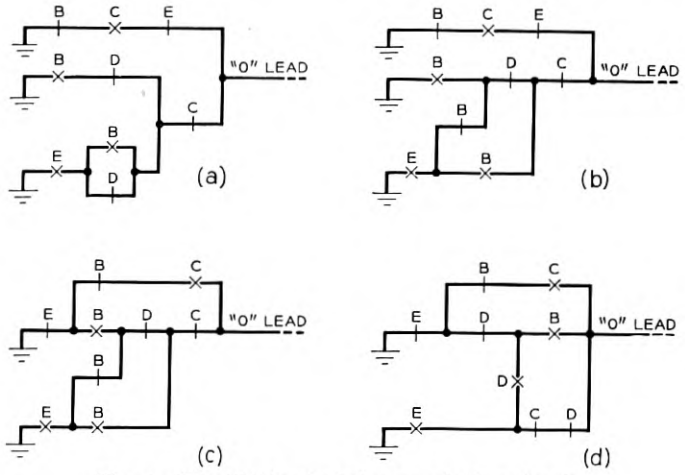


Fig. 4—The "0" lead of the translating circuit.

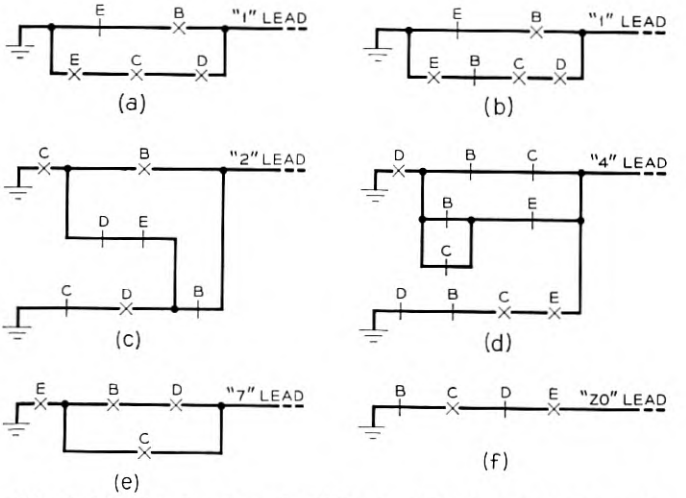


Fig. 5—The "1, 2, 4, 7, and Z0" leads of the translating circuit.

exemplified by the use of Fig. 5b rather than Fig. 5a. Although various forms of all of the output leads were tested for efficient combination only the form used is shown for outputs other than the "0" and "1" leads.

It is essential to scrutinize the final network for possible sneak paths. Sometimes to avoid these sneak paths it is necessary to add one contact on one relay to allow savings on others. Here again the inspection techniques go hand in hand with switching algebra and the need for both is obvious. The algebra obtains the various forms which are capable of

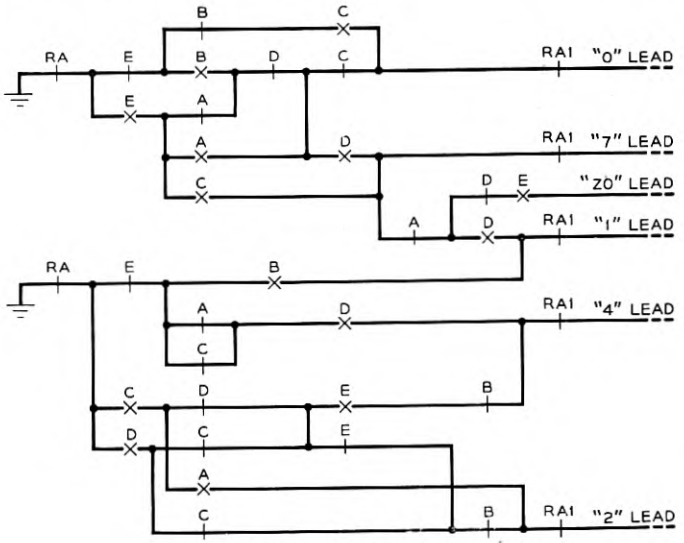


Fig. 6—Combined translating circuit.

different degrees of combination very quickly and efficiently. The inspection method is then necessary for the actual combination of these forms.

The additional RA relay contact is necessary to assist in avoiding interconnections between the output leads until after the second pulse is received. The final assignments of either A or B relay contacts are chosen to equalize the load on these relays.

THE COMPLETED CIRCUIT

The final form of the counting and translating circuit is shown on Fig. 7. The relays are all double wound to gain the benefits of current drain reduction. One additional advantage of using double windings is the relay code reduction made possible since now only two codes are necessary. One code serves the A relay and one other code serves the

B, C, D and E relays. In comparison to this total of five relays and two codes the circuit in present use in the latest crossbar system requires ten relays and seven codes.

AN ALTERNATE DESIGN OF THE PULSE DIVIDER

To illustrate the application of algebra where the apparatus contemplated puts less premium on contact minimization but more on standardization and winding minimization, certain modifications of the proposed circuit are considered.

In the event that new apparatus developments make possible the construction of relays that meet the necessary speed requirements even though winding impedance is increased, it appears possible (if the pulse divider is redesigned) to use only one code having a single winding for all five relays. The use of added contacts might be allowable if the new type of relay carries more springs than the present relay.

The redesign of the pulse divider to use single windings can be accomplished by manipulation of the basic algebraic expressions derived earlier for the pulse divider.

Thus for the A relay

$$\begin{aligned} f(A) &= (L' + B')(A + LB') \\ &= [L' + B'][A + B'(L + B)] \qquad (11a) \end{aligned}$$

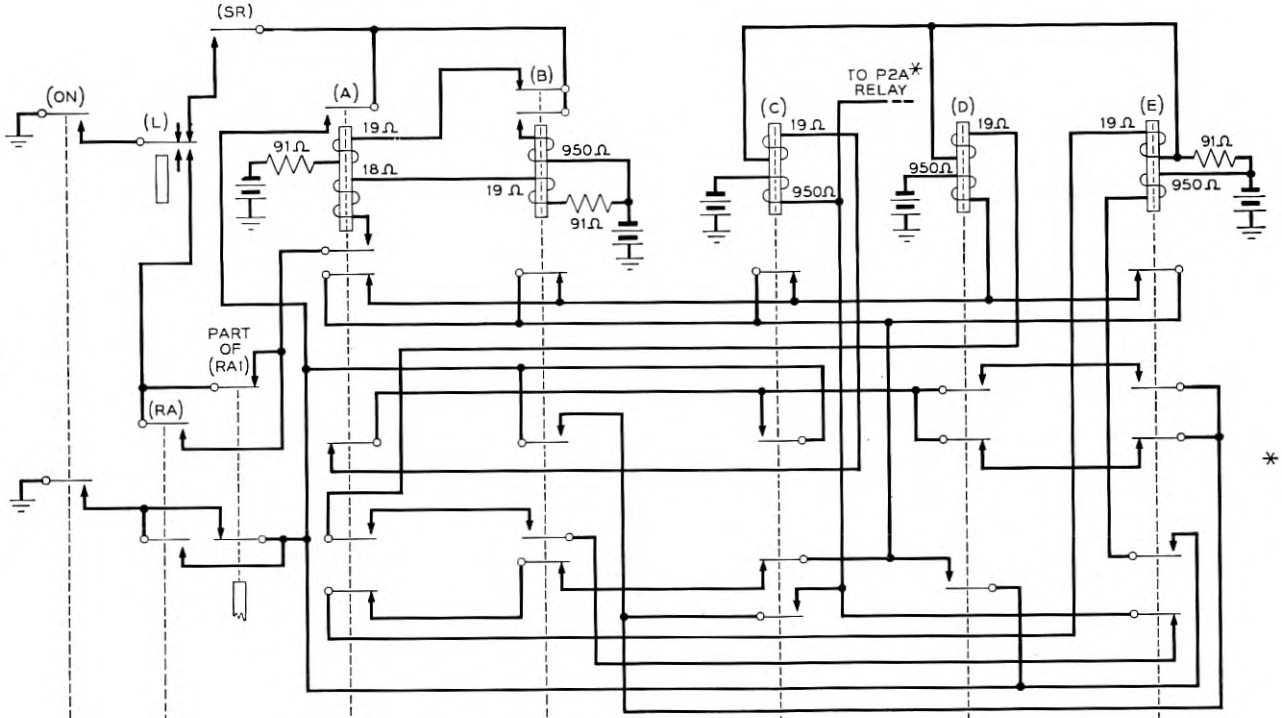
By attempting to manipulate the B relay control circuit into the same form, one obtains

$$\begin{aligned} f(B) &= (L + A)(B + L'A) \\ &= (L + A)(L' + B)(A + B) \qquad (3b) \\ &= [L' + B][A + BL] \qquad (3b) \\ &= [L' + B][A + B(L + B')] \qquad (11a) \end{aligned}$$

The schematics represented by the above algebraic expressions are shown in Fig. 8a and 8b. The circuit of Fig. 8c is obtained by combining the first two circuits so that only a single transfer is needed on the L relay. Note however that it is necessary to make the lower two B transfers have continuity action to insure proper functioning. Fig. 8d shows the pulse divider drawn in conventional form.

CONCLUSION

As far as is known, the dial pulse counting and translating circuit described herein requires fewer relays than any other circuit with similar



CONTROL CIRCUIT

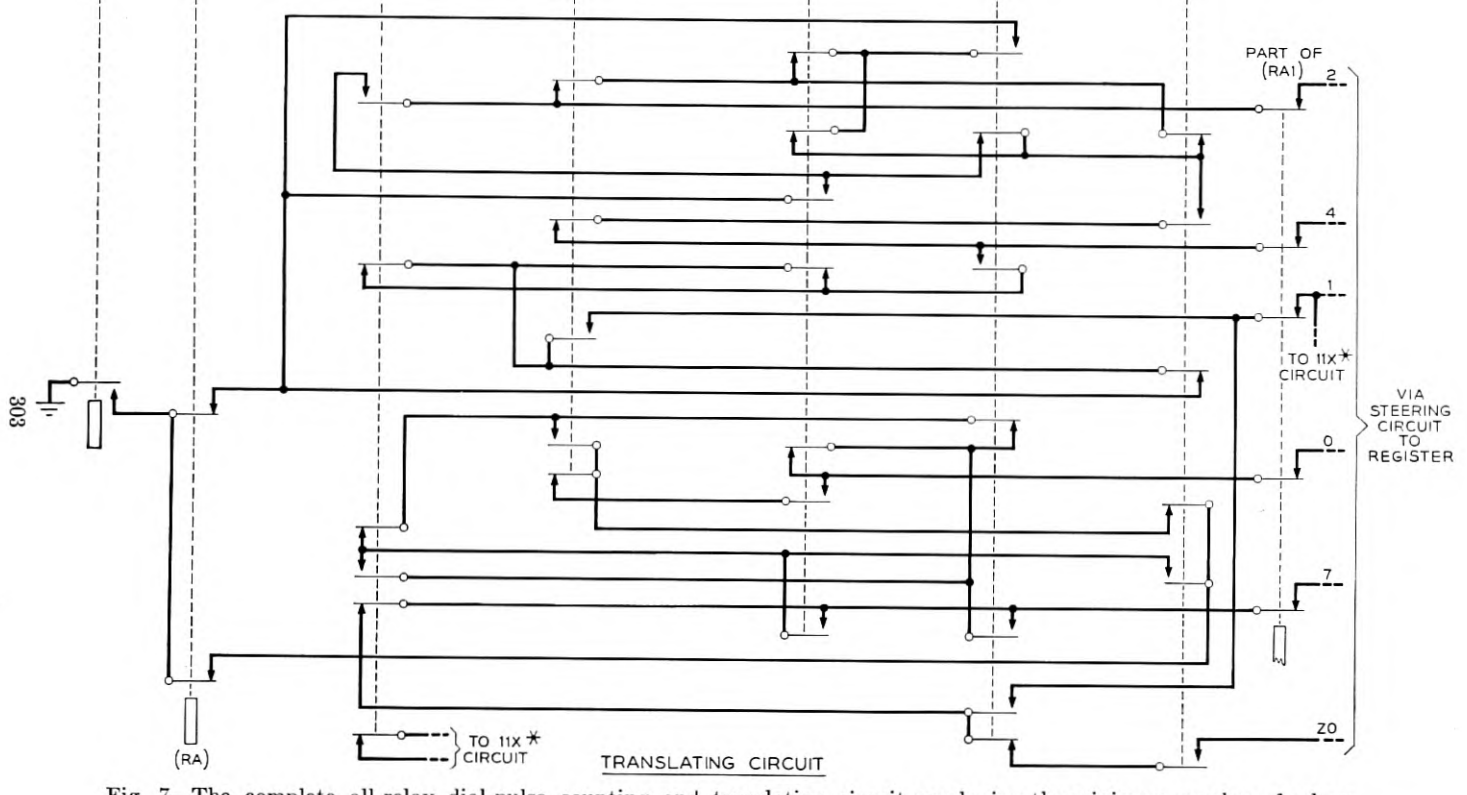


Fig. 7—The complete all-relay dial-pulse counting and translating circuit employing the minimum number of relays.

functions at present employed in Bell System standard switching equipment. The previous dial pulse counter used in the latest crossbar system required a total of ten relays. Thus the present design represents a considerable saving in cost and space. To a certain extent this result can be ascribed to the use of switching algebra during the circuit development.

Relay circuits designed on the basis of utilizing a large proportion of the possible combinations permitted by the component relays usually require heavy spring pile-ups. Since general purpose relays are limited in the number of springs which they can carry, this type of circuit usually

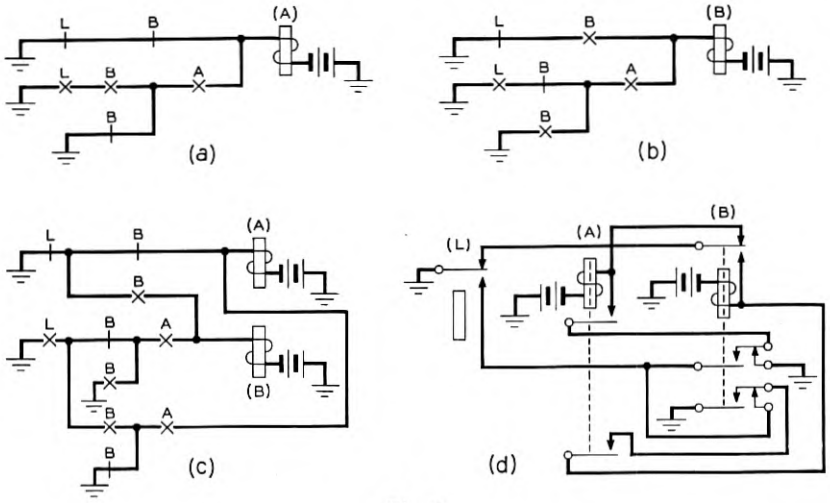


Fig. 8

entails considerable design effort to make most effective use of the available springs. Application of switching algebra to this aspect of the design problem can often provide crucial assistance.

It is recognized that switching algebra, in its present state of development, does not permit complete mathematical statement and manipulation of multi-terminal networks as represented by the counting and translating circuits. It does provide, however, facilities in manipulating two-terminal networks into a variety of forms from which can be selected those that combine most readily. This can result not only in a saving of time, but also in improved circuits which might not be realized by other design techniques. Unfortunately the algebra in its present state does not indicate when the optimum circuit has been attained. To some extent this is caused by apparatus or circuit considerations to which, since it is

concerned solely with contact networks, the algebra does not apply. Thus, there is still considerable room left for the ingenuity and judgment of the switching circuit designer.

As a result of the experience in designing the dial pulse counter and translator, certain observations on the use of the algebra are believed to be valid. Although switching algebra may be used in the design of the simplest circuits, the most noticeable benefits are obtained by the application of the algebra to the design of those circuits in which the control and output paths are complex and interrelated. The particular minimum relay counting and translating circuit under discussion is an excellent example of this type of circuit. A secondary advantage of the algebra is its compact notation and its value as an efficient circuit "bookkeeping" method.

ACKNOWLEDGMENTS

The author is indebted to Joseph Michal who made invaluable contributions as co-designer of this circuit and also verified all the algebraic manipulations contained in this paper. The author also wishes to acknowledge the suggestion made by G. R. Frost as to the possibility of using the bridge network in the control circuit. This work was carried on under the supervision of L. J. Stacy and F. K. Low whose many valuable suggestions were incorporated into this development.

Interaction of Polymers and Mechanical Waves

BY W. O. BAKER AND J. H. HEISS

(Manuscript received October 19, 1951)

New techniques of Mason, McSkimin, Hopkins and co-workers for generation of shear waves over the frequency range 2 × 10^2 to 2.4 × 10^7 cps have been used to study mechanical properties of chain polymers. Polymer solids, melts and dilute solutions, representing the main states in which plastics and rubbers are fabricated or used, were explored to find the characteristic relaxation times, rigidities and viscosities of various chemical structures. Polyisobutylene, hevea rubber, polydimethyl siloxane, vinyl chloride-acetate copolymers and plasticized nitrocellulose were compared with polyethylene and polyamides as examples of the range of solid properties encountered.

As melts, several polyisobutylenes, polybutadiene, polypropylene, polypropylene sebacate and poly-α-methyl styrene were investigated as models for varying degrees of chain substitution. Chain rigidity in, for instance, polyisobutylene, seemed to reflect visco-elastic over-all configurational changes up through the kilocycle range, but nearest neighbor interactions took over in the megacycle region, leading to moduli of 10^9 dynes/cm^2 even for syrupy fluids.

In dilute solution, polyisobutylene, polystyrene, natural rubber and polybutadiene microgel exhibited characteristic dynamic viscosities and rigidities depending linearly on concentration. Presumably, this reflects mechanical properties of isolated chains. Some possible models were suggested for the frequency dependence of such properties.

INTRODUCTION

The "equilibrium" mechanics of polymers, the giant molecules of plastics and rubbers, have been quite elegantly developed in the range of high strains ("kinetic theory" of elasticity—Meyer,[1] *et al.*). However, the molecular displacements as these strains, and, indeed, much smaller ones, occur, are little understood.[2] Nevertheless, it is essential to know about detailed motions in connecting chemical structure with physical properties. Only in this way can there be obtained from the chemical industry compositions which will serve properly in telephone apparatus.

Other studies have treated one way of getting at these mechanisms by relating stress relaxation, creep, viscosity, etc. to a distribution of molecular relaxation times (and energy barriers), as originated by Kuhn.[3, 4] Another approach is to strain polymers with periodic waves over a very wide spectrum of wavelengths, eventually going to frequencies comparable with those of the thermal vibrations of significant groups or segments in the macromolecules. The resulting dispersion or resonance phenomena can then be examined. Hence a mechanical radiation field can interact with the masses of elementary structural units, as the usual electromagnetic field interacts with atomic and group charges. In general, direct interpretations of this kind must be done with shear waves, and, at least, not *only* with longitudinal or ultrasonic waves.

This kind of study is now proceeding using waves generated and followed by piezoelectric crystals connected in as actual electromechanical circuit elements (A. M. Nicolson, 1919). Recent schemes of Mason and co-workers cover the frequency range from 10×10^3 to 60×10^6 cps, as reported in the paper by Mason and McSkimin in the last issue, while a tuning fork method used by I. L. Hopkins has been applied to "soft" polymers (rubbers) over the range 10^2 to 10^4 cps (the general range of J. D. Ferry's work at Wisconsin on concentrated polymer solutions).

The relation of these studies to the scientific and technical exploitation of plastics and rubbers is in knowing what a particular chemical composition does to strength, stiffness, ease of molding, impact toughness, etc. That is, are there qualities of the interaction of saturated aliphatic groups that make polyethylene or polyisobutylene have some glass-like as well as liquid-like, or rubbery, nature even at room temperature? If so, conditions causing brittle failures must be watched for. How is the storage of molecular strains in injection molded plastics reduced by increasing molding temperature (when the kinetic theory stiffness per chain actually increases)? These and many similar problems may be generalized under the headings below; in each case the chemical structure of the macromolecule appears to be reflected in relaxation times which combine in different ways to give flow or rigidity, toughness or brittleness.

Extrusion and Molding

Non-Newtonian flow leading to "frozen-in" stresses, subsequent distortion and irregular shapes of plastics[5] and rubbers,[6] implies energy

storage in the sheared molecules. The dynamic shear studies will confirm this. Also dispersion of carbon black and other pigments is restrained by elastic qualities of "liquid" polymers (i.e., instead of "mixing", compounds just microscopically deform and later re-form.) Likewise, the efficiency of compounding[7] and extrusion[8] depend on how quickly the molecules relax after straining.

Impact Strength, Brittleness and Tenacity

Toughness, mechanical shock resistance, ultimate elongation and strength reflect the facility with which the polymer molecules can be displaced without breaking the piece. Thus, they accommodate to the stress by motions presumably similar to those described above. (The situation is complicated when crystallites are also displaced.[9]) In any case, time sensitivity in the range 10^{-5} sec upward exists.[10, 11] The discussion by Morey[11] is a valuable survey of these ideas, and explicitly notes the significance of multiple relaxation processes on damping of shock waves. Evidence of the relation of simple changes in chemical structure to the principle relaxation times effective in these physical properties of plastics and rubbers is thus another part of the dynamics studies. The "brittle point", or volume-temperature transition of amorphous polymers,[12, 13] apparently reflects directly the correspondence of the time of experiment with dominant relaxation time of the polymer.[14, 15] A few measurements on plasticized polymethyl methacrylate (from which, however, no actual rigidities were calculated) indeed indicate abrupt stiffening as a function of frequency at a given temperature.[16] However, the changes measured were too small and indefinite to indicate any particular molecular relaxation. Other work[17] with plasticized polymers is nevertheless concordant with the current findings that molecular relaxations and not long range order determine embrittlement. The converse of this is, of course, that as some "transition" is approached, hysteresis, heat build up, flex cracking and fatigue are greatest.

Creep, Stress Relaxation and Recovery

Even these "long time" qualities of plastics, such as found in cold flow, apparently result from integrated displacements of rapidly oscillating segments of the chains. A most interesting analysis of stress relaxation in rubbers employs Kuhn's suggestion of a particular distribution of relaxation times.[18] The present point is that, again, these relaxation times reflect processes which should appear directly in reaction of the polymer with high frequency shear waves.

From these aspects above the current results of dynamics studies will be reviewed.

POLYMER SOLIDS: OVER-ALL MECHANICS

Solid polymers will denote rubbers and plastics in the state in which they are technically used. This is usually their most complex form, with inter- and intra-molecular factors undistinguished. Thus, separation and identification of the main relaxation processes are difficult or impossible. However, it is interesting to consider typical values of modulus and viscosity as related to chemical structure, in the range of frequencies corresponding to extrusion rates, and stresses in actual use.

These values of dynamic modulus and viscosity are distinct from the usual quantities in the literature. The usual expressions are for longitudinal (sound) waves, and give dynamic Young's modulus[19]

$$E^* = E_1 - iE_2$$

E_2 measures the out of phase part of the force-displacement relation, and $E_2 = \omega \cdot$ ("effective viscosity coefficient"). Now, the general elastic constants are $\lambda + 2\mu$, with λ = Lamé's constant and μ = shear modulus. Here,

$$\lambda + 2\mu = K + \tfrac{4}{3}\mu,$$

with K = bulk modulus. Alternately,

$$E_1 = \frac{3K}{\lambda + \mu}\,\mu = \frac{3\lambda + 2\mu}{\lambda + \mu}\,\mu.$$

However, in general the present results lead to the simpler shear modulus μ. Further the energy losses studied are expressible directly as the usual shear viscosity

$$\mu' = \eta.$$

Previous comprehensive studies of the dynamics of rubbers over significant frequency ranges have yielded loss factors either written as E_2/E_1 (see above),[19] or as a function of the shear viscosity based on Stoke's assumption that the compressional (dilatational) viscosity is zero.[20] But as Nolle[19] and Ivey, Mrowca and Guth[20] clearly recognize, recent work has strongly manifested the presence of compressional viscosity in simple liquids[21] as well as polymeric ones.[22, 23] Hence, the present understanding relating molecular structure to viscosity, plasticity and visco-elasticity is unsuitable for interpreting mechanical wave motion more complex than in shear, unless shear constants are also known.

This sums up to mean that the *chemical* interpretation of basic polymer mechanics requires shear wave measurements. Nevertheless, fascinating evidence of the existence of fine-structure relaxations in polymer solid has come from longitudinal wave investigations.[19, 20, 24, 25, 26] Also, the pioneering shear wave studies of Ferry and collaborators[27, 28] on concentrated solutions of polymers have suggested intrinsic relaxations of the chain molecules in a highly plasticized "semi-solid" state.

The more simplified findings cited below will be seen to unify approaches in this field. Comment must first be made, however, on formulation of experimental results in dynamics of polymers.

Expression of Dynamic Properties

Alternate and equivalent expressions have been thoroughly surveyed;[29] all represent combinations of either Maxwell (series) springs and pistons (elasticity and viscosity) or Voigt (parallel) springs and pistons. Obviously, there is no physical separation of elastic and viscous elements in a polymer molecule, so the irrelevance of the *detail* of the model need not be emphasized. However, the models lead to convenient formulation of *relaxation times* which dielectric studies, in particular, have shown have clear connections with chemical structure. In this chapter, sometimes one and sometimes the other model, or combination, will be used, with the symbols shown on the next page.

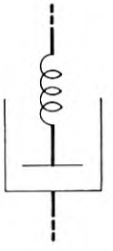
Other symbols are sometimes used,[30] but should be easily identified in terms of the above.

Rubbers and Soft Plastics

In Table I, the shear moduli of rigidity, μ, and of viscosity, μ', are shown as calculated for the Kelvin-Voigt model, for polymers having the indicated units of structure. The frequencies are from a few hundred to a few thousand cycles, hence, in the range of much technical use, (flexing of tires $\sim$300 cps) and rates of shear during processing.[31, 32] Data are from a general study by I. L. Hopkins[33] of the Bell Laboratories, based on a tuning fork transducer introduced by Rorden and Grieco.[34] The strains employed were always small, in the range 0.3 to 1.5 per cent; μ and μ' were essentially independent of strain, except for some loaded rubber stocks. The μ values clearly trace the magnitudes to be expected in going from the most typical rubber (hevea) to the semi-rigid plastics (vinyl chloride-acetate copolymer and plasticized cellulose nitrate). As anticipated from steady-stress observations the "plastics" have $\mu > 10^7$ dynes/cm^2. Increase of μ with frequency is also greater as

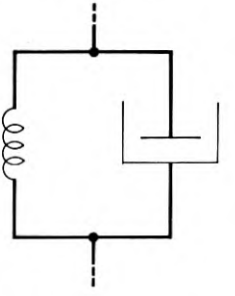
the "plastics" range is approached; a relaxation region is implied. Figs. 1 to 4 show the dispersion of rigidity with frequency in more detail. Especially striking in Figs. 1 and 2 is the small temperature dependence (at least between 27° and 66°C) of μ. Because of experimental uncertainty, μ cannot be said to be actually higher at the higher temperatures in accord with straight kinetic theory, but at least it is strongly tending that way, as also noted for lower frequencies studies on natural rubber.[19] Nothing like this appears for the plastics; in plasticized nitrocellulose the 100-cycle rigidity decreases 10-fold from 27° to 66°C. This is, then, the second general dynamic quality which reflects the low van der Waals' (dipole, dispersion and induction) forces in hevea rubber and polydimethyl siloxane, as well as their intrachain flexibility. Interchain forces in polyisobutylene (Butyl rubber) are low too, but barriers to flexibility because of sterically hindered-CH_3 groups come in. Table I and Fig. 3

MAXWELL



σ = strain
S = stress
t = time
τ = relaxation time
τ' = retardation time
$\mu = G$ = modulus
$\mu' = \eta$ = viscosity

KELVIN-VOIGT



$$\frac{d\sigma}{dt} = \frac{1}{\mu}\frac{dS}{dt} + \frac{S}{\eta}$$

$$S = S_0 e^{-\frac{\mu t}{\eta}} = S_0 e^{-\frac{t}{\tau}}$$

$$\tau = \frac{\eta}{\mu}$$

For const. S, $\frac{d\sigma}{dt} = \frac{S}{\eta}$ or $\eta = \frac{S}{\frac{d\sigma}{dt}}$

There is same stress on each element; the total strain = sum of single strains.

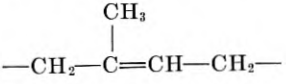
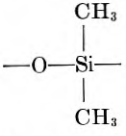
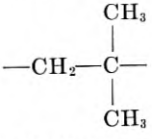
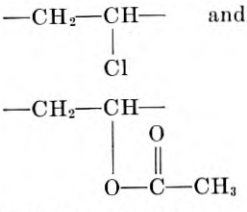
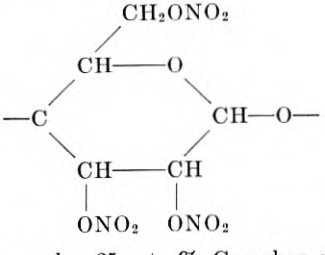
$$\frac{d\sigma}{dt} = \frac{S}{\eta} - \frac{\mu}{\eta}\sigma$$

$$\sigma = \frac{S}{\mu}\left(1 - e^{-\frac{\mu t}{\eta}}\right) = \sigma_0 e^{-\frac{t}{\tau'}}$$

$$\tau' = \frac{\eta}{\mu}$$

There is same strain in each element; the total stress = sum of single stresses.

TABLE I

Polymer Unit	Shear Modulus, μ, dynes/cm² 27°C		Shear Viscosity Poises, μ', 27°C	
	100 cycles	5000 cycles	100 cycles	5000 cycles
Hevea rubber 	3×10^6	5.5×10^6	350	40
Polydimethyl siloxane 	0.7×10^6	1×10^6	300	30
Polyisobutylene 	5×10^6	30×10^6	8,000	1,500
Polyvinyl chloride (~92%)-acetate (~8%) plasticized by ~36% di-octylphthalate 	13×10^6	80×10^6	25,000	2,000
Cellulose nitrate  and ~25 wt. % Camphor plasticizer.	60×10^6	250×10^6	80,000	4,500

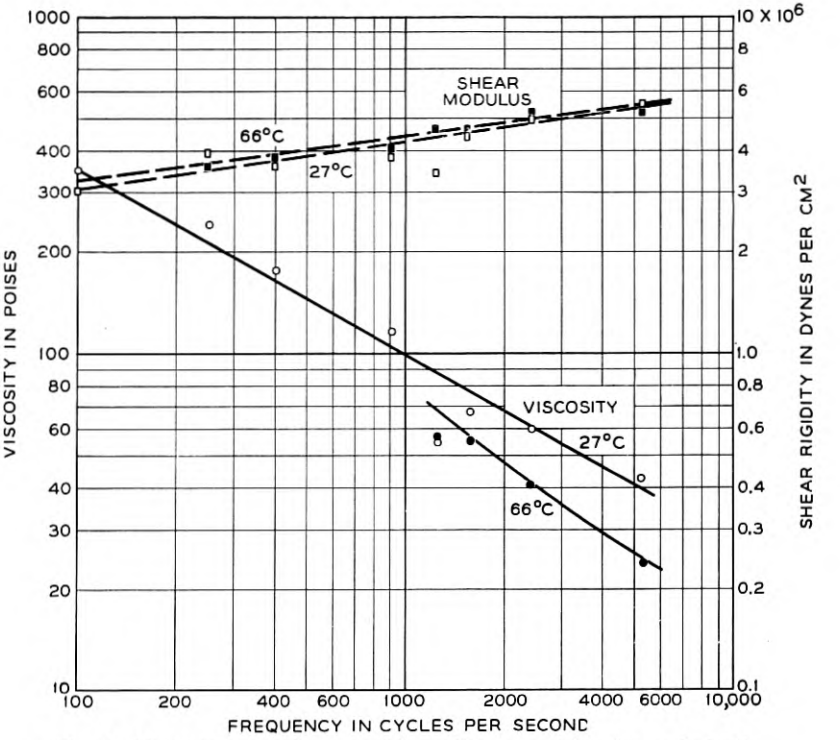


Fig. 1—Viscosity and shear modulus of hevea rubber (cross-linked).

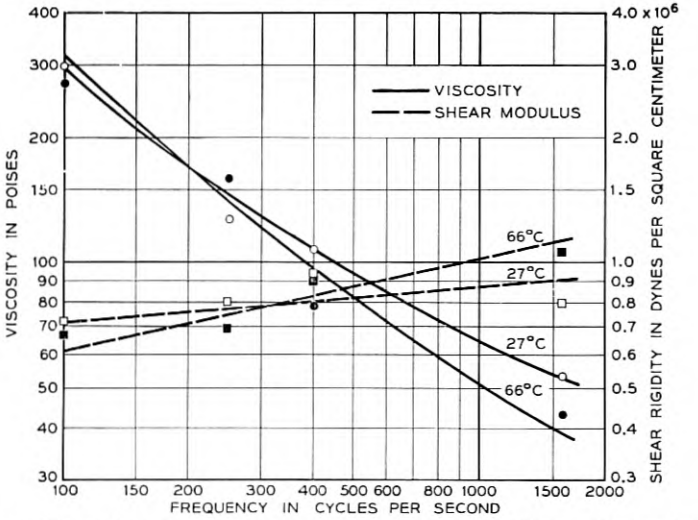


Fig. 2—Viscosity and shear modulus of polydimethyl siloxane (cross-linked).

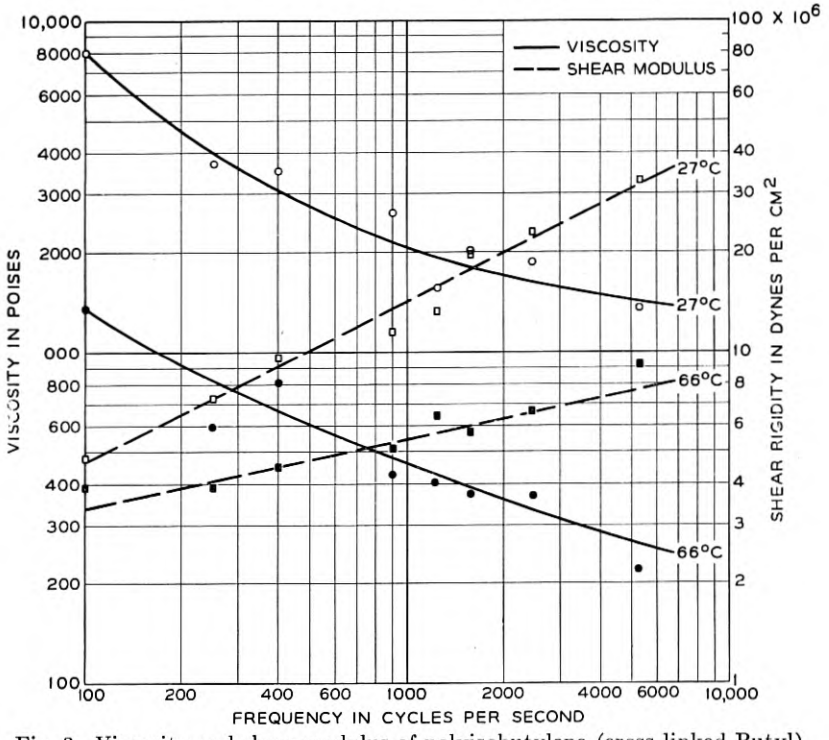


Fig. 3—Viscosity and shear modulus of polyisobutylene (cross-linked Butyl).

emphasize that thinking about the mechanics of a particular chemical structure must include the spatial relationships of groups within the chains, as well as between them.

The dynamic viscosities in Table I are also in accord with the sequence of structures. Their frequency dispersion again connotes varying relaxation processes. Natural rubber's low inner friction, for both compressional[35] and shear waves[36] is famous in its low hysteresis heating. (This unique property is geopolitically crucial, because adequate truck and bus tires cannot yet be made of any other rubber.) Indeed, it is striking that at 100 cycles, a piece of gum rubber has a local viscosity of only 350 poises. The silicone rubber gum also has high elastic efficiency, and its temperature coefficient of viscosity is very low (see Fig. 2), like the thermal coefficients for familiar silicone liquids. It is exciting to speculate in Figs. 1 and 2, whether more precise measurements which Hopkins is now undertaking will confirm the apparently *negative* temperature coefficients of viscosity at some frequencies. "Kinetic theory

viscosity" arising from transfer of momentum among thermally agitated chain segments, does not seem to have been considered in the theory of perfect rubbers. As in gases, it would require an increase of viscosity with temperature.

In polyisobutylene, however, the dynamic viscosity leaps upward in both magnitude and temperature dependence. It should be emphasized that this is, again, for a cross-linked (Butyl) gum—an infinite network like the hevea gum, with presumably infinite macroscopic viscosity. The striking thing is that this internal viscosity is not greatly dependent on the network, at the degrees of "cure" used in rubber technology. For instance, recent studies over the frequency range 20–600 cycles, on high molecular weight, $\bar{M}_\eta = 1.2 \times 10^6$, linear polyisobutylene,[37] give, at 25°C and 100 cycles, $\mu' = 4800$ poises, although the steady flow viscosity of this polymer at this temperature is greater than 3×10^9 poises.[38] Then, the infinite network ($\eta_{\text{steady flow}} \rightarrow \infty$) Butyl polymer of Fig. 3 has at 27°C and 100 cycles $\mu' = 8000$ poises. At 1000 cycles agreement appears to be about the same, and is tolerable considering the several

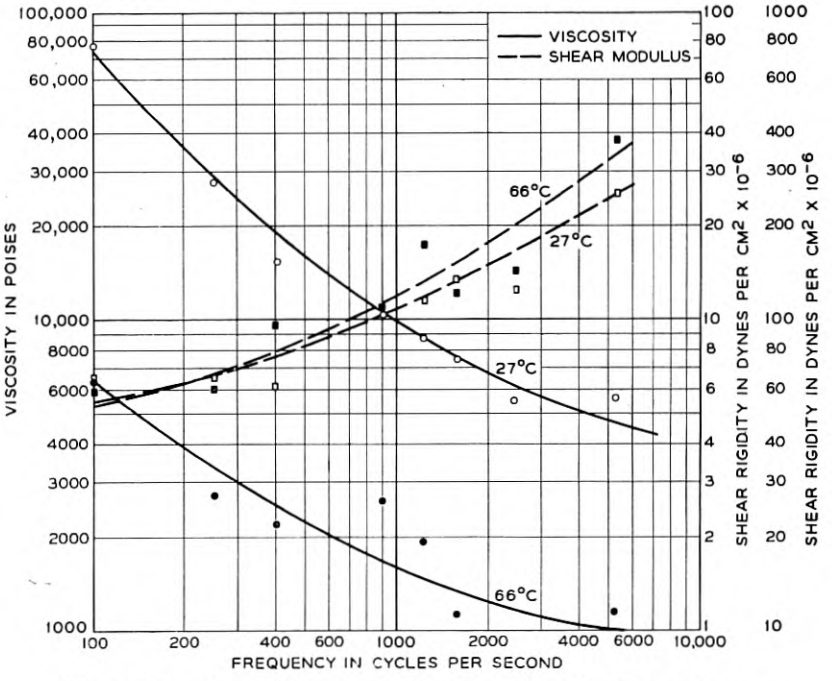


Fig. 4—Viscosity and shear modulus of plasticized cellulose nitrate.

per cent of compounding ingredients in the Butyl gum, and possibility of a small fraction of low molecular weight uncured polymer in it.

Also, wide variations in the degree of cure of Butyl gums were studied without large changes in μ'.[33] In this regard, the particular sample of Fig. 3 had an equilibrium swelling ratio (= volume swollen polymer in cyclohexane at 25°C/volume insoluble part of dry vulcanizate) of 4.84. This indicates an M_c value (average molecular weight between cross-links) of <20,000.[39] Actually many of the dynamic properties can probably be found in individual chain units or segments even smaller than this. This is a significant point in engineering applications where plastics may be cured to reduce creep but where it is desired to retain typical "chain" properties to increase impact toughness. That is, usually some optimum condition for this compromise can be found. The later section on liquids will suggest that physical properites typically associated with chain polymers can indeed reside in even shorter chain sections than the M_c's observed in usual gum rubbers.

Filled Polymers

Marked effects of carbon black and other pigments are of course familiar in both steady and alternating mechanics of rubbers.[19, 24, 35, 36] Brief comment on their influence on dynamic shear properties and thus relaxation mechanisms involved may be directed toward plastics, also, however. Thus, technologically it would be desirable to load thermoplastics with considerable volumes of "inert" fillers, just as is done with rubber. But, almost invariably strength and toughness decline, instead of improving, as in the rubber case. A reason for this appears in investigations by Hopkins when carbon black (a standard type of reinforcing black) was added to Butyl rubber of the sort described in Table I. It is that stiffness seems to rise more rapidly than internal viscosity—i.e., a given strain results in proportionately higher stress than the accompanying internal viscosity provides means for dissipating the stress (as on impact). Hence, the brittleness which fillers normally engender in thermoplastics may represent this change in μ vs. μ' balance. Table II illus-

TABLE II

Wt. Per Cent of Carbon Black in Butyl Stock	Shear Modulus, μ, dynes/cm² at 27°C		Shear Viscosity, μ', Poises, at 27°C	
	100 cycles	5000 cycles	100 cycles	5000 cycles
15.2	8×10^6	60×10^6	11,000	2,000
28.6	45×10^6	150×10^6	35,000	3,000

trates some values for Butyl compounds. The swelling ratio (SR) for the compound containing 28.6 wt. per cent filler has dropped to 3.2, implying also considerable reductions in M_c (since theoretically $(1/SR)^{5/3} = \frac{a}{M_c} - 2bM_c$[39]). Thus, the apparent chain segment between cross-links is shorter than in the unfilled stock (the two were cured to give closely similar degrees of primary valence cross-linkage) and correspondingly the steady-pull modulus is higher. Yet, the internal friction, while also higher, seems to reflect new relaxations from interaction with the filler, and total shock-absorbing power has declined.

Microcrystalline Polymers

The preceding studies at comparatively low frequencies indicated (1) magnitudes of shear rigidity and internal viscosity characterizing rubbers and soft plastics. By familiar shifts of temperature or frequency, they would also apply to polymers known as hard, amorphous plastics at room temperature such as polystyrene and polymethyl methacrylate. (2) Dispersion of μ and μ' with frequency affirm that the intrinsic or fine structure relaxations have times $<10^{-3}$ to 10^{-4} sec, and so refer to chemical units much smaller than the average molecules in the usual technical rubbers and plastics. A way to get at what sizes and habits these units might have will be by investigation of low molecular weight polymer liquids. But, while still in the section on solids, it is recalled that microcrystalline polymers such as polyethylene, polyesters (Terylene), polyamides (nylons), cellulose esters, polyvinylidene chloride, polyacrylonitrile etc., have mechanical properties dominated by their crystalline-amorphous ratios.[9, 26, 40, 41] The amorphous volumes are clearly those which donate the flexibility, toughness and shock-resistance of these plastics and textile fibers.[9, 40] An interesting point is, how "viscous" are the disordered chain segments? In an over-all sense, all kinds of dissipation including crystallite friction, analogous to grain friction in metals, scattering of longitudinal waves, and stiffening by low temperatures can occur in these polyphase systems. Thus, effects of chain orientation as well as lateral order (crystallinity) have been detected in dynamics studies.[26, 41] The intrinsically liquid-like or amorphous components of this behaviour—and the things which will correlate most simply with dipole concentration and other chemical features—are most accessible to study at very high frequencies. For, in these polymer solids, unlike the essentially continuous and homogeneous amorphous ones first discussed, the mechanics reflect small regions having widely divergent properties. Thus, methods developed by H. J. McSkimin of Bell Tele-

phone Laboratories (described in the last issue), have been used to probe for elemental reactions at the upper end of the frequencies presently available. Both longitudinal and shear waves were used. In polyethylene, a wavelength for the shear waves was 0.0074 cm., at $f = 8.55 \times 10^6$ cycles, and in polyhexamethylene adipamide (the usual 6–6 textile nylon), the shear wavelength was 0.0125 cm., for $f = 8.67 \times 10^6$ cycles, all at 25°C.

The important consequence of these experiments so far has been that, despite the small strains involved, the viscosity appears to be a "polymer" viscosity, rather than an inner friction involving just a few liquid-like atoms per unit. Thus, polyethylene of "equilibrium" crystallinity and average molecular weight corresponding to an intrinsic viscosity in xylene of $[\eta] = 0.89$ (at 85°C), was measured over the range from 0 to 50°C. The results from both longitudinal and shear wave measurements

TABLE III

Temp., °C.	f, cycles/sec	Viscosity Poises		
		μ'	$\lambda' + 2\mu'$	λ'
0	8×10^6	15	38	8
0	25×10^6	5	14	4
30	8×10^6	15	34	4
30	25×10^6	5	13	3

are given in Table III. These viscosities are expressed in this case for a Kelvin-Voigt model, of rigidity and viscosity *in parallel*. The rigidities associated with these viscosities are about 3×10^9 dynes/cm^2, or not far from the value under steady pull of about 1×10^9.

Now this suggests that the rigid plastic polyethylene retains, even under mechanical impulse of microsecond duration, a shock-absorbing capacity reflected in a shear viscosity of 5–15 poises, and a compressional viscosity of 3–8 poises. The former, μ', may roughly correspond to the liquid viscosity of a paraffin-like chain of from 50 to 65 c-atoms in length. Thus, the dynamics measurements seem to relate to basic premises of polymer structure. These are that the amorphous regions (whose existence is shown quite independently by x-ray scattering, density, heat-capacity, etc.) indeed take up and dissipate sudden stresses which the microcrystallites, despite their great strength, would be too brittle to sustain.

These results give hope that further probing of the dynamics of liquid-like elements in rigid plastics will eventually lead to precise adjustment

of molecular weight, chemical structure (degree of branching in polyethylene), crystallinity, etc. These quantities, when fitted to a given pattern of μ, λ, μ' and λ' at proper frequencies would yield plastics of optimum serviceability under the multitude of stresses encountered in use.

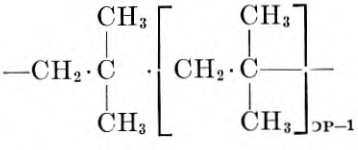
A similar liquid-like structure-even where the (crystalline) rigidity is much higher[40] and mobile chain segments smaller—apparently occurs in polyamides. Presumably the hydrogen bonding and dipole interactions are very imperfect in the disordered regions,[40] and there the chain interaction is reminiscent of polyethylene. For instance, in polyhexamethylene adipamide, measurements in the 8 to 30 megacycle range do indicate that the Lamé elastic constant λ is about 5.6×10^{10} dynes/cm^2, but only about 3×10^{10} for polyethylene. This reflects over-all stiffness dominated by crystallites. Nevertheless, the compressional viscosity, λ' is 17–6 poises (going from 8 to 30 mc) for the polyamide, but only 5–2 poises for polyethylene. Of course, since there is dispersion in both cases, these relative magnitudes might be quite different at some other frequency or temperature (all above are at 25°C). Yet it remains that the nylon, despite its hardness, also has a liquid-like component more viscous than that of polyethylene. Similar relations appear in the shear viscosities, μ', also determined for these two systems. For the 6–6 polyamide, μ' goes from 19 to 7 poises over the 8 to 30 mc interval while polyethylene changes from 15 to 5. These quantities indicate again, as with the polyethylene, that "polymer liquids" rather than just a few small groups of atoms are the important mechanical elements even at frequencies of 10^7. Now polystyrene, an amorphous polymer, also has rigidities of about 10^{10} dynes/cm^2 but the μ' and λ' values at room temperature are far below 5 to 20 poises, and glass-like brittleness (although not so bad as silica glass) results.

So far, then, the two characteristic extremes of polymer mechanics have been discussed: (1) the soft rubbers, whose dynamics at low kilocycle frequencies imply, at ordinary temperatures, predominantly overlapping combinations of relaxation processes whose relaxation elements involve many segments per molecular chain; and (2) the hard, microcrystalline plastics, whose behaviour is predominated by relaxation processes having times of 10^{-6} to 10^{-7} sec because the longer period (slower) displacements have been relaxed out at the temperatures of normal use. (Likewise, interconvertability by temperature[19] between these two extremes is presumed. Also, a certain correspondence between dielectric and dynamic relaxations in these classes is indicated.[41a]) Next, it is in-

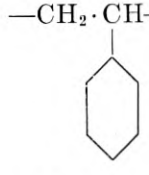
teresting to see what are the simplest structures (particularly in terms of molecular weight) yielding these effects. In other words, what kind of liquid really exhibits "polymer mechanics?" No detailed answer to this can be given below, but results on some polymer liquids of low average molecular weight will indicate that the mechanisms in rubbers and plastics are probably more general than previously supposed.

POLYMER LIQUID MECHANICS

By techniques described in detail elsewhere,[22,23] a series of polyisobutylene liquids have been investigated. These polymers were made by ionic catalyzed mass polymerization at reduced, but not very low, temperatures. While no great care to purify the monomer was used, such polymerizations require fair purity to go at all. Seemingly, the resulting liquids do represent a polymer homologous series, although head-to-tail sequence of the monomer units, some single ethyl rather than paired methyl side groups, etc., may differ slightly from the higher molecular weight forms in Butyl rubber and polyisobutylene gum. Whatever are these details, it appears that the polymers represent a linear hydrocarbon chain, with essentially two methyl groups on every other chain atom:

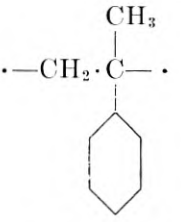


By contrast, polyethylene, with the nominal chain $—CH_2 \cdot CH_2 \cdot CH_2 \cdot CH_2—$ and to a lesser extent polystyrene,



have chains in which rotation about the bonds is less sterically hindered. The final section, on isolated polymer chains (in dilute solution), will consider this aspect further. However, some results will be reported below on a low molecular weight poly-α-methyl styrene, which may be considered structurally a cross between the rubber, polyisobutylene, and

the plastic, polystyrene,



Other studies in progress on liquid polybutadiene, polyisoprene, polypropylene, and polypropylene sebacate from which further information about intra-chain stiffness may be derived, will also be noted.

Properties of the polyisobutylenes studied are summarized in Table IV, some additional molecular weights in this range appear as extra points in some of the high-frequency graphs. The molecular weights $\bar{M}_\eta$ are "intrinsic viscosity" averages[42, 42a] and, with reasonable estimations of the $\frac{\bar{M}_n}{\bar{M}_\eta}$ ratio, check with cryoscopic number average, $\bar{M}_n$, values on such materials, which are in turn listed in the table as expressed by melt viscosity relations of Fox and Flory.[42a] These molecular weights repre-

TABLE IV

Polyisobutylene Polymer	$\overline{DP}_\eta$	$\bar{M}_\eta$	$\bar{M}_n$	25°C viscosity Poises	Maxwell μ	Voigt μ' Poises	Maxwell μ' Poises	Freq. Cycles
A	10	565	318	0.37	3×10^8	0.6	0.6	14×10^6
A″	30	1660	697	39.6	6.2×10^6	16.5	18.8	2×10^4
					1.7×10^9	7.9	10.0	14×10^6
B	45	2520	1070	216	3×10^9	15.2	24.2	14×10^6
C	56	3350	1720	737	3.6×10^9	20.2	47.9	14×10^6
D	74	4170	2530	1840	4.5×10^9	23.4	78.9	14×10^6
E	147	8240	4850	4600	5.3×10^9	27.2	92.3	14×10^6

sent reasonable averages rather than absolute values for these heterogeneous polymers. The $\overline{DP}$ values are just the number of isobutylene units per average chain. The η values are the steady flow viscosities at low rates of shear—usually determined by a falling ball.

Rigidity and Viscosity Magnitudes

The properties of these liquids ranging from polymer A having only forty times the viscosity of water to E, which begins to approach fluidities of technical polymer melts (polyamides, for instance), were explored

in the kilocycle range with shear waves generated by torsional crystals and in the megacycle region by shear waves with the reflectance method and by longitudinal (ultrasonic) waves from a pulse propagation technique.[23] The results have been expressed in two ways. First, in earlier reports,[22, 23] a trend corresponding with experiment was given by two Maxwell elements arranged in parallel. This result is too simple compared to the distributions of relaxation times previously proposed for high molecular weight polymers to reproduce *detailed* observation. Nevertheless, perhaps because of the smaller molecules involved, there seems to be clear indication that two *principal* relaxations predominate the mechanical reactions of these liquids over the range of frequencies of present interest, 10^2 to 10^7 cps. For example, for polymer D, these are:

First Relaxation	Second Relaxation
$f_c \sim 4 \times 10^3$ cycles $\mu \sim 4 \times 10^7$ dynes/cm^2	$f_c \sim 5 \times 10^6$ cycles $\mu \sim 6 \times 10^9$ dynes/cm^2

(In accounting for the second main relaxation, a hysteresis component had to be introduced whose significance has been suggested.[22])

Second, specific values of shear rigidity μ (Maxwell) and μ (Voigt), shear viscosity μ' (Maxwell) and μ' (Voigt) as well as the constants for related compressional wave systems, $\lambda + 2\mu$ (elastic) and $\lambda' + 2\mu'$ (viscous) have been calculated for particular frequencies. Unlike in the first way of expression, these latter quantities are all highly frequency dependent. However, they describe conditions at various frequencies of interest, and are thus often worthwhile.

Both ways of looking at the data lead, as implied by the figures above, to the proposal that typical polymer *stiffness* (shear rigidity of $\sim 10^7$ dynes/cm^2) is present at $\bar{M}_\eta \sim 1600$, with $\overline{DP}_\eta \sim 30$, or an average chain length of about 60 carbon atoms. This appears when the straining is done in 10^{-3} to 10^{-4} sec. In the 10^{-6} to 10^{-8} sec range, rigidity occurs for even an average chain length of 20 atoms as shown in Table IV.

STRUCTURAL FACTOR IN LIQUID MECHANICS

The main relaxations in the kilocycle range in polyisobutylene liquids seem to lead to quasi-configurational elasticity. This is where the kinetic theory tendency for a most probable separation of chain ends is retarded by viscous interaction of segments between and within the chains. Hence, the middle dashed curves of Fig. 5, showing shear elasticity for some of the polymers of Table IV, decrease exponentially with increasing tem-

perature. While pure kinetic theory elasticity would give a modulus *increasing* linearly with increasing temperature, these systems, like all practical rubbers and plastics, actually grow softer with rising temperature when deformed dynamically. It is striking, nevertheless, that a modulus of $\sim 10^7$ dynes/cm^2 seems characteristic of the visco-elastic energy storing of these simple polymer structures. As noted below this is 10^3 less than the crystal-like, close-packed, stiffness found for these same molecular frequencies above their second principal relaxation time.

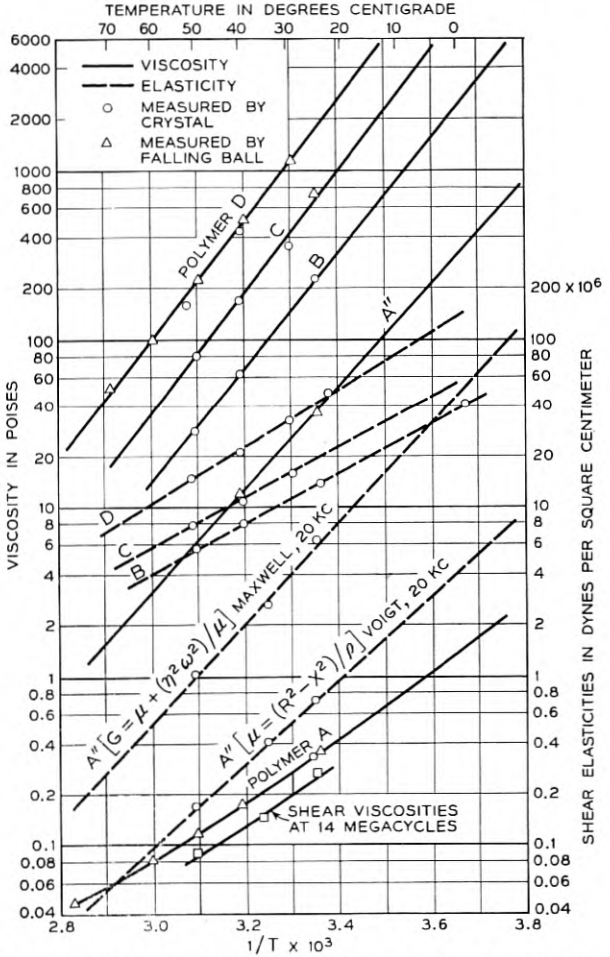


Fig. 5—Shear elasticities and viscosities of polyisobutylene liquids assuming a Maxwell model with relaxation frequencies 10^3–10^4. (Lower dashed curves for frequency-dependent models.)

Thus, it seems that the former, 10^7, modulus is typical of the structural arrangements in polymers said to be above their second order transition temperatures[12, 13] while the second, 10^9, modulus reflects interactions below the freezing-in.

These conclusions obtain regardless of the particular expression of the data. But, for comparison, curves are shown on Fig. 5 for a polyisobutylene A″ in which the dynamic modulus μ at *20 kc* is computed for both Maxwell and Voigt elements. The two points denoting the steady flow viscosity of polymer A″ rank it with respect to the others in the series.

Apparently even very fluid polymer melts, chain molecule plasticizers, and small segments of long molecules must be expected to show appreciable rigidity when stressed rapidly. Referring to the introduction it is reasonable that rough extrusions, frozen-in molding stresses and the like are so easily produced. The lines of Fig. 5 are not, of course, implied to be linear over any considerable temperature range. In the region represented, the temperature coefficient for viscous flow is about 16 kcal for the B, C and D liquids (about 12 for A). This agrees roughly with the steady flow values found for very high molecular weight polyisobutylene.[38, 42a] The temperature coefficient for the rigidity is less, as would be expected, since the whole center of gravity of the chain need not be displaced, but only local segments.

This quasi-configurational elasticity is increased by molecular weight (although kinetic theory elasticity of chain segments in a network is decreased by increasing segment length). The log μ vs density at 25°C plotted in Fig. 6 indicates that the chief influence is the number of chains per cc, since the points for all the molecular weights now lie on a single line. It should be repeated that the elasticity modulus plotted, μ, is again for a roughly frequency-*independent* or "absolute" model.[22, 23] The same is true for the three solid lines on Fig. 6, showing μ in the second, or 10^7 cycle, relaxation range. Here effects of detailed liquid structure come out; the three average molecular weights no longer lie so nearly on a single line. This elasticity is presumably from the crystal-like interaction of nearest-neighbor segments. If temperature is adjusted so that densities are the same, it is seen that the *lower* average molecular weight liquid has the *higher* elasticity modulus. This difference is not large, and should not be interpreted as showing an equal segment interaction, for a polymer of lower specific volume (B compared to D). Rather, it emphasizes in this relaxation range, approaching the "glass" behaviour, that the relaxation rate is vastly more temperature dependent than the specific volume change alone, and structural variations in the

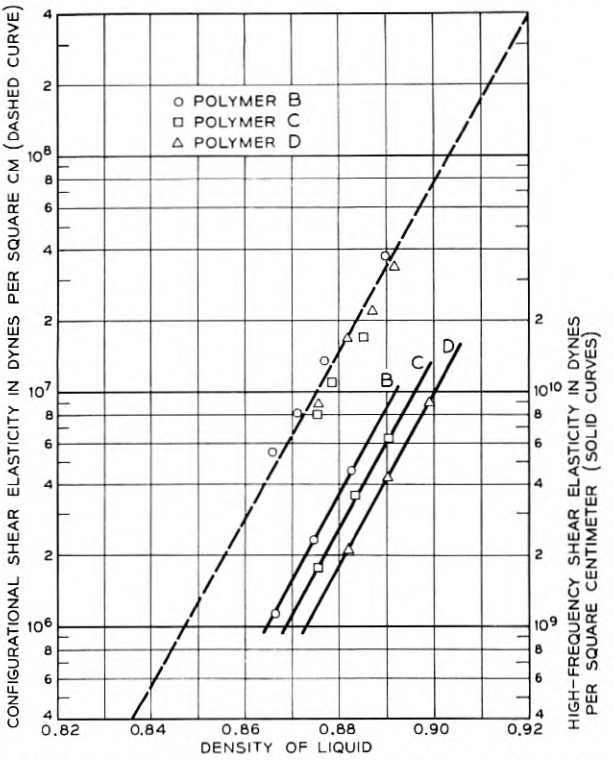


Fig. 6—Principal shear elasticities of polyisobutylenes as related to their densities.

packing of segments in the liquid (coordination number, etc.) become important.

"Soft" and "Hard" Liquid States; Second Order Transitions

A recent noteworthy study[43] of volume-temperature and viscosity-temperature changes in polystyrene (with a note on polyisobutylene) brings out many points in common with ideas of polymer liquid structure indicated by the dynamics work.[22, 23] Particularly, the fact that according to steady state measurements, the "local configurational arrangement of the polymer segments"[43] below T_g remains fixed accords with the postulations from dynamics work. That is, above the second main relaxation, it seems to be just the interactions in these fixed arrangements which cause the glassy (or "crystalline") dynamic modulus of 10^9 to 10^{10} dynes/cm^2. Further, the point that T_g is *not* an isoviscous state for polymers[43] agrees with the dynamics result that macroscopic

viscosity of the polymer has relatively little to do with the actual values of dynamic viscosities. These would be at frequencies where the response of the polymer liquid to the mechanical field is determined only by motions within the local fixed arrangements mentioned above.

Fig. 7 illustrates this, where on one scale the macroscopic viscosity is plotted according to the familiar log-log relation with molecular weight. Two extremes of average molecular weight, $\bar{M}_n$ and $\bar{M}_\eta$ are used for the liquids, to show that the molecular weight distribution does not alter the general conclusions. ($\bar{M}_\eta$ is an upper limit weight average figure.) On the other scale, the dynamic viscosity μ', in this case for a single element *frequency-dependent* Voigt model, shows low values and marked curvature. These betoken the relaxation in which molecular weight, through its effect on free volume and other structural factors, is signif-

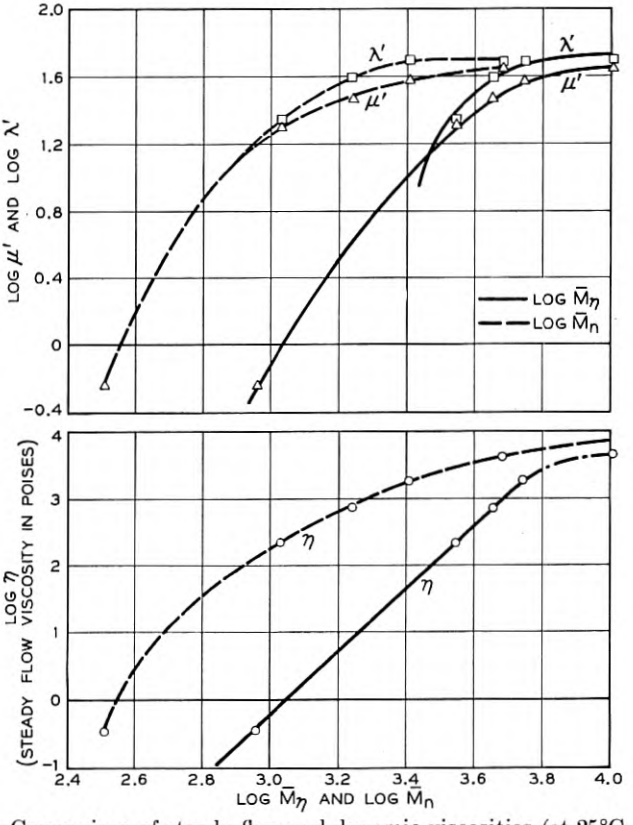


Fig. 7—Comparison of steady flow and dynamic viscosities (at 25°C and 8 mc) of polyisobutylene liquids of different molecular weights.

icant for displacements superficially quite different from those in macroscopic viscosity.

The compressional viscosity, λ' is also plotted, for the same model, in Fig. 7. It is, within experimental error, zero for polymer A′, as determined by shear and compressional wave studies at 8 mc frequency.[22] This is a rare case, then, where the attenuation of sound waves through a liquid has been quantitatively accounted for by the shear viscosity. But, as soon as the average molecular weight rises to 1000 or so, λ' comes in clearly, and the new mechanism for dissipating compressional or dilatational stresses is developed. As this presumably represents directly free volume or coordination number changes in liquid structure,[44, 45] its detailed study near T_g,[43] and in connection with brittle points of rubbers, may eventually be especially fruitful.

Another depiction of influence of average molecular weight in these liquids on dynamic viscosities occurs in Fig. 8. Here, the λ_c' curve is

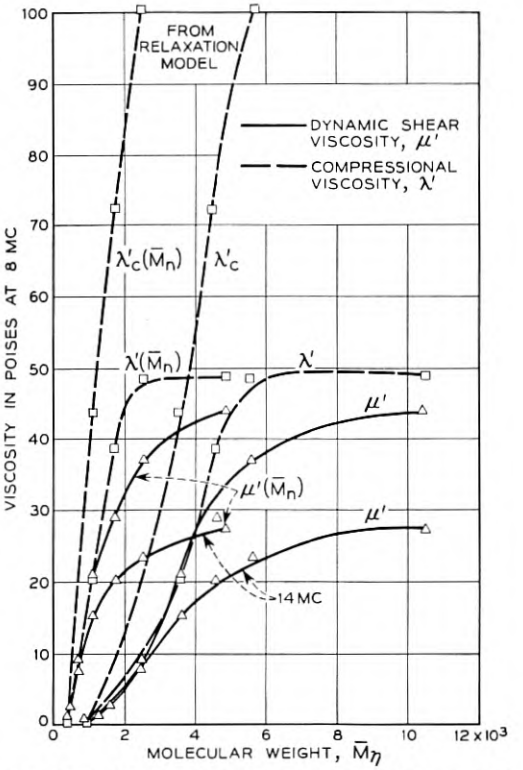


Fig. 8—Dynamic viscosities of polyisobutylene liquids as function of average molecular weight (25°C).

for, again, a crude model attempting to show compressional viscosity over the whole frequency range, while the other viscosities are Voigt expressions at 8 (or 14) mc. Extremes of molecular weight averages are shown.

Comparison of the "soft" or quasi-configurational rigidities, expressed, like the μ of Fig. 5 as relatively frequency independent μ_c, with the "hard" or glassy rigidities is given in Fig. 9. The λ and μ values are for the Voigt model at 8 mc. The graph does not show the bend-over of the "soft", μ_c, curve with molecular weight, but that happens more gradually. The "hard" rigidities λ and μ quite readily show this inflection. As before, the relaxing segments must be <100 chain atoms, according to the behaviour of the molecules at room temperature.

Concerning influence of molecular weight on engineering "brittle points" of such importance in rubber technology, the present studies agree with earlier proposals. Thus, although the T_g or v-T second order transition point always decreases with decreasing molecular weight,[43]

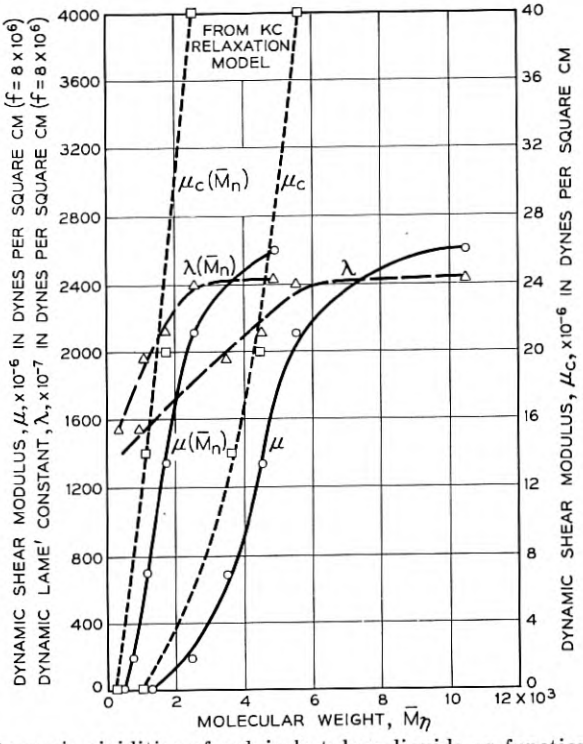


Fig. 9—Dynamic rigidities of polyisobutylene liquids as function of average molecular weight (25°C).

the brittle point tends to increase as the molecules get smaller. This was supposed to be because the ultimate elongation (doubtless due to viscoelastic or quasi-configurational elasticity and not kinetic theory elasticity as sometimes said) declined with chain length, so that the specimen broke at lower and lower strains,[13] even though it was not really in the glassy state. Now, the results above[22, 23] demonstrate that the shear modulus at a given temperature does fall off for average molecular weights of polyisobutylene below ~5000. Hence, the mechanical embrittlement of the low molecular weight samples is not because they are stiffer, but because they are weaker.

POLY-α-METHYL STYRENE: A "PLASTIC" LIQUID

Most of the liquid studies have been on polyisobutylene polymers, made "hard" or "soft" by temperature or frequency, but under use

TABLE V

Poly-α-methyl styrene	η Poises	μ', 14 mc, Poises		μ, 14 mc dynes/cm²	
		Maxwell	Voigt	Maxwell	Voigt
I	242	111.4	23.6	5.1×10^9	4×10^9
II	4340	502.7	14.3	7.6×10^9	7.4×10^9

conditions, considered rubbery. If one of the methyls in polyisobutylene is replaced by a phenyl, poly-α-methyl styrene, a hard plastic is produced. Low molecular weight polymers of this composition are, however, liquids at room temperature. Hence, it is interesting to compare their reaction to mechanical waves with that of polyisobutylene liquids of similar macroscopic viscosity. Table V lists a few properties at 25°C.

Polymer I has roughly the steady flow η of polyisobutylene B. Also, the Voigt μ' at 14 mc is similar: 15.2 poises compared to the 23.6 poises of the poly-α-methyl styrene. However, the Voigt μ is already 50 times higher for the phenyl substituted chain. (This shows the shift of the *second* relaxation range, of course, where nearest neighbor interactions rule.) Even more striking, the temperature coefficient of a μ_c', calculated as the second principal shear viscosity in a frequency independent model,[23] as before, is about 24 kcal, compared to about 12 for μ_c' for polyisobutylene in its similar model. Thus, although there was no apparent difference between the mechanical properties of these two polymer liquids in their room temperature state, their dynamics diverged remarkably. This was when they were studied with shear waves whose

frequencies approached the glass-into-rubber relaxation times. Clearly, again, individual interaction of chain-like chemical units and not any micellar or other special aggregation of them, predominates polymer mechanics.

It still remains, however, to separate interactions of the basic units within and between chains. Most likely, the model plastic vs rubber liquids just discussed differ in the high frequency region substantially *only* because of *inter*-chain forces between phenyl vs methyl groups. However, especially in the high-frequency region, questions of *intra*-chain structure, such as the steric hindrance of adjacent pairs of methyl groups in polyisobutylene, restricted rotations about bonds, etc., come in. Obviously where configurational or quasi-configurational displacements are important, as in all cases of elongation >20 per cent (this is certainly an *upper* limit), flexibility of single chains needs to be understood. This is built deeply into chemical structure; plasticizers presumably may change over-all configuration as well as modify interaction, but they are impotent to vary flexibility. Accordingly, problems of rubber, usable in the Arctic, and of wire and cable insulation bendable at low temperatures always come back to whether the polymer chain bonds have free rotation. Some examples of the combinations of effects within and between chains can indeed be shown in several other polymer liquids which are rubber models.

This influence of small changes in chemical structure is compactly illustrated by comparing a few other hydrocarbon polymer liquids with polyisobutylene. Also, rather dilute dipolar groups have been introduced in the linear polyester liquid polypropylene sebacate, whose structure is otherwise like that of hydrocarbons.[45a] In Table VI, liquids of the given structure with some (unknown) distribution of molecular weights, were studied with shear waves at 77 and 142 kc at a temperature where each had the same steady flow viscosity. The figure chosen was 700 poises, and the temperature range required to adjust to it in the series was 10.9° to 85°C, meaning that the liquids had comparable consistencies at ordinary temperatures.

Despite these similarities under steady stress, the retardation times, τ', vary three-fold, with the highly substituted hydrocarbon chains, polyisobutylene and polypropylene, the highest. Despite the intermolecular action of the dipoles in polypropylene sebacate, the low polymer has a short retardation time, although its "brittle point" with decreasing temperature is far above that of polybutadiene or even polyisobutylene. Presumably the flexibility around C—O—C bonds rather compensates for increased dipole interaction. Where both low polarity

TABLE VI—*Dynamic Properties of Polymer Liquids of Varying Structure.*

Polymer	°C	Density, g/cc	η Poises	Voigt					Maxwell				
				77 kc		142 kc		$\tau' = \frac{\eta}{G}$ sec	77 kc		142 kc		$\tau = \frac{\eta}{G}$ sec
				η Poises	G, dynes/cm²	η Poises	G, dynes/cm²	142 kc	η Poises	G, dynes/cm²	η Poises	G, dynes/cm²	142 kc
Polyisobutylene C	25.4	0.8857	700	165.1	3.5×10^7	134.2	6.4×10^7	21×10^{-7}	197	2.2×10^8	173	3×10^8	6×10^{-7}
Polypropylene	85.0	.8248	700	23.6	6.5×10^6	22.0	9.9×10^6	22×10^{-7}	31.3	2.6×10^7	27.5	5×10^7	6×10^{-7}
Polybutadiene	10.9	.8767	700	7.2	5.5×10^6	5.2	6.3×10^6	8×10^{-7}	25.1	7.7×10^6	14.5	1×10^7	15×10^{-7}
Polypropylene Sebacate	41.7	1.0421	700	12.7	9×10^6	8.5	11.6×10^6	7×10^{-7}	40	1.3×10^7	28.6	1.7×10^7	17×10^{-7}

and chain flexibility obtain, as in polybutadiene and the silicones, dynamic properties apparently accord with brittle points in implying small temperature coefficients of relaxation times. In fact, the temperature coefficient for dynamic viscosity of polybutadiene is only about 1.5 kcal, whereas a comparable figure for polyisobutylene and polypropylene is 12 kcal.

The frequency range in which the structural comparisons above were made, resides, as discussed earlier, in the zone of configurational viscoelasticity. That is, over-all shape changes, rather than just nearest neighbor interactions, are predominant even at these comparatively short average chain lengths. Now, other recent studies of polyisobutylene liquids, at 5 to 100 cps frequency, exhibit no rigidity at 25°C and above, although they become non-Newtonian rapidly as temperature is

TABLE VII. *Shear Dynamics of Polyisobutylene A''.*
($\overline{M}_\eta$ = 1660; $\eta_s^{25°}$ = 39.6 Poises)

T, °C	Freq., cps	Voigt		Maxwel	
		μ dynes/cm²	η Poises	μ dynes/cm²	η poises
25	266	3.8×10^3	39.2	1.2×10^6	39.3
25	1601	4.8×10^4	38.4	3.1×10^6	39.0
27	25300	5.4×10^5	19.9	1.9×10^7	20.5
27	41390	1.5×10^6	19.9	1.9×10^7	21.5
27	53060	1.5×10^6	18.1	2.6×10^7	19.3

reduced.[45b] The questions are, where does the configurational elasticity drop out, as frequency is reduced at 25°C; and does it seem reasonable that this dispersion correlates with a shift in frequencies at lower temperatures. Partial answers are given by very recent studies of I. L. Hopkins of Bell Telephone Laboratories. He has equipped the tuning fork vibrator described earlier[33] with two parallel vanes filled in between with a film of polymer liquid. Pure shear properties can be derived from the response of this system. Table VII lists a few typical figures obtained on polyisobutylene polymer A''. These indeed show that the kilocycle relaxation zone (some new data by McSkimin's torsional pulse method are given for it) extends smoothly down to where dynamic and steady stress viscosities are equal. Seemingly there are no new "extra long time" relaxation mechanisms; probably the slow relaxation times sometimes indicated for high molecular weight rubbers are just displacements of this configurational relaxation to long times because of high molecular weight and internal viscosity.

By contrast to the conclusions associated with the data of Table VII,

some observations at low frequencies on isoviscous properties of polyisobutylenes A″ and C indicate nearly *identical* retardation times. Thus A″ at 25°C and C at 61°C have $\eta_s = 39.6$ poises. The η values at 266 and 1600 cps are also about 39 poises, μ at 266 cps is 3800 dynes per cm^2 for both liquids, and at 1600 cps is from 3.5×10^4 to 4.8×10^4 dynes per cm^2.

In the final section, mechanical waves have been used to explore dilute polymer solutions, to see how isolated molecules behave, free of interaction with each other.

DILUTE POLYMER SOLUTIONS

Physical Principles in Measurements

Precise information on dynamics of solutions approaching infinite dilution (and thus complete separation of the polymer chains) is desired here. Again these must be shear dynamics; bulk rigidity of ordinary liquids is so high that a few polymer molecules added cause little effect. Dilution is emphasized because even at 1 per cent by volume, high polymer molecule coils frequently interact, especially in "good" solvents. Thus, several workers have detected shear rigidity in polymer solutions, in one case for polymethyl methacrylate of average molecular weight 320,000, at 1 per cent concentration in *o*-dichlorobenzene.[46] Very low frequencies used (~10 cycles) there and in an earlier study[47] suggest, however, that even here, appreciable entangling of the molecules created a temporary network such as studied by Ferry.[27, 28] Such was certainly present in the 5 to 18 per cent solutions of cellulose acetate in dioxane measured in one of the earliest observations of shear rigidity in polymer solutions.[48]

Accordingly, since strictly linear, and hence non-interacting, mechanics are sought for the macromolecules in dilute solution, careful evaluation of experiments is essential. Since already it appears that important over-all (quasi-configurational) relaxations occur for, say, polyisobutylene in the kilocycle range, and it is suspected that not *all* of the interactions involved are between chains, the torsional crystal techniques are attractive. The absolute viscosity of these solutions is very low, so the ammonium dihydrogen phosphate crystal whose piezoelectric qualities are appropriate for polymer liquids in the circuits previously noted[22, 23, 49] is advantageously replaced by quartz.

Detailed electromechanical behaviour of such crystals in the pure liquids cyclohexane and benzene is of first concern. The electric field applied to electrodes on the suspended crystal produces mechanical

torsion generating pure shear waves. These waves may be modified by the environment around the crystal (vacuum, gas, liquid, solid) and react back. Thereby a mechanical resistance, R_M, and a mechanical reactance, X_M, are imposed on the electrical properties of the crystal element in the circuit. This connection comes out as:

$$\Delta R_E = K_1 R_M$$

$$\Delta f = -K_2 X_M ,$$

where ΔR_E is the increase in measured electrical resistance of the crystal element in the medium compared to in vacuum (or practically in *dry* air or nitrogen). The decrease in resonant frequency of the crystal element under these conditions is Δf. Thus K_1 and K_2 are electromechanical constants, which fundamentally may be calculated from the dimensions and piezoelectric constants of the crystals.[49] Now, in simple, Newtonian liquids,

$$R_M = X_M = \sqrt{\pi f \eta \rho}$$

Thus, by carefully measuring ΔR_E (or Δf) on a liquid of accurately known density ρ and viscosity η, at a given frequency f, and a given temperature, the constants K_1 and K_2 may be evaluated without assumptions and approximations of deriving them. Their constancy will then reflect the electromechanical stability of the system. Their behaviour under various conditions will be illustrated below. One further point is that when a liquid or solution does exhibit shear rigidity, or, in other words, if the single large molecules in a dilute solution are able to store energy, then $R_M > X_M$. Hence, in this case, the observed quantities ΔR_E, and especially Δf require particular precision.

In this regard, typical magnitudes of change of f_R between *dry* air and pure cyclohexane, at various temperatures, appear in Fig. 10. Questions often arise as to the arbitrariness of suspension of the radiating crystal, by the fine supporting and lead wires. The effects with the plain wires, in the solid curves of Fig. 10, are somewhat, but not radically, changed when a metal bead is put on, heavily loading vibrations in the wires, as shown by the dashed curves. In Fig. 11, a somewhat larger influence of the loaded support wires is shown for the R_E values, but both curves, by their smoothness and shape over a temperature range where the thermal expansion and other elastic constants of the metal support wires are quite different from those of the quartz crystal, affirm reliability of mounting and electromechanical coupling.

Fig. 10 shows, even for an 80-kc crystal, that Δf for an organic liquid

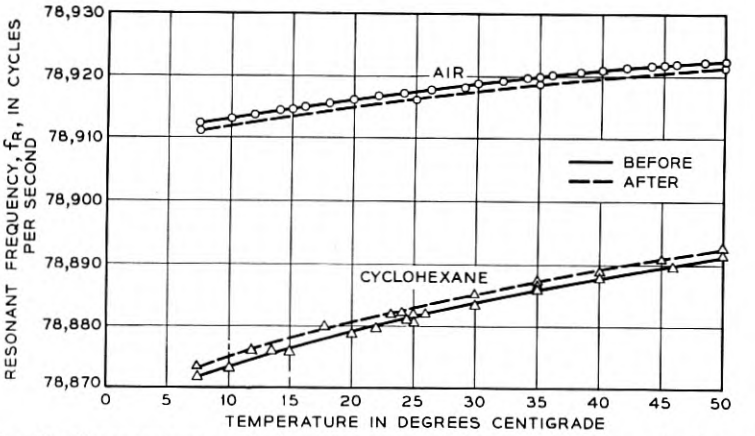


Fig. 10—Temperature variation of resonant frequency before and after adding weights to mounting wires.

(or dilute polymer solution) is bothersomely small. An excellent oscillator at 20 kc can hardly be expected to drift less than ±2 cycles, but at 20 kc the Δf like that between the sets of curves on Fig. 10 might be only 10 cycles, so 20 to 35 per cent error could come in. Hence, a different scheme for measurement of f_R than that in earlier systems[23, 49] was evolved. The tenth harmonic of the (say 80 kc) resonant frequency was beat against the 79th harmonic of a controlled standard 10-kc frequency. An interpolation oscillator accurately readable to 1 cycle then supplies the many hundred (roughly 1000) difference between these two high

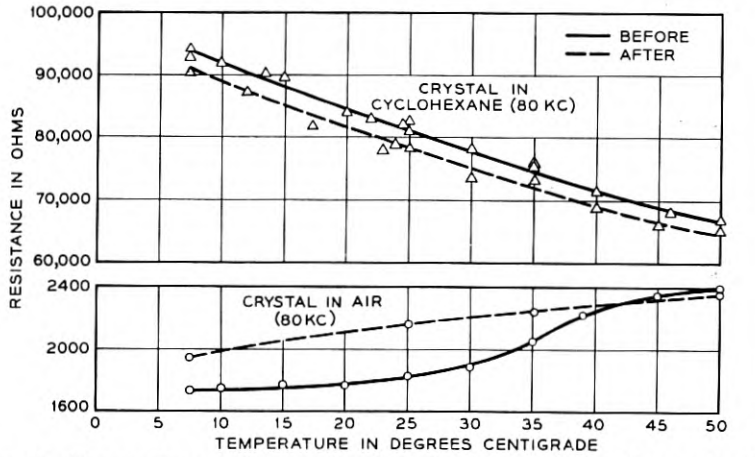


Fig. 11—Temperature variation of resistance at resonance before and after adding weights to mounting wires.

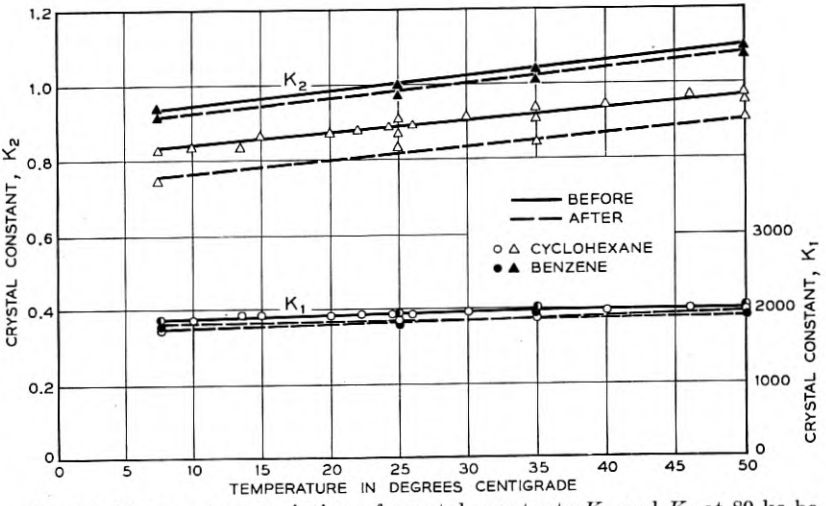


Fig. 12—Temperature variation of crystal constants K_1 and K_2 at 80 kc before and after adding weights to mounting wires.

harmonics. In this way, and in about 30 sec a balance can be conveniently achieved and the required ten-fold gain in accuracy attained.

By these means, and with best literature values of viscosity and density (which were checked in the laboratory at several temperatures) for purified solvents, curves for K_1 and K_2 were obtained as exhibited in Fig. 12 for 80 kc. Behaviour of K_1 at different frequencies over a tem-

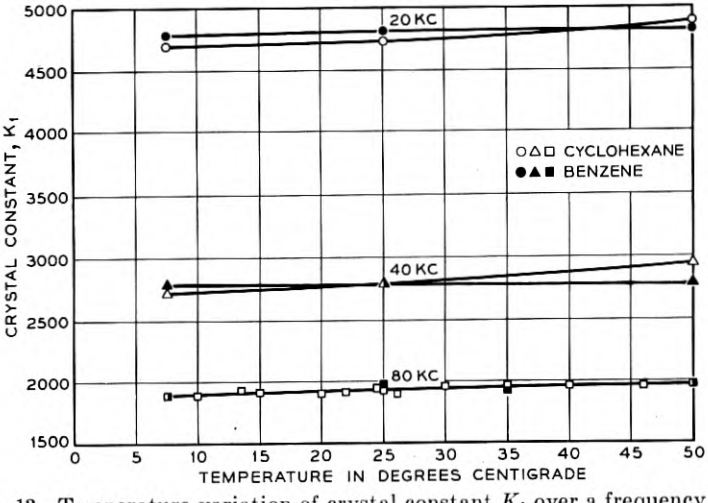


Fig. 13—Temperature variation of crystal constant K_1 over a frequency range with benzene and cyclohexane as standard fluids.

perature range is shown in Fig. 13, and of K_2, in Fig. 14. Fig. 14 brings out the significant point that in the present arrangement, where the oscillating crystal is immersed in the liquid studied, the *dielectric* properties of the liquid are important. Apparently the dielectric losses even of these purified hydrocarbons are different enough so that K_2 at 80 kc is quite different for benzene and cyclohexane. (Dielectric studies have previously indicated difficulty in preparing benzene having theoretically expected loss.) It is also possible that slight differences in wetting the crystal cause K_2 to vary with the liquid used.

The K_1 and K_2 values determined for all the various conditions above were then used under these conditions for measurements on the polymer

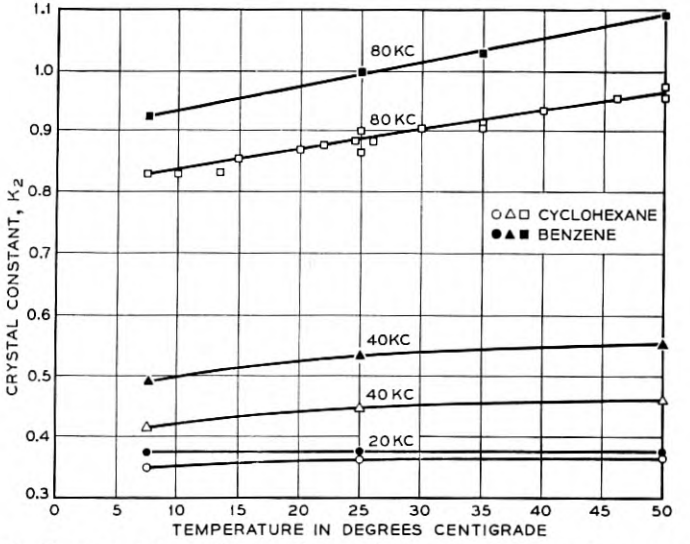


Fig. 14—Temperature variation of crystal constant K_2 over a frequency range with benzene and cyclohexane as standard fluids.

solutions in the kilocycle range. In the megacycle range, the balanced shear wave reflectance technique[23] gave satisfactory results over certain concentration zones which could fairly well be extrapolated to high dilutions. Thus, over the whole spectrum, there seems to be no doubt about the reality of the effects described below. That is, their magnitude far exceeds experimental uncertainty, as demonstrated in this section.

POLYISOBUTYLENE SOLUTIONS; DYNAMICS OF SEPARATE CHAINS

Solutions of polyisobutylene of $\bar{M}_\eta = 1.2 \times 10^6$ from about 0.1 to 1.0 wt. per cent concentration in cyclohexane yield R_M and X_M curves as shown in Fig. 15. The points coincide for the pure solvent, as they

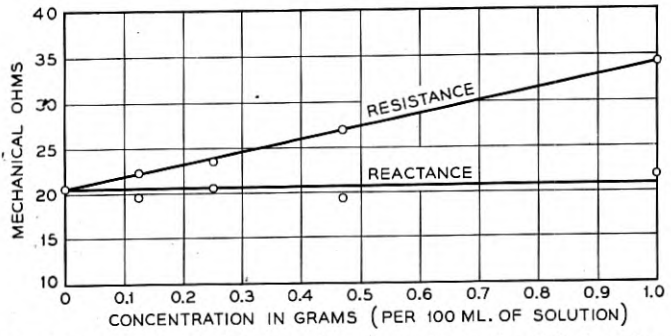


Fig. 15—Electromechanical interaction of solutions of polyisobutylene ($\overline{M}_\eta = 1.18 \times 10^6$) in cyclohexane with crystal vibrating torsionally at 20 kc.

should for a liquid having only viscosity. But, apparently as soon as any polymer chains are added, the curves diverge. A stiffness coming from separate chain molecules is being displayed.[50] Qualitatively, theoretical expectations of Kuhn[51, 52] and others seem justified, at least that there is a relaxation mechanism for isolated chains.

The usual question of how best to express the dynamical results arises. The procedure of earlier sections for polymer solids and liquids will be followed. In general, a frequency dependent modified *Maxwell* element as sketched on Fig. 16 will be used. However, a frequency-independent analysis has also been carried out for one sample system, and, from this, basic mechanical constants of single "average" molecules are obtained, if it is reasonable to relate the mechanical models for the liquid continuum to the discrete chains dissolved in it.

Fig. 17 shows typical results from the simple scheme of Fig. 16, where the pure solvent viscosity, η_A, has been considered to be in parallel with a Maxwell element. The total shear rigidity of the solution (at a given concentration) is represented by μ_B. The viscosity of the polymer molecule coils in solution with the solvent streaming through them is

$$\mu_B = \frac{(R^2 - X^2)\,\omega\,\eta_B}{\omega\rho\eta_S - 2RX}$$

$$\eta_A + \eta_B = \eta_S$$

$$\eta_A = \frac{2RX}{\omega\rho} - \frac{(R^2 - X^2)^2}{\omega\rho} \cdot \frac{1}{\omega\rho\eta_S - 2RX}$$

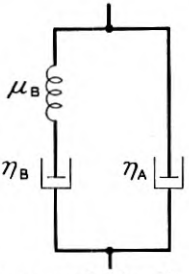


Fig. 16—Relations for calculation of shear stiffness and viscosity of dilute polymer solutions.

taken to be η_B . Thus, the steady flow viscosity, $\eta_s = \eta_A + \eta_B$. Also,

$$\frac{\eta_A + \eta_B}{\eta_A} = \eta_r \qquad \text{or} \qquad \frac{\eta_B}{\eta_A} = \eta_{sp}$$

under steady flow, or, alternatively, approximately a "dynamic intrinsic viscosity"

$$\left[\frac{\eta_B}{\eta_A} \cdot \frac{1}{c}\right]_{c \to 0}$$

can be written for any given frequency.

The curves in Fig. 17 are frequency dependent, however, although it turns out that η_B is only slightly so. Nevertheless, the considerable rise of η_A above the pure solvent viscosity, as the concentration is increased, indicates other mechanisms are being lumped into η_A . As usual, some

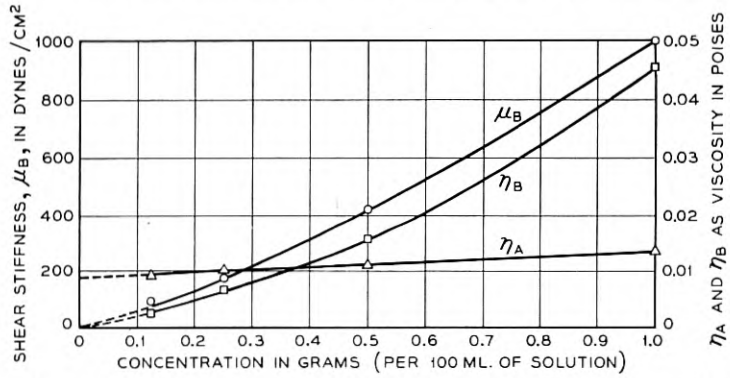


Fig. 17—Rigidity and viscosities of polyisobutylene ($\overline{M}_\eta = 1.18 \times 10^6$) in cyclohexane, at 25°C and 20 kc.

extensive distribution of relaxation times is probably responsible. However, from the chemical point of view, it is best to see if some principal mechanisms related to known structures can be identified. If so, they could be associated with new ideas about the details of polymer intrinsic viscosities, as well as the form of isolated molecules.[52, 53, 54, 55]

First, the frequency dependence of the μ_B of the model of Fig. 16 is as shown on Fig. 18. Striking regions of dispersion appear, although more points are needed to define the 10^5 cycle zone. Actually, many sets of data have been obtained in the 10^4 cycle zone. Recently, an immersed quartz tuning fork has given the approximate value shown for 2300 cycles. The experiments of Fig. 18 were on a polyisobutylene having $\overline{M}_\eta = 3.9 \times 10^6$, dissolved in cyclohexane. Values of η_A and η_B were, of course, also obtained. The results were then analyzed for a system of

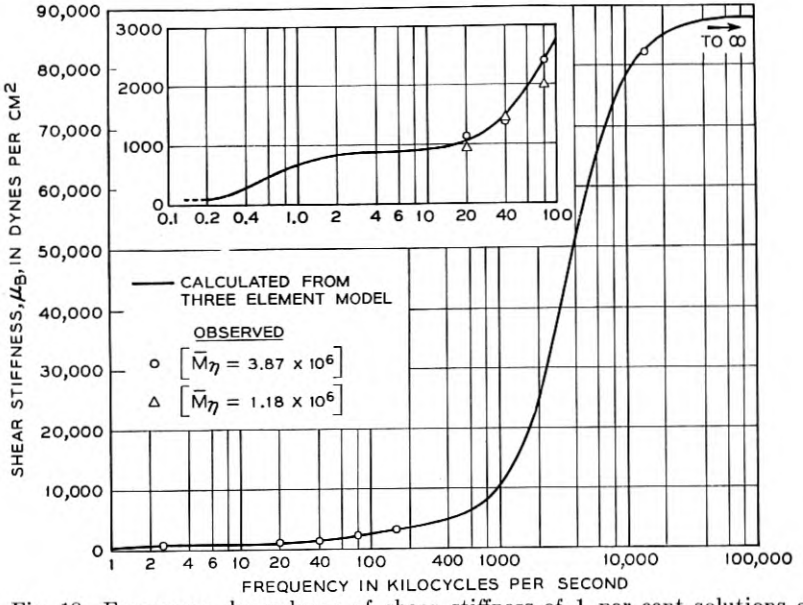


Fig. 18—Frequency dependence of shear stiffness of 1 per cent solutions of polyisobutylene in cyclohexane at 25°C.

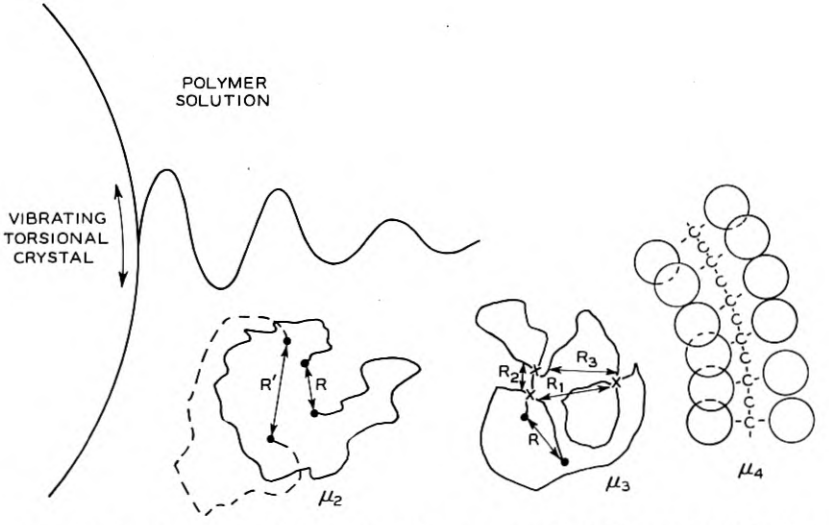


Fig. 19—Schematic diagram of possible sources of rigidity of single chain molecules in solution.

three Maxwell elements in parallel with again, as in Fig. 16, a solvent viscosity, this time called η_1 (truly absolute solvent viscosity) in parallel with them. The μ_B curve of Fig. 18, running through the observed points could then be calculated, by some special trial methods, with which Messrs. H. T. O'Neil and O. J. Zobel of Bell Telephone Laboratories kindly helped.

This analysis, identifying three principal relaxation regions for the motions of polyisobutylene chains in cyclohexane, gave for a 1 per cent solution (taken as linear part of concentration curve and hence equivalent to high dilution).

$$\begin{aligned}
\text{Principal rigidities } \mu_2 &= 890 \text{ dynes/cm}^2 \\
\mu_3 &= 3{,}190 \text{ dynes/cm}^2 \\
\mu_4 &= 84{,}000 \text{ dynes/cm}^2 \\
\text{Principal viscosities } \eta_1 &= 0.0082 \text{ poise (pure cyclohexane)} \\
\eta_2 &= 0.255 \text{ poise} \\
\eta_3 &= 0.006 \text{ poise} \\
\eta_4 &= 0.004 \text{ poise} \\
\text{Principal relaxation frequencies } f_2 &= 550 \text{ cycles} \\
f_3 &= 8.45 \times 10^4 \text{ cycles} \\
f_4 &= 3.52 \times 10^6 \text{ cycles}
\end{aligned}$$

Tentatively, these mechanisms may be schematically described as on Fig. 19. Here, the polymer coil, subjected to shear waves in dilute solution, exhibits rigidities μ_2, μ_3 and μ_4, all shown on different scales. μ_2 is the configurational elasticity because of actual changes in root mean square separation of chain ends, as from R to R'. It is retarded by viscous drag through the solvent, η_2, which is presumably the main source of characteristically high η_r of chain polymer solutions. The relaxation frequency for this mechanism is low—a few hundred cycles. It may come in significantly in work on more concentrated solutions at low frequencies,[28, 45, 48, 57] where chain entanglement is nevertheless the *dominant* factor.

μ_3 is when segments of the same chain in the molecular coil temporarily entangle with each other. Striking evidence has recently been given by Fox and Flory[58] that because of mutual interference, the theoretical random flight configuration of a chain gives very much too small a

molecular coil volume, V_e. This suggests that thermal agitation tending toward a smaller V_e, and excluded volume or repulsions forcing a large one, will cause collisons or entanglements which might last long enough to give *a van der Waals* cross-bond as denoted by crosses on the μ_3 sketch. (The actual forces in these would somewhat resemble those between different molecules in the concentrated solutions of Ferry.[28, 37]) This mechanism has the reasonable (based on Ferry's and others' work) relaxation frequency of 8.45×10^4. A small viscosity, η_3, may comprise friction of slippage at the entanglement points, with both the polymer and associated solvent molecules.

In Fig. 19, μ_4 is a relatively high stiffness presumed to be some average hindrance to rotation of one segment with respect to another. In the sketch, close-packed spheres representing methyl groups in polyisobutylene are portrayed. Their force fields overlap more in some places than others, in the meandering of the chain to form the molecular coil (of course, some tail-to-tail structures may be important here; they have all been shown head-to-tail in the sketch). Thus, this total internal steric restraint on chain flexibility, with a relaxation frequency of 3.5×10^6, contributes greatly to the large dispersion of rigidity in the megacycle range noted in Fig. 18. The related viscosity, η_4, is again low.

There is no doubt a considerable distribution of relaxation characteristics associated with each and all of these mechanisms.

Physical Properties Per Molecule

Since the viscosities and rigidities in the dilute solutions indeed seem to be additive with the number of molecules present, values of these properties, for the hypothetical mechanisms, can be expressed per average chain. Of course, the measured quantities are expressed as constants per cc of solution, but it may be useful to think of in terms of one average chain in each cc. Then, the shear deformation of this chain could be denoted by a force constant. The associated viscosities remain, however, dependent on solvent surroundings. Thus, for the polyisobutylene of $\bar{M}_\eta = 3.9 \times 10^6$, in cyclohexane solution, at 25°C the molecular quantities are:

$$[f_2] = 17 \times 10^{-13} \text{ dyne cm} \qquad [\eta_2] = 1.6 \times 10^{-16} \text{ poise}$$

$$[f_3] = 6 \times 10^{-12} \text{ dyne cm} \qquad [\eta_3] = 3.9 \times 10^{-18} \text{ poise}$$

$$[f_4] = 16 \times 10^{-11} \text{ dyne cm} \qquad [\eta_4] = 2.4 \times 10^{-18} \text{ poise}$$

In the section on polymer liquids, the high-frequency modulus μ was attributed to a nearest-neighbor glass or crystal-like interaction (since the actual values were indeed typical of the hardest organic solids).

However, in polyisobutylene (and to some degree in poly-α-methyl styrene), it is especially difficult to distinguish *inter*-chain from *intra*-chain crowding of methyl groups. Thus, while average center-to-center separation of methyls is ~4 Å in adjacent chains,[40] it is <2.5 Å within chains, in polyisobutylene. This crowding is apparently strong; the observed ΔH_{pzn} is only 12.8 kcal per mole instead of the 19.2 expected.[55, 56] The energy of steric hindrance thus amounts to almost half of the actual heat of polymerization. It is reasonable that a large part of the hardness of a mass of polyisobutylene chains, such as in the liquids, should therefore reflect the same mechanism as that for μ_4 (Fig. 19) in the dilute solutions. A rough check on this can be made. A polyisobutylene having considerably lower molecular weight than 3.9×10^6 and thus intermediate between the "liquid" and "solid" ranges, had a Maxwell shear modulus in the megacycle region (14 mc) of $\mu = 5.3 \times 10^9$, at 25°C. The number of molecules/cc, with individual $[f_4]$ given above, necessary to give the observed density of this polymer was multiplied by $[f_4]$, giving $\mu = 2.8 \times 10^9$ dynes/cm^2. Accordingly, about half of the observed high frequency rigidity of polyisobutylene, at 25°C, may be calculated from a "molecular constant" embodying intra-chain stiffness.

Much more refined and detailed treatments are required to generalize these "molecular constants" which are after all, as shown below, dependent on using a thermodynamically "inert" solvent. However, much as structurally significant dipole moments can be derived from measurements in dilute solutions, it seems hopeful that macromolecular mechanics can be so elucidated. Also additional structures, such as polypropylene and polydimethyl siloxane compared to polyisobutylene, are currently being studied.

Temperature Variation

Some further behaviour at different temperatures and solubilities of separate chains in dilute solution may now be considered against this background of possible mechanisms. Practically, these studies will bear on processing and properties, lacquers, paints, and casting solutions of polymers, as well as on the other qualities outlined in the introduction. Results may be conveniently discussed in terms of the modified Maxwell single element, with factors η_A, η_B, and μ_B (Fig. 16). Mostly, the kilocycle range, reflecting molecular coil changes, will be of interest. For comparison, it may be noted that at 20 kc, the polyisobutylene whose $\mu_B = 1061$ dynes/cm^2 in 1 per cent solution in cyclohexane receives 889 dynes/cm^2 of this from μ_2, the retarded configurational mechanism; 169

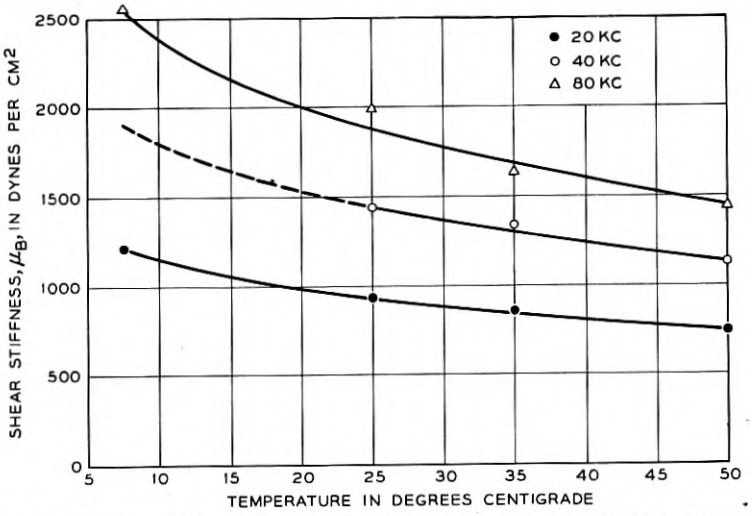


Fig. 20—Temperature variation of rigidity of 1 per cent solution of polyisobutylene ($\bar{M}_\eta = 1.18 \times 10^6$) in cyclohexane.

dynes/cm^2 from the entangled segments stiffness, μ_3 ; and only 3 dynes/cm^2 from the intra-chain stiffness, μ_4 . Thus, chain configuration mechanics, including the associated viscosities, can be well enough thought of in the following paragraphs, in terms of μ_B , η_B and η_A .

The exponential decrease of μ with temperature familiar for polymer solids and liquids is much suppressed in the μ_B vs T curve of Fig. 20. While the μ_4 , internal rotation, mechanism for single polyisobutylene molecules probably has a considerable activation energy, that for the μ_2 , configurational, rigidity should be very small. Then, without retardation, the intrinsic chain modulus would rise with rising temperature. These influences seem to combine to give the modest decline of μ_B appearing in Fig. 20. If these rigidities are plotted against $1/T$, the temperature coefficient is 2.3 kcal. This is much less than the familiar values for the stiffening of rubbery solids, and emphasizes that *inter*-chain action reigns then.

Solvent Variation

Effects of solvents of different (mostly positive) heats of mixing on state of polyisobutylene molecules in solution have been nicely established by Fox and Flory.[58] Especially, this work has clarified principal factors in the intrinsic viscosity expression

$$[\eta] = \frac{V_e}{M} \cdot \frac{\varphi}{100} = K_0 M^{1/2} \alpha^3 \varphi.$$

Here, V_e = effective volume per molecule (and hence as determined by chain configuration), M = molecular weight, α represents change in linear extent of molecule because of mutual interference of segments[58] and φ expresses the hydrodynamics interaction of solvent and molecular coil (including varying degrees of "straining through" the coil).[61, 62, 63, 64] Interpretation of the mechanical properties of chains in dilute solution, with reference to the rough concepts of Fig. 16, arouses particular interest in the factor α^3. For a high molecular weight polyisobutylene, intrinsic viscosity theory[58] indicated that α^3 the ratio for volume of actual coil divided by volume for ideal random flight coil was 3.81 in cyclohexane but only 1.42 in benzene, both at 30°C. This striking alteration in equilibrium chain configuration, a variable which is not readily introduced into polymer liquids or solids, appears in the inherent viscosity vs c curves in cyclohexane, Fig. 21, and benzene, Fig. 22. The large difference in $[\eta]$ at 25°C, 6.00 in cyclohexane vs ~1.5 in benzene, indeed emphasizes the different solvent powers.[65] Likewise, the large increase of $[\eta]$ with temperature in Fig. 22 accents the poor solvent qualities of benzene.[66] Too, empirically, polymer molecules which are either tight coils or are actually chemically cross-linked to form microgel molecules characteristically show *positive* slopes of inherent viscosity vs c plots.[67] Accordingly, all this evidence for large changes in the conformation of chain molecules in "good" vs "poor" solvents should show up in dynamics of dilute solutions. Also, technically, quite different physical properties are found for polymer-plasticizer compounds where compatibility is high (good solvent) than where it is low (poor solvent). Here,

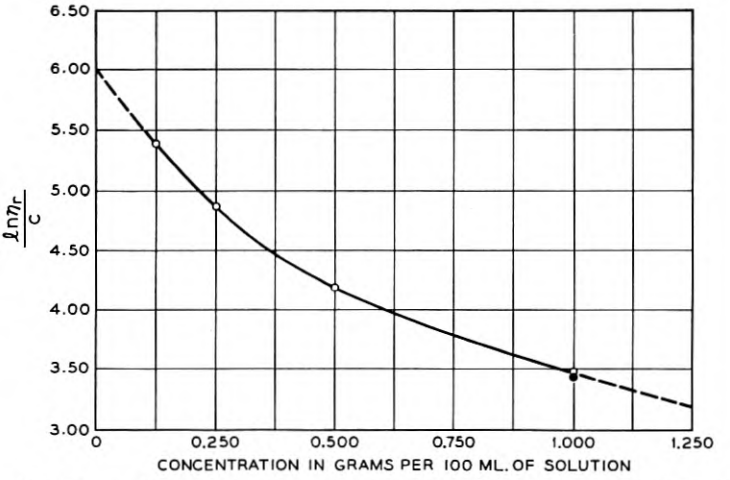


Fig. 21—Inherent viscosity of polyisobutylene ($\overline{M}_\eta$ = 3.87 × 10⁶) in cyclohexane at 25°C.

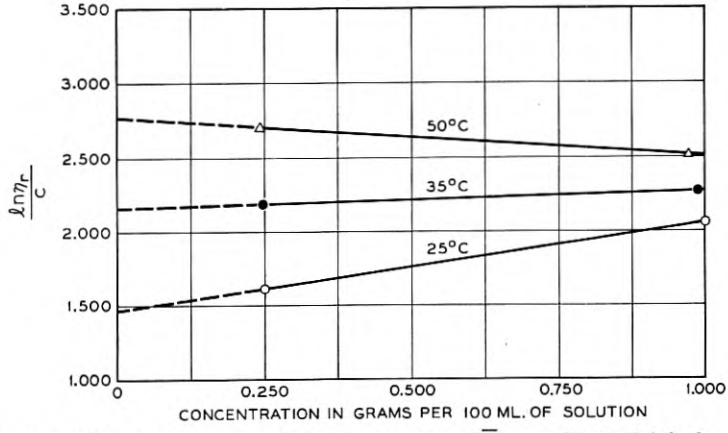


Fig. 22—Inherent viscosity of polyisobutylene ($\bar{M}_\eta = 3.87 \times 10^6$) in benzene, at various temperatures.

more flexible compositions are often produced with low compatibility plasticizers—indeed, sometimes with those on the verge of phase separation than with those with highly favorable heats of solution.[68] This would mean that the bad solvents would compress the chains so that they would be more easily strained than if they were in a "free chain" or even extended configuration. If single chain, visco-elastic stiffnesses are acting this way, the dynamic μ_B would then actually decline as heat of mixing become more positive.

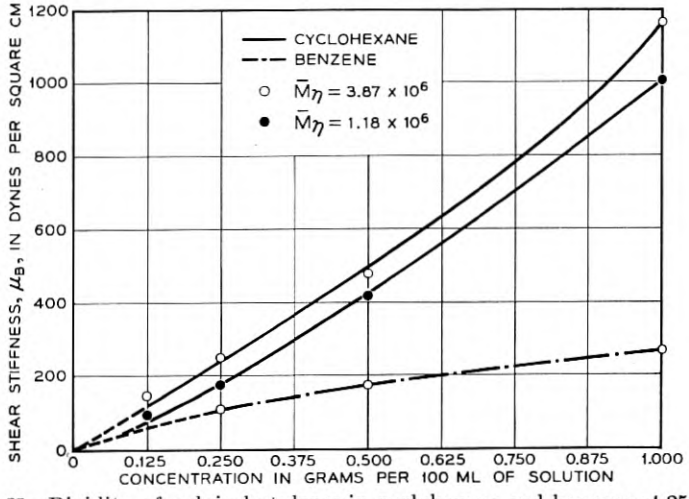


Fig. 23—Rigidity of polyisobutylenes in cyclohexane and benzene at 25°C and 20 kc.

This seems to take place, as indicated by the lower compared to the upper curve on Fig. 23. Here, the μ_B of the usual modified Maxwell model, at 20 kc, is plotted against c for the polyisobutylene of $\bar{M}_\eta = 3.9 \times 10^6$. Also, the middle curve shows μ_B for a polymer of about a third of this molecular weight; while there is a small reduction in μ_B with $\bar{M}_\eta$ in this range, it is much less than the reduction caused by tightening up the polymer coil.

The μ_B values per average molecule, $[f_B]$, fall from 18×10^{-13} dyne cm in cyclohexane to 7×10^{-13} in benzene. (Of course, $[f_B]$ for the intermediate molecular weight polymer in cyclohexane is only 5×10^{-13} because so many more molecules are present in solution.)

The temperature dependence of μ_B also becomes nearly zero at least over the narrow range from 25 to 50°C, in benzene compared to cyclohexane. This seems to accord with the indications previously, from Fig. 20, for a lower molecular weight polymer, that different mechanisms are competing. These may be the configurational, with $\mu \propto T$, and relaxation, with μ varying in some complicated way with T. Thus $[\eta]$ increases markedly with T, and presumably denotes an expanding molecular coil tending toward the "normal" configuration in cyclohexane. At the same time, the relaxation processes with rising temperature tend to cause the decrease in μ_B typical of the upper, solid, curves on Fig. 24. In engineering use, often times poorly compatible plasticizers give compounds which stiffen more gradually with temperature than do "solvent" plasticized ones.

For similar reasons, the dynamic molecular coil viscosity, η_B, ought to vary less with temperature in thermodynamically poor than in good solvents. This is indeed seen in Fig. 25. On the other hand, η_A for the modified Maxwell element has been described as the solvent viscosity with segment hindrance and restricted rotation terms from the polymer molecules lumped in with it. These latter terms are presumably little affected by over-all configuration (μ_2 term; the μ_3 mechanism will be somewhat affected, but not the μ_4, on Fig. 19). Thus, η_A should have comparable temperature dependence in both good and bad solvents, as seems to be indicated by Fig. 26.

Microgel Molecule Solutions

The statistical coil of linear polymer molecules may be replaced by a chemically fixed, cross-linked network in microgel molecules.[67] These may be made completely rigid, like Einstein spheres, or highly swellable. The latter are hybrids between rigid spheres and coiled chains. In

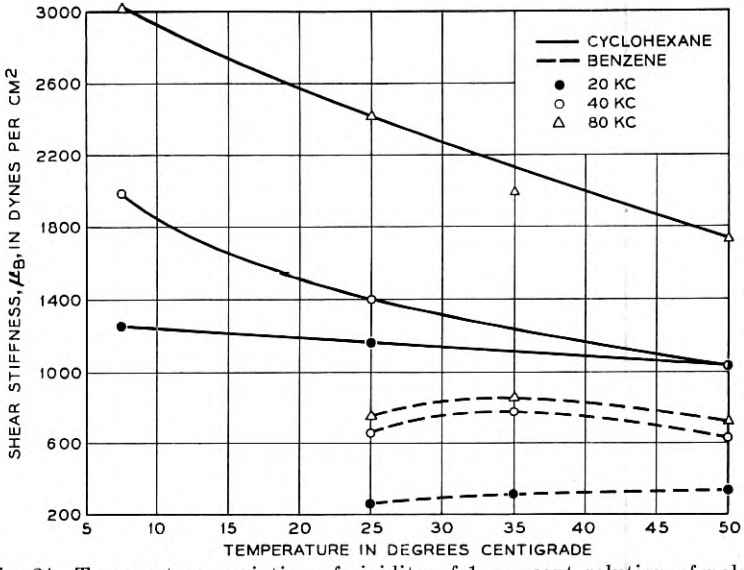


Fig. 24—Temperature variation of rigidity of 1 per cent solution of polyisobutylene ($\bar{M}_\eta = 3.87 \times 10^6$) in cyclohexane and benzene.

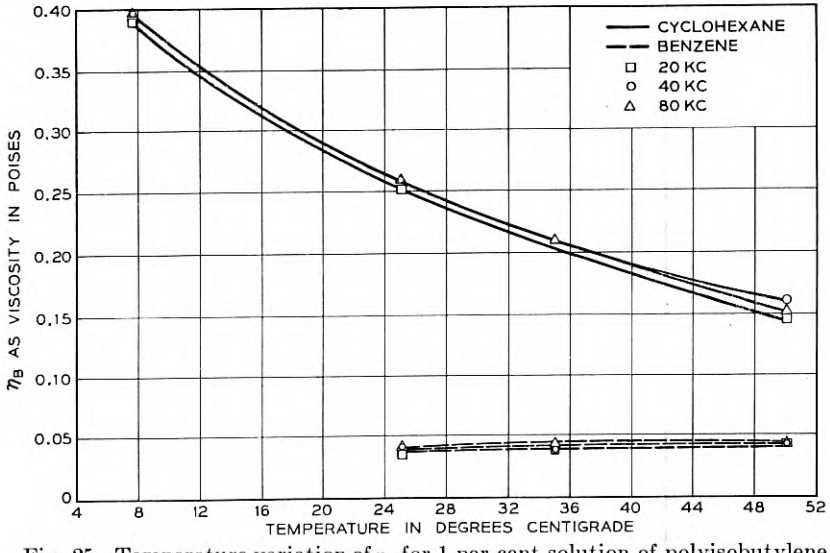


Fig. 25—Temperature variation of η_B for 1 per cent solution of polyisobutylene ($M_\eta = 3.87 \times 10^6$) in cyclohexane and benzene.

synthetic rubber, they confer unique flow properties, causing the excellent processibility of GR-S 60. However, dynamic tenacity, such as in flex crack growth, is degraded by their presence. Now presumably the excellent extrusion qualities of synthetic rubber composed of from 60 to 80 per cent microgel molecules are because of their individual shear stiffness. Thus, if a wire coating, for instance, is extruded at high rates of shear, chain molecules are deformed, and store energy just as discussed in the earlier sections on liquids. After emerging from the extrusion die, they relax, and cause the gross retraction, shrinkage and roughness shown in the wire insulation of the upper photograph of Fig. 27. A polymer with about 70 per cent microgel molecules gives the smooth covering shown in the lower specimen of Fig. 27. Here, the shearing stresses of extrusion seem insufficient to distort the tiny networks of the microgel molecule; in any case, the covering does not roughen or relax. Similar effects have been found for microgel plastics. Nevertheless, unlike gross or macro gelation, the whole melt can flow.

On this basis, dilute solutions of microgel molecules ought to indicate high shear rigidity per molecule. The mechanism μ_3 of Fig. 19, in which now the junction points are not temporary, but are primary valence cross-links, should be predominant. Fig. 28 shows, for a polybutadiene microgel in cyclohexane,[67] that μ_B has indeed risen, compared to equal

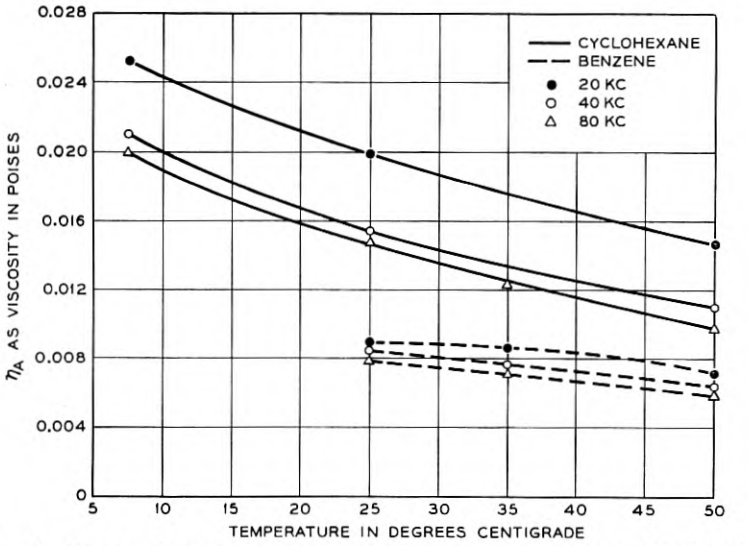


Fig. 26—Temperature variation of η_A for 1 per cent solution of polyisobutylene ($\bar{M}_\eta = 3.87 \times 10^6$) in cyclohexane and benzene.

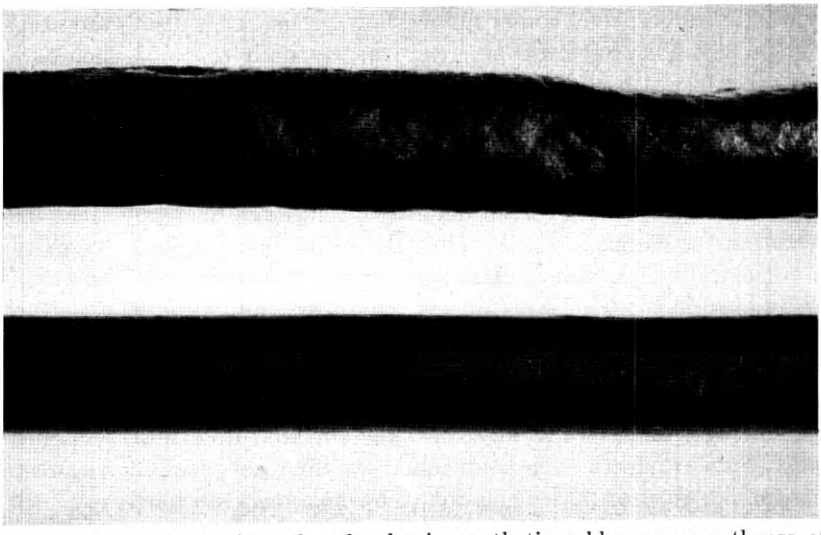


Fig. 27—Effect of microgel molecules in synthetic rubber on smoothness of extruded wire insulation. Rough covering is from high-speed extrusion of GR-S without microgel.

weights of chain molecules. Further, accompanying the extremely high average molecular weight of the microgel (18.6×10^6), the $[f_B]$ per average molecule is 42×10^{-12} dyne cm or about twenty-five times that of the polyisobutylene with $\bar{M}_\eta = 3.9 \times 10^6$. Also, the temperature coefficient for μ_B of polybutadiene microgel is low.

Of course, polybutadiene, as chains or as microgel molecule segments, has many double bonds. These will surely influence the μ_4, or internal rotation mechanism. Further work remains to show just what is their effect in the microgel case. But, it is interesting to compare μ_B values for Hevea rubber chains with those for, say, polyisobutylene, which has only single bonds in the chain.

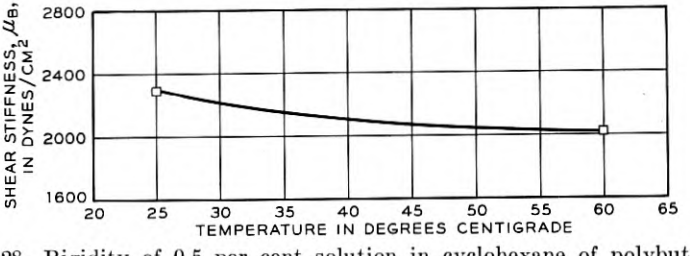


Fig. 28—Rigidity of 0.5 per cent solution in cyclohexane of polybutadiene microgel ($\bar{M}_W = 18.6 \times 10^6$) at 20kc.

Hevea Rubber Solutions

The comparison of equal weight concentrations of natural rubber in cyclohexane with polyisobutylene in cyclohexane is surprising:

Hevea rubber $\bar{M}_\eta = .23 \times 10^6$ $\mu_B = 1350$ dynes/cm^2, 1 per cent solution (corr.).

Polyisobutylene $\bar{M}_\eta = 1.2 \times 10^6$ $\mu_B = 1000$ dynes/cm^2, 1 per cent solution (corr.).

Both results are at 20 kc. The higher value for natural rubber may be because of the double bonds causing stiffening of the chain. On the other hand, maybe easy rotation around single bonds raises the μ_3 part. Certainly the *viscous* retardation *within* natural rubber chains is very low, as noted in the section on solids. However, its interaction with, or configuration, in cyclohexane may be peculiar. The $[f_B]$ per average molecule is, however, low, being 15×10^{-14} dyne cm at 25°C.

Polystyrene Solutions

Much work, on light scattering and other properties, has indicated appreciable intra-chain stiffness for polystyrene,[69] but still much freedom compared to polyisobutylene.[69a] However, this work, as well as ΔH_{pzn} of 17 kcal compared to ~19 kcal calculated for no steric hindrance, suggests comparatively small restraints on ideal flexibility. This needs to be checked by a frequency analysis of dilute solution mechanics, but polystyrene seems to be a reasonable example of "plastic" behaviour at room temperature because of interaction *between* the chains. (It is recalled that, earlier, α-methyl styrene polymer was cited as plastic model showing both intra- and inter-chain stiffness. Unlike in polystyrene, the intra-chain factor shows up in a ΔH_{pzn} of 9–10 kcal, a third less than that calculated if there were no steric hindrance.) Thus, no evidence of unusual stiffness appears in Fig. 29, when, indeed, the μ_B values are considerably lower, for equal weight concentrations, than those for natural rubber. The highly milled rubber studied had $\bar{M}_\eta$ very nearly that of $\bar{M}_\eta = 0.234 \times 10^6$ of the polystyrene, so the $[f_B]$ per average polystyrene chain, 4.5×10^{-14} dyne cm is less than a third that of the rubber. No wonder that at high temperatures, where the phenyl group interaction between chains is much reduced, polystyrene makes a good rubber. Also, in Fig. 29 are shown data for a polymer of $\bar{M}_\eta = 1.2 \times 10^6$, made in emulsion and having $[\eta] = 4.350$ in benzene at 25°C.

The polystyrene solutions discussed above were in benzene, a good solvent. Here, the situation is converse to that for polyisobutylene; for polystyrene, cyclohexane is a poor solvent and benzene, good. Hence, if the previous interpretation of reduced single chain *quasi-configurational* (μ_2) stiffness is general for solvents of more endothermic mixing, the "plastic" molecule polystyrene should show it in cyclohexane. This is indeed evident in Fig. 30, showing one of the same polystyrenes of Fig. 29, measured at 20 kc (normalized to 1 per cent concentration). Also, on Fig. 30 are shown the inherent viscosity (practically, the intrinsic viscosity, in this case) and the absolute viscosity of the 1 per cent solution

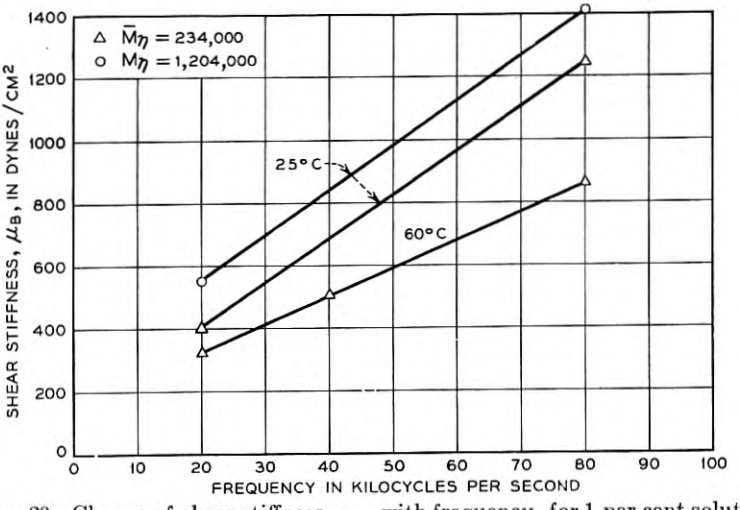


Fig. 29—Change of shear stiffness, μ_B, with frequency, for 1 per cent solutions of polystyrene in benzene.

under steady flow, η_s. These are all plotted against temperature down to phase separation, at about 26° to 27°C.

The marked positive slope of the $\ell n \eta_r/c$ curve denotes the large contraction in molecular coil volume preceding phase separation or insolubility. The absolute viscosity, η_s, however, rises with declining temperature because it is dominated by solvent viscosity, but when the polymer phase comes out, η_s abruptly falls off.

The μ_B values are consistent with this steady flow behaviour, except that the *rise* of μ_B at the turbidity point seems to be because a layer of swollen polymer-rich phase forms on the torsional crystal surface. This condition is seen in Fig. 30 to coincide nicely with the abrupt changes in steady flow viscosity.

The slight maximum in the μ_B curve at about 35°C may not be real.

It does come near the point of minimum interaction for the whole system. In any case, as discussed before, the average temperature coefficient of μ_B in the poor solvent is very low compared to the good solvent. The values of μ_B are roughly $\frac{2}{3}$ to $\frac{1}{2}$ those in benzene.

GENERAL THEORY OF SINGLE CHAIN MECHANICS; KUHN AND KIRKWOOD

As noted before, much of the present understanding of stress-strain properties of polymer chains, in dilute solutions, liquids or solids, has come from W. Kuhn's long interest in this subject. Many supplementary contributions have been stimulated by Kuhn's work, and new points of view have been introduced by others. For instance, recently new and different proposals have been made about the flow birefringence and non-Newtonian viscosity of solutions of deformable spheres.[70] These ideas could be tested on suitable microgel solutions.

Recently, moreover, an especially significant general theory of viscoelastic behaviour of polymer in solution has been constructed by Kirkwood.[71] It explicitly considers the hydrodynamic conditions leading to the rigidity now observed for high-frequency shear waves. It formulates definitely the configurational changes of isolated chains in solution when strained in shear. As this theory is advanced to forms where simpler calculations can be made, it may answer many of the questions raised by the new experiments on single chain properties.

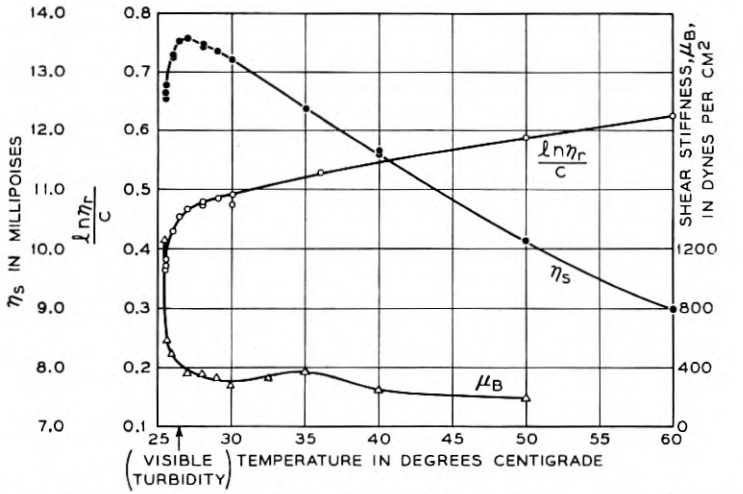


Fig. 30—Temperature dependence of absolute viscosity, η_S, inherent viscosity, $\ell n \eta r/c$, and shear stiffness, μ_B of, 1 per cent solution of polystyrene in cyclohexane down through turbidity point.

CONCLUSIONS

To leave some impression of the elemental chemical structures which move around when wood, rubber, plastics, textiles and finishes are used mechanically—that has been the aim of this study. Polymer viscosities have been found in a variety of "solids"; rigidities have been demonstrated for very fluid "liquids" and solutions. Studies of these solid and liquid extremes have given some chemical reality to the classical spring and dashpot models.

Existence of compressional viscosity has been shown for polymer liquids and solids. It may comprise a new quality for investigation of polymer structure. At present, too little is known of its origin to interpret further the effects of intense ultrasonic irradiation of polymer solutions. Experiments of Schmid and co-workers[72] early indicated degradation of molecular weight of polystyrene, so irradiated, but whether this is chemical, from local heating in the solvent, or actual physical coupling with the wave field, is still unsettled. However, these workers also considered a compressional stiffness of the polymer molecules in the solutions,[73] and showed that if there was coupling, it was not inertial (by dissolving polystyrene in solvents of exactly the same density, no reduction in effect was observed). A point of general interest arises here; impact fractures of plastics presumably actually fracture some primary valence bonds. This is certainly true for many thermoset materials, and probably for chain compounds. Hence, if the detailed mechanism of how compressional waves move and perhaps rupture polymer segments were known, information on the baffling problems of ultimate strength would be gained. The observations above on dependence of λ and λ' on molecular weight and structure provide only the barest start on this but a new goal is in view. Too, basic questions of how rapidly molecules being formed in a polymerization equilibrate in temperature with their surroundings are elucidated by compressional wave propagation constants. For instance, absolute rate measurements on velocity of chain growth cannot be said to be isothermal if they seem to be faster than the thermal relaxation times which the ultrasonic measurements indicate can be $\sim 10^{-6}$ to 10^{-5} sec.

Likewise, more thorough understanding of velocity and dispersion of compressional waves in polymer solutions would clear up anomalies in velocity measurements for a wide variety of polymers,[74] some of which have been tentatively attributed to chain branching.

ACKNOWLEDGEMENT

Besides the extensive collaboration of W. P. Mason and H. J. McSkimin, we should like to attest to the help of T. G. Kinsley.

BIBLIOGRAPHY

1. Meyer, von Susich and Valko, *Kolloid-Z.*, **59,** p. 208 (1932); Meyer and Ferri, *Helv. chim. Acta*, **18,** p. 570 (1935).
2. Meyer and van der Wyk, *J. Polymer Sci.*, **1,** p. 49 (1946).
3. Kuhn, *Zeit. physik. Chem.*, **B42,** p. 1 (1939).
4. Simha, *J. Appl. Phys.*, **13,** p. 201 (1942).
5. Gilmore and Spencer, *Mod. Plastics*, **27,** p. 143 (Apr. 1950); *Ibid.*, **27,** (Dec. 1950).
6. Mooney, *J. Colloid Sci.*, **2,** p. 69 (1947).
7. Vila, *Ind. Eng. Chem.*, **36,** p. 1113 (1944).
8. Dexter and Dienes, *J. Colloid Sci.*, **5,** p. 228 (1950).
9. Hopkins, Baker and Howard, *J. Appl. Phys.*, **21,** p. 206 (1950).
10. Haward, *Trans. Far. Soc.*, **34,** p. 267 (1943).
11. Morey, *Ind. Eng. Chem.*, **37,** p. 255 (1945).
12. Jenckel, *Zeit. f. Elektrochem.*, **43,** p. 796 (1937).
13. Boyer and Spencer, *J. Appl. Phys.*, **16,** p. 594 (1945).
14. Richards, *J. Chem. Phys.*, **4,** p. 449 (1936).
15. Baker, *India Rubber World*, **110,** p. 543 (1944).
16. Alexandrov and Lazurkin, *Acta Physicochimica USSR*, **12,** p. 647 (1940).
17. Boyer and Spencer, *J. Polymer Sci.*, **2,** p. 157 (1947). An apt survey of much study of plasticizers and mechanical properties.
18. Andrews, Hofman-Bang and Tobolsky, *J. Polymer Sci.*, **3,** p. 669 (1948); Dunell and Tobolsky, *Textile Res. J.*, **19,** p. 63 (1949); Brown and Tobolsky, *J. Polymer Sci.*, **6,** p. 165 (1951).
19. Nolle, *J. Polymer Sci.*, **15,** p. 1 (1950); *J. Appl. Phys.*, **19,** p. 753 (1948).
20. Ivey, Mrowca and Guth, *J. Appl. Phys.*, **20,** p. 486 (1949).
21. Hall, *Phys. Rev.*, **71,** p. 318 (1947).
22. Mason, Baker, McSkimin and Heiss, *Phys. Rev.*, **73,** p. 1074, p. 1873 (1948).
23. Mason, Baker, McSkimin and Heiss, *Ibid.*, **75,** p. 936 (1949).
24. Sack, Motz and Work, *J. Appl. Phys.*, **18,** p. 451 (1947).
25. Lyons and Prettyman, *J. Appl. Phys.*, **19,** p. 473 (1948).
26. Ballou and Smith, *J. Appl. Phys.*, **20,** p. 493 (1949).
27. Ferry, Sawyer and Ashworth, *J. Polymer Sci.*, **2,** p. 593 (1947) for general review.
28. Ferry, *Jour. Res., NBS.*, **41,** p. 53 (1948).
29. Alfrey and Doty, *J. Appl. Phys.*, **16,** p. 700 (1945).
30. Leaderman, *J. Colloid Sci.*, **4,** p. 193 (1949).
31. Kelsey and Dillon, *J. Appl. Phys.*, **15,** p. 352 (1944).
32. Wilson and Smith, *Ind. Eng. Chem.*, **41,** p. 770 (1949).
33. Hopkins, *Trans. A.S.M.E.*, **73,** p. 195 (1951).
34. Rorden and Grieco, *J. Appl. Phys.*, **22,** p. 842 (1951).
35. Gehman, Woodford and Stambaugh, *Ind. Eng. Chem.*, **33,** p. 1032 (1941).
36. Dillon, Prettyman and Hall, *J. Appl. Phys.*, **15,** p. 309 (1944).
37. Marvin, Fitzgerald and Ferry, *J. Appl. Phys.*, **21,** p. 197 (1950).
38. Fox and Flory, *J. Am. Chem. Soc.*, **70,** p. 2384 (1948).
39. Flory, *Ind. Eng. Chem.*, **38,** p. 417 (1946).
40. Baker, Chapter 8 in *High Polymers*, edited by Twiss, Reinhold Publishing Corp., New York, 1945.
41. Work, *Textile Res. Journal*, **19,** p. 381 (1949).
41a. Tuckett, *Trans. Far. Soc.*, **40,** p. 448 (1944); Würstlin, *Kolloid-Z.*, **120,** p. 84 (1951).
42. Flory, *J. Am. Chem. Soc.*, **65,** p. 372 (1943).

42a. Fox and Flory, *J. Phys. and Colloid Chem.*, **55,** p. 221 (1951).
43. Fox and Flory, *J. Appl. Phys.*, **21,** p. 581 (1950).
44. Debye, *Z. Elektrochem.*, **45,** p. 174 (1939).
45. Galt, *Phys. Rev.*, **73(2),** p. 1460 (1948).
45a. Baker, *Rubber Chem. Tech.*, **18,** p. 632 (1945).
45b. Harper, Markowitz and DeWitt, *Abstracts, XII Int. Congress Chemistry*, p. 274 (1951).
46. Van Wazer and Goldberg, *J. Appl. Phys.*, **18,** p. 207 (1947).
47. Kendall, *Rheol. Bull.*, **12,** p. 26 (1941).
48. W. Philipoff, *Physik. Z.*, **35,** p. 884, p. 900 (1934).
49. Mason, *Trans. A.S.M.E.*, **69,** p. 359 (1947).
50. Baker, Mason and Heiss, *Bull. Am. Phys. Soc.*, **24,** p. 29 (1949).
51. Kuhn and Kuhn, *Helv. chim. Acta*, **29,** p. 609, p. 830 (1946); *J. Colloid Sci.*, **3,** p. 11 (1948).
52. Flory and Fox, *J. Am. Chem. Soc.*, **73,** p. 1904 (1951).
53. Fox, Fox and Flory, *Ibid.*, **73,** p. 1901 (1951).
54. Fox and Flory, *Ibid.*, **73,** p. 1909 (1951).
55. Fox and Flory, *Ibid.*, **73,** p. 1915 (1951).
56. Kuhn and Grün, *J. Polymer Sci.*, **1,** p. 183 (1946).
57. Carver and Van Wazer, *J. Phys. and Colloid Chem.*, **51,** p. 751 (1947).
58. Fox and Flory, *Ibid.*, **53,** p. 197 (1949).
59. Evans and Tyrrall, *J. Polymer Sci.*, **2,** p. 387 (1947).
60. Roberts, *Jour. Res., NBS.*, **44,** p. 221 (1950).
61. Kuhn, *Kolloid-Z.*, **68,** p. 2 (1934); Kuhn and Kuhn, *Helv. chim. Acta*, **26,** p. 1394 (1943).
62. Huggins, *J. Phys. Chem.*, **43,** p. 439 (1939).
63. Debye, *Phys. Rev.*, **71,** p. 486 (1947).
64. Brinkman, *Appl. Sci. Res.*, **A1,** p. 27 (1947).
65. Huggins, *J. Appl. Phys.*, **10,** p. 700 (1939).
66. Alfrey, Bartovics and Mark, *J. Am. Chem. Soc.*, **64,** p. 1557 (1942).
67. Baker, *Ind. Eng. Chem.*, **41,** p. 511 (1949).
68. Boyer, *J. Appl. Phys.*, **20,** p. 540 (1949).
69. Zimm, *J. Chem. Phys.*, **16,** p. 1099 (1948); Outer, Carr and Zimm, *Ibid.*, **18,** p. 830 (1950).
69a. Kunst, *Rec. trav. chim.*, **69,** p. 125 (1950).
70. Cerf, *Compt. rend.*, **226,** p. 1586 (1948); *Ibid.*, **227,** p. 1221 (1948).
71. Kirkwood, *Rev. trav. chim. Pays-Bas*, **68,** p. 649 (1949).
72. Schmid and Rommel, *Z. Physik. Chem.*, **A85,** p. 97 (1939).
73. Schmid and Beuttenmüller, *Z. Elektrochem.*, **49,** p. 325 (1943); **50,** p. 209 (1944).
74. Natta and Baccaredda, *Gazz. chim. Ital.*, **79,** p. 364 (1949).

The Reliability of Telephone Traffic Load Measurements by Switch Counts

BY W. S. HAYWARD, JR.

(Manuscript received October 15, 1951)

The switch count method of telephone traffic measurement is subject to sampling errors. The nature of these errors is discussed and formulas are derived which describe the extent of the errors under normally encountered traffic conditions.

INTRODUCTION

Of prime importance to the telephone traffic engineer is the determination of the busy season busy hour load carried by groups of trunks or other circuits of a telephone switching system. Three direct methods of measuring such loads are found in the field today. These are:

a. Peg Count and Holding Time Method

The number of calls carried by the circuit group during the observation period is counted. This number multiplied by the average holding time per call (in hundreds of seconds) and divided by the length of the observation period (in hours) gives an estimate of the group load in units of hundred-call-seconds per hour (CCS). The major drawback to this peg count method is that it requires a separate determination of the average holding time per call for the group under observation. R. I. Wilkinson[1] has analyzed the sources of errors of holding time measurements. In addition, correlation between load and holding time introduces an error which has not been studied.

b. Switch Count Method

At fixed intervals the circuit group is scanned and the number of busy circuits is counted. The total number of busy conditions counted divided by the number of scans is, then, an estimate of the load on the group in units of average simultaneous calls or erlangs*. This estimate is generally converted to CCS (1 erlang = 36 CCS) by traffic engineers since the

* The name "erlang" for average simultaneous call was adopted at a plenary meeting of the CCIF at Montreux in October, 1946.

load entries of most traffic tables are in terms of CCS. For theoretical studies the erlang is a more convenient unit and will be retained here.

c. Continuous Method

The busy condition of each circuit is represented by a fixed increment of electrical current through an ampere-hour meter. The instantaneous current is then analogous to the calls simultaneously present so that the meter, which integrates the current, may be calibrated to indicate hundred-call-seconds or erlang-hours directly. Although this method is potentially the most accurate, practical difficulties have limited its use.

In addition to these direct methods, there are several methods of indirect load measurement which, relying more heavily on traffic theory, make use of partial load indications, such as duration of group busy or the number of calls finding the group busy. Such measurements are less reliable than the direct measurements particularly when applied to underloaded groups.

This paper is concerned with the reliability of switch count load measurements since this method appears to have prospects of considerably wider adoption in the future. Main emphasis will be placed, both qualitatively and by the application of error formulas, on the relative effects of various measurement and traffic parameters on the accuracy of switch count measurements. Where long derivations of formulas are required they are deferred to the Appendix.

SOURCES OF ERROR

As has been described, switch count measurements yield the average number of calls found present when a group of circuits is scanned at fixed intervals during an observation period. Usually only that period of the day during which the load is greatest is of interest to the traffic engineer. Because the load during such periods also fluctuates from day to day, measurements of the loads for several days must be averaged to provide a useful load estimate.

There are two main sources of error, therefore, in switch count estimates of telephone traffic loads:

1. Each individual count of busy circuits is separated from the next by a time interval during which changes in load are not detected. Consequently, the load indicated by measurement may differ appreciably from the actual load carried. This difference can be decreased by decreasing the interval between scans.

2. Even if the load carried during a measurement period were known very accurately, it is still only a sample of the many loads that might be offered by the same source of traffic under statistically identical conditions. Therefore, the average of several load readings may be expected to be somewhat in error as an estimate of the true average of the traffic source. The latter will be referred to as the *source* load to distinguish it from the *carried* load.

Mechanical and human errors are likely to be present as well but, since they are not inherent in the switch count method, they will be neglected here.

SWITCH COUNT ERROR

As shown in the Appendix, for periods of observation which are relatively long with respect to average holding time made on traffic with certain assumed characteristics, the average error of switch counts in estimating traffic load *carried* in the same period is zero. The coefficient of variation of the error, which is the standard deviation of the error expressed in per cent of the traffic load carried, is given by:

$$V_x \doteq 100 \sqrt{\left[rc\,\mathrm{ctnh}\left(\frac{r}{2}\right) - 2\right]\frac{\bar{t}}{a'NT}} \qquad (1^*)$$

$$\mathrm{ctnh}\left(\frac{r}{2}\right) = \frac{1 + e^{-r}}{1 - e^{-r}} = \text{hyperbolic cotangent of } \frac{r}{2}$$

$$rc = T/\bar{t} > 20$$

where r = ratio of scan interval to holding time

$\bar{t}$ = average holding time

a' = *carried* load in erlangs

c = number of switch counts

T = length of observation period

N = number of observation periods

and where the following assumptions are made:

a. Calls originate individually and collectively at random.†

b. Holding times are exponentially distributed.

c. Congestion loss from the group is negligible.

* I have recently learned that these *carried* load formulas have been published by Conny Palm in *Tekniska Medelanden från Kungl. Telegrafstyrelsen*, 1941. nr. 7–9.

† See T. C. Fry, *Probability and Its Engineering Uses*, D. van Nostrand Co. Inc., New York, p. 216, for a definition of this condition.

As shown in the Appendix, this formula simplifies, when $r \leq 2$, to

$$V_x \doteq \frac{100}{c/T}\sqrt{\frac{1}{6a'NT\bar{t}}} \tag{2}$$

where c/T = rate of scan in cycles per time unit. From equation (2) it is apparent that if the scan interval is of the order of a holding time, the error of an estimate of traffic *carried* is inversely proportional to the rate of scan and inversely proportional to the square root of average load, holding time and hours of observation. For example, take the case where switch counts are made during the busy hour, five minutes apart on a trunk group carrying calls with an average holding time of 3 minutes and an average load of 5 erlangs (180 CCS). What is the error in the estimated load *carried* if the readings for ten days are averaged? (Assume conditions (a), (b) and (c) are met.) We have

N = 10 observation periods

T = 1 hour

$\bar{t}$ = 1/20 hour

a' = 5 erlangs

c = 12 scans per observation period

$rc = T/\bar{t}$ = 20 average holding times per observation period

From equation (2) since $T/\bar{t} = 20$ and $r = T/c\bar{t} = 1.7$

$$V_x = \frac{100}{12}\sqrt{\frac{1}{6\cdot 5\cdot 10\cdot 1\cdot 1/20}} = 2.15\%$$

If, as proposed in the Appendix, it is assumed that the error has a normal distribution, there is 90 per cent assurance that observed values will fall within $1.64V_x$, or in the example within 3.52 per cent, of the true average*. Note that this error limit would be halved if the rate of scan were doubled or if four times as many hours of observation were taken.

The coefficient of variation of the switch count error for constant values of $T/\bar{t}$ as a function of r is plotted on Fig. 1 for one observation period of a one erlang load. For loads other than one erlang the coefficient of variation is found by dividing by $\sqrt{a'N}$. Thus in the example we have, using the dotted curve,

$$V_x = \frac{15}{\sqrt{5\cdot 10}} = 2.1\%$$

* This assumes that a sufficient number of observations are taken so that a priori information may be neglected in making an estimate of the universe.

or more accurately using the solid curve,

$$V_x = \frac{16}{\sqrt{5 \cdot 10}} = 2.3\%$$

The error of using equation (2) is seen to be negligible for most purposes even when T/t is less than 20. The probability of an observation occuring within a given number of standard deviations is widely published for the normal curve. A few values are given below:

z	P_z Probability of exceeding $\pm z\sigma$ or $\pm zV$
0.6745	0.50
1.44	0.85
1.64	0.90
2.00	0.9545
3.00	0.9973

Fig. 2 is a plot for 40 observations of measured load vs carried load. Each observation was made for a half hour period on a panel line finder group with switch counts made at the start and middle of the period.

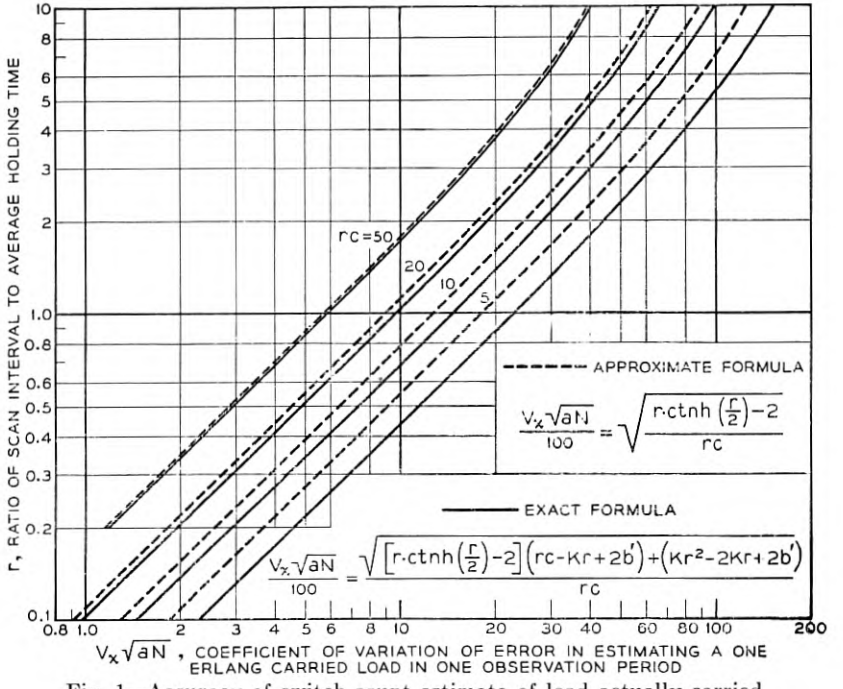


Fig. 1—Accuracy of switch count estimate of load actually carried.

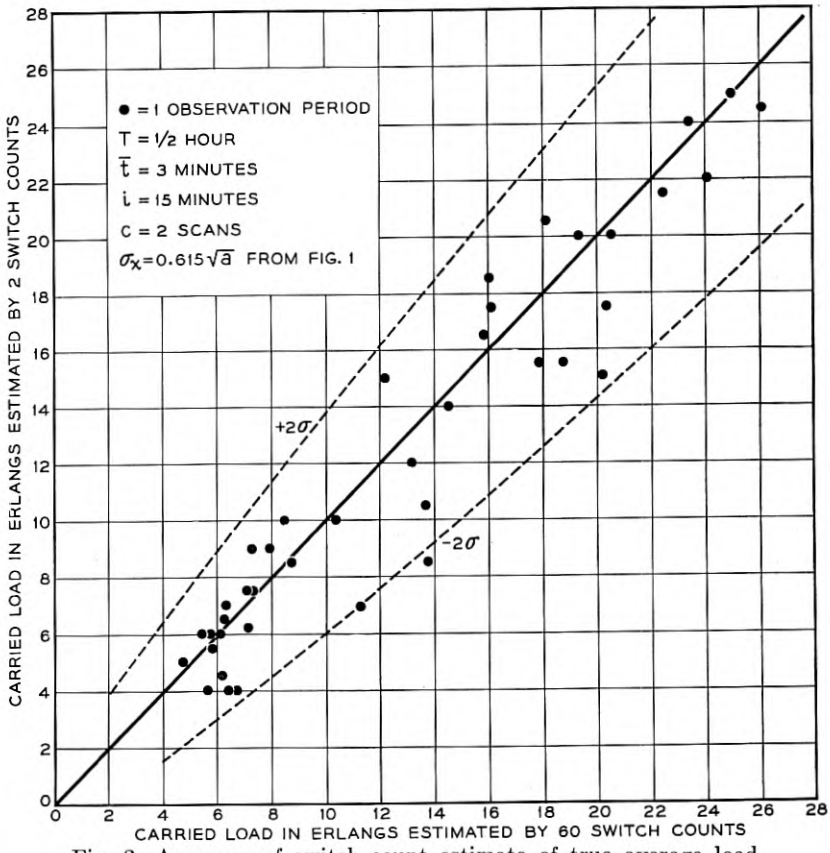


Fig. 2—Accuracy of switch count estimate of true average load.

This is compared with the average of switch counts made every 30 seconds which has a relatively negligible error. The average holding time per call for the group was 176 seconds. The accuracy of only two counts is surprisingly good and the observations are seen to lie satisfactorily between the 2σ limits.

ERROR OF TRAFFIC IN A GIVEN PERIOD AS AN ESTIMATE OF THE SOURCE LOAD

The average traffic carried in two different periods but generated by the same traffic source is subject to statistical variation. As a result, any measurement of load, even if measurement errors are eliminated, is only a sample of the wide range of traffic loads that might have been generated by the same source of traffic under identical circumstances.

J. Riordan has shown[2] that the standard deviation of the average traffic load for any one period is given by

$$\sigma_y = \sqrt{\frac{2a\bar{t}^2}{T^2}\left(\frac{T}{\bar{t}} - 1 + e^{-T/\bar{t}}\right)} \tag{3}$$

where a = the average source load

$\bar{t}$ = average holding time per call

T = length of observation period

(Assumptions are as before with an additional one that all periods are in statistical equilibrium)

When $\bar{t}/T \ll 1$ this reduces to the form also given by F. W. Rabe[3]

$$\sigma_y = \sqrt{\frac{2a\bar{t}}{T}} \tag{4}$$

or expressed in a per cent of the average

$$V_y = 100\sqrt{\frac{2\bar{t}}{aT}} \tag{5}$$

When N periods of length T are observed the coefficient of variation is reduced further to:

$$V_y = 100\sqrt{\frac{2\bar{t}}{aNT}} \tag{6}$$

In the example of the previous section,

$$N = 10$$

$$T = 1$$

$$\bar{t} = 1/20$$

$$a = 5$$

$$V_y = \sqrt{\frac{2 \cdot 1/20}{5 \cdot 10 \cdot 1}} = 4.47\%$$

COMBINATION OF ERRORS

Evidently if switch count readings are used to estimate the average which may be expected in other periods, the two errors described above should both be taken into account. The errors are probably correlated but this correlation is weak and at present no method of allowing for it

is evident. Such a refinement would probably change the equation for standard deviation only slightly from that derived for the independent case; therefore independence will be assumed. The standard deviation of the sum of two independent variables is the square root of the sum of the squares of the component standard deviations:

$$\sigma_s = \sqrt{\sigma_x^2 + \sigma_y^2} \tag{7}$$

$$= \sqrt{\left[\text{rctnh}\left(\frac{r}{2}\right) - 2\right]\frac{\bar{t}a'}{NT} + \frac{2\bar{t}a}{NT}} \tag{8}$$

Assuming $\frac{a'}{a}$ is approximately unity, that is, that carried load is approximately equal to source load,

$$V_s = 100\sqrt{\frac{\bar{t}}{anT}\,\text{rctnh}\left(\frac{r}{2}\right)} \tag{9}$$

In the example given,

$$V_s = 4.96\%$$

There is, then, 90 per cent assurance that the source average is within $1.64 \times 4.96 = 8.1$ per cent of the observed average. Note that doubling the switch count rate (which halves the switch count error) reduces the total error only to 7.6 per cent (about 6.7 per cent improvement), while doubling the number of hours of observations reduces the error to 5.9 per cent (about 30 per cent improvement). Plots of the coefficient of variation of a one hour observation of a one erlang load versus scan rate for various average holding times are given in Fig. 3 for a wide range of holding times. The coefficient of variation of error in estimating other loads may be found from Fig. 3 by dividing the unit load coefficient by $\sqrt{aNT}$. In the example, the unit load coefficient is found, by entering Fig. 3 with $\bar{t} = 3$ minutes and rate of scan $= c/T = 12/1$ scan cycles per hour, to be 35.0 per cent. Dividing by $\sqrt{5 \cdot 1 \cdot 10}$ gives a coefficient of variation of 4.96 per cent as before. It is evident from Fig. 3 that increasing scan rates is not a universal way to improve the accuracy of *source* load estimates.

CHOICE OF SCAN RATES

What then governs the choice of scan rate? Evidently increasing the rate increases the accuracy of *carried* load estimates to any point de-

sired. This is far from true if *source* load is being estimated. If the cost of making a scan is constant, increasing the number of observation periods and decreasing the scan rate will improve accuracy of *source* load estimates without changing measurement costs. The number of hours available for measuring, of course, limits this procedure, while the increase in accuracy becomes negligible as r becomes large. On the other hand, if the cost of each observation is only slightly affected by the cost of making additional scans, a high scan rate might be justified.

In applying the above relationships to traffic measurements, the usual question raised by the traffic engineer will be either how many hours of data need he take to be reasonably sure of his estimate or, conversely, how sure is he of an estimate based on available data. Assuming as before that the error distribution is normal, the per cent plus or minus error limits within which a proportion, P_z, of the estimates will fall is given by zV_s; the value of z corresponding to any selected P_z may be found from tables of the normal probability distribution. "Reasonably sure" is often taken to mean that there is 90 per cent assurance that the error does not exceed 5 per cent. When P_z is 0.90, z is 1.64, so that under this condition $1.64V_s = 0.05$, or $V_s = 0.0305$. Given scan rate and holding time, V_s is proportional to $1/\sqrt{aNT}$ according to equation (9) or Figure 3. When V_s is held constant, aNT is constant so that the plot of log NT against log a is linear, as shown in Figs. 4 and 5. The number of hours needed to meet any chosen reliability

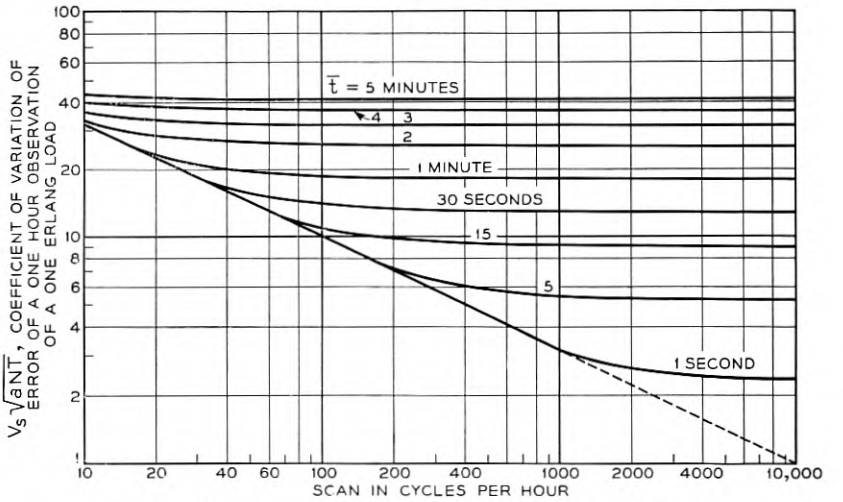


Fig. 3—Efficiency of switch counts for usage measurement.

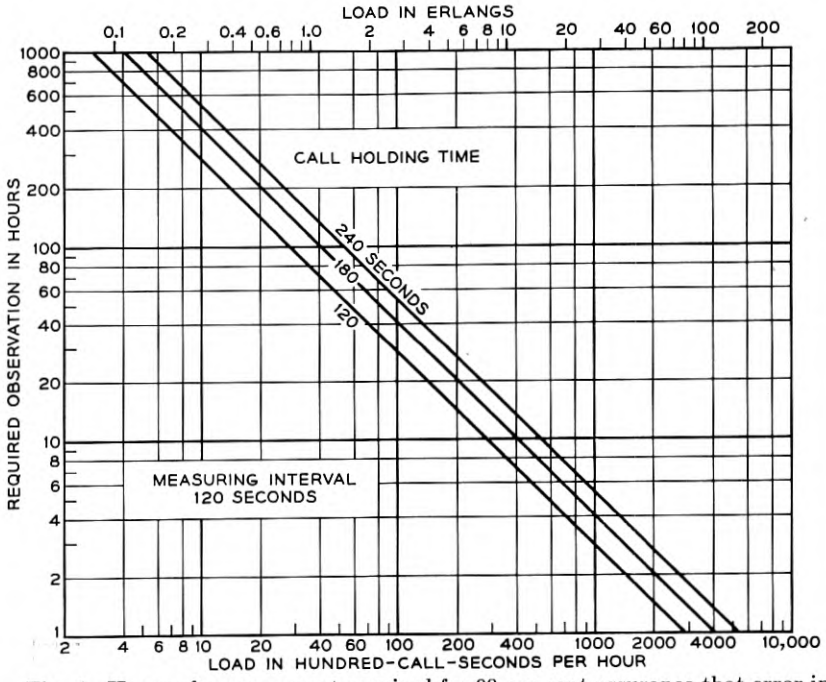


Fig. 4—Hours of measurement required for 90 per cent assurance that error in estimating *source* load does not exceed plus or minus 5 per cent when measuring interval is 120 seconds.

requirements may then be read directly from such graphs. In the second type of question, z, NT, scan rate and holding time are fixed so that zV_s is proportional to $1/\sqrt{a}$. Plotting log zV_s against log $\sqrt{a}$ again gives a linear plot as shown on Fig. 6.

In the numerical example above, the limits of error corresponding to 90 per cent assurance may be read from Fig. 6 which is plotted for the appropriate assurance, average holding time and scan interval. Reading the error limits at the point where the 10 hours measured line crosses 180 CCS (5 erlangs) gives ± 8.1 per cent as before. Fig. 5 may be entered to find the total number of hours required to reduce this error to 5 per cent. Reading at the point where the 180 second holding time line crosses 180 CCS gives 26 hours.

QUALITATIVE EXTENSION OF THEORETICAL APPROACH

The original traffic assumptions made in deriving the theoretical results above are:

a. Calls originate collectively and individually at random.

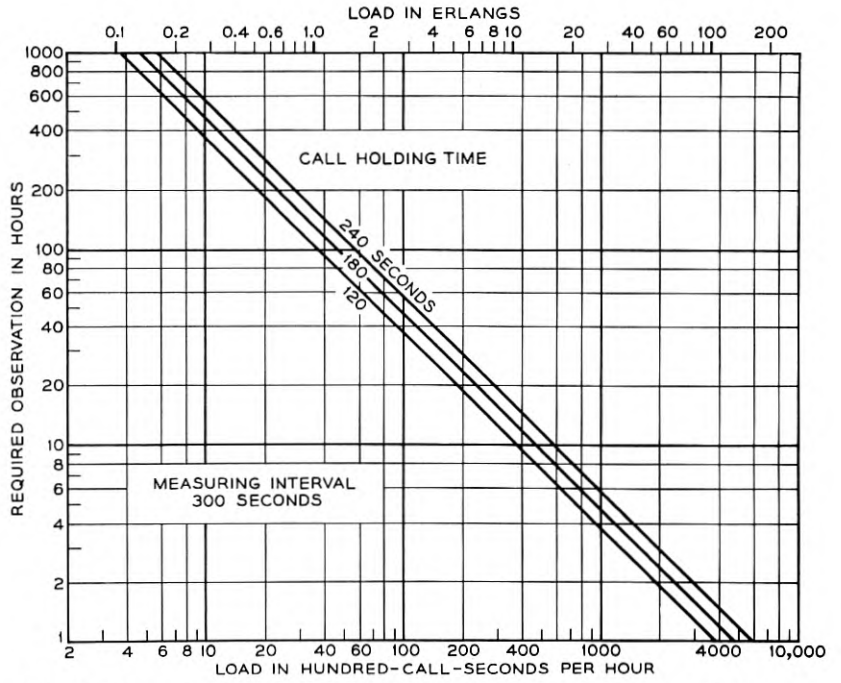


Fig. 5—Hours of measurement required for 90 per cent assurance that error in estimating *source* load does not exceed plus or minus 5 per cent when measuring interval is 300 seconds.

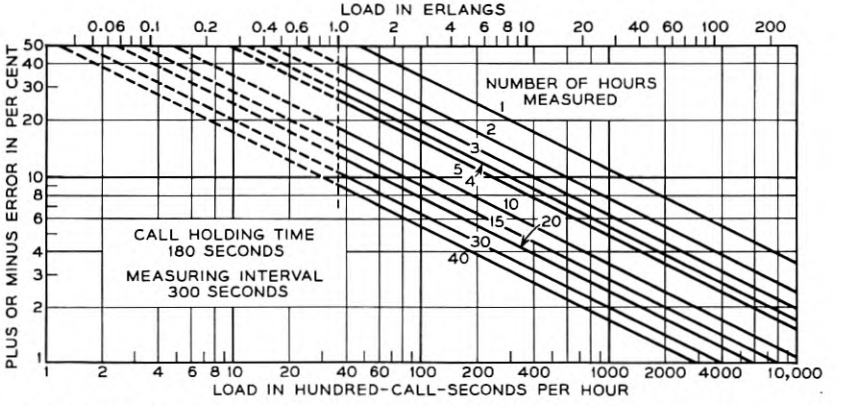


Fig. 6—Limits of error reached with 90 per cent assurance in estimating *source* load.

b. Holding times are exponentially distributed.
c. Congestion loss from the group is negligible.
d. Observation periods are in statistical equilibrium.

How do departures from these assumptions affect the reliability of usage measurements?

a. Holding Time Distribution

Experience in application of delay and loss formulas has shown that theories based on exponential holding times are often applicable to other holding time distribution cases which have a wide range. However, for a constant holding time distribution special theories often are called for. The average and standard deviation of switch count estimates of *carried* load when holding time is constant, are given in part 2 of the Appendix. It is shown there that for estimates of *carried* load,

$$\bar{x} = 0$$

$$r \geq 1 \qquad V_x = 100 \sqrt{\frac{\bar{t}}{a'NT}(r-1)} \tag{10}$$

$$r \leq 1 \quad \text{minimum } V_x = 0 \quad (r = 1, \tfrac{1}{2}, \tfrac{1}{3}, \tfrac{1}{4}, \text{ etc.})$$

$$\text{maximum } V_x = 100 \frac{r}{2} \sqrt{\frac{\bar{t}}{a'NT}} \quad (r = \tfrac{2}{3}, \tfrac{2}{5}, \tfrac{2}{7}, \text{ etc.}) \tag{11}$$

Since constant holding times found in practice are often very short, the case of $r \geq 1$ is the most likely to be met. For all values of r greater than one, the error given by formula (1) for exponential holding times is somewhat greater than the error given by formula (10) for constant holding times, so use of formula (1) for the constant holding time case is conservative. For values of r less than 1, the error is an oscillating function of r. The coefficient of variation varies from zero to 23 per cent above that for exponential holding times. Where r may not be accurately known the formula for exponential holding times again seems appropriate.

In making estimates of the *source* load when the holding time is constant, if $r \geq 1$, each scan is uncorrelated with any other, since no call can be counted twice, and may be considered a random sample of traffic. There are a total of Nc scans which have an average scan of a and standard deviation $\sqrt{a}$. The average error in estimating a is, therefore:

$$\bar{s} = 0$$

with coefficient of variation

$$V_s = 100 \sqrt{\frac{1}{aNc}} = 100 \sqrt{\frac{\bar{t}}{aNT} r} \tag{12}$$

Equation (12) may also be derived with the procedure used for equation (9) using $\sigma_s^2 = \sigma_x^2 + \sigma_y^2$. For values of r large enough to make $\operatorname{ctnh}\left(\frac{r}{2}\right) \doteq 1$ equation (12) is approached by equation (9). For smaller values of r (but with r still greater than 1), V_s for constant holding times is less than V_s for exponential holding times. When $r = 1$, there is no *carried* load error. For values of r less than 1, the coefficient of variation of error in estimating *source* load average will vary from

$$\sqrt{\frac{\bar{t}}{aNT}} \quad \text{to} \quad \sqrt{\frac{\bar{t}}{aNT}\left(1 + \frac{r^2}{2}\right)}$$

depending on the exact value of r. It is interesting to note that V_s for $r = 0.5$ is the same as for $r = 1.125$.

b. Loss

The effect of loss in the group depends upon the disposition of the lost calls. In general, accuracy in measuring *carried* load increases with increased loss because under these circumstances fewer load changes occur between scans. This is evident in the extreme case of a group which is 100 per cent loaded; a single switch count gives a correct reading for any length period. Obviously load readings at 100 per cent occupancy are not very useful in estimating *offered* loads since the amount of lost load cannot even be guessed at. However, in the cases of lost calls held (Poisson) or cleared (Erlang B), the *offered* load may be estimated from the *carried* load (less and less accurately as occupancy increases) and in the case of lost calls delayed the *offered* and *carried* loads are likely to be the same even at high occupancies. With high loss, therefore, estimates of *source* load are subject to errors not considered in deriving equation (99); however, switch count error in estimating *carried* load will be materially less than predicted by equation (1).

c. Random Call Origination

On trunk groups which are alternate routes, calls may no longer be considered as originating at random. The resultant grouping of call originations will tend to decrease the accuracy of switch count measurements in estimating *carried* load; however, there is a corresponding decrease in accuracy in estimating the *source* load from the *carried* load so that accuracy in estimating *carried* load may be less worthwhile.

d. Statistical Equilibrium

Statistical equilibrium may be thought of as the absence of trends in subscriber calling rates or holding times with the passage of time. The effect of trends on switch count accuracy in measuring *carried load* is very small except where the changes in traffic level are frequent and abrupt with respect to the scan frequency. Such traffic behavior is rare.

Trends within the busy hour complicate the problem of estimating the average *source load*. However, it can be shown that if the trends are small (in the order of 10 per cent to 20 per cent) little error is introduced by assuming that no trend exists. Large trends (in the order of 100 per cent), however, may indicate that the traffic source is so unstable that more hours of traffic data should be taken in order to insure that the sample is representative.

Trends from day to day do not affect the *source* load estimates in the same way as within hour trends. The *source* loads are seldom exactly the same on any two days although in most offices a load pattern is repeated from week to week. The traffic engineer may be interested in the average *source* load of either a typical week day in the busy season or, sometimes, of the average of the two highest days in the week. As long as the *source* load of each particular day remains close to the average for that day of the week, the general average for several different days of the week, will be known with about the same accuracy as if they had all come from a common source. If, however, there is no stable pattern in the *source* load, a third error in estimating the average is generated. There is some difficulty in determining whether or not variations in load, as indicated by measurements, are due to sampling variations or to an unstable source. Quality control methods might be used to detect instability but gathering and processing sufficient data for such an analysis might prove uneconomical. In general, if a traffic engineer feels that his *source* load is unstable he will need more hours of data than indicated by formula (9) to meet a given criterion of reliability.

CONCLUSIONS

A theoretical approach to the problem of the accuracy of switch count measurements in estimating *carried* load and average *source* load has been explored. It is believed that the assumptions made are satisfied sufficiently often in practice to enable fairly wide application of the results of this exploration to traffic measurements. However, it should be kept in mind that where the assumptions are clearly not valid, special allowances will need to be made. In any case, the confidence placed in

usage measurements by a traffic engineer is a function of his experience and judgment. It is hoped that the results of this study will add to the knowledge essential to sound traffic engineering.

APPENDIX

DERIVATION OF SWITCH COUNT ERROR IN ESTIMATING CARRIED LOAD—WITH EXPONENTIAL HOLDING TIMES

This derivation is based on a similar derivation by R. I. Wilkinson[1]. However, since load rather than holding time is of interest here, the emphasis has been somewhat shifted.

Assume that switch count measurements are being taken on traffic with:

a. Calls originated individually and collectively at random
b. Exponentially distributed holding times
c. Negligible loss

Let i = interval between scans

$\bar{t}$ = average holding time

a' = traffic carried, in erlangs

T = length of observation period

$r = \frac{i}{\bar{t}}$ = number of holding times in a scan interval

$c = \frac{T}{i}$ = number scans in observation period

$rc = \frac{T}{\bar{t}}$ = number of holding times in observation period

N = number of observation periods.

Consider that the observation period begins with the first scan and ends i time units after the last scan. It is desired to find the error in estimating the true load carried by averaging the number of circuits found busy on each scan. Following Wilkinson's method we will first estimate the error of the switch count method in measuring the contribution of a single call to total usage and then modify it to take account of n calls. Calls of two types must be considered, those originating outside the interval and extending into it, Type I, and those originating within the interval, Type II. Both types may be subdivided depending on whether or not they extend beyond the end of the observation period. These are

indicated in Fig. 7. Only that part of a call which falls within the observation period contributes to the usage of that period. First the error made by switch counts in measuring this contribution will be derived.

Type I

Consider a call which is already in progress at the start of the observation period. Its duration beyond that point, according to theory, will be exponentially distributed about an average of $\bar{t}$.

If this duration, t, is between 0 and i, the call will be counted once (a measured contribution of i erlang hours) and a positive error of $x = i - t$ will be made. The same error will be made if $t = 2i - x$ so that the call

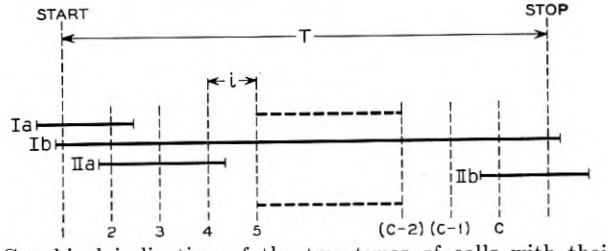


Fig. 7—Graphical indication of the two types of calls with their two subdivisions.

is counted twice and so forth. Summing all the ways of making an error x, we have:

$$P(x)\,dx = f(i - x) + f(2i - x) + \cdots f(ci - x) \tag{1}$$

where $f(i - x)$ is the probability of $t = i - x$ and

$$f(t) = \frac{1}{\bar{t}}\, e^{-t/\bar{t}}\, dx$$

Calls lasting beyond ci neither start nor end in the observation period so that their contribution is measured without error. For these:

$$P(0) = P(t \geq ci) = e^{-rc} \tag{2}$$

Therefore:

$$P_{x>0}(x)\,dx = \frac{1}{\bar{t}}\, e^{-\frac{i-x}{\bar{t}}}\, dx + \frac{1}{\bar{t}}\, e^{-\frac{2i-x}{\bar{t}}}\, dx + \cdots + \frac{1}{\bar{t}}\, e^{-\frac{ci-x}{\bar{t}}}\, dx \tag{3}$$

Letting

$$y = \frac{x}{\bar{t}}; \qquad e^{-cr} = b = 1 - b'; \qquad K = \sum_{n=0}^{c-1} e^{-nr} = \frac{b'}{1 - e^{-r}}$$

$$P_{x>0}(x)\,dx = e^{y}e^{-r}K\,dy \tag{4}$$

The moment generating function $M_I(\alpha)$ of y is:

$$\begin{aligned} M_I(x) &= \int_0^r P(y)\,dy\,e^{\alpha y} + e^{-rc} \\ &= b + K\frac{e^{r\alpha} - e^{-r}}{1 + \alpha} \end{aligned} \tag{5}$$

Neglecting terms of order higher than α^2,

$$M_I(\alpha) = 1 + \alpha(rK - b') + \frac{\alpha^2}{2}(Kr^2 + 2b' - 2rK) \tag{6}$$

Type II

Calls of Type II may have either positive or negative errors given by:

$$\begin{aligned} P_{x\le 0}(x)\,dx = \frac{i + x}{i}\,[&f_0(-x) + f_1(i - x) + f_2(2i - x) \\ &+ \cdots + f_{c-1}((c - 1)i - x)] \\ + g_0(-x) + g_1(i - x) + g_2(2i - x) + \cdots + g_{c-1}&[(c - 1)i - x] \end{aligned} \tag{7}$$

$$P_{x\ge 0}(x)\,dx = \frac{i - x}{i}\,[f_1(i - x) + f_2(2i - x) + \cdots + f_c(ci - x)]$$

where $f_n(ni - x)$ = probability that a Type II call has length $ni - x$ and ends before the end of the observation period.

$$= \frac{1}{\bar{t}}\,e^{-\frac{ni-x}{t}} \cdot \frac{c - n}{c}\,dx$$

$g_n(ni - x)$ = probability that a Type II call starts $ni - x$ before the end of the observation period and ends after the end of the observation period.

$$= \frac{1}{T}\,e^{-\frac{ni-x}{t}}\,dx$$

Equation (7) becomes:

$$P_{x\leq 0}(x)\,dx = \left[\frac{i+x}{i}\sum_{n=0}^{c-1}\frac{1}{\bar{t}}\,e^{-\frac{ni-x}{\bar{t}}}\cdot\frac{c-n}{c}+\sum_{n=0}^{c-1}\frac{1}{T}\,e^{-\frac{ni-x}{\bar{t}}}\right]dx$$

$$P_{x\geqq 0}(x)\,dx = \left[\frac{i-x}{i}\sum_{n=1}^{c}\frac{1}{\bar{t}}\,e^{-\frac{ni-x}{\bar{t}}}\cdot\frac{c-n}{c}\right]dx \tag{8}$$

Letting

$$\frac{x}{\bar{t}} = y, \qquad K = \sum_{n=0}^{c} e^{-nr} \text{ as before}$$

and noting that

$$\sum_{n=0}^{c-1} ne^{-nr} = \frac{K-cb}{1-e^{-r}}$$

$$P_{x\leq 0}(x)\,dx = e^{y}\left[\left(1+\frac{y}{r}\right)\left(K-\frac{1}{c}\,\frac{e^{-r}K-cb}{1-e^{-r}}\right)+\frac{K}{rc}\right]dy$$

$$P_{x\geq 0}(x)\,dx = e^{-r}e^{y}\left[\left(1-\frac{y}{r}\right)\left(K-\frac{1}{c}\,\frac{K-cb}{1-e^{-r}}\right)\right]dy \tag{9}$$

The moment generating function of this pair of equations is the sum of their separate m.g.f.'s:

$$M_{II}(\alpha) = \int_{-r}^{0} P_{y\leq 0}(y)e^{\alpha y}\,dy + \int_{0}^{r} P_{y\geq 0}(y)\,dy\,e^{\alpha y}\,dy$$

$$= \frac{\begin{aligned} &rc+K+c-2\,\frac{c-Ke^{-r}}{1-e^{-r}}+\frac{c-K}{1-e^{-r}}\,e^{\alpha r}+\frac{(c-K)e^{-r}}{1-e^{-r}}\,e^{-\alpha r} \\ &\qquad\qquad +\alpha(K+rc-e^{-r}Ke^{-\alpha r})\end{aligned}}{rc(1+\alpha)^2} \tag{10}$$

Neglecting terms of order higher than α^2,

$$M_{II}(\alpha) = \frac{1}{rc}\left\{rc+\alpha(b'-rK)+\frac{\alpha^2}{2}\left[\left(rc\operatorname{tnh}\left(\frac{\alpha}{2}\right)-2\right)(rc-rK+2b')\right]\right\} \tag{11}$$

Now the number of Type I calls present in an observation is a variable—with average "a" and a Poisson distribution. Similarly the number of Type II calls is a variable, independent of the number of Type I calls, with an average of "$a\,\frac{T}{\bar{t}}$" or "arc" and a Poisson distribution. Ac-

cording to the laws governing the compounding of variables the moment generating function of the sum of n variables y, when n is also variable with generating function $G(t)$, is $G(M(\alpha))$ where $M(\alpha)$ is the moment generating function of y.

The generating function of a Poisson variable with average "a" is e^{-a+at} so that

$$\begin{aligned} G(M_I(\alpha)) &= e^{-a+aM_I(\alpha)} \\ G(M_{II}(\alpha)) &= e^{-arc+arcM_{II}(\alpha)} \end{aligned} \tag{12}$$

These independent variables may be added by multiplying their moment generating functions to give the m.g.f. of the total measurement error of the *carried* load

$$M(\alpha) = e^{-(a+arc)+aM_I(\alpha)+arcM_{II}(\alpha)} \tag{13}$$

From (6), (11) and (13) the following parameters are found:

$$\bar{y} = 0 \tag{14}$$

$$\sigma_y^2 = arc\left[\left(r\operatorname{ctnh}\left(\frac{r}{2}\right) - 2\right)\left(1 - \frac{K}{c} + \frac{2b'}{rc}\right) + \left(\frac{rK}{c} - 2\frac{K}{c} + \frac{2b'}{rc}\right)\right]$$

If, now, rc is sufficiently large

$$\sigma_y \doteq \sqrt{arc\left[r\operatorname{ctnh}\left(\frac{r}{2}\right) - 2\right]} \tag{15}$$

It is more convenient to deal with the standard deviation expressed as per cent of the *carried* erlang load, the coefficient of variation. This is done by multiplying both sides of equation (15) by $\bar{t}$ to convert the time dimension from holding times to hours, dividing by T to convert from erlang-hours to erlangs, dividing by a' to convert to proportion of *carried* load, and multiplying by 100 to convert to per cent. Assuming $\frac{a}{a'^2}$ is approximately $\frac{1}{a'}$:

$$V_x \doteq 100\sqrt{\frac{\left[r\operatorname{ctnh}\left(\frac{r}{2}\right) - 2\right]\bar{t}}{a'T}}$$

When N observations are made this reduces further to

$$V_x \doteq 100\sqrt{\frac{\bar{t}}{a'NT}\left[r\operatorname{ctnh}\left(\frac{r}{2}\right) - 2\right]}$$

$$\text{Now } \operatorname{ctnh}(x) = \frac{1}{x} + \frac{x}{3} - \frac{x^3}{45} + \frac{2x^5}{945} - \cdots \qquad (x^2 < \pi^2)$$

and for $r \leq 2$

$$\frac{r/2}{3} \gg \frac{(1/2)^3}{45}$$

Therefore

$$r\text{ctnh}\left(\frac{r}{2}\right) \doteq 2 + \frac{r^2}{6}$$

$$V_x \doteq 100 \sqrt{\frac{\bar{t}}{a'NT}\frac{r^2}{6}} = \frac{100}{c/T}\sqrt{\frac{1}{6a'NT\bar{t}}} \qquad (r \leq 2) \quad (16)$$

The error in *carried* load may be considered as the sum of a large number of independent errors. Its distribution may, therefore, be expected to approach the normal distribution. Comparison of the third and fourth moments of the normal distribution with those of the error distribution (which may be obtained from equation (13)) show good agreement for values of a' greater than 1.

DERIVATION OF SWITCH COUNT ERROR IN ESTIMATING CARRIED LOAD WITH CONSTANT HOLDING TIMES

Wilkinson has shown[1] that, for constant holding time, switch count error in measuring the holding time of one call has an average

$$\bar{x} = 0$$

and standard deviation

$$\sigma_x = \sqrt{-x_1 x_2}$$

where $T \gg \bar{t}$

$T \gg i$

x_1 = negative error

x_2 = positive error

Divide the problem into two parts:

1. For $r > 1$

$$x_1 = -\bar{t}$$

$$x_2 = i - \bar{t}$$

$$\sigma_x = \sqrt{\bar{t}i - \bar{t}^2} = \bar{t}\sqrt{r - 1} \qquad (17)$$

2. For $r \leq 1$

$$\text{Min. } \sigma_x = 0 \qquad \text{for} \qquad r = 1, \tfrac{1}{2}, \tfrac{1}{3}, \tfrac{1}{4}, \text{ etc.}$$

$$\text{Max. } \sigma_x = \sqrt{\frac{i}{2} - \frac{i}{2}} = \frac{i}{2} = \bar{t}\,\frac{r}{2} \tag{18}$$

$$\text{for} \qquad r = \tfrac{2}{3}, \tfrac{2}{5}, \tfrac{2}{7}, \text{ etc.}$$

Expressing this error in terms of carried load and proceeding as in Part I of the Appendix

1. $r > 1$ $\qquad V_x = 100 \sqrt{\dfrac{\bar{t}}{a'NT}\,(r-1)}$ (19)

2. $r \leq 1$ Min. $V_x = 0$

$$\text{Max. } V_x = 100\,\frac{r}{2}\sqrt{\frac{\bar{t}}{aNT}} \tag{20}$$

$$= \frac{100}{c/T}\sqrt{\frac{1}{4a'NT\bar{t}}}$$

Equation (20) compares favorably with the exponential holding time coefficient of variation of error of

$$\frac{100}{c/T}\sqrt{\frac{1}{6a'NT\bar{t}}}$$

REFERENCES

1. R. I. Wilkinson, "The Reliability of Holding Time Measurements," *Bell System Tech. J.*, **20**, pp. 365–404, October, 1941.
2. J. Riordan, "Telephone Traffic Time Averages," *Bell System Tech. J.*, **30**, pp. 1129–1144, October, 1951.
3. F. W. Rabe, "Variations of Telephone Traffic," *Electrical Communications*, **26**, pp. 243–248, 1948.

The following books contain descriptions of the use of generating functions in solving probability problems:

4. A. C. Aitken, *Statistical Mathematics*, Oliver and Boyd, Ltd., Edinburgh, 1947, pp. 16–23.
5. W. Feller, *An Introduction to Probability Theory and Its Applications*, John Wiley & Sons, Inc., New York, 1950, Chap. 11.

Network Representation of Transcendental Impedance Functions

BY M. K. ZINN

(Manuscript received November 5, 1951)

The purpose of the paper is to show that the admittance or impedance of certain continuous structures, such as, for example, a finite length of transmission line of any sort, or resonant cavity, can be represented exactly at all frequencies by a network comprising lumps of constant resistance R, inductance L, conductance G and capacitance C. The network will contain an infinite number of branches, in general, although a finite number may be used if it is desired to represent only certain modes.

The procedure is based upon a proposition known to students of function theory as "Mittag-Leffler's theorem," which amounts, roughly, to an extension of rational functions to apply to transcendental functions of the type encountered in the theory of continuous structures.

Several illustrative examples of the network synthesis are given.

GENERAL

Students of network theory are familiar with the fact that the impedance at a pair of terminals in a linear network comprising a finite number of resistors, inductors and capacitors, connected in any manner, is a rational function of the frequency having, in general, the fractional form of one polynomial divided by another. They are also familiar with the partial fraction rule whereby the function can be broken up into a series of elementary fractions, each of which exhibits one of the poles of the original function. This form is sometimes useful in the problem of network synthesis, where the impedance function is given and the object is to find a network having this impedance.

The purpose of the present paper is to show how a similar procedure can be carried out for certain transcendental impedance functions pertaining to structures having distributed constants, such as, for example, a resonant cavity or a piece of transmission line. The method employs a well-known proposition of function theory, which is usually referred to as Mittag-Leffler's theorem. This theorem provides a tool for breaking up a transcendental meromorphic function into an infinite series of simple fractions in much the same way as the partial fraction rule is used to break up a rational meromorphic function. The series representation

provides a means of determining a network of resistors, inductors and capacitors that will have an impedance equal to the specified transcendental impedance function. This process will be referred to as obtaining a "network representation" of the function. If the given function is the impedance of some continuous (i.e., non-lumped) electric structure, the result will be an equivalent network for the structure. For other purposes, such as, possibly, analogue methods of computing, the given function may not arise from any electrical structure. In either case, the network representations to be derived are possible only if the function satisfies certain restrictions, which are stated in the section immediately following.

The discussion is confined to transcendental impedance functions because of the technological interest in the electromagnetic structures with which they are associated and because they have not received as much attention as rational functions in the literature dealing with network synthesis. The problem with which this paper is concerned can then be stated as follows: given, a transcendental impedance function satisfying certain conditions: to determine a network comprising elements of constant resistance, inductance and capacitance whose driving-point impedance function, at a pair of terminals, will equal the given function at all frequencies, real and complex (except at the poles).

For illustration of the procedure, three examples are given. The first is the impedance of a short-circuited or open-circuited transmission line in which the distributed primary constants, R, L, G and C are assumed to be invariable with frequency. The second and third examples are the impedances of resonant cavities driven in two different modes. In these examples the variation of resistance with frequency, due to "skin-effect," is taken into account.

IMPEDANCE FUNCTIONS

The functions under discussion will be referred to as "impedance functions" with the understanding that the term is meant to include "admittance functions" as well. By reason of the duality principle that runs through all electric circuit theory, any general proposition developed for one must apply to the other. The functional designation, $F(p)$, will be used to denote either an impedance or an admittance function. When a distinction is necessary, the impedance will be designated by $Z(p)$ and the admittance by $Y(p)$. The independent complex variable p is the generalized radian frequency. (For sustained sinusoidal currents and voltages, $p = i\omega = 2\pi if$ where f is the real frequency.)

For the applications contemplated, $F(p)$ is a transcendental mero-

morphic function, which term implies that the function is given by the ratio of two entire functions, one or both of which is transcendental, and that the singularities of the function are ordinary poles, except for the point at infinity, which is an essentially singular point. In order to realize the particular network developments to be given, it will be supposed that the function satisfies the further restrictions given below:

(1) All the poles lie in the left half of the p-plane with none on the imaginary axis.

(2) $F(\bar{p}) = \bar{F}(p)$. (The superbar denotes the complex conjugate of the unbarred symbol.)

(3) Real part $[F(i\omega)] \geq 0$ for all real values of ω.

These three conditions are necessary to insure that the function is the impedance of a possible linear, passive electric circuit structure. Interpreted physically in terms of this possible equivalent structure, the first condition specifies that the structure shall be stable; that is, every natural mode of oscillation dies away exponentially. The second condition specifies that the natural oscillations are real functions of time. The third condition specifies that if a sinusoidal current flows at the driving-point terminals of the equivalent structure, the average real power delivered to it will be positive. Since these three conditions, or their equivalents, are frequently mentioned in discussions of network theory, it is assumed that they are understood without more detailed explanation.

In addition to the above restrictions on the form of the impedance function, the following two conditions, while not necessary, will be imposed to limit the scope of the discussion:

(4) All the poles of $F(p)$ are simple.

(5) $F(p) = 0(1)$, exactly, as $| p | \rightarrow \infty$ everywhere except at the poles.

Condition (4), while limiting the scope of the exposition required, does not restrict the application of the results in any important way, because most impedance functions for which a network representation may be required have only simple poles.

Condition (5) implies that as p increases along any straight line drawn through the origin and not passing through any pole of $F(p)$, the modulus of $F(p)$ either approaches a limit or oscillates between finite limits. The physical implication of this condition is that the response of the network as a function of time to a suddenly applied cause begins with a discontinuity of the same degree as that of the cause. For example, the current response of the network to an applied step of voltage begins with a finite discontinuity. This behavior is a characteristic of continuous (non-lumped) electromagnetic structures, which furnish the principal application of the network developments to be described.

MITTAG-LEFFLER'S THEOREM[7]

Let the poles of the given function $F(p)$ be $p_1, p_2, p_3 \cdots$, where

$$0 < |p_1| \leq |p_2| \leq |p_3| \cdots$$

and let the residues at the poles be $A_1, A_2, A_3 \cdots$, respectively. Suppose that it is possible to draw a sequence of closed contours, C_n, such that C_n encloses $p_1, p_2, \cdots p_n$ but no other poles and such that the minimum distance of C_n from the origin tends to infinity with n. Suppose also that $F(p)$ satisfies conditions (2), (4) and (5) above. Then Mittag-Leffler's theorem gives the following series development for $F(p)$:

$$F(p) = F(0) + \operatorname*{Limit}_{N\to\infty} \sum_{n=-N}^{N} \left(\frac{A_n}{p - p_n} + \frac{A_n}{p_n} \right) \tag{1}$$

The notation here used employs the convention that

$$p_{-n} = \bar{p}_n \qquad \text{and} \qquad A_{-n} = \bar{A}_n,$$

since, by virtue of condition (2), the poles occur in conjugate complex pairs. The value, $n = 0$, then allows for a pole on the negative real axis.

Given any suitable function, the procedure is to determine its value for $p = 0$ and the location of its poles. The residues are next determined by

$$A_n = \operatorname*{Limit}_{p\to p_n} (p - p_n)F(p).$$

Then the Mittag-Leffler expansion can be written down at once.

NETWORK REPRESENTATION

In the series (1) the terms occur in pairs with conjugate complex poles and residues. The object is to obtain a network representation of each such pair of terms. If $F(p)$ is taken as an admittance, the branches representing the pairs of terms will all be connected in parallel; if $F(p)$ is taken as an impedance, they will all be connected in series.

Methods for obtaining a network representation for a rational function, such as the one comprising a pair of terms in the series (1), are well known. It is only necessary to describe certain procedures of particular application to the present problem. Brune[5] has stated that the necessary and sufficient condition for a network representation of a rational function of p to be realizable is that it be a "positive real function," that is, a function that is real for real values of p and whose real part is positive,

or zero, when the real part of p is positive, or zero. In view of conditions (1) and (2) above, only one test[12] need be applied to each pair of terms of the series (1): the sum of a pair of terms will be a positive real function if, and only if, the real part of their sum is greater than, or equal to, zero for all purely imaginary values of p.

The general term pair for which a network representation is sought is

$$F_n(p) = \frac{A_n}{p - p_n} + \frac{\bar{A}_n}{p - \bar{p}_n} + \frac{A_n}{p_n} + \frac{\bar{A}_n}{\bar{p}_n} = P_n(p) - P_n(0) \qquad (2)$$

Evidently two cases can be distinguished at the outset, depending upon whether $P_n(0)$ is positive or negative. If $P_n(0)$ is positive, the network branch, in order to be realizable, should be designed to represent $P_n(p)$. The left-over negative term, $-P_n(0)$, then can be absorbed in the positive first term, $F(0)$, of the series (1); more will be said of this later. If, on the other hand, $P_n(0)$ is negative, the network branch should represent the whole term, $P_n(p) - P_n(0)$. This procedure insures that the real part of the branch impedance will be positive, or zero, at zero and infinite frequencies. To guarantee that the resistance is positive at all other frequencies requires further tests now to be specified.

Let the real and imaginary coefficients of the poles and residues of the n^{th} term be

$$p_n = -\alpha_n + i\beta_n, \qquad \bar{p}_n = -\alpha_n - i\beta_n$$
$$A_n = a_n + ib_n, \qquad \bar{A}_n = a_n - ib_n$$

(With this notation, α_n and β_n are always positive; a_n and b_n can be either positive or negative.) Then (dropping the subscripts)

$$P(p) = \frac{2(a\alpha - b\beta) + 2ap}{\alpha^2 + \beta^2 + 2\alpha p + p^2}$$

$$R[P(i\omega)] = \frac{2(a\alpha - b\beta)(\alpha^2 + \beta^2) + 2\omega^2(a\alpha + b\beta)}{(\alpha^2 + \beta^2)^2 + 2\omega^2(\alpha^2 - \beta^2) + \omega^4} \qquad (3)$$

$$P(0) = \frac{2(a\alpha - b\beta)}{\alpha^2 + \beta^2}$$

$$R[P(i\omega) - P(0)] = \frac{-2(a\alpha^3 - 3\alpha^2 b\beta - 3a\alpha\beta^2 + b\beta^3)\omega^2 - 2(a\alpha - b\beta)\omega^4}{(\alpha^2 + \beta^2)[(\alpha^2 + \beta^2)^2 + 2(\alpha^2 - \beta^2)\omega^2 + \omega^4]}$$

The necessary and sufficient conditions[12] for the real part of a rational function of p to be positive, or zero, for purely imaginary values of p are that the function be positive for $p \to \pm i\infty$ and have no imaginary roots of odd multiplicity. When this test is applied to the functions $P(p)$ and

$P(p) - P(0)$, as given by (3), the following conditions are obtained: *$P(p)$ will be a positive real function if, and only if,*

$$a\alpha - b\beta > 0; \qquad \text{i.e.} \qquad P(0) > 0 \tag{4}$$

and

$$a\alpha + b\beta > 0$$

$P(p) - P(0)$ will be a positive real function if, and only if,

$$a\alpha - b\beta < 0; \qquad \text{i.e.} \qquad P(0) < 0 \tag{5}$$

and

$$a\alpha^3 - 3\alpha^2 b\beta - 3a\alpha\beta^2 + b\beta^3 < 0.$$

If all terms of the series satisfy one or the other of these conditions, network branches can be devised to represent all the terms and all the R, L, G, C elements of the branches will be positive.

In case all the terms are of the type where $P_n(0)$ is positive, so that the network branches are made to represent $P_n(p)$, the left-over constant terms can be collected and added to the first term, $F(0)$, of the series. This collection of terms then must be represented by a final branch of pure resistance, or conductance, of value,

$$F(0) - \sum_{n=0}^{\infty} P_n(0)$$

If the sum of the variable terms approaches zero for $p \to \pm i\infty$, the final constant term supplies the high frequency resistance of the function $F(p)$ and since this must be positive, if condition (3) is satisfied, the final resistive element will be positive. If the series converges non-uniformly, the sum of the variable terms can have a value other than zero as $p \to \pm i\infty$ in spite of the fact that every term approaches zero individually. In that case (see example 1) all or part of the high frequency resistance may be supplied by the sum of the variable terms.

In case all the terms are of the type where $P_n(0)$ is negative, so that the network branches are made to represent the sum, $P_n(p) - P_n(0)$, of the variable and constant terms and the series is uniformly convergent, all the high frequency resistance is provided by the branches representing these terms. The first term, $F(0)$ then supplies the dc resistance, which is positive by condition (3). Non-uniform convergence can modify this division of high- and low-frequency resistance, however.

Cases can arise in which the series contains terms of both types. In such a case the dc resistance, or high frequency resistance, or both, of

the given function might be less than the sum of the variable terms for these frequencies, with the result that the final resistance branch would be negative for either the series or parallel type of network development.

To make the procedure as concrete as possible, particular forms of networks are described in the section following with explicit formulas for computing their elements.

NETWORK FORMULAS

Simple forms of network branches are shown in Figs. 1 and 2. Those of Fig. 1, referred to as branches of "the first kind" are suitable for connection in parallel where the given function $F(p)$ is an admittance, $Y(p)$, while networks of "the second kind," shown in Fig. 2, are suitable for connection in series to represent an impedance, $F(p) = Z(p)$. The networks of Figs. 1a and 2a apply where the value $P_n(0)$ of the general term is positive, while Figs. 1b and 2b apply where $P_n(0)$ is negative. Figs. 3 and 4 illustrate, respectively, networks of the types of Figs. 1a and 2a

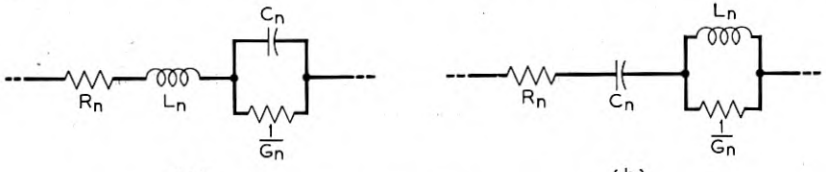


Fig. 1—General branches of the first kind.

Fig. 1a (use where $F(p) = Y(p)$ and $Y_n(0) > 0$)	*Fig. 1b* (use where $F(p) = Y(p)$ and $Y_n(0) < 0$)
$L_n = \dfrac{1}{2a_n}$	$L_n = \dfrac{\beta_n^2(\alpha_n^2 + \beta_n^2)^2(a_n^2 + b_n^2)}{2M^3}$
$\dfrac{1}{L_nC_n} = \beta_n^2\left(\dfrac{b_n^2}{a_n^2} + 1\right)$	$\dfrac{1}{L_nC_n} = \dfrac{M^2}{\beta_n^2(a_n^2 + b_n^2)}$
$\dfrac{G_n}{C_n} = \dfrac{1}{a_n}(a_n\alpha_n - b_n\beta_n)$	$G_nL_n = -\dfrac{a_n\alpha_n - b_n\beta_n}{M}$
$\dfrac{R_n}{L_n} = \dfrac{1}{a_n}(a_n\alpha_n + b_n\beta_n)$	$R_nC_n = \dfrac{N}{M(\alpha_n^2 + \beta_n^2)}$
$G_o = Y(0) - \sum_{n=o}^{\infty} Y_n(0)$	$G_o = Y(0)$

(6)

connected to form the completed network with the final non-reactive branch, G_o or R_o, in place.

Formulas for the network elements are obtained by equating the poles and residues of the network impedance function to the given poles and residues of the general term of the series. Since both poles and residues occur in conjugate complex pairs, and since equality of real and imaginary parts is involved, there are four equations, which are necessary and sufficient to determine the four constants, R, L, G, C, of the network. The formulas that are obtained by solving these equations are given beneath Figs. 1 and 2.

The values given for G_o and R_o in each case assume that all the terms of the series are of the type specified for that case.

Fig. 2—General branches of the second kind.

Fig. 2a (use where $F(p) = Z(p)$ and $Z_n(o) > 0$)	*Fig. 2b* (use where $F(p) = Z(p)$ and $Z_n(o) < 0$)
$C_n = \dfrac{1}{2a_n}$	$C_n = \dfrac{\beta_n^2(\alpha_n^2 + \beta_n^2)^2(a_n^2 + b_n^2)}{2M^3}$
$\dfrac{1}{L_nC_n} = \beta_n^2\left(\dfrac{b_n^2}{a_n^2} + 1\right)$	$\dfrac{1}{L_nC_n} = \dfrac{M^2}{\beta_n^2(a_n^2 + b_n^2)}$
$\dfrac{R_n}{L_n} = \dfrac{1}{a_n}(a_n\alpha_n - b_n\beta_n)$	$R_nC_n = -\dfrac{a_n\alpha_n - b_n\beta_n}{M}$
$\dfrac{G_n}{C_n} = \dfrac{1}{a_n}(a_n\alpha_n + b_n\beta_n)$	$G_nL_n = \dfrac{N}{M(\alpha_n^2 + \beta_n^2)}$
$R_o = Z(0) - \sum_{n=0}^{\infty} Z_n(0)$	$R_o = Z(0)$

(7)

where $M = a_n(\beta_n^2 - \alpha_n^2) + 2\alpha_n\beta_nb_n$

$N = -a_n\alpha_n^3 + 3\alpha_n^2b_n\beta_n + 3a_n\alpha_n\beta_n^2 - b_n\beta_n^3$

In the case of the parallel-type networks (Figs. 1a and 1b), $p_n = -\alpha_n + i\beta_n$ is a pole of the admittance, $Y(p)$, and $A_n = a_n + ib_n$ is the corresponding residue. In the case of the series-type network, the same symbols represent a pole and residue of the impedance, $Z(p)$.

The networks specified by Figs. 2a and 2b are duals of the networks of Figs. 1a and 1b, respectively, and are obtained from the latter merely by replacing L_n by C_n, R_n by G_n, and vice versa.

The formulas are intended to apply to complex poles. They can be applied to real poles by taking b_n and β_n equal to zero and doubling the residue, a_n, but this procedure is unnecessary, because the network rep-

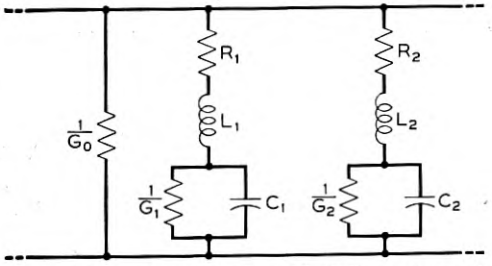


Fig. 3—Network of the first kind (branches 1a).

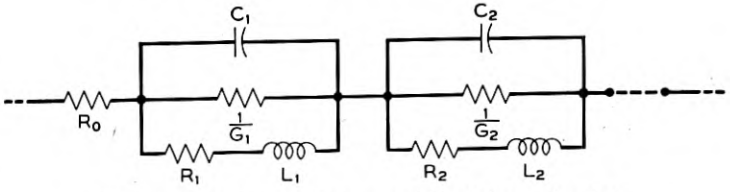


Fig. 4—Network of the second kind (branches 2a).

resentation of a real pole can be found readily enough by inspection of the impedance terms involved. (See Example 1.)

The above discussion is intended to sketch a general picture of the procedure. Individual cases may involve considerable detail that can be understood more readily by reference to the next section.

APPLICATIONS

Example 1a: A transmission line with its far terminals short-circuited affords a simple illustration of the equivalent network theory. Let it be assumed that the parameters, R, L, G and C of the line are constants. In the more advanced examples to follow, the variation of these parameters with frequency for a particular kind of line will be taken into consideration.

The impedance of the short-circuited line (Fig. 5) is

$$Z = Z_0 \tanh \Gamma \tag{1-0}$$

where Z_0 is the characteristic impedance and Γ is the total propagation constant of the line. We have

$$Z_0 = \left(\frac{R + pL}{G + pC}\right)^{1/2} \tag{1-1}$$

$$\Gamma = [(R + pL)(G + pC)]^{1/2} \tag{1-2}$$

R, L, G and C being given for the *total length* of line.

To obtain a development in terms of network branches of the kind shown in Fig. 1, we consider the admittance function,

$$Y = Y_0 \coth \Gamma \tag{1-3}$$

where $Y = 1/Z$ and $Y_0 = 1/Z_0$. Our first task is to find the poles of this function and the residues. Since the complex frequency variable p occurs

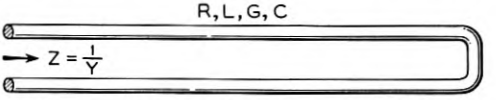


Fig. 5—Short-circuited transmission line.

under square roots in both Z_0 and Γ, it might be suspected, offhand, that the singularities of the function are branch points rather than poles. Such is not the case, however. There are no branch points and all the poles are simple.

The singularities of Y are to be found among the zeros of $\tanh \Gamma$, which occur at

$$\Gamma = i\pi n, \qquad n = 0, \pm 1, \pm 2, \pm 3, \cdots \tag{1-4}$$

To determine them, we solve

$$\Gamma^2 = (R + pL)(G + pC) = -\pi^2 n^2 \tag{1-5}$$

and find these roots:

$$p_n = -\alpha_n + i\beta_n, \qquad p_{-n} = \bar{p}_n = -\alpha_n - i\beta_n$$

where

$$\alpha_n = \frac{G}{2C} + \frac{R}{2L}$$

$$\beta_n = \left[\frac{\pi^2 n^2}{LC} - \left(\frac{G}{2C} - \frac{R}{2L}\right)^2\right]^{1/2} \qquad (n > 0) \tag{1-6}$$

For $n = 0$, the above would give

$$p_0 = -\frac{R}{L}, -\frac{G}{C}$$

But if we let $\Gamma \to 0$, so that $\tanh \Gamma \to \Gamma$, we find that only the point, $-R/L$, is a singularity of Y; the other point, $-G/C$, is a regular point. Therefore Y has only one real singularity.

To find the nature of the singularities of Y, we next calculate

$$\underset{p \to p_n}{\text{Limit}} \left[\frac{p - p_n}{Z_0(p) \tanh \Gamma(p)} \right] = A_n \tag{1-7}$$

and find that at each p_n the limit exists and has the value

$$A_n = \frac{1}{Z_0(p_n)\Gamma'(p_n)} = \frac{1}{L} - \frac{i\left(\alpha_n - \frac{R}{L}\right)}{\beta_n L} = a_n + ib_n \tag{1-8}$$

where $\Gamma'(p_n) = \dfrac{d}{dp}\Gamma(p)$, evaluated at $p = p_n$. The fact that this limit exists shows that all the singularities are simple poles. The values of A_n are then the residues at these poles.

When we now apply formulas (6) to determine the elements in the general branch of the equivalent network of Fig. 1a, we obtain, for $n > 1$,

$$L_n = \frac{L}{2}, \qquad \frac{1}{L_n C_n} = \frac{\pi^2 n^2}{LC}, \qquad \frac{G_n}{C_n} = \frac{G}{C}, \qquad \frac{R_n}{L_n} = \frac{R}{L}. \tag{1-9}$$

The network then comprises an infinite number of such branches in parallel. Each branch has the same elements R_n and L_n, equal, respectively, to half the total resistance and inductance of the transmission line, but the elements G_n and C_n decrease from one branch to the next in inverse proportion to the squares of the integers.

The Q of the n^{th} branch, which can be regarded as the Q of the associated resonance of the short-circuited line, is

$$Q_n = \frac{\omega_n}{2\alpha_n} = \frac{\omega_n}{\frac{G_n}{C_n} + \frac{R_n}{L_n}} = \frac{\omega_n}{\frac{G}{C} + \frac{R}{L}} \tag{1-10}$$

where

$$\omega_n = \sqrt{\frac{1}{L_n C_n} - \frac{G_n^2}{C_n^2}} = \sqrt{\frac{\pi^2 n^2}{LC} - \frac{G^2}{C^2}} \tag{1-11}$$

Thus, for small dissipation, the resonances would become sharper in direct proportion to the frequency (if the parameters R, L, G, C, were invariable with frequency, as assumed).

The above described branches of the equivalent network account only for the complex poles ($n > 1$) of the admittance function. Two more branches remain to be calculated. One is for the real pole ($n = 0$), which occurs at $p_0 = -R/L$, with residue, $A_0 = \frac{1}{L}$. The required branch for this pole is

$$\frac{A_0}{p - p_0} = \frac{1}{R + pL} \tag{1-12}$$

The other is the final conductance branch, which is calculated as follows:

$$G_0 = Y(0) + \sum_{n=-\infty}^{\infty} \frac{A_n}{p_n} = \sqrt{\frac{G}{R}} \coth \sqrt{GR} - \frac{1}{R} - 2G \sum_{n=-\infty}^{\infty} \frac{1}{\pi^2 n^2 + GR} = 0 \tag{1-13}$$

so that, for this example, the conductance branch vanishes. The network is drawn in Fig. 6.

A series type of network, as shown in Fig. 7, can be determined by

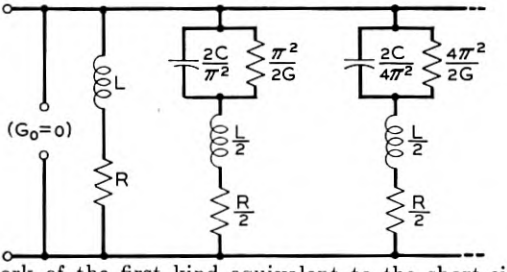


Fig. 6—Network of the first kind equivalent to the short-circuited line of Fig. 5.

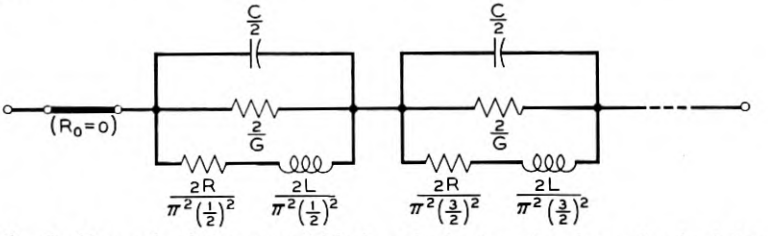


Fig. 7—Network of the second kind equivalent to the short-circuited line of Fig. 5.

similar means. Since, however, it is a dual of the parallel network of Fig. 9 for the open-circuited line, next to be discussed, it can be drawn immediately, without further calculation, once the latter has been found.

Example 1b: We now calculate a network for the same line with its far terminals open (Fig. 8). To obtain a network of the first kind, with branches in parallel, we deal with the admittance function,

$$Y = Y_0 \tanh \Gamma \tag{1-14}$$

The singularities of Y are found among the zeros of $\coth \Gamma$, which occur at

$$\Gamma = i\pi(n + \tfrac{1}{2}), \qquad n = 0, \pm 1, \pm 2, \pm 3, \cdots \tag{1-15}$$

The points $p = -R/L$ and $-G/C$ are *both* regular points this time. ($-G/C$ is a zero of Y.) The singularities are simple poles, as before, with residues,

$$A_n = \frac{1}{Z_0(p_n)\Gamma'(p_n)} \tag{1-16}$$

as before.

The network branches for the complex poles are therefore obtained merely by putting $n + \frac{1}{2}$ in place of the n in all formulas of the short-circuit network. There is no branch corresponding to the branch $R + pL$ of the other network and the conductance branch is again found to be zero. The complete parallel network is drawn in Fig. 9 and the series network, in Fig. 10.

It will be observed that the series network of Fig. 10 is the dual of the

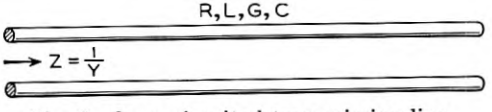


Fig. 8—Open-circuited transmission line.

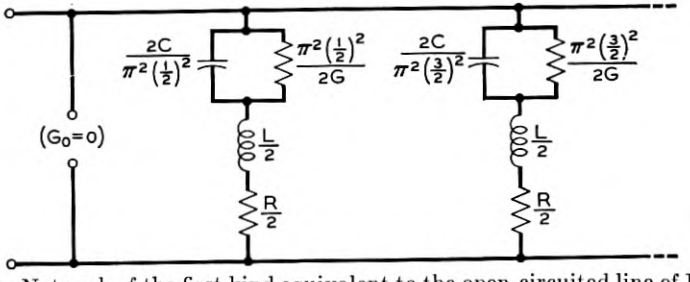


Fig. 9—Network of the first kind equivalent to the open-circuited line of Fig. 8.

parallel network of Fig. 6 and the series network of Fig. 7 is the dual of the parallel network of Fig. 9. These dual relationships are of course a result of the fact that the impedance of an open-circuited line is the dual of the impedance of the same line when short-circuited.

Example 2: Short-circuited Concentric Line (or Toroidal Cavity with E Radial). The preceding example considered a fictitious transmission line of invariable parameters, R, L, G, C, having a perfect short circuit at one end. The present example has to do essentially with the same problem but considers it from a more practical point of view. The variation of R and L with frequency is taken into account and the impedance of the "short-circuit" is no longer neglected.

Let the line be the piece of coaxial cable plugged at both ends with conducting material as illustrated in Fig. 11. Considered from an alternative point of view, our line is now a toroidal cavity oscillating in the

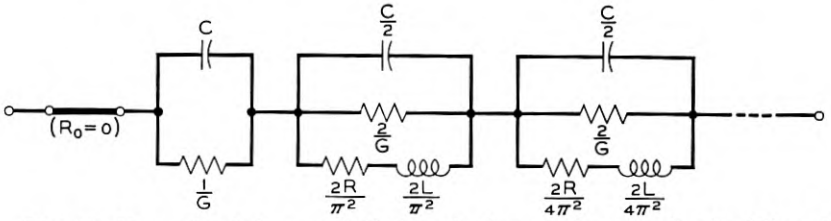


Fig. 10—Network of the second kind equivalent to the open-circuited line of Fig. 8.

mode where the electric force E is directed radially and the magnetic force H lies in planes at right angles to the axis. If we assume the cavity to be excited, or "driven," from one end,* the impedance that is effective in defining the selective characteristic of the cavity with respect to frequency is the total impedance at that end, that is, the sum of the impedance Z_1, viewed into the cavity, and the impedance, Z_2, of the adjacent end-plug. Therefore, we have to deal with the impedance,

$$Z = Z_1 + Z_2. \tag{2-1}$$

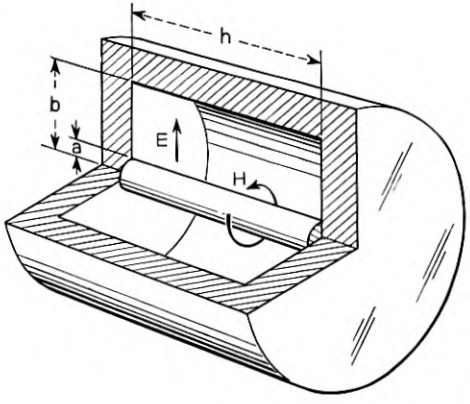
By "impedance" is here meant the same thing that one considers in looking at the problem from the point of view of transmission line theory, namely, the complex ratio, for exponential oscillations, of the voltage between the inside and outside cylindrical surfaces to the total current

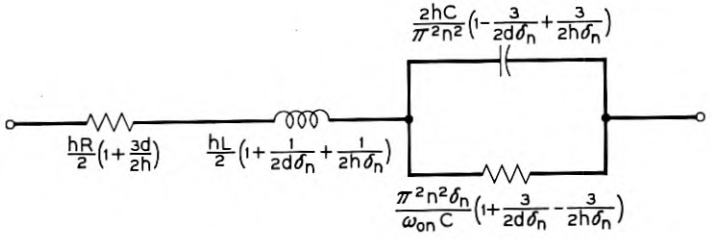
* For determining the "natural frequencies" of oscillation of the cavity, it is immaterial at what point along it the impedance is taken; the total impedance at every point has the same roots. The impedance is, nevertheless, not the same at all points so that the behavior of the cavity, when driven, will depend to some extent on the driving point.

flowing axially in the inner conductor at the same point. The zeros of Z define the natural frequencies of oscillation of the cavity and their associated damping constants, or Q's. Our task is to develop an equivalent network for this Z.

We have

$$Z = Z_1 + Z_2 = Z_0 \left(\frac{1 + \rho e^{-2\gamma h}}{1 - \rho e^{-2\gamma h}} + \frac{1 + \rho}{1 - \rho} \right) \tag{2-2}$$





$$R = \frac{1}{2\pi}\sqrt{\frac{\omega_{on}\mu}{2g}}\left(\frac{1}{a}+\frac{1}{b}\right) \qquad \omega_{on} = \frac{\pi n v}{h}$$

$$L = \frac{\mu_0}{2\pi} \log \frac{b}{a} \qquad v = \frac{1}{\sqrt{\mu_0 \epsilon_0}} = 3\,(10^8)$$

$$C = \frac{2\pi\epsilon_0}{\log \frac{b}{a}} \qquad \delta_n = \sqrt{\frac{\omega_{on}\mu g}{2}}$$

(e.g.) $\mu_0 = 4\pi\,(10^{-7})$, $\epsilon_0 = \frac{10^{-9}}{36\pi}$, FOR AIR IN M.K.S. UNITS

$\mu = \mu_0$, $g = 5.8\,(10^7)$, FOR COPPER IN M.K.S. UNITS

($n = 1$ FOR FUNDAMENTAL MODE)

Fig. 11—Toroidal cavity, E radial.

where

$$Z_0 = \left(\frac{R + pL}{G + pC}\right)^{1/2}, \qquad \gamma = (R + pL)^{1/2}(G + pC)^{1/2} \tag{2-3}$$

$$\rho = \frac{Z_2 - Z_0}{Z_2 + Z_0} \tag{2-4}$$

$$Z_2 = \frac{\eta}{2\pi} \log \frac{b}{a} \tag{2-5}$$

$$R + pL = \frac{\eta}{2\pi a} \frac{I_0(\sigma a)}{I_1(\sigma a)} + \frac{\eta}{2\pi b} \frac{K_0(\sigma b)}{K_1(\sigma b)} + \frac{p\mu_0}{2\pi} \log \frac{b}{a} \tag{2-6}$$

$$G + pC = \frac{2\pi}{\log \frac{b}{a}} (g_0 + p\epsilon_0) \tag{2-7}$$

h, a, b = cavity length, inner radius, outer radius, as shown in Fig. 11, all measured in meters

$$\eta = \left(\frac{p\mu}{g}\right)^{1/2}$$
$$\sigma = (p\mu g)^{1/2} \tag{2-8}$$

μ, g are permeability, conductivity of the conducting material of the walls (for copper: $\mu = 4\pi(10^{-7})$, $g = 5.8(10^7)$ in M.K.S. units).

μ_0, g_0, ϵ_0 are permeability, conductivity, dielectric constant of the dielectric material occupying the cavity (for air: $\mu_0 = 4\pi(10^{-7})$, $g_0 = 0$, $\epsilon_0 = (10^{-9})/36\pi$ in M.K.S. units), p = generalized frequency variable.

$I_0(z)$, $I_1(z)$ are Bessel functions of the first kind for imaginary argument and of order 0, 1.

$K_0(z)$, $K_1(z)$ are Bessel functions of the second kind for imaginary argument and of order 0, 1.

Except for ignored small deviations of the field around the corners of the cavity, the above formulas are exact. To arrive at results that are sufficiently compact to be useful, we make these approximations, at the start:

$$Z_0 = K_0 = \left[\frac{L}{C}\right]^{1/2} = \frac{\eta_0}{2\pi} \log \frac{b}{a}, \tag{2-9}$$

where

$$\eta_0 = \left[\frac{\mu_0}{\epsilon_0}\right]^{1/2} = 120\pi \text{ ohms} \tag{2-10}$$

From this,

$$\rho = \frac{\eta - \eta_0}{\eta + \eta_0} \tag{2-11}$$

Having in mind microwave applications, where the moduli of the arguments of the Bessel functions are >3000, we take

$$\frac{I_0(z)}{I_1(z)} = \frac{K_0(z)}{K_1(z)} = 1$$

so that

$$R + pL = \frac{\eta}{2\pi}\left(\frac{1}{a} + \frac{1}{b}\right) + \frac{p\mu_0}{2\pi}\log\frac{b}{a} \tag{2-12}$$

Also, we have in mind only air dielectric and assume any loss therein to be negligible; that is, we assume $G = 0$.

All further approximations that are made are either

$$\frac{1}{1-\Delta} \doteq 1 + \Delta \qquad \text{or} \qquad (1 + 2\Delta)^{1/2} \doteq 1 + \Delta$$

where, for an air-space enclosed by copper walls, and for frequencies on the order of 30,000 megacycles, Δ is on the order of 10^{-4}. For cavities made of other materials, the results obtained may not be sufficiently accurate and the problem would have to be reviewed from the start. In particular, the results do not hold for a cavity having walls of magnetic material, because we assume here that the permeability of the metal walls is the same as that of air; i.e., $\mu = \mu_0$.

To obtain an equivalent network of the first kind, we deal with the admittance, which is, from (2-2),

$$Y = \frac{1}{Z} = H_0 \frac{(1-\rho)(1-\rho e^{-2\gamma h})}{2(1-\rho^2 e^{-2\gamma h})} \tag{2-13}$$

where $H_0 = 1/K_0$.

The poles of Y are then the zeros of $1 - \rho^2 e^{-2\gamma h}$, which are obtained by successive approximations. We first make a close estimate of the zeros by assuming that the impedance of the short-circuiting plugs is zero; that is, we assume, $Z_2 = 0$, whence $\rho = -1$. To obtain this estimate, we have to solve

$$\gamma h = \frac{ph}{v}\left(1 + \frac{2}{d\sigma}\right)^{1/2} = \pi i n \qquad (n = \pm 1, \pm 2, \pm 3 \cdots) \tag{2-14}$$

where

$$d = \frac{2ab \log (b/a)}{a + b}$$

and $v = 3(10^8)$ meters per second. The approximate solution is

$$p_{1n} = p_{0n}\left(1 + \frac{1}{d\sigma_{0n}}\right)$$

where

$$p_{0n} = \frac{i\pi nv}{h} \qquad \text{and} \qquad \sigma_{0n} = (p_{0n}\mu g)^{1/2}$$

Next we improve our estimate of the zeros by the well-known method involving the derivative of the function, $1 - \rho^2 e^{-2\gamma h}$, with respect to p, evaluated at p_{1n}. This now takes account of the actual impedance of the end-plugs. The values of the zeros, so obtained, are

$$p_n = -\alpha_n + i\beta_n, \qquad p_{-n} = \bar{p}_n = -\alpha_n - i\beta_n$$

where

$$\begin{aligned} \alpha_n &= \omega_{0n}\left(\frac{1}{2d\delta_n} + \frac{1}{h\delta_n}\right) \\ \beta_n &= \omega_{0n}\left(1 + \frac{1}{2d\delta_n} - \frac{1}{h\delta_n}\right) \end{aligned} \tag{2-15}$$

where δ_n* is the real part of σ_{0n}. That is,

$$\delta_n = (\omega_{0n}\mu g/2)^{1/2}$$

where

$$\omega_{0n} = \frac{\pi nv}{h}.$$

As an incidental matter of interest, the above gives the Q of the cavity at any resonance, namely

$$Q_n = \frac{\beta_n}{2\alpha_n} = d\delta_n \frac{1}{1 + \frac{2d}{h}} \tag{2-16}$$

For example, the dimensions, $a = .5$ cm., $b = 1.0$ cm., $h = .5$ cm. provide a cavity that resonates at about 30,000 megacycles. Then the Q's at the first three resonances would be as follows:

n	$\frac{\omega_{0n}}{2\pi}$	Q
1	$30{,}000 \times 10^6$	4250
2	$60{,}000 \times 10^6$	6010
3	$90{,}000 \times 10^6$	7360

* For any frequency, $\delta = (\omega\mu g/2)^{1/2}$ is sometimes referred to as the "skin depth" because it is the depth of metal at which the current density falls to $1/e$ times its value at the surface of the metal.

The importance of including the effect of the end-plugs in determining Q is shown by the fact that, if they were assumed to have zero impedance, Q at the first resonance would be 12,120 instead of 4250.

To determine the residues at the poles, we write

$$Y = H_0 \frac{(1-\rho)(1-\rho e^{-2\gamma h})}{2(1-\rho^2 e^{-2\gamma h})} = \frac{F(p)}{G(p)} \tag{2-17}$$

and then the residue at a simple pole p_n is

$$A_n = \frac{F(p_n)}{G'(p_n)} \tag{2-18}$$

This limit is found to exist, showing that the poles are, in fact, simple. The value found for the residue, A_n, is

$$\begin{aligned} A_n &= a_n + ib_n, \qquad A_{-n} = \bar{A}_n = a_n - ib_n \\ a_n &= \frac{H_0\omega_{0n}}{\pi n}\left(1 - \frac{1}{2d\delta_n} - \frac{1}{2h\delta_n}\right) \\ b_n &= \frac{H_0\omega_{0n}}{\pi n}\left(\frac{1}{2d\delta_n} + \frac{1}{2h\delta_n}\right) \end{aligned} \tag{2-19}$$

When formulas (6) are applied to determine the elements of the tuned branches of the equivalent network of the first kind, the results are, for the n^{th} branch,

$$\begin{aligned} L_n &= \frac{K_0\pi n}{2\omega_{0n}}\left(1 + \frac{1}{2d\delta_n} + \frac{1}{2h\delta_n}\right) \\ \frac{1}{L_nC_n} &= \omega_{0n}^2\left(1 + \frac{1}{d\delta_n} - \frac{2}{h\delta_n}\right) \\ \frac{G_n}{C_n} &= \frac{\omega_{0n}}{2h\delta_n} \\ \frac{R_n}{L_n} &= \omega_{0n}\left(\frac{1}{d\delta_n} + \frac{3}{2h\delta_n}\right) \end{aligned} \tag{2-20}$$

In terms of the R, L and C of the piece of coaxial line, the elements of the n^{th} branch are as follows:

$$\begin{aligned} L_n &= \frac{hL}{2}\left(1 + \frac{1}{2d\delta_n} + \frac{1}{2h\delta_n}\right) \\ R_n &= \frac{hR}{2}\left(1 + \frac{3d}{2h}\right) \\ C_n &= \frac{2hC}{\pi^2n^2}\left(1 - \frac{3}{2d\delta_n} + \frac{3}{2h\delta_n}\right) \\ G_n &= \frac{\omega_{0n}C}{\pi^2n^2\delta_n}\left(1 - \frac{3}{2d\delta_n} + \frac{3}{2h\delta_n}\right) \end{aligned} \tag{2-21}$$

where

$$L = \frac{\mu_0}{2\pi} \log \frac{b}{a}, \qquad R = \frac{1}{2\pi} \left(\frac{\omega_{0n}\mu}{2g} \right)^{1/2} \left(\frac{1}{a} + \frac{1}{b} \right)$$

$$C = \frac{2\pi\epsilon_0}{\log \frac{b}{a}}, \qquad \omega_{0n} = \frac{\pi n v}{h}, \qquad \delta_n = \left(\frac{\omega_{0n} n \mu g}{2} \right)^{1/2}$$

The network is shown in Fig. 11.

It will be found that a "leakage" element, G_n, appears in the equivalent network, although the air dielectric in the cavity was assumed to have no leakage ($G = 0$). This element arises from the end-plugs and is necessary to account for the dissipation in them.

To obtain a network exactly equivalent to the cavity at all frequencies, we should add a branch corresponding to $n = 0$, as was done in example 1. This branch would make the equivalence hold down to and including zero frequency. But, inasmuch as the approximations that have been made hold only for the high frequencies, where the resonances occur, it would be inconsistent to add this branch. What has been arrived at, then, is a partial network representation that gives a close approximation to the impedance of the cavity at high frequencies, only.

Example 3: Toroidal Cavity with E Axial. For further illustration, we consider another mode of oscillation of the short-circuited concentric transmission line investigated in the previous example. This time it is assumed that the radial electric force vanishes while the axial electric force between the end-plugs exists. The magnetic force is directed in circles concentric with the cylindrical central conductor, as before. This situation is illustrated in Fig. 12, which is the same as Fig. 11, except for the new disposition of the E-vector.

For the new mode of oscillation, where the wave is a cylindrical one propagated back and forth between the inner and outer conducting cylinders, the oscillatory space is naturally thought of as a "toroidal cavity," while, in the previous example, where the wave was propagated axially back and forth between the terminal discs, the space was called a "concentric line." Actually, the cavity itself has the same geometric form in the two cases. A practical distinction may exist, however, in that the axial mode of oscillation could be more easily excited in a cavity whose axial length is large compared to its radius, while the cylindrical mode would arise more easily in a flat "pillbox" cavity whose radius is large compared to its axial dimension.

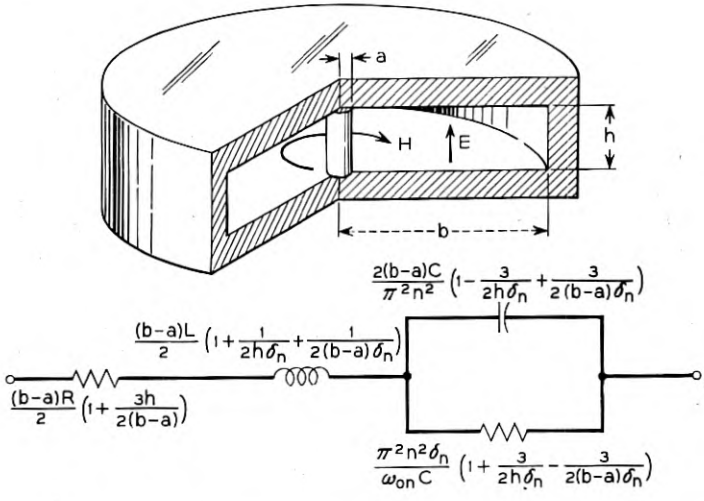
The approach to the problem will be that of transmission line theory, as before. This time, the "line" comprises two circular discs between

which the cylindrical wave is propagated. The series impedance and shunt admittance of such a line are functions of the radius and so will be designated $Z(r)$ and $Y(r)$, respectively. Their values are given below:

$$Z(r) = \frac{2\eta + i\omega\mu h}{2\pi r} \tag{3-1}$$

$$Y(r) = \frac{i\omega 2\pi r \epsilon_0}{h} \tag{3-2}$$

These formulas take into account the losses in the flat walls but assume the conductance of the air between them to be negligible. Losses in the inner and outer "short-circuiting" cylinders will be taken into account by the boundary conditions.



$$R=\frac{1}{\pi a}\sqrt{\frac{\omega_{on}\mu}{2g}} \qquad \omega_{on}=\frac{\pi n v}{b-a}$$

$$L=\frac{\mu_0 h}{2\pi a} \qquad v=\frac{1}{\sqrt{\mu_0\epsilon_0}}=3\,(10^8)$$

$$C=\frac{2\pi a\epsilon_0}{h} \qquad \delta_n=\sqrt{\frac{\omega_{on}\mu g}{2}}$$

SEE FIGURE 11 FOR $\mu_0, \epsilon_0, \mu, g$

(n = 1 FOR FUNDAMENTAL MODE)

Fig. 12—Toroidal cavity, E axial.

If V is the voltage between the flat faces of the cavity at a radius r and I the total current in the lower face at this radius, we have

$$\begin{aligned} \frac{dV}{dr} &= -IZ(r) \\ \frac{dI}{dr} &= -VY(r) \end{aligned} \tag{3-3}$$

By differentiating,

$$\frac{d^2V}{dr^2} = -I\frac{dZ}{dr} - Z\frac{dI}{dr} = \left(\frac{1}{Z}\frac{dZ}{dr}\right)\frac{dV}{dr} + VZY \tag{3-4}$$

But

$$ZY = (2\eta + i\omega\mu h)\frac{i\omega\epsilon_0}{h} = \gamma^2$$

which is a squared propagation constant, *independent of* r, and

$$\frac{1}{Z}\frac{dZ}{dr} = -\frac{1}{r}$$

Therefore,

$$\frac{d^2V}{dr^2} + \frac{1}{r}\frac{dV}{dr} - \gamma^2 V = 0 \tag{3-5}$$

is the differential equation for the voltage. The usual solution of this equation is a linear combination of $I_0(\gamma r)$ and $K_0(\gamma r)$ but since, in this case, the arguments will be almost purely imaginary, it is more convenient to employ the pair of functions, $J_0(-i\gamma r)$ and $N_0(-i\gamma r)$.

The solution for the voltage between the upper and lower surfaces at radius r is

$$V(r) = AJ_0(-i\gamma r) + BN_0(-i\gamma r) \tag{3-6}$$

and, from this, the total radial current in the lower surface, at that radius, is

$$I(r) = -\frac{1}{Z}\frac{dV(r)}{dr} = -iY_0(r)[AJ_1(-i\gamma r) + BN_1(-i\gamma r)] \tag{3-7}$$

where

$$Y_0(r) = 1/Z_0(r) = [Y(r)/Z(r)]^{1/2}$$

The impedance at the inner radius a, looking outward, is then

$$Z_1(a) = \frac{V(a)}{I(a)} = iZ_0(a)\frac{AJ_0(-i\gamma a) + BN_0(-i\gamma a)}{AJ_1(-i\gamma a) + BN_1(-i\gamma a)} \tag{3-8}$$

The total impedance at a (inward + outward) for which we require an equivalent network is

$$Z = Z_1(a) + Z_a$$

where Z_a is the impedance of the central plug to axial current, viz.,

$$Z_a = \frac{\eta h}{2\pi a} \frac{I_0(\sigma a)}{I_1(\sigma a)} \tag{3-9}$$

To evaluate the constants A and B, the following boundary conditions are imposed at radii a and b:

$$\text{at } a\text{: } V = V(a), \text{ a given voltage}$$

$$\text{at } b\text{: } V = I(b)Z_b$$

where Z_b is the impedance of the other "short circuit," comprising the outer cylindrical wall. It is given by

$$Z_b = \frac{\eta h}{2\pi b} \frac{K_0(\sigma b)}{K_1(\sigma b)} \tag{3-10}$$

Except for ignored small deviations of the field around the corners of the cavity, the above expressions are exact. The process of finding the singularities of Z by successive approximations results in expressions that are too long to write down here. To obtain results sufficiently compact for engineering use, we resort to the following asymptotic approximations for the Bessel functions:

$$\begin{aligned} J_0(z) &\sim \left(\frac{2}{\pi z}\right)^{1/2} \cos\,(z - \pi/4) \\ J_1(z) &\sim \left(\frac{2}{\pi z}\right)^{1/2} \cos\,(z - 3\pi/4) \\ N_0(z) &\sim \left(\frac{2}{\pi z}\right)^{1/2} \sin\,(z - \pi/4) \\ N_1(z) &\sim \left(\frac{2}{\pi z}\right)^{1/2} \sin\,(z - 3\pi/4) \\ \frac{I_0(z)}{I_1(z)} &\sim 1, \qquad \frac{K_0(z)}{K_1(z)} \sim 1 \end{aligned} \tag{3-11}$$

Also, with an error on the order of 10^{-4},

$$Z_0(r) \sim \frac{h\eta_0}{2\pi r} = K_0(r) = 1/H_0(r)$$

These substitutions result in the following asymptotic formula for the total impedance Z at radius a

$$Z = K_0(a) \frac{\frac{2\eta}{\eta_0} \cos kx + i\left(1 + \frac{\eta^2}{\eta_0^2}\right) \sin kx}{\cos kx + \frac{i\eta}{\eta_0} \sin kx} \tag{3-12}$$

where $k = \frac{b}{a} - 1$ and $x = -i\gamma a$.

To find an equivalent network of the first kind to represent Z, we deal with the admittance, $Y = 1/Z$. It is instructive and saves much work to put Y in the form of exponential functions, with the substitution

$$\rho = \frac{\eta - \eta_0}{\eta + \eta_0}$$

which is the reflection coefficient at both inside and outside cylindrical surfaces of the cavity. By this means we obtain

$$Y = H_0(a) \frac{(1 - \rho)(1 - \rho e^{-2ikx})}{2(1 - \rho^2 e^{-2ikx})} \tag{3-13}$$

This is now identical in form to the formula (2-13) of example 2, where the E-vector was radially, instead of axially, directed. In fact, since

$$ikx = \gamma(b - a)$$

and

$$\gamma = \frac{i\omega}{v}\left(1 + \frac{2}{h\sigma}\right)^{1/2}$$

comparison with the similar formulas of example 2 shows that all the results of that example can be made to apply to the present one merely by changing the dimensional parameters as follows:

Example 2 (E *radial*)		Example 3 (E *axial*)
h	goes into	$b - a$
$d = \frac{2ab \log (b/a)}{a + b}$	goes into	h

The first result of interest is the value of Q, which is

$$Q_n = h\delta_n \frac{1}{1 + \frac{2h}{b - a}} \tag{3-14}$$

where, as before, δ_n is the "skin depth" equal to the real part of σ_n. That is,

$$\delta_n = \sqrt{\frac{\omega_{0n}\mu g}{2}}$$

To gain an idea of numerical magnitudes, consider the same cavity used in example 2. The dimensions are, as before, $h = .5$ cm., $b - a = .5$ cm. For the square cross-section chosen, the first resonance again occurs at 30,000 megacycles, very nearly, and we can make the following direct comparison of the Q's for the two modes of oscillation:

n	$\omega_{0n}/2\pi$	Q_n	
		Ex. 2 (E radial)	Ex. 3 (E axial)
1	$30{,}000 \times 10^6$	4250	4370
2	$60{,}000 \times 10^6$	6010	6180
3	$90{,}000 \times 10^6$	7360	7560

Due to the asymptotic approximations used, the results for example 3 are not as accurate as those for example 2; the two sets of results show only that the Q of the cavity is substantially the same for the two different modes of oscillation.

The poles of Y are given by

$$\begin{aligned} p_n &= -\alpha_n + i\beta_n, \qquad p_{-n} = \bar{p}_n = -\alpha_n - i\beta_n \\ \alpha_n &= \omega_{0n}\left[\frac{1}{2h\delta_n} + \frac{1}{(b-a)\delta_n}\right] \\ \beta_n &= \omega_{0n}\left[1 + \frac{1}{2h\delta_n} - \frac{1}{(b-a)\delta_n}\right] \end{aligned} \tag{3-15}$$

and the residues are

$$\begin{aligned} A_n &= a_n + ib_n, \qquad A_{-n} = \bar{A}_n = a_n - ib_n \\ a_n &= \frac{H_0(a)\omega_{0n}}{\pi n}\left[1 - \frac{1}{2h\delta_n} - \frac{1}{2(b-a)\delta_n}\right] \\ b_n &= \frac{H_0(a)\omega_{0n}}{\pi n}\left[\frac{1}{2h\delta_n} + \frac{1}{2(b-a)\delta_n}\right] \end{aligned} \tag{3-16}$$

Applying formulas (6) gives the following values for the n^{th} branch of

the network of the first kind:

$$\begin{aligned} L_n &= K_0(a)\,\frac{\pi n}{2\omega_{0n}}\left[1+\frac{1}{2h\delta_n}+\frac{1}{2(b-a)\delta_n}\right] \\ \frac{1}{L_nC_n} &= \omega_{0n}^2\left[1+\frac{1}{h\delta_n}-\frac{2}{(b-a)\delta_n}\right] \\ \frac{G_n}{C_n} &= \frac{\omega_{0n}}{2(b-a)\delta_n} \\ \frac{R_n}{L_n} &= \omega_{0n}\left[\frac{1}{h\delta_n}+\frac{3}{2(b-a)\delta_n}\right] \end{aligned} \tag{3-17}$$

in all of which $\omega_{0n} = \pi nv/(b-a)$ and $v = 1/(\mu_0\epsilon_0)^{1/2} = 3(10^8)$ meters per second.

The results can be put in the same form as those obtained for the other cavity mode, dealt with in example 2, by employing the "primary constants" of the cylindrical transmission line, viz.:

$$R(a) = \frac{1}{\pi a}\left[\frac{\omega_{0n}\mu}{2g}\right]^{1/2} \qquad L(a) = \frac{\mu h}{2\pi a}$$

$$G(a) = 0 \qquad C(a) = \frac{2\pi a\epsilon_0}{h}$$

In terms of these constants, the elements of the n^{th} branch of the equivalent network of the first kind are

$$\begin{aligned} L_n &= \frac{(b-a)L(a)}{2}\left(1+\frac{1}{2h\delta_n}+\frac{1}{2(b-a)\delta_n}\right) \\ R_n &= \frac{(b-a)R(a)}{2}\left(1+\frac{3h}{2(b-a)}\right) \\ C_n &= \frac{2(b-a)C(a)}{\pi^2n^2}\left(1-\frac{3}{2h\delta_n}+\frac{3}{2(b-a)\delta_n}\right) \\ G_n &= \frac{\omega_{0n}C(a)}{\pi^2n^2\delta_n}\left(1-\frac{3}{2h\delta_n}+\frac{3}{2(b-a)\delta_n}\right) \end{aligned} \tag{13-8}$$

The network is shown in Fig. 12.

As in the preceding example, a leakage element arises, in spite of the fact that we assumed initially that g_0 of the air in the cavity is zero. This element accounts for the losses in the inner and outer cylindrical walls.

A number of people with whom the above material has been discussed have given helpful comments and criticisms. I wish to acknowledge my debt in this respect to H. Nyquist, S. A. Schelkunoff, R. M. Foster, S. O. Rice, J. Riordan and W. H. Wise.

BIBLIOGRAPHY

1. R. M. Foster, "A Reactance Theorem," *Bell System Tech. J.*, 1924, **3,** p. 259; also, "Theorems on the Driving Point Impedance of Two-Mesh Circuits," *Bell System Tech. J.*, 1924, **3,** p. 651.
2. W. Cauer, "Die Verwirklichung von Wechselstromwiderst ständen vorgeschriebener Frequenzabhängigkeit," *Archiv für Electrot.*, 1926–27, **17,** p. 355.
3. G. A. Campbell and R. M. Foster, "Fourier Integrals for Practical Applications," *Bell System Tech. J.*, Oct. 1928, pp. 639–707; *Bell System Monograph* B-584.
4. T. C. Fry, "The Use of Continued Fractions in the Design of Electric Networks," *Bull. of Am. Math. Soc.*, 1929, **35,** p. 463.
5. O. Brune, "Synthesis of a Finite Two-terminal Network Whose Driving-point Impedance Is a Prescribed Function of Frequency," *Journal of Math. and Physics* (*M.I.T.*), Oct. 1931, **I,** p. 191.
6. Sidney Darlington, *Journal of Math. and Physics* (*M.I.T.*), Sept. 1939, **4,** pp. 257–353.
7. E. C. Titchmarsh, "The Theory of Functions," Oxford Univ. Press, 2nd ed., 1939, p. 110.
8. Gustav Doetsch, *Theorie und Anwendung der Laplacetransformation*, 1st Am. Ed. 1943, p. 139.
9. S. A. Schelkunoff, "Representation of Impedance Functions in Terms of Resonant Frequencies," *Proc. of the I.R.E.*, 1944, **32,** No. 2, p. 83.
10. H. W. Bode, *Network Analysis and Feedback Amplifier Design*, D. Van Nostrand Co., 1945.
11. P. I. Richards, "A Special Class of Functions with Positive Real Part in a Half-plane," *Duke Math. J.*, 1947, **14,** pp. 777–786. Also "General Impedance Function Theory," *Quart. Appl Math.*, 1948, **6,** pp. 21–29.
12. E. A. Guillemin, *The Mathematics of Circuit Analysis*, John Wiley and Sons, Inc., New York, 1950, pp. 409–422.

Abstracts of Bell System Technical Papers* Not Published in This Journal

Universal Equalizer Chart. D. A. ALSBERG[1]. *Electronics*, **24**, pp. 132, 134, Nov., 1951.

Modification of familiar Smith chart consolidates on one time-saving plot all positive-value solutions to the two general equations for series, shunt, and bridged-T audio equalizers.

Limits on the Energy of the Antiferromagnetic Ground State. P. W. ANDERSON[1]. *Phys. Rev.*, **83**, p. 1260, Sept. 15, 1951.

Post-War Achievements of Bell Laboratories, I. O. E. BUCKLEY[1]. *Bell Tel. Mag.*, **30**, pp. 163–173, Autumn, 1951.

Filamentary Growths on Metal Surfaces—"Whiskers". K. G. COMPTON[1], A. MENDIZZA[1], and S. M. ARNOLD[1]. *Corrosion*, **7**, pp. 327–334, Oct., 1951. (Monograph 1885).

Filamentary growths have been found on metal surfaces of some of the parts used in telephone communications equipment, particularly on parts shielded from free circulation of air. The growths are of the same character as those known as "whiskers," which developed between the leaves of cadmium plated variable air condensers and caused considerable trouble in military equipment during the early part of World War II. An investigation has been under way in an attempt to determine the mechanism of growth of the whiskers, found not only on cadmium plated parts but also on other metals. This paper summarizes the findings to date as revealed by the study of approximately one thousand test specimens of different metals, solid and plated, exposed under various environmental conditions. The study is being extended in the light of the findings which have developed during the course of the work.

An Unattended Broad-Band Microwave Repeater for the TD-2 Radio Relay System. R. W. FRIIS[1] and K. D. SMITH[1]. *Elec. Eng.*, **70**, pp. 976–981, Nov., 1951. (Similar article in The Bell System Technical Journal,

---

* Certain of these papers are available as Bell System Monographs and may be obtained on request to the Publacation Department, Bell Telephone Laboratories, Inc., 463 West Street, New York 14, N. Y. For papers available in this form, the monograph number is given in parentheses following the date of publication, and this number should be given in all requests.

[1] B. T. L.

October, 1951, Part II, entitled *The DT-2 Microwave Radio System* reprinted as Monograph 1921).

To meet the stringent requirements of the 4,000-mile transcontinental microwave relay system, a number of new developments had to be included in the design of the repeater stations. The circuits of these uuattended stations, and how they are maintained, are the subject of this article.

The Bell System's Part in Defending the Nation. F. R. KAPPEL[2]. *Bell Tel. Mag.*, **30**, pp. 141–152, Autumn, 1951.

Quickly and accurately checks performance of private or common-carrier p-m or f-m mobile telephone transmitters, such as those used in 30 to 44 -mc highway and 152 to 175-mc urban service. Measures r-f power output, audio sensitivity, signal-to-noise ratio and harmonic distortion and gives speech intelligibility check in few minutes.

Mobile Transmitter Testing Set. G. J. KENT[3]. *Electronics*, **24**, pp. 106–109, Nov., 1951.

A New Electrolysis Switch for Underground Lead Sheath Cable Drainage Systems. V. B. PIKE[1]. *Corrosion*, **7**, p. 1, Oct., 1951.

A High Temperature Stage for the Polarizing Microscope. E. A. WOOD[1]. *Am. Mineral.*, **36**, pp. 768–772, Sept.–Oct., 1951.

A Precise Sweep-Frequency Method of Vector Impedance Measurement. D. A. ALSBERG[1]. *Proc. I.R.E.*, **39**, pp. 1393–1400, Nov., 1951. (Monograph 1911).

The impedance of a two-terminal network is defined completely by the insertion loss and phase shift it produces when inserted between known sending and receiving impedances. Recent advances in precise wide-band phase and transmission measuring circuits have permitted practical use of this principle. Reactive and resistive impedance components are read directly from a simple graphical chart in which frequency is not a parameter. The basic principle described promises attractive possibilities in many cases of impedance measurements where present methods are inadequate.

Electron-Vibration Interactions and Superconductivity. J. BARDEEN[1]. *Revs. Modern Phys.*, **23**, pp. 261–270, July, 1951. (Monograph 1912).

The Copper Oxide Rectifier. W. H. BRATTAIN[1]. *Revs. Modern Phys.*, **23**, pp. 203–212, July, 1951.

It is shown that the conductivity in the ohmic part of the cuprous oxide layer can be explained with the usual band picture of semiconductors only by assum-

[2] A. T. & T. Co.
[3] W. E. Co.

ing the presence of some donor-type impurities in addition to the usual acceptor type. The energy difference between the acceptors and the filled band is 0.3 electron volt, and the total number of impurity atoms is about 10^{14} to 10^{16} per cm^3, the number of donors being less than but of the same order as the number of acceptors. One finds that the density of ion charge in the rectifying layer is of the same order of magnitude as the difference between the donors and acceptors found from the conductivity. The field at the copper-cuprous oxide interface is about 2×10^4 volts/cm; the height of the potential at the surface as compared with the oxide interior is about 0.5 volt; and the thickness of the space charge layer about 5.0×10^{-5} cm. The diffusion equation for flow of current through this space charge region can be integrated to give the current in terms of the field at the interface and the applied potential across the space charge layer. Two currents are involved, one from the semi-conductor to the metal (I_s) and one from the metal to the semiconductor (I_m) which is similar to a thermionic emission current into the semiconductor. The net current is, of course, $I = I_m - I_s$. One can get this "emission" current (I_m) by dividing the true current by the factor $1 - \exp(-eV_a/kT)$, where V_a is the applied potential. This emission current depends on the absolute temperature and on the field at the copper-cuprous oxide interface. At high fields the logarithm of the current is proportional to the square root of the field, and at low fields the current decreases more rapidly indicating a patchy surface having small areas of low potential maximum from which all the emission comes when the field is large.

Effect of Packaging on Corrosion of Zinc Plated Equipment. K. G. COMPTON[1], S. M. ARNOLD[1], and A. MENDIZZA[1]. *Corrosion*, **7**, pp. 365–372, Nov., 1951.

Physics as a Science and an Art. K. K. DARROW[1]. *Phys. Today*, **4**, pp. 6–11, Nov., 1951. (Monograph 1914).

The last of six invited papers presented on October 25th during the symposium on "physics today" which keynoted the 20th Anniversary Meeting of the American Institute of Physics in Chicago.

Ionization by Electron Impact in CO, N_2, NO, and O_2. H. D. HAGSTRUM[1]. *Revs. Modern Phys.*, **23**, pp. 185–203, July, 1951. (Monograph 1916).

Ionization by electron impact in diatomic gases has been studied in this work with a mass spectrometer designed to measure m/e, appearance potential, and initial kinetic energy for each ion observed. Results have been obtained for the gases CO, N_2 NO, and O_2 with some confirmatory work in H_2. Discussion is included of the nature and identification of dissociative ionization processes and of the retarding potential and appearance potential measurements. Values of important quantities such as the dissociation energies of CO, N_2 , and NO; the sublimation energy of C; the electron affinity of O; and the excitation energy of O^- are determined again by electron impact in this work.

[1] B. T. L.

Equivalent Temperature of an Electron Beam. M. E. HINES[1]., Letter to the Editor., *J. Appl. Phys.*, **22**, pp. 1385–1386, Nov., 1951.

Bell System Cable Sheath Problems and Designs. F. W. HORN[1] and R. B. RAMSEY[1]., *Elec. Eng.*, **70**, pp. 1070–1075, Dec., 1951. (Monograph 1917).

Engineering Planning. H. S. OSBORNE[2]. *J. Eng. Educ.*, **42**, pp. 121–125, Nov., 1951.

Acceptance Inspection of Purchased Material. J. E. PALMER[3] and E. G. D. PATERSON[1]. *Ind. Quality Control*, **8**, pp. 23–27, Nov., 1951.

A Note on the Partial Differential Equations Describing Steady Current Flow in Intrinsic Semiconductors. R. C. PRIM[1]. Letter to the Editor. *J. Appl. Phys.*, **22**, pp. 1388–1389, Nov., 1951.

General Theory of Symmetric Biconical Antennas. S. A. SCHELKUNOFF[1]. *J. Appl. Phys.*, **22**, pp. 1330–1332, Nov., 1951. (Monograph 1922).

This paper presents the input admittance of a symmetric biconical antenna of an arbitrary angle as the limit of a certain sequence of functions. The first term of this sequence approaches the exact expressions for the input admittance as the cone angle approaches either zero or 90°. For this reason our conjecture is that this term represents a good first approximation for all angles.

Artificial Dielectrics for Microwaves. W. M. SHARPLESS[1]. *Proc. I. R. E.*, **39**, pp. 1389–1393, Nov., 1951. (Monograph 1923).

This paper presents a procedure for measuring the dielectric properties of metal-loaded artificial dielectrics in the microwave region by the use of the short-circuited line method. Formulas, based on transmission-line theory, are included and serve as guides in predicting the approximate dielectric properties of certain loading configurations.

[1] B. T .L.
[2] W. E. Co.

Contributors to this Issue

W. O. Baker, B.S., Washington College, Maryland, 1935; Ph.D., Princeton University, 1938; Bell Telephone Laboratories, 1939–. Dr. Baker has carried on investigations of the molecular structure and physical properties of polymers, particularly the fundamental constitution of synthetic rubbers and plastics. Harvard Fellowship, 1936–37 and Proctor Fellowship, 1938–39. Member of American Chemical Society, American Physical Society, and American Society for Testing Materials.

J. H. Heiss, B.S. in Ch.E., Newark College of Engineering, 1942; Bell Telephone Laboratories, 1934–. Mr. Heiss has devoted his time to studying experimental wire coating procedures and the test methods involved, the experimental production of high polymers (plastics) and the examination of their physical properties, and the properties of high polymers in solution. Member of American Chemical Society.

W. S. Hayward, Jr., A.B., Harvard University, 1943; S.M., Harvard Graduate School of Engineering, 1947; Aircraft Radio Laboratory, June–December 1943; U. S. Navy, Aviation Electronics Officer, 1944–46; Bell Telephone Laboratories, 1947–. Mr. Hayward has taught telephone switching circuit design at the Laboratories and is currently making probability studies of telephone traffic.

Brockway McMillan, B.S., Massachusetts Institute of Technology, 1936; Ph.D., Massachusetts Institute of Technology, 1939; Instructor of Mathematics, Massachusetts Institute of Technology, 1936–39; Proctor Fellow and Henry B. Fine Instructor in Mathematics, Princeton University, 1939–42; U.S.N.R., 1942–46, studying exterior ballistics of guns and rockets; Los Alamos Laboratory, Spring 1946; Bell Telephone Laboratories, 1946–. Dr. McMillan has been engaged in mathematical research and consultation work. Member of American Mathematical Society, Institute of Mathematical Statistics, and A.A.A.S.

R. E. Staehler, B.E.E., College of the City of New York, 1947; M.E.E., Polytechnic Institute of Brooklyn, 1948; U. S. Army 1943–46, Communications Officer; Instructor in Electrical Engineering, Polytechnic Institute of Brooklyn, Fall, 1950; Bell Telephone Laboratories, 1948–. After completing the Communications Development Training Program,

with rotational assignments in the transmission, switching, and apparatus departments, Mr. Staehler became a member of a switching group concerned with both local and toll signaling. He is working at present on a new voice-frequency toll signaling development. Member of Tau Beta Pi and Eta Kappa Nu.

M. K. ZINN, B.S. in E.E., Purdue University, 1918; U. S. Army, 1918–19. Amer. Tel. and Tel., 1919–1934; Bell Telephone Laboratories, 1934–. As a transmission engineer, Mr. Zinn has been concerned with land and submarine loaded cables, telephone instruments, buried cable and submarine cable with amplifiers and special problems. Member of A. I. E. E. and Tau Beta Pi.