ELEMENTARY THEORY OF TRANSMISSION
AND REFLECTION
Fundamental Relations and Geometry

R. M. REDHEFFER

TECHNICAL REPORT NO. 24

NOVEMBER 20, 1946

RESEARCH LABORATORY OF ELECTRONICS

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

ELEMENTARY THEORY OF TRANSMISSION AND REFLECTION
Fundamental Relations and Geometry

By

R. M. Redheffer

Abstract

The concept of linearity and the relations fundamental to trans-
mission and reflection theory are derived from the properties at the termi-
nals of an object, without reference to its interior. The so-called vector
diagram of conventional theory is interpreted as a conformal mapping, as a
polar plot, and as a linear transformation. Conditions for its exact valid-
ity are obtained, as are the errors when the diagram is known to be only
approximate. When no approximations are made, different interpretations of
the vector diagram lead to different geometrical figures, each of which is
investigated in detail. For ease of reference the results are summarized in
numbered paragraphs distributed throughout the text.
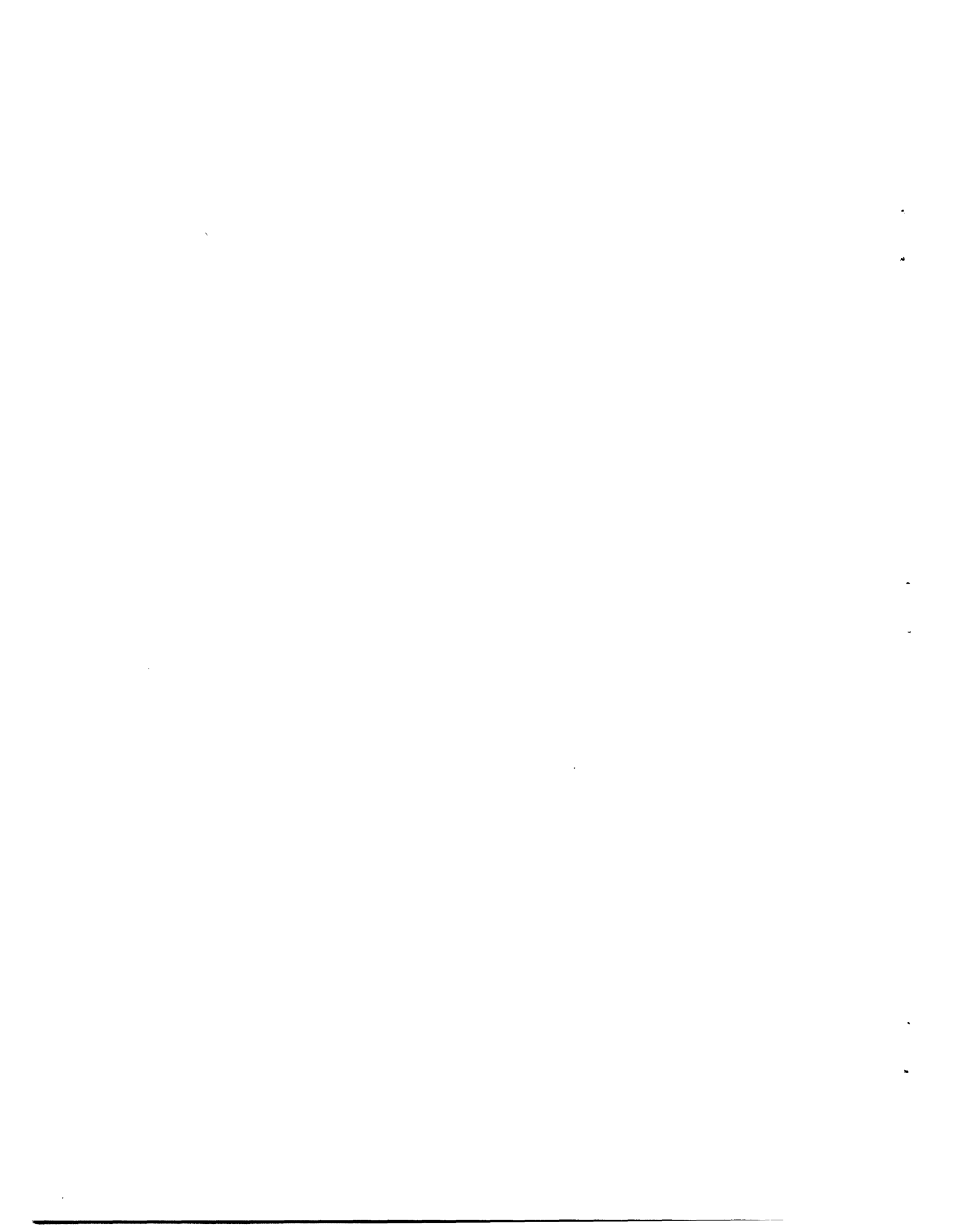
1. Fundamental Relations

1.1. Introduction. In this article, transmission and reflection coefficients are used to investigate the behavior of an object or an array of objects in a transmission line at microwave frequencies. The questions considered are of the type which, at least for the lower frequencies, are almost always treated by means of impedance, rather than by means of transmission and reflection coefficients as here described. Since impedance is very common in engineering practice, while transmission and reflection are less often employed, it is natural to inquire whether the advantages of the new approach are sufficient to offset its unfamiliarity. Such a question is particularly relevant in that impedance can be used for microwave problems as well as for problems at the lower frequencies, and many of its general properties are preserved in the two cases. Moreover, the reflection coefficient r is related to the impedance z by

$$r = \frac{1 - z}{1 + z} \tag{1}$$

with a similar equation for transmission in terms of transfer impedances; and hence it might be supposed that the new method, which differs only in this rather trivial change of variable, could offer no advantages not already inherent in the old. These and other arguments in favor of the impdedance approach are certainly very cogent, and have induced a number of authors to write whole books on microwave theory, in which the notion of impedance is used almost exclusively (Reference 2, 3). The greater part of Reference 3, in particular, is devoted to the establishment of equivalences between the familiar impedance relations of circuit theory and the equations of an electromagnetic field.

There is a large class of problems, nevertheless, in which the alternative method used here offers many advantages. Reflection and transmission properties frequently have a simple physical interpretation, for example, and results which are intuitively evident with this point of view may be rather difficult to derive on the basis of impedance. Because of this difference in physical meaning, the change of variable indicated in (1) is actually less trivial than one would at first suppose. At least for microwave frequencies, moreover, the derivation of results from first principles is somewhat easier with the present approach than with the methods of circuit theory, and the final form is often simpler. An example is given by the equation for moving the reference point down the line, which becomes

$$z(x) = z_0 \frac{z + z_0 \tanh(ikx)}{z_0 + z \tanh(ikx)} \tag{2}$$

for impedance, while we have the simpler (and more easily derived) relations

$$r(x) = r(0)e^{2ikx}$$

$$t(x) = t(0)e^{ikx} \tag{3}$$

for transmission and reflection. In such cases the impedance procedure is attractive, not because it is really better suited to the problem at hand, but because it is more familiar. An illustration of this idea is found in the fact that reflection methods are easier to explain to a person unacquainted with either theory.

The present investigation was prompted partly by these considerations and partly by the fact that impedance methods have been fully described in many works, while only the most cursory consideration appears to have been given to the corresponding theory of reflection. The subject is here presented as a field of study in its own right, and no attempt is made to correlate the results with those of circuit theory.

1.2. **Definitions.** Following Reference 1 we may write the equation of an electromagnetic wave in the form

$$E(x,t) = Ae^{i(kx - \omega t)} \qquad (4)$$

where $E(x,t)$ represents the complex electric field at point x and time t, both quantities being measured from some fixed origin. The wave may be travelling down a transmission line in the general sense, which includes any arrangement with cylindrical symmetry; for example, the line may be a waveguide of arbitrary cross section, or free space. For our present purposes the nature of the line is sufficiently specified by the so-called propagation constant k of Eq. (4), which will be assumed independent of field strength over the range of fields under consideration. With such an assumption, which is always made in work of this kind, we find that k is simply a constant, complex if the transmission line is dissipative, real if it is lossless. In the latter case the material filling the line must be a perfect dielectric (zero conductivity) and the line itself must be made up of perfect conductors (infinite conductivity). We may then write

$$k = 2\pi/\lambda_g \qquad (5)$$

where λ_g is the wavelength, in the transmission line, of the particular mode under discussion.

In all relations with which we shall be concerned here, the time dependence $e^{-i\omega t}$ of Eq. (4) plays no part, and need never be explicitly stated. This is the reason for our choice of Eq. (4), which leads to slightly simpler relations than the corresponding engineering form obtained by writing $i = -j$. With the time dependence omitted, the complex amplitude of the wave at a particular point is defined as the coefficient of $e^{-i\omega t}$. By means of Eq. (4) one may then write

$$A(x) = e^{ikx}A(0) \qquad (6)$$

where $A(0)$ is the complex amplitude at some point of the line, while $A(x)$ is the complex amplitude at a distance x down the line in the direction of propagation. If there are several waves all travelling in the same direction, the complex amplitude of the resultant wave is defined to be the sum of the individual complex amplitudes. This property will be explicitly used only at points of the transmission line, not at points interior to the reflecting objects presently to be introduced. Its justification for this case follows from the well-known superposition principle for plane waves, which in turn is valid whenever k is, as here assumed, independent of field strength. Similarly, the resultant field at a point in the line is obtained by adding the complex amplitudes of all waves at that point, whether these waves are travelling toward the generator or toward the load.

The complex amplitude, as thus defined, is just a complex number, and hence it may be written in the form

$$A = |A|e^{iA'} \tag{7}$$

The quantity $|A|$ in this equation is called the <u>amplitude</u> (as opposed to the complex amplitude) while A' is the <u>phase</u>. For the power in the wave one may write

$$\text{power} = (\text{constant})|A|^2, \tag{8}$$

an expression which is more convenient for our present purposes than explicit use of the Poynting vector. Thus, it is desirable not to introduce the field vector H, but to express all results in terms of E. This requirement is met by Eq. (8), and the proportionality constant, which depends on the characteristics of the transmission line, will cancel out of the relations with which we shall be concerned here. If the line is lossless, so that no power is absorbed by it, one would expect the power at any one point to be the same as any other point. That such is indeed the case is clear from Eqs. (5), (6) (8); and conversely, we see that if this condition is satisfied, then the line must necesesarily be lossless.

The foregoing discussion applies to an infinite line or, which is nearly the same thing, to a line terminated by a matched load. More generally, however, the line may contain an obstruction, as shown in Fig. 1. The wave from the generator now gives rise
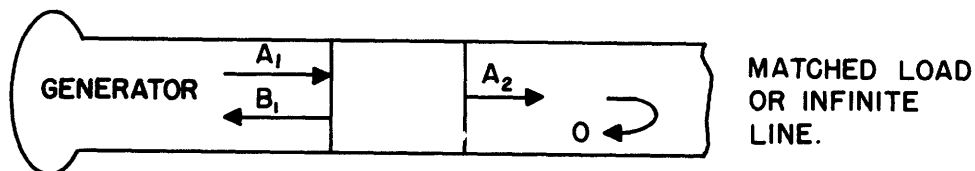

Figure 1. Transmission line containing a single reflecting object. The wave A_1 is the incident wave; A_2 and B_1 are the transmitted and reflected waves, repectively.

to two other waves, one being transmitted through the object, the other being reflected back toward the generator. Over a limited range of values for the field it may happen that the amplitudes of these two waves (called the <u>transmitted</u> and <u>reflected</u> waves) are proportional to that of the original wave (called the <u>incident</u> wave). Such a situation leads to the following definition:

> An object is said to be <u>linear from the left</u> if the complex amplitudes of the transmitted and reflected waves are both proportional to that of the incident wave, whenever this wave is incident from the left. Linearity from the right is similarly defined, and an object is said to be <u>linear</u> if it is linear both from the left and from the right. The transmission coefficient is then defined, in corresponding cases, to be the constant ratio of complex transmitted to incident amplitude, and similarly for the reflection coefficient.

As suggested above, it is clear that a physical object can be linear only for certain values of the field; if the field is indefinitely increased, thermal and other

effects will destroy the linearity of any real object. Whenever we speak of a linear object, therefore, it is implied that the field is suitably restricted -- an assumption which is standard in this kind of work, although not always stated explicitly. It is similar to the assumption previously made concerning the constancy of k. In fact we see from the definition that a section of line of length x will be linear if and only if k is independent of field strength; in this case the two transmission coefficients are both equal to e^{ikx}, and the reflection coefficients are zero.

Over the given range of the field, it is implied in the definition that the ratios are constant, with a given incident wave, even when another wave is traversing the object in the opposite direction. Similarly, the coefficients are defined only for a single mode, the result for several modes being obtained by superposition. For simplicity, however, we shall assume here that energy is propagated in the transmission line at one mode only, although the object itself may excite other modes, and it may have any number of junctions. A so-called magic T, for example, fulfills all requirements for a linear object in this sense.

It is worth noting that the above definition of linearity differs from the one used in the circuit theory, in that we do not here assume linearity throughout the interior of the object. The condition is stated in terms of the terminals only, a point of view which is characteristic of transmission and reflection methods. The present concept of linearity is slightly more general than the usual one; for example, an object may be linear in this sense without satisfying the reciprocity theorem. In other words, the transmission from the left and right need not be equal. It may be noted too that the distinction between complete linearity and uni-directional linearity is not entirely academic. At sufficiently low fields, for example, certain vacuum tube circuits may be linear in one direction only; an illustration for r-f frequencies, which does not contain vacuum tubes, is shown in Fig. 2.
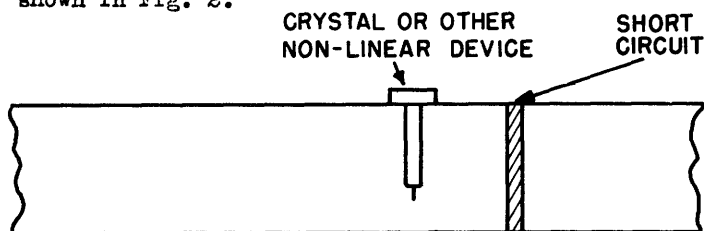

Figure 2. Example of an object linear from the right, but not from the left.

1.3. Transmission and Reflection of Two Objects. Throughout the foregoing discussion it was assumed that the transmission line contained at most one reflecting object. Proceeding now to the case in which two or more objects are present, we find that this new situation introduces no additional complications, but may be solved on the basis of the original assumptions and definitions. Thus, for the two objects of Fig. 3a let us denote the resultant amplitudes at the indicated points by A_i, B_i. That the field can be thus resolved into plane waves follows from the assumed properties of the transmission line and the two objects. The wave A_4 is equal to the resultant wave A_3 moving from left to right at 3, multiplied by the left-hand transmission coefficient of the second object, so

that one may write

$$A_4 = A_2 e^{ikx} t_1$$ (9)

after using Eq. (6) to transfer A_2 to the point 3. Similarly the wave A_2 is equal to the
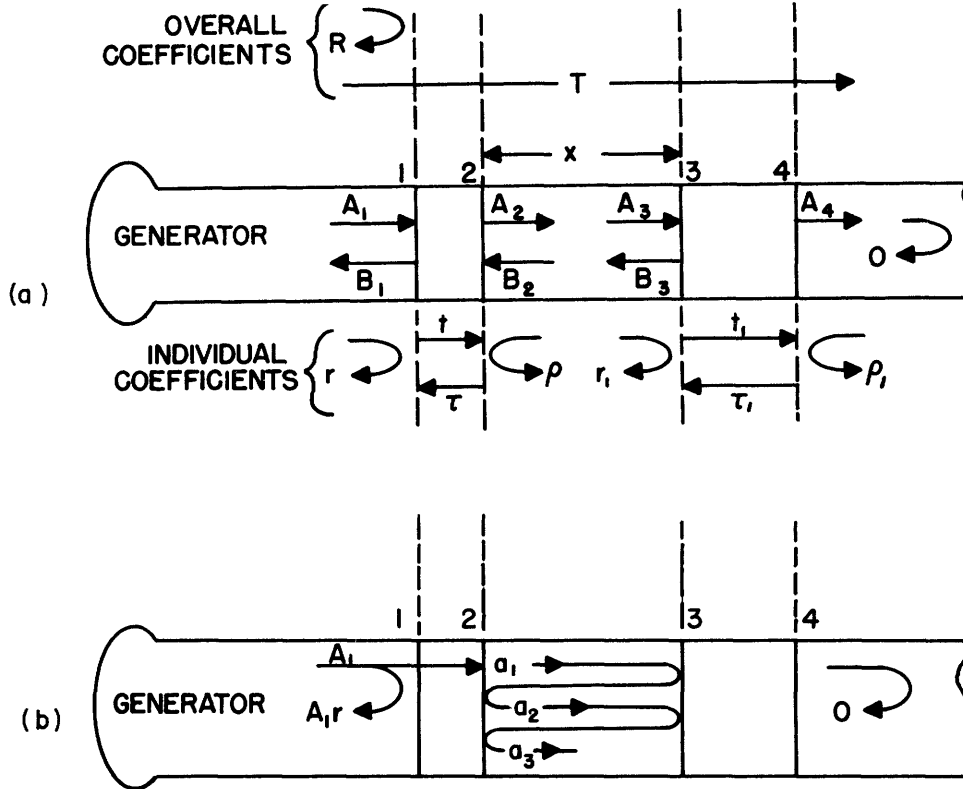

Figure 3. Two methods of computing the over-all transmission and reflection coefficients for two objects in terms of those for the individual objects. a) Solution by simultaneous equations. b) Solution by multiple reflections.

portion of A_1 which is transmitted through the first object, plus the portion of B_2 which is reflected at that object:

$$A_2 = A_1 t + B_2 \rho$$ (10)

This relation requires use of the superposition principle at point 2 of the transmission line. The wave B_2 is obtained by transferring A_2 to the point 3 by Eq. (6), then reflecting this wave at the second object to fina B_3 and finally transferring the reflected wave B_3 back to point 2. We thus obtain

$$B_2 = A_2 e^{ikx} r_1 e^{ikx}.$$ (11)

And finally, by using superposition at the point 1 we find that the wave B_1 is equal to the part of A_1 which is reflected at the first object, plus the part of B_2 which is transmitted:

$$B_1 = A_1 r + B_2 \tau .$$ (12)

After solving these four equations one obtains

$$T = \frac{A_4}{A_1} = \frac{t t_1 e^{ikx}}{1 - \rho r_1 e^{2ikx}} \qquad (13)$$

$$R = \frac{B_1}{A_1} = r + \frac{r_1 t \tau e^{2ikx}}{1 - \rho r_1 e^{2ikx}} \qquad (14)$$

which, by definition, give the left-hand transmission and reflection coefficients of the pair of objects, this pair being itself regarded as an object. Expressions for the right-hand coefficients can be obtained by the same calculation, with A_1 incident from the right. By symmetry it is clear that the results would be equivalent to the above, except that corresponding Greek and Latin letters, for each object, would be interchanged.

In many cases the behavior of a system of objects can be determined by inspection provided one has a clear physical interpretation of the reflection mechanism. Such an interpretation is given by the following alternative derivation of Eq. (13). In Fig. 3b, the resultant wave moving from left to right at point 2 is given by the sum of the waves. Each of these individual waves in turn is obtained from its predecessor by the relation

$$a_{i+1} = a_i e^{ikx} r_1 e^{ikx} \rho \qquad (15)$$

which may be used repeatedly to give

$$a_1 = A_1 t, \qquad a_2 = A_1 t r_1 \rho e^{2ikx}, \qquad a_3 = A_1 t (r_1 \rho e^{2ikx})^2, \ldots$$

For the amplitude A_2 one may therefore write

$$A_2 = a_1 + a_2 + a_3 + a_4 + \ldots$$
$$= A_1 t [1 + (r_1 \rho e^{2ikx}) + (r_1 \rho e^{2ikx})^2 + (r_1 \rho e^{2ikx})^3 + \ldots]$$
$$= \frac{A_1 t}{1 - r_1 \rho e^{2ikx}} \qquad (16)$$

whenever $|r_1 \rho e^{2ikx}| < 1$. A similar process of summation will give the amplitudes A_4 and B_1, but these may be more simply obtained in terms of (16). Thus, the wave at point 4 is found, as above, by letting the wave A_2 proceed to the point 3 and then through the second object. This procedure gives Eq. (13). Similarly, the wave B_1 is obtained by letting A_2 go to the point 3, be reflected, return to the point 2, and go through the first object. We thus obtain Eq. (14), after adding the term r arising from initial reflection of A_1 by the first object.

Although, in the general case, the procedure suggested by Fig. 3b is less concise than that of Fig. 3a, there are many simple situations in which it can be profitably used. As an example let us derive Eqs. (3), the realtions for travelling down the line. These can be obtained directly from (13) and (14) by the substitutions

$$t = \tau = e^{ikx} \qquad r = \rho = 0 \qquad (17)$$

or by the alternative substitutions

$$t = \tau = 1 \qquad r = \rho = 0 \qquad x = x. \qquad (18)$$

Physically, Eq. (17) means that the first object is a section of line x units long, and

that the spacing of the two objects is zero(Fig. 4a with 2 coinciding with 3). The second
equation, (18), means that the first object is a section of line with zero length, sepa-
rated a distance x from the second object (Fig. 4a with 2 coinciding with 1). The two
situations are clearly equivalent; and by actual substitution one finds, in each case,that
Eqs. (13), (14) reduce to (3). To obtain a more immediate derivation one may construct
the analogue of Fig 3b for this particular case. From such a diagram the desired result
can be obtained by inspection (see Fig. 4a). Reasoning of this type entails no loss of
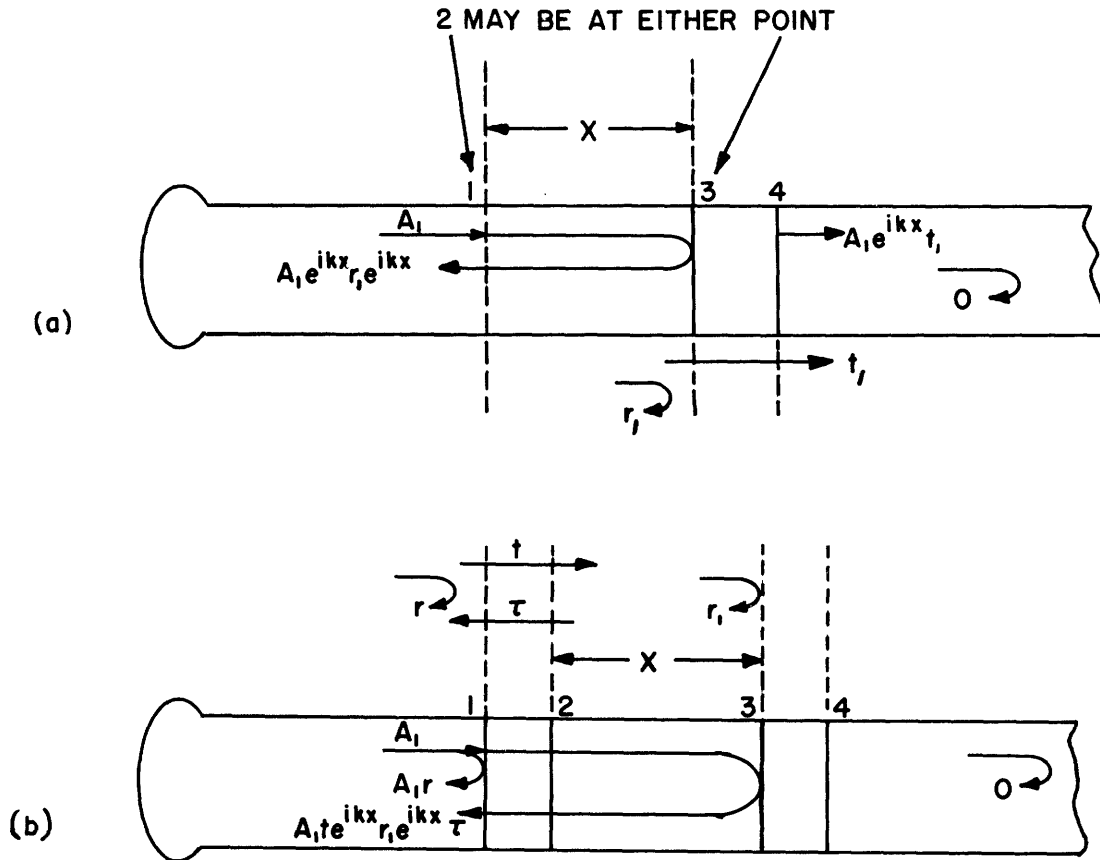rigor and, as in the present case, it can often be used with advantage.

Figure 4. The use of figures for deriving simple results by inspection. a) Derivation
of the equations for travelling down a line, Eqs. (3). b) Derivation of the
equation leading to the approximate vector diagram, Eq. (29).

1.4. Linearity of Several Objects. If each of two objects is linear from the left only,
then the two objects together will usually not be linear in either direction. More
generally, if one object of a series is linear from the left only, all the others being
linear in both directions, then the whole series will ordinarily be linear neither from
the left nor from the right. If the exceptional object happens to be at the right-hand
end, however, then the series will always be linear from the left. These and similar

observations may be readily proved by use of diagrams as above described.

Uni-directional linearity of each object, then, does not imply uni-directional linearity of the series. On the other hand if each of two objects is linear in both directions (the word _linear_ being always understood in the sense of Sec. 1.2), then the two objects will also be linear in both directions. This result is clear from either of the above derivations of Eqs. (13), (14). By regarding two objects as themselves forming an object, one can obtain the corresponding result for three; by treating the three as a single object, the result with four is obtained, and so on. Proceeding in this way can easily give an inductive proof of the following, which summarizes the chief results of this section:

1. If each object of a series is linear, then the whole series, considered as an object, is also linear. The over-all transmission and reflection coefficients for two objects are given, in terms of the individual coefficients, by Eqs. (13), (14). These equations are valid for any linear objects, and hence they may be applied repeatedly to obtain the transmission and reflection coefficients of the whole series.

From now on, all the objects with which we shall be concerned will be linear in the sense of Sec 1.2. The result just stated shows that every combination of these objects will also be linear, and hence that this property need not be separately verified in any given situation.

2. Reflection as a Transformation

2.1. Separation of Amplitude and Phase. By the foregoing expressions we obtain the complex coefficients T, R. The absolute value and phase, which are frequently required in practice, can be computed in terms of these complex coefficients. As above, we use primed letters to denote the phase associated with a given complex number, so that for example $t = |t| e^{it'}$, with similar relations for the other quantities. If the line is lossless, as we shall assume throughout this section, then the transmission satisfies

$$|T|^2 = \frac{|t t_{,}|^2}{1 - 2|\rho r_{,}| \cos \phi + |\rho r_{,}|^2} \tag{19}$$

$$T' = t' + t_{,}' + kx + \tan^{-1}\left[\frac{|\rho r_{,}| \sin \phi}{1 - |\rho r_{,}| \cos \phi}\right] \tag{20}$$

where all variables are now real and ϕ is defined by Eq. (26). These equations are so simple that their chief properties can be determined by inspection. If graphical methods are to be used, moreover, one can obtain complete representation on a single page by plotting $\left|\frac{T}{t t_{,}}\right|$ or $T'-t'- \tau_{,} -kx$ versus ϕ with $|\rho r_{,}|$ as a parameter.

Before proceeding to the corresponding expressions for reflection, which are more complicated, let us introduce new parameters d and A defined by the equations

$$d = |d| e^{id'} = t\tau - r\rho \tag{21}$$

$$A = t' + \tau' - r' - \rho' \tag{22}$$

Because of its importance in determining the general character of a reflecting object, we shall refer to d as the _discriminant_. Its chief properties are discussed in RLE Technical Report No. 25; for the present, these results will be used without proof as required.

In terms of d, Eq. (14) takes the form

$$R = \frac{r + dr_, e^{2ikx}}{1 - \rho r_, e^{2ikx}} \tag{23}$$

which leads to

$$|R|^2 = \frac{|r|^2 + 2|drr_,| \cos \theta + |dr_,|^2}{1 - 2|\rho r_,| \cos \phi + |\rho r_,|^2} \tag{24}$$

$$R' = r' + \tan^{-1} \frac{|dr_,| \sin \theta}{|r| + |dr_,| \cos \theta} + \tan^{-1} \frac{|\rho r_,| \sin \phi}{1 - |\rho r_,| \cos \phi} \tag{25}$$

for the absolute value and phase. The quantities ϕ and θ, which depend on the spacing of the two objects, are given by

$$\phi = 2kx + r_,' + \rho' \tag{26}$$

$$\theta = 2kx + r_,' + d' - r' = \phi + d' - \rho' - r'. \tag{27}$$

The relations (24), (25) are inconveniently complicated, and for this reason it is desirable to present the same information in other ways. There are several alternative methods of representation, and a simple introduction to each of them is given by the so-called _vector diagram_ discussed below.

2.2. Vector Diagram. In practical work it frequently happens that the reflection coefficients are small. When this is the case one may sometimes neglect terms involving the product $\rho r_,$, so that Eq. (13) takes the simplified form

$$T \cong tt_, e^{ikx} \tag{28}$$

while (14) becomes

$$R \cong r + r_, t\tau e^{2ikx}. \tag{29}$$

This latter equation may be written

$$R \cong r + r_, e^{2ikx} \tag{30}$$

when $t\tau \cong 1$, which is often true for small reflections. Physically, Eq. (28) means that the over-all transmission is approximately equal to the product of the individual coefficients for the first object, for the second object, and for the section of line between them. Equation (30) says that the over-all reflection is the sum of the individual reflections, if that for the second object is duly modified to account for the length of line. Either (28) or (29) can be obtained from first principles by inspection, if we simply neglect interaction between the two objects (see Fig. 4b).

According to (30), the complex numbers r and $r_, e^{2ikx}$, which are equivalent to two-dimensional vectors, are added in the usual way to give a new complex number or vector R. The quantity r represents the reflection of the first object, while re^{2ikx} represents that of the second, as measured at a point immediately behind the first. In this

-9-

sense the vector R, representing the over-all reflection, is simply the sum of the vectors representing the individual reflections, a property which is the origin of the term **vector diagram** generally applied to the construction of Fig. 5b. If we keep the first object fixed and change x by moving the second, then the first vector remains unchanged while the second rotates as shown, the angle through which it moves being exactly proportional to x. These and similar results, which are standard in work of this kind, follow easily from (30) and from well-known properties of complex numbers.

We have regarded Eq. (30) as representing the sum of two vectors. Many other interpretations are possible. For example, one may regard it as a conformal transformation in the complex plane. Thus, when x is changed the reflection at a point just behind the first object (point 2 of the figures) will move around a circle with center at the origin and with radius $|r_1|$. If we now measure the reflection at point 1 instead of at 2, this circle, approximately, becomes the one shown in Fig. 5. The process may be interpreted as a complex transformation because Fig. 5, besides representing vector addition, is also equivalent to the so-called **Argand diagram**: it gives the imaginary part of R versus its real part.

From another point of view, Fig. 5 is found to give a plot in ordinary polar coordinates with origin at the point r. In other words the quantity $|R-r|$ is plotted as radius while $2kx$ is taken as the angle. This interpretation is satisfactory because the angle through which the second vector turns is, as noted above, exactly proportional to x.

A final interpretation, quite different from any of the foregoing, is to regard (30), not as a complex transformation, but as a simple linear transformation of the quantity $r_1 e^{2ikx}$, which represents the reflection coefficient of the second object as measured from point 2. A similar interpretation can be given to (28), and more generally, either pair of equations (28), (29) or (28), (30), when taken together, give a linear transformation of the transmission and reflection.

Because the vector diagram has so many desirable properties, one is naturally led to inquire whether it ever gives an exact, rather than an approximate, representation of Eq. (14). By inspection of (14) we find the following result:

ii. Equation (30) is exactly true, for some finite range of values of x, if and only if we have $r_1(|\rho| + |1-t\tau|) = 0$. In this case it is valid for all x. When the object is passive in the normal sense (i.e. contains no source of power), the condition becomes $r_1(|r| + |\rho| + |1-t\tau|) = 0$. Hence in this case, for all practical purposes either one or the other of the two objects must not be present at all. Similarly, Eq. (29) is valid for some range of values of x if and only if $t\tau\rho r_1 = 0$, and in this case it is valid for all x. The corresponding condition for (28) is $t\rho t_1 r_1 = 0$.

Hence these approximations are exact only in trivial cases. We have discussed them at some length, nevertheless, because the errors are often small, and many of the general properties are retained. Thus, the simple equations, (28),(29),(30) give an intro-
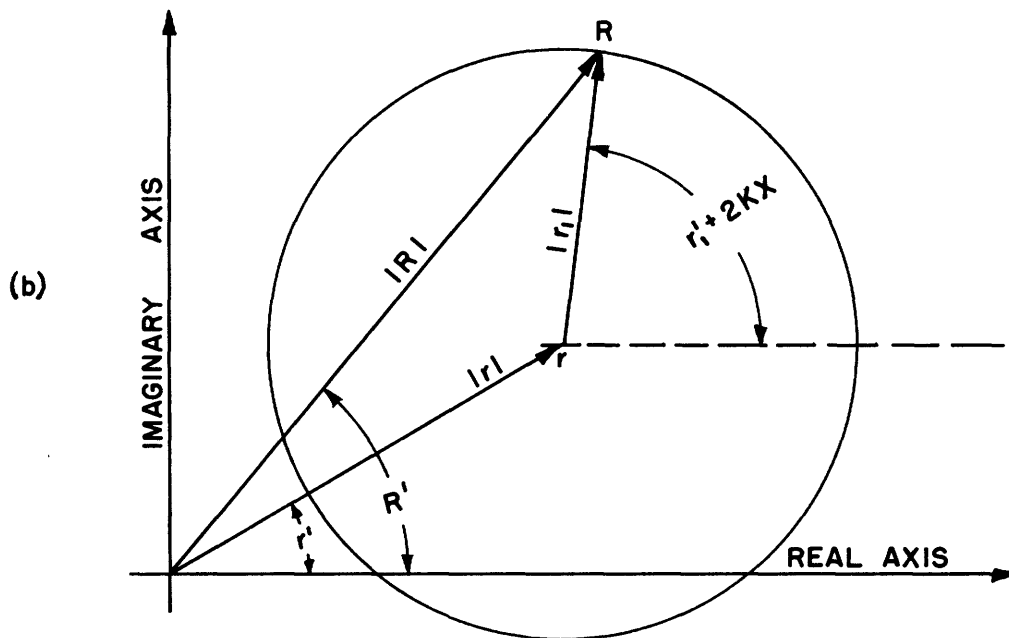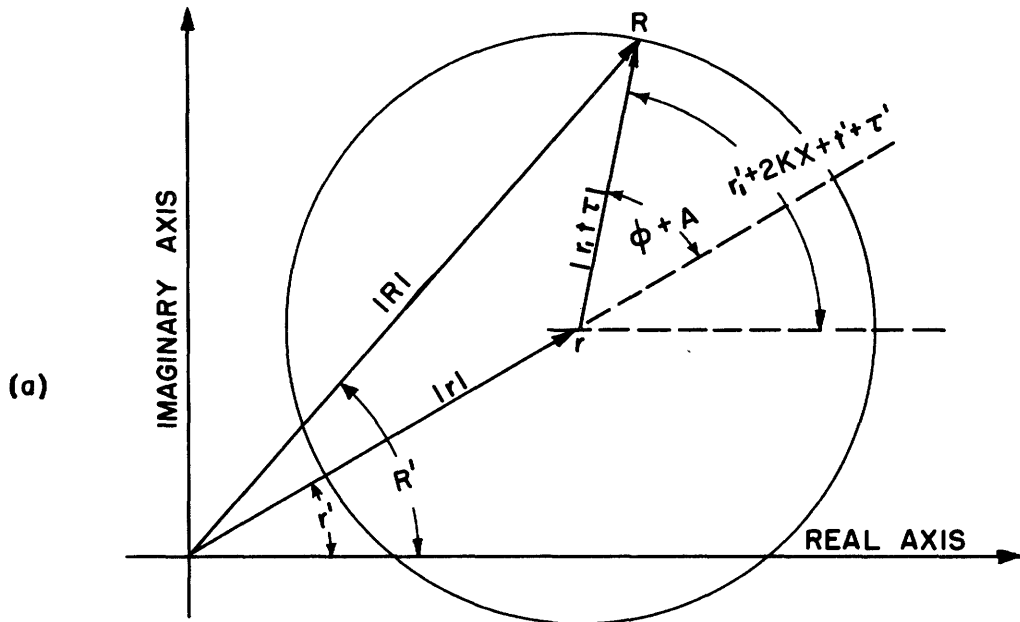
Figure 5. Two forms of the vector diagram for addition of reflections. Figure 5a is based on Eq. (29), and takes account of the transmission of the first object. Figure 5b is based on Eq. (30), and neglects the transmission of the first object.

-11-

duction to the exact equation (14) treated in detail below. Except for the vector representation, each of the above interpretations has its analogue in the exact case, although they then lead to different geometrical figures.

2.3. Conformal Representation. One may plot the real part of R versus its imaginary part, or the absolute value $|R|$ versus its phase, to obtain the customary representation in the complex plane. This procedure, which corresponds to the second of the above interpretations of Fig. 5, is by far the most important method of representation. In principle the radius for a polar plot is given by (24) and the angle is given by (25); but in practice it would be too complicated to use these relations directly. Instead, let us solve Eq. (23) for $r_i e^{2ikx}$, then multiply both sides by ρ, and finally take the absolute value to obtain

$$\left| \frac{R - r}{R + d/\rho} \right| = \left| \rho r_i \right| .$$ (31)

The reflection R must vary in such a way, therefore, that the absolute value on the left of Eq. (31) is equal to a constant, independent of x. It is well known that such a relation, which implicitly determines $|R|$ as a function of R', represents a circle with center at

$$c = |c|e^{ic'} = e^{ir'} \frac{|r|+|\rho dr_i|^2 | e^{i(d'-r'-\rho')}}{1 - |\rho r_i|^2}$$ (in general) (32)

$$= r \frac{1 - |r_i|^2}{1 - |rr_i|^2}$$ (lossless) (33)

and with radius

$$a = \frac{|t\tau r_i|}{1-|\rho r_i|^2}$$ (34)

$$= |r_i| \frac{1-|r|^2}{1-|rr_i|^2} .$$ (35)

The points

$$r_i - d/\rho$$ (in general) (36)

$$e^{ir'}|r|, \ e^{ir'}/|r|$$ (lossless) (37)

are inverse points. The maximum and minimum values of $|R|$ are quite difficult to derive directly from (14) or (24), but by using the relations just obtained we easily find

$$\text{extremes} = |c| \pm a = \frac{1}{1 - |\rho r_i|^2} \left[\sqrt{|r|^2+2|\rho r dr_i|^2|\cos(d'-r'-\rho')+|\rho dr_i|^2|^2} \pm |t \tau r_i| \right]$$ (38)

Results accompanied by the word lossless are valid whenever the first object absorbs no energy; they follow from the equations $|t|^2+|r|^2 = |\tau|^2+|\rho|^2 = 1$, $|\rho| = |r|$, $d = -e^{i(r' + \rho')}$ established for this case in RLE Technical Report No. 25.

Let us compare these exact relations with the familar vector diagrams discussed above. By combining the relation, fractional error = (approximate minus correct)/correct,

-12 -

with Eqs. (34), (35) we obtain the following, which describes the behavior of the radius:

iii. A complex plot of reflection is a circle in the exact case, just as in the approximate cases. For Fig. 5a, the fractional error in radius is $-|rr_i|^2$ or $-|\rho r_i|^2$ according as the first object is lossless or not; and hence the radius in this figure is never greater than the true radius. Similarly, the fractional error in Fig. 5b is

$$|r|^2 \frac{1 - |r_i|^2}{1 - |r|^2} \quad \text{or} \quad (1 - |t\tau| - |\rho r_i|^2)/|t\tau|, \text{ and hence,}$$

for the lossless case, the radius in this figure is never less than the true radius. As far as the radius is concerned, Fig. 5a or 5b is the more accurate, in the lossless case, according as $|r_i|^2(2-|r|^2)$ is greater or less than one.

Having considered the radius, let us investigate the location of the center. In the approximate diagram, the center coincides with the point r, which point may be termed the **fixed mismatch** because it represents the mismatch prevailing before the second, or moving, object was added to the system. As we see by Eq. (23), this simple property -- coincidence of center and fixed mismatch -- is not generally obtained in the exact case. For quantitative investigation let us use Eq. (32) to obtain

$$c - r = \frac{|t\tau\rho r_i^2|}{1 - |\rho r_i|^2} \; e^{i(t'+\tau'-\rho')} \tag{39}$$

while from (14) we have

$$R - r = \frac{r_i \, t\tau \, e^{2ikx}}{1 - \rho r_i \, e^{2ikx}} . \tag{40}$$

These relations lead to results which may be summarized as follows (see Fig. 6):

iv. A necessary and sufficient condition that the circle be centered at the origin is that $|r|+|\rho r_i^2 d| = 0$ or that simultaneously $d' = \pi + r' + \rho'$, $|r| = |\rho d r_i^2|$; the corresponding condition for coincidence of the center and the fixed mismatch is $t\tau\rho r_i = 0$. By (ii) it follows that coincidence of the center with the fixed mismatch, and exact validity of Fig. 5a, are equivalent conditions. In terms of the radius a, the distance $|r-c|$ from the center to the fixed mismatch is given by $a|\rho r_i|$.

If we draw a line from the origin to the fixed mismatch, and a second line from the center to the fixed mismatch, then the angle between these lines will depend only on the first object, not on the second; in fact this angle equals the angle

Figure 6. Reflection of two objects, as plotted in the complex plane without approximation. a) General case. b) First object lossless.

A of Eq. (22) whenever the equation is well defined. The result remains valid for $r=0$ if $r'=0$ is assumed for the now indeterminate value of r'. For a lossless object we have $A=\pi$ (RLE Technical Report No. 25) and hence the center, in this case, will always lie on the line joining the fixed mismatch to the origin, as shown in Fig. 6b.

Finally, the distance $|R-r|$ from a variable point R on the circle to the fixed mismatch is related to the over-all transmission T of the two objects by $|R-r| = \left| \dfrac{\tau \, r_i \, T}{t_i} \right|$.

In practical work it is frequently desirable to adjust the reflection of one or the other object in such a way that the over-all reflection is zero. This process is sometimes termed _matching_ the line. If the origin (that is, the point $R=0$ in Fig. 6) is inside the circle, then one must decrease $|r_i|$ or increase $|r|$ before a match can be obtained, the contrary procedure being required if the origin is outside. When it is actually on the circle, then a perfect match can always be obtained by suitable adjustment of x. Because of these results, which are evident from the figure, it is of interest to investigate the circumstances in which the origin is inside, on, or outside the circle. By Eqs. (32),(34) one finds the following, after noting that the origin will be inside if and only if $|a|>|c|$:

v. The point $R=0$ will be interior to the circle, on the circle, or exterior to the circle, according as $|r/d|$ is less than, equal to, or greater than $|r_i|$. For zero loss we have $|d| = 1$ and hence the condition becomes $|r|$ less than, equal to, or greater than $|r_i|$. This condition is the same as that for the simplified vector diagram of Fig. 5b; in other words, if the first object is lossless, then the origin will be in, on, or outside of the circle in the approximate case, according as it is in, on, or outside of the circle in the exact case.

With the vector diagram of Fig. 5 the angle determining the position of the radius vector R-r is such that the initial vector measured at point 2 and the transformed vector measured at point 1 move with the same angular velocity; in other words, uniform linear motion of the second object will lead to uniform angular rotation of the radius vector. With the exact diagram, however, this situation no longer prevails. For quantitative investigation let ψ represent the angular position of the radius vector, the angle being measured from the line joining c and r (Fig. 6a). This choice of origin leads to considerable simplification. By applying the law of cosines to the triangle formed by the three points c, r, R we obtain

$$| R-r|^2 = |r-c|^2 + a^2 - 2a|r-c|\cos\psi \tag{41}$$

which gives

$$\frac{\sin\psi}{\sin\phi} = - \frac{(1-|\rho r_i|^2)\sin\phi}{1 - 2|\rho r_i|\cos\phi + |\rho r_i|^2} \tag{42}$$

-15-

by virtue of Eqs. (39), (40). This result is the desired relationship between ϕ and ψ, and when combined with results previously established, it leads to the following:

vi. If the angle ψ determines the position of the transformed vector (Fig. 6) while ϕ determines that of the original vector, then the ratio $\sin \psi / \sin \phi$ has the same form whether the loss is zero or finite, and does not depend explicitly on any of the quantities $|t|, |T|, t', T', \rho', r'$, or r_i'. It may be expressed as $-|t_\ell|^2$, where t_ℓ is the transmission of two lossless objects which are an electrical distance ϕ apart and each of which has a reflection coefficient $\sqrt{|\rho r_i|}$.

The two vectors will rotate with the same angular velocity if and only if $\rho r_i = 0$; hence uniformity of the angular velocity, and exact validity of Fig. 5a with non-zero radius, are equivalent conditions.

In terms of $|R-r|$, which represents the distance from a variable point on the circle to the fixed mismatch, we have

$$-\frac{\sin \psi}{\sin \phi} = \frac{|R - r|^2}{|c - r|} \left| \frac{\rho}{t\tau} \right| = \frac{|R - r|^2}{a|r_i t\tau|} \quad \text{where } |r-c| \text{ is the}$$

fixed distance from center to mismatch, while $\underline{a}$ is the radius. The variation of $|R-r|^2$ as one traverses the circle, then, gives a measure of uniformity of the angular velocity. In terms of the overall transmission T for the two objects, this last relation becomes

$$\frac{\sin \psi}{\sin \phi} = -\left| \frac{r_i \tau T^2}{at^3} \right| \quad \text{by virtue of iv.}$$

If the first object is lossless one may use Eq. (24) to obtain $\underline{a}$, the radius, as a function of $|r|$, the distance to the fixed mismatch. For this lossless case a few simple properties are also found when, instead, we obtain $\underline{a}$ in terms of $|c|$, the distance to the center. Such a relation is readily found by combining Eqs. (33), (35); we have in fact

$$a = \left[1 + a_o^2 - \sqrt{(1 - a_o^2)^2 + 4|c|^2 a_o^2}\right] / 2 \, a_o \qquad (43)$$

where $a_o = |r_i|$ is the radius before transformation, that is, the radius corresponding to $r = 0$. A simple geometrical interpretation of this equation is given by Fig. 7a, which is due to Rieke, and another is given by Fig. 7b, which was suggested to the author by S. Silver.

Equation (43), like most of the others, may be derived by impedance methods, although the calculations for the general case are quite laborious. There is an interesting special case, however, in which impedance methods lead directly to this result. If the mismatch r is introduced by varying a single stub of a lossless tuner, then the circle in the impedance plane is translated parallel to the imaginary axis. By well-known properties of the transformation (1) relating the impedance and reflection planes, one immediately obtains Fig. 7b; and from this in turn Eq. (43) may be derived. Such reasoning can be generalized as described in Reference 4, where a more detailed discussion is
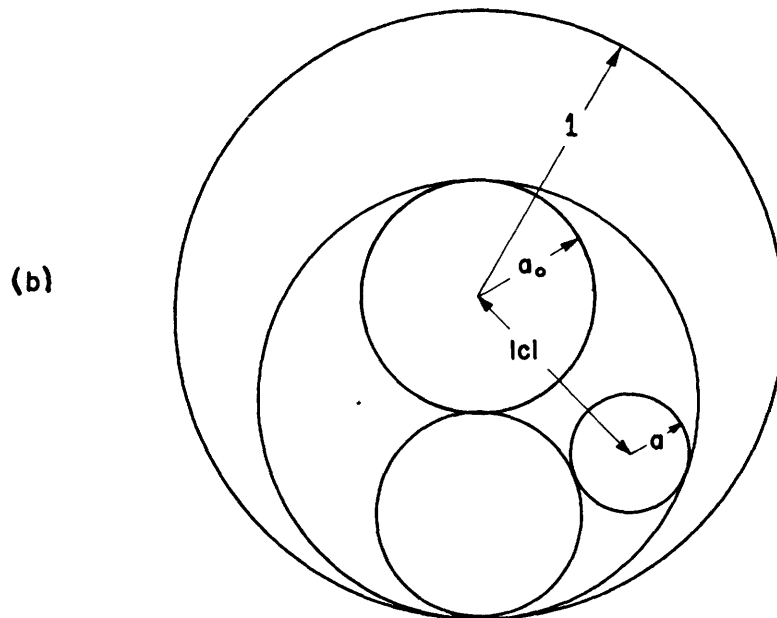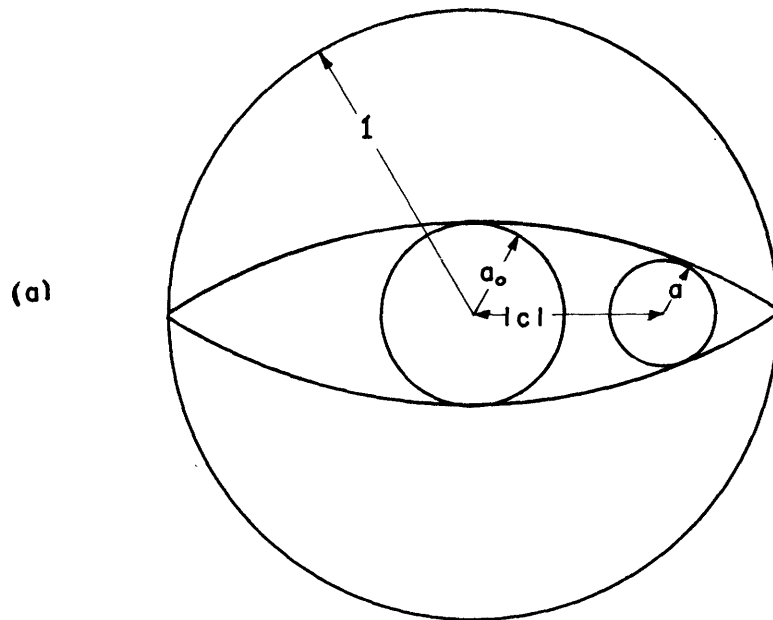
Figure 7. Two geometrical constructions for obtaining the final radius
a in terms of the initial radius a_o and the distance to the
center $|c|$. The first object is assumed to be lossless.

given; for the present it suffices to note that Fig. 7b, even as to details, admits a simple physical interpretation.

2.4. Polar Coordinates. Up to this point we have assumed that the complex reflection coefficient R is to be plotted in the complex plane or Argand diagram, the variable ϕ being represented only implicitly. This procedure is an extension of the well-known vector diagrams for small reflections, Fig. 5, to give an exact, instead of an approximate, portrayal of the function. Such an extension may be made in other ways, as noted above; for example, instead of retaining the property that the real and imaginary parts of R be represented explicitly, we may insist that the angle remain equal to a constant plus ϕ. In other words, the length of the line from the fixed mismatch to a given point on the 'circle' will be unchanged, but the angle determining its position will now be proportional to the distance x separating the two objects.

With this point of view, which corresponds to the third of the above interpretations for the vector diagram, one finds that the resulting curve will no longer be a circle in the general case. For quantitative investigation let us combine Eqs. (26) and (40) to obtain

$$|R - r| = \frac{|r_1 t \tau|}{[1 - 2|\rho r_1| \cos \phi + |\rho r_1|^2]^{\frac{1}{2}}} \tag{44}$$

which, when plotted in polar coordinates with ϕ as angle, gives the desired representation. We observe that (44) is unchanged when we replace ϕ by $-\phi$, and hence that $\phi = 0$ gives an axis of symmetry. The length of this axis is the sum of the maximum and minimum distances to the fixed mismatch, while the width of the curve is given by the expression $|R - r| \sin \phi$ when maximized with respect to ϕ (see Fig. 8). By these results, and by comparison of Eq. (44) squared with

$$r = \frac{\rho}{1 + e \cos \theta} ,$$

the standard form of an ellipse in polar coordinates, one obtains the following, which on account of iv, may also be applied to transmission:

vii. If the angle $\phi + A$ of Fig. 5a is kept equal to $\phi + A$, the radius being determined by its exact value (44), then the curve so obtained will be a circle or a point if and only if $t \tau \rho r_1 = 0$. In this case the exact figure (8) will be identical with the approximate figure (5a), as we see by ii; that is, degeneration of the curve to a circle or point, and exact validity of Fig. 5a, are equivalent conditions.

The maximum diameter of the curve in Fig. 8 is always equal to the diameter 2a obtained for the circle in the exact case [Fig. 6a and Eq. (34)]. The minimum diameter or width of the curve, on the other hand, is always equal to the diameter $2|t \tau r_1|$ obtained for the circle in the

approximate case (Fig. 5a). The width of the curve, in particcular, does not depend explicitly on any of the variables $|r|, |\rho|$, r', ρ', t', τ', or r_I'.

If we replace the radius of this polar plot by its square, other things being the same, then the curve becomes an ellipse with the fixed mismatch at one focus. The eccentricity depends only on $|\rho r_I|$ and is equal to $2|\rho r_I|/(1+|\rho r_I|^2)$. The ratio of axes is equal to $(1+|\rho r_I|^2)/(1-|\rho r_I|^2)$, the voltage standing-wave ratio for a power reflection $|\rho r_I|^2$; the ratio of maximum to minimum distance from the fixed mismatch is equal to $(1+|\rho r_I|)^2/(1-|\rho r_I|)^2$, which is the power standing-wave ratio for an amplitude reflection $|\rho r_I|$.
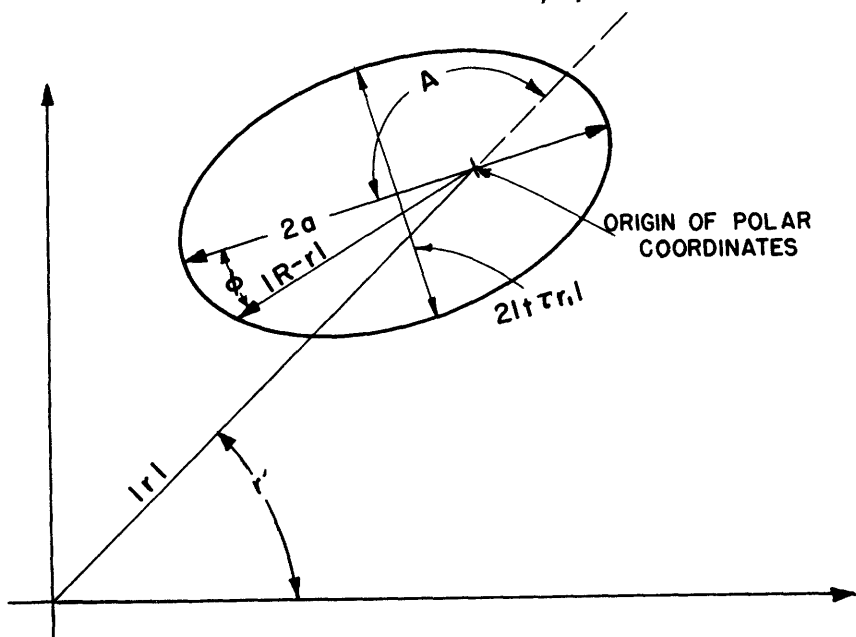

Figure 8. Radius vector of Fig. 5a, as plotted in
polar coordinates without approximation.

2.5. Linear Transformation. Proceeding now to the fourth of the above interpretations of Fig. 5, we attempt to find a linear transformation which will represent the effect of the first object. Neither Eq. (13) nor (14) is linear as it stands, and hence one cannot expect to express the transmission itself, or the reflection itself, in the desired manner. Following Reference 5 we therefore assume $t \tau t_I \neq 0$ and write Eqs. (13), (23) in the form

$$\frac{1}{T} = \frac{1}{t} \frac{1}{t_I} e^{-ikx} - \frac{\rho}{t} \frac{r_I}{t_I} e^{ikx}$$

$$\frac{R}{T} = \frac{r}{t} \frac{1}{t_I} e^{-ikx} + \frac{d}{t} \frac{r_I}{t_I} e^{ikx} . \tag{45}$$

For any fixed value of x these equations represent a simple linear transformation of the

-19-

variables $1/t_i$, r_i/t_i into new variables $1/T$, R/T. It is clear that such variables characterize the object just as well as the original ones t_i, r_i, in the sense that, if either set is known, the other set can be found. Aside from the trivial restriction that no transmission be zero, the chief shortcoming of these new parameters is that they do not lend themselves readily to physical interpretation.

   With matrix notation the transformation represented by (45) can be written in the form (Reference 5)

$$\begin{pmatrix} \dfrac{1}{T} \\[2mm] \dfrac{R}{T} \end{pmatrix} = \begin{pmatrix} \dfrac{1}{t}\, e^{-ikx} & -\dfrac{\rho}{t}\, e^{ikx} \\[2mm] \dfrac{r}{t}\, e^{-ikx} & \dfrac{d}{t}\, e^{ikx} \end{pmatrix} \begin{pmatrix} \dfrac{1}{t_i} \\[2mm] \dfrac{r_i}{t_i} \end{pmatrix} \tag{46}$$

and hence calculation of the reflection for a series of objects is equivalent to calculation of a matrix product. This question will be considered in detail subsequently; it is mentioned here to illustrate one of the advantages of the otherwise rather arbitrary substitution (45).

References

1. J. A. Stratton, "Electromagnetic Theory", McGraw-Hill, New York, 1941.

2. J. C. Slater, "Microwave Transmission", McGraw-Hill, New York, 1942.

3. "The Elements of Wave Propagation Using the Impedance Concept", TRE Mathematical Group, Report No. 26, December 1, 1941.

4. R. M. Redheffer, "Radome Bulletin Number 18: The Interaction of Microwave Antennas with Dielectric Sheets", RL Report 483-18, March 1, 1946.

5. P. M. Marcus, "The Interaction of Discontinuities on a Transmission Line", RL Report 930, February 6, 1946.