CRG REPORT NO. 84

THEORY OF MICROWAVE COUPLING SYSTEMS

by

E. T. Jaynes

Approved by ____________________
Dr. C. E. Cleeton
Technical Head

Released by ____________________
Capt. H. C. Owen
Commanding Officer

COMBINED RESEARCH GROUP
NAVAL RESEARCH LABORATORY
ANACOSTIA STATION
WASHINGTON, D. C.

29 August 1945

THEORY OF MICROWAVE COUPLING SYSTEMS

Contents

List of Illustrations

Explanation of Symbols Used

Γ = Reflection Coefficient

K = Coefficient of Coupling

α = Attenuation constant of a transmission line (Chap. 3)

β = Phase-shift constant of a transmission line (Chap. 3)

$\gamma = \alpha + j\beta$ = Propagation constant (Chap. 3)

ω = Applied frequency

ω_0 = Resonant frequency of a system

$\delta = \frac{\omega - \omega_0}{\omega_0}$ = Fractional frequency deviation from resonance.

S = Standing wave voltage ratio

ρ = SWR seen looking into an unloaded cavity at resonance

γ = SWR seen looking into a loaded cavity at resonance

η = SWR seen looking into a cavity off resonance

Q_u = Unloaded Q of a cavity

Q_c = Coupled Q of a cavity

Q_L = Loaded Q of a cavity

θ Electrical length of a line or phase angle of a load impedance.

ϕ = Phase angle of the parallel combination of a load impedance and a normalizing impedance.

ψ = Electrical length of line between a cavity and the probe on a standing-wave detector.

$\alpha = 2 Q_L \delta$ = Detuning parameter relative to Q_L (Chap. 8)

$\epsilon = 2 Q_c \delta$ = Detuning parameter relative to Q_c (Chap. 8)

THEORY OF MICROWAVE COUPLING SYSTEMS

I. Introduction

At the present stage of the CRG development program it is becoming clear that the work of the past year has resulted in a number of useful by-products in addition to the main development work. Valuable experience has been gained in many different fields, and a considerable amount of fundamental original work has been done which at present exists largely in the minds and notebooks of a few persons in the Combined Research Group. It is, of course, highly desirable that this experience and informati be made available to those who will be responsible for conducting and judging the future tests on various units, and to workers in related fields.

One of the fields in which there is need for a more wide-spread understanding is the theory of microwave coupling systems. As tests on more and more RF units are conducted and reported, it has been necessary to repeat the various theoretical relationships in each such report, in order to convey the reasons for and the significance of the various measurements which are made. This is due to the non-existence of any general account of this theory to which reference might be made. In order to correct this situation and at the same time fill a need which has long existed for an exposition of this theory for the use of those who are not directly concerned with the design of microwave coupling systems, but have occasion to make them in the course of other work, it is proposed here to present an account of the relationships which follow from linear circuit theory, in the form which has been found most useful at microwave frequencies.

In the past few years the intensive work in microwaves has resulted in the development of a new field of techniques. A number of books have appeared treating the theory of transmission lines and resonant cavities at microwave frequencies, but there is as yet no generally available account of the theory of the equally necessary coupling systems which connect these components. Techniques for the measurement and design of such systems are well known by those who have had considerable experience with ther and there are a number of general circuit theorems which have special value as applied to coupling systems at microwave frequencies, but which seem never to have been stated in ways which made their general validity apparent. Fragments of a general theory of microwave coupling systems have appeared in a large number of technical reports, but they are usually stated with reference to a special case, and often derived from a highly artificial equivalent circuit. A number of extremely general propositions have been stated in this way, but with no hint that they apply beyond the specific case being considered, so that it is only after a perusal of a large number of such reports and considerable exercis of physical intuition that one can acquire a clear picture of the field.

There is probably no proposition in this paper which is not well known to those who have had much experience with microwave coupling systems and have had access to classified technical report and such people may find a large part of the discussion quite trivial. That is because an effort has been made to work out in some detail the answers to several questions which prove to be stumbling-blocks to those who are in the process of learning about microwave systems, and who often entertain grave doubts regarding the validity of many standard measurement techniques, because these have never been explained and justified in a general, explicit way. In a field such as this, where there is a wide variation of initial understanding of the subject, the only safe way of writing an exposition is to include a discussion of the fundamentals upon which the later developments are based. In fact, the necessary "groundwork" of initial concepts will occupy nearly as much space as the actual study of coupling systems. However, this discussion is concentrated in separate chapters which may be omitted.

II. Fundamental Principles

Use of Impedances

In RF work we are concerned almost all of the time with three types of components, namely resonant cavities, transmission lines (including waveguides), and arrangements for coupling energy between these systems. Of course, the distinction between these types of components is not sharp, and it is often impossible to say exactly where a cavity or line ends and a coupling system begins. The properties of transmission lines and resonant cavities are well known, as they have been treated in a number of excellent recent books. It is the purpose of this discussion to develop the relations which must hold in a linear coupling system regarded as a four-terminal network connecting two lines, two cavities, or a line and a cavity, and to point out how these relations may be used to find values of circuit parameters, such as "Q" or coefficient of coupling, in terms of measurements which can be made with the most commonly available laboratory instruments. Such measurements form the basis for any logical design or development procedure.

The analysis will be carried out largely in terms of impedance relations rather than in terms of the more fundamental viewpoint starting from Maxwell's equations and leading to relations between electric and magnetic fields for three reasons:

1) The formal method of solution using Maxwell's equations in boundary-value problems requires very advanced mathematics and usually yields detailed solutions of special cases rather than the general relations sought here, while these general relations are expressible in terms of circuit and impedance concepts using much simpler mathematics.

2) Laboratory measurement equipment which has been developed to date almost invariably measures impedances directly rather than field intensities.
3) Circuit and impedance concepts are familiar to more people than are relations between field components.

It is well known that impedance relations form a rigorous basis for the analysis of transmission lines such as coaxial lines and waveguides carrying a single wave type, and it is quite natural and common to represent a resonant cavity by a resonant lumped-constant circuit, although many people are uncertain as to how valid this representation is. Accordingly, it may be well to point out here that is has been shown by Hansen* that two quantities L and C may be defined in terms of integrals of the vector potentials of the fields throughout the volume of a cavity of arbitrary shape, such that the ordinary circuit equations for the equivalent circuit involving L and C as inductance and capacitance are identical with the Lagrangian relations which insure that Maxwell's equations are satisfied. A separate set of L and C is defined in this way for each of the modes of oscillation of the cavity, except for a constant which adjusts the ratio L/C according to the way the current in the resonator is defined for each mode. The equivalent circuit then consists of an infinite number of lumped-constant resonant circuits, and since the ordinary circuit equations for these determine a solution of Maxwell's equations which satisfies all the conditions imposed, it follows from the uniqueness theorem that this equivalent circuit is the complete one. This conclusion could also be deduced from Foster's reactance theorem, as applied to a circuit coupled into the cavity. The result is that a single lumped-constant equivalent circuit may be used with confidence as long as we restrict ourselves to frequencies far from the resonant points of other modes of oscillation.

---

*Hansen, W.W. "A Type of Electrical Resonator," J.App. Phys., Vol. 9, No. 10, (October, 1938) p. 654.

Graphical Representation of Impedances

Probably the most useful mental tool available in RF work is the device of representing impedances graphically. There are a number of charts which have been used in this way, of which only two, the rectangular plot and the "Smith" chart, have become generally used. The reason for the usefulness of these methods is that the human mind perceives geometrical relationships much more readily than relationships expressed in the form of mathematical equations. Once the correspondence between position on some chart and a mathematical quantity has been established, one can let the equations fade into the background and record measurements, do one's reasoning, and even derive new results in terms of the geometry of the chart.

Any type of chart has, of course, an infinite number of geometrical properties, and from each of these some fact about networks and impedances may be found. No matter how much experience one has had with impedance charts, he can still learn new useful facts about them at a rate limited only by his ambition.

Rectangular Impedance Charts

The rectangular impedance plot, in which the abscissa and ordinate represent resistance and reactance, the real and imaginary components of an impedance, is well known. Its chief unique property is that if one interprets impedances as vectors on this chart, the angle between a vector and the real axis is equal to the electrical angle between voltage and current in the corresponding impedance, and the length of the vector is equal to the ratio of the magnitudes of the voltage and current. The result of connecting impedances in series is equivalent to vector addition by the parallelogram law in a rectangular impedance chart.

Rectangular charts of admittance, the reciprocal of impedance are in general use for discussions involving elements connected in parallel. The coordinates are the conductance G and susceptance B, which for an element of impedance Z = R + jX are given by:

$$G = \frac{R}{R^2 + X^2} \quad , \qquad B = \frac{-X}{R^2 + X^2} \qquad (1)$$

In this chart the roles of voltage and current are interchanged and the operation of vector addition corresponds to connecting elements in parallel.

The Smith Chart

In rectangular charts the location of all physically available impedances (those with a positive resistance component) is the right half-plane. Since the area of this plane is infinite it is impossible to have a rectangular chart on which all possible impedances may be represented. There is, however, a method of mapping this half-plane onto a circle of finite size so that all such impedances may be represented as points in the circle. This is the Smith Chart,*which was described in the reference below as a device for making transmission line calculations. It is illustrated in Figure 1. Instruments based on this chart, and called "transmission line calculators" are commercially available. However, the association of this device exclusively with transmission line problems is unfortunate, as it applies only to a special type of line, namely one having a characteristic impedance equal to a pure resistance, and it is extremely useful in studying many general types of networks that bear no resemblance to transmission lines. Accordingly, we shall develop the theory of this chart in terms of a much simpler physical concept than the impedance relations along a transmission line.

*P.H. Smith, Electronics, January 1939.

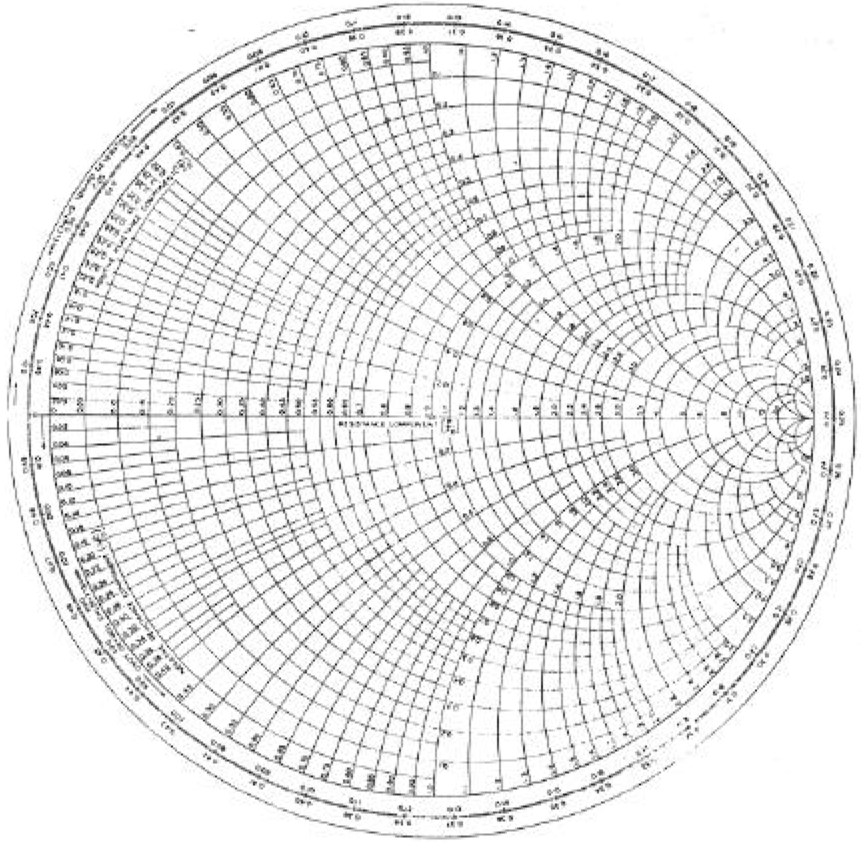


FIGURE 1

A1386

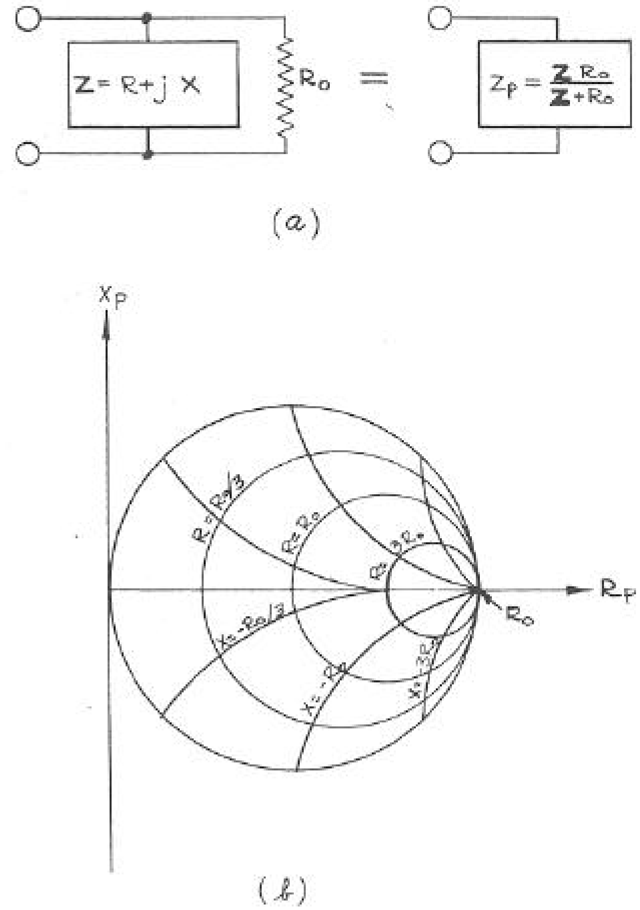


FIGURE 2

A1386

In Figure 2a is shown a resistance R_o connected in parallel with an impedance $Z = R + jX$. Regarding R_o as fixed, we wish to find how the resultant impedance Zp of the combination varies with Z. It is apparent that if Z = 0, then ZP = 0, if $Z = R_o$, then $Zp = R_o/2$, and if $Z = \infty$, then $Zp = R_o$. If Z is a pure reactance, Zp must lie somewhere on the largest circle in the rectangular impedance plot of Figure 2b. In general, as R varies keeping x constant, Zp moves along the circle given by:

$$(R_p - R_o)^2 + \left(X_p - \frac{Ro^2}{2X}\right)^2 = \frac{R_o^4}{4\,X^2} \qquad (2)$$

while if X is varied for constant R, the locus of Zp is the orthogonal set given by:

$$\left[R_p - R_o\,\frac{2R + Ro}{2(R+Ro)}\right]^2 + X_p^2 = \frac{Ro^4}{4(R+R_o)^2} \qquad (3)$$

A few of these circles are shown in Figure 2b. Note that they are identical in form and labeling with the coordinate circles of the Smith Chart in Figure 1. If any physically attainable impedance is connected in parallel with a resistance R_o, the impedance of the combination will lie on or within a circle of diameter R_o, tangent to the reactance axis at the origin. Since corresponding to each possible value of Z there is one and only one value of Z_p, we see that the operation of placing a fixed resistance R_o in parallel with the variable impedance Z is equivalent to mapping the infinite domain of Z onto a finite circle. As remarked above, the center of this circle represents Z = Ro, and we can therefore make the center represent any value of resistance in which we are especially interested by choosing, Ro equal to that value. We then say that the Smith Chart is normalized with respect to Ro.

The use of the Smith Chart in connection with general networks will be considered in Chapter III of this report, but it will be expedient to discuss here the application to transmission lines, which is the basis of its present widespread use. To establish the connection with transmission lines, we note that:

$$Z_p = \frac{Z\,Ro}{Z+R_o} \quad , \quad 2\,\frac{Zp}{Ro} - 1 = \frac{Z - Ro}{Z + Ro} = \Gamma \qquad (4)$$

the quantity (Z-Ro)/(Z+Ro) is recognized as the reflection coefficient Γ at the end of a transmission line of characteristic impedance Ro, when terminated in a load impedance Z.

The transformation from the quantity Zp/Ro to this reflection coefficient is seen to consist of a change of scale by a factor of two, and a translation of the origin one unit to the right. This is illustrated in Figure 3., and it is seen that the origin from which the complex number Γ is measured is the center of the Smith Chart, while the points for which the absolute magnitude of Γ is unity constitute the periphery of the Chart. Therefore, an alternative way of looking at a Smith Chart is to regard it as a plot of the reflection coefficient of a transmission line in the circular region $|\Gamma| \leq 1$. The convenience of the chart in transmission line problems then follows from the fact that if Γ is expressed in polar coordinates:

$$\Gamma = |\Gamma| e^{j2\theta}$$

then $|\Gamma|$ depends only on the standing wave ratio $S = \frac{E_{max}}{E_{min}}$ in the line, through the relations

$$|\Gamma| = \frac{S - 1}{S + 1} \quad , \quad S = \frac{1 + |\Gamma|}{1 - |\Gamma|} \qquad (5)$$

and θ is the distance in electrical degrees from the load to the nearest voltage maximum. Thus the measured values of SWR and position of the standing waves are simply the polar coordinates locating the load impedance on a Smith Chart. In a transmission line with no losses, the impedance seen at any point looking toward the load is located on the Smith Chart by rotating the point representing the load impedance thru an angle equal to twice the electrical length of line between the load and the point of measurement. As one moves along the line, the impedance seen looking toward the load moves at a uniform rate along a circle of constant SWR, concentric with the outer circle.

A physical interpretation of position on the Smith Chart in terms of voltage and current vectors may be based on the properties of the reflection coefficient. If the line is analyzed in terms of two waves traveling in opposite directions, the ratio at any point of the voltage vector in the reflected wave to the voltage in the incident wave is by definition the reflection coefficient at that point. If, as in Figure 4 we represent the incident voltage vector by unity, the reflection coefficient Γ is equal to the reflected voltage vector. The total voltage on the line at that point is the vector sum of these components, and is represented by a vector from the left edge of the chart to the point representing the load impedance. The variation of total voltage along the line can then be visualized immediately as the tip of the voltage vector moves along a circle of constant SWR. Similarly, we may put the current in the incident wave equal to unity, in which case the current vector in the reflected wave is $(-\Gamma)$, since the direction of current flow for a given voltage in a wave reverses when the direction of propagation of the wave reverses. This is represented by a vector equal and opposite to the reflected voltage vector, and the total

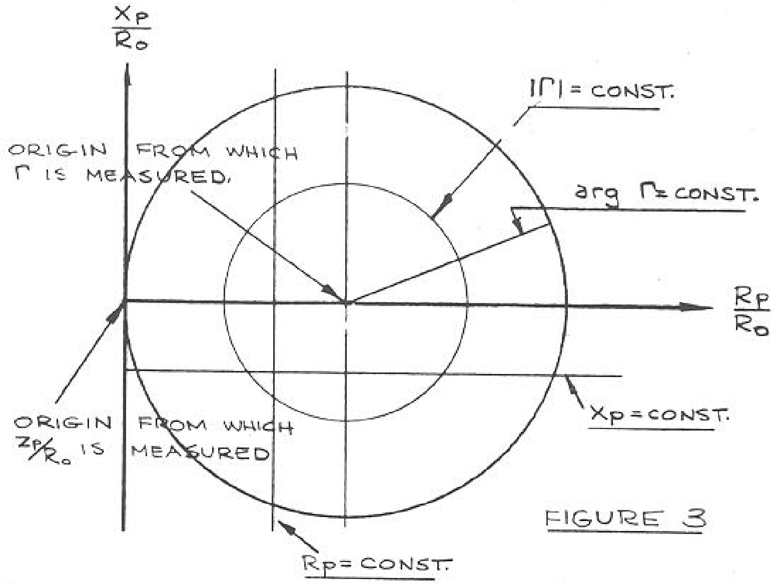


FIGURE 3

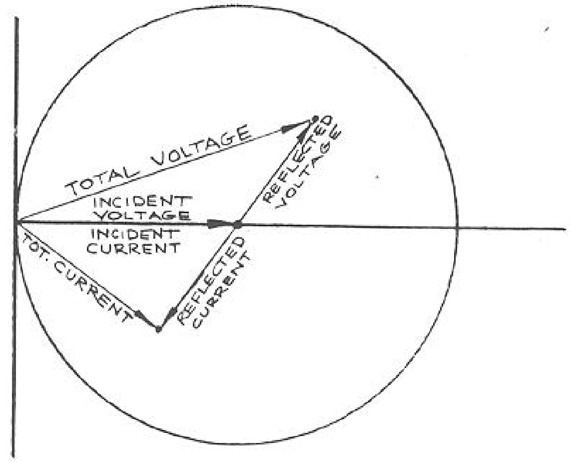
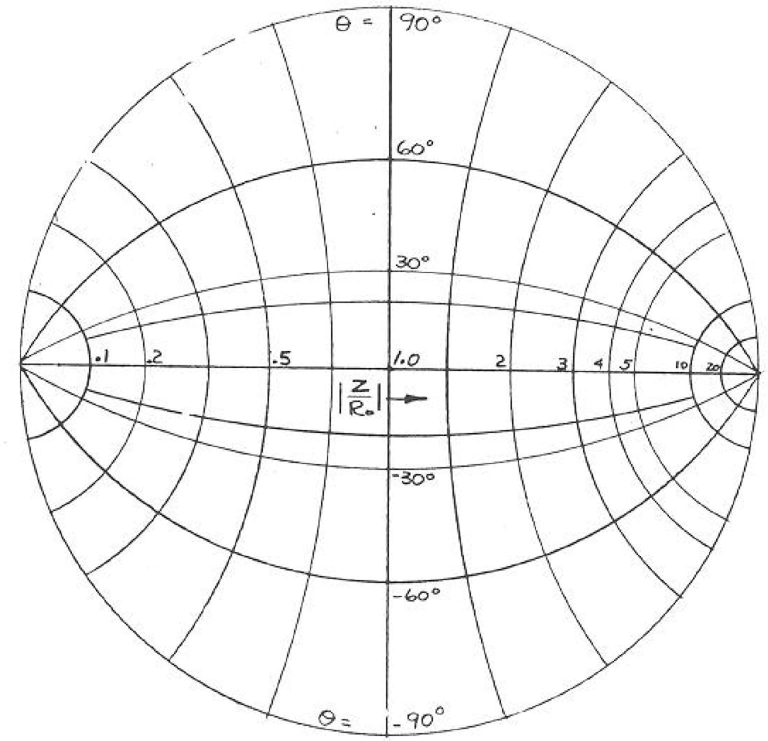


FIGURE 4

A 1386



$$\frac{Z}{R_o} = \left|\frac{Z}{R_o}\right| e^{j\theta}$$

FIGURE 5

A 1386

current is the vector sum of the current in the incident and reflected waves. It can easily be seen that as one moves along the line, a voltage minimum occurs at the same point as a current maximum, and vice versa.

The coordinate circles drawn on the Smith Chart of Figure 1 are lines of constant resistance component or constant reactance component in the impedance Z, or in the load impedance of the transmission line. They may be regarded as a distortion of the straight lines used as coordinates on a rectangular chart, such as would occur if a rectangular chart were drawn on a sheet of rubber, and then stretched and warped until they were bent into the Smith Chart pattern. Mathematically, this transformation is called a conformal map, and it has the property that angles between lines are preserved in going from one chart to the other. Therefore, all the coordinate circles in the Smith Chart intersect at right angles, as do the straight lines on a rectangular chart. A further property that is of great importance in reasoning with these charts is that circles are preserved; that is, if the locus of some variable impedance is a circle on one chart, it is also a circle on the other (a straight line being of course, a special case of a circle)* Therefore, if one can find intuitively three facts about such an impedance locus (such as two points through which it must pass and the slope of the tangent at one of them) it is at once completely determined without the necessity of using any mathematics. In general, the center of one circle will not transform to the center of the other circle, and the distribution of points along the periphery will be altered in a transformation, but this seldom causes difficulty if one keeps a sufficient number of geometrical properties of the charts in mind.

One frequently has occasion to think of an impedance in terms of absolute magnitude and phase angle rather than resistance and reactance, so it is well to know what the lines of constant magnitude and constant phase look like on a Smith Chart. These circles are shown in Figure 5. Note that they too always intersect at right angles. Impedance charts on which these circles are printed as coordinates instead of the constant resistance and reactance lines, are called "Hemisphere Charts". It is clear that they possess the advantage of symmetry and more uniform spacing of the coordinate circles, which makes them preferred to the regular Smith Charts in some cases, although they really amount to the same thing.

---

*This is a property possessed by conformal transformations of a special type, known as a linear fractional transformations, of which the above is an example. The general linear fractional transformation between the complex numbers W and Z is given by:

$$W = \frac{AZ + B}{CZ + D}$$

where A, B, C and D are any complex numbers.

The above explanations should serve as a working introduction to the use of the Smith Chart, although they naturally can only "scratch the surface" of the subject. The properties described in this section have been chosen because each of them plays an essential role in some later development.

III. Properties of a General Linear Four-Terminal Network

As has been explained above, we shall regard any arbitrary coupling system between two units as a linear four-terminal network connected between them. The most general statements which we can make about this network will then be true for each of its possible specific forms such as coupling loops, capacity probes, irises, etc. It will be shown that due to the fact that distributed constant transmission lines are almost invariably connected to a microwave coupling system, the distinction between various arrangements nearly disappears, and consists only in a shift of a certain reference point along the line.

The theory of general four-terminal networks has been treated by a number of authors, perhaps most comprehensively by Guillemin*. Since our interest here is in the application to microwaves, our analysis will differ somewhat from usual expositions, in order to keep results in an appropriate form for graphical interpretation.

It is customary in conventional network analysis to focus attention on the transfer properties of the network; that is, given the input signal, to find the output signal. Transfer properties are our fundamental interest in microwaves too, but if a theory is to be useful, it must be concerned with experimentally measurable quantities, and we cannot usually be so direct at the present stage of our techniques. However, impedances are quite easy to measure at microwave frequencies, so the logical procedure is to develop a theory in which the transfer characteristics of a network may be deduced from its impedance relations. Therefore we will study the properties of a network in terms of the input impedance as a function of load impedance, and it will then turn out that the transfer properties may be obtained from impedance relations as needed in a very simple way.

In recent years there has been a growing tendency to analyze circuits in terms of admittances instead of impedances. The reason for this is often artistic, although many cases exist in which a saving of algebra is effected. We will therefore find it expedient to present both forms of analysis here, so that they may be mixed indiscriminately in later applications. However, since it is quite confusing to carry on two developments simultaneously, we shall consider first only impedance relations, and then repeat the most important formulas in terms of admittances.

---

* Guillemin, E.A. Communication Networks, Vol. 2

Turning now to the general four-terminal network, Fig. 6 illustrates the current and voltage conventions. Linear circuit theory shows that regardless of whether its elements have lumped or distributed parameters, and regardless of the complexity of its interconnections, the behavior of any four-terminal linear network may be completely specified by three independent functions of frequency. Accordingly, wherever the geometry of an RF system becomes simple enough so that an impedance may be defined, this impedance will depend on any other impedance in the system, regarded as a load impedance, through some relation involving three independent functions of frequency which, for a non-resonant intermediate network and in a narrow frequency range, become three constants. These parameters may be chosen in many different ways, the most useful one for our purpose being the scheme of two open-circuit driving-point impedances Z_{11} and Z_{22}; and one mutual impedance Z_{12}, connected by:

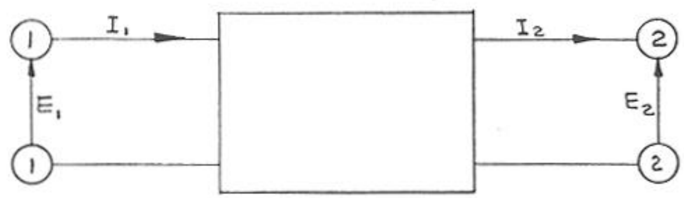
$$\begin{aligned} E_1 &= Z_{11}I_1 - Z_{12}I_2 \\ E_2 &= Z_{12}I_1 - Z_{22}I_2 \end{aligned} \qquad (6)$$

A T-section network with parameters Z_{11}, Z_{12}, and Z_{22} is shown in Fig. 7a. This is the usual form in which a general network is visualized, and it is one whose elements often correspond roughly to elements physically present in RF components. Another form in which the general network is often visualized is the transformer shown in Fig. 7b. The usefulness of this arrangement as a mental tool is limited by the fact that it can not be applied in the useful practical cases where one of the driving-point impedances becomes zero, or the coupling exceeds the amount corresponding to K = 1.

It follows from equations (6) that if a load impedance Z_2 is connected to terminals 2, the corresponding input impedance seen at terminals 1 is:

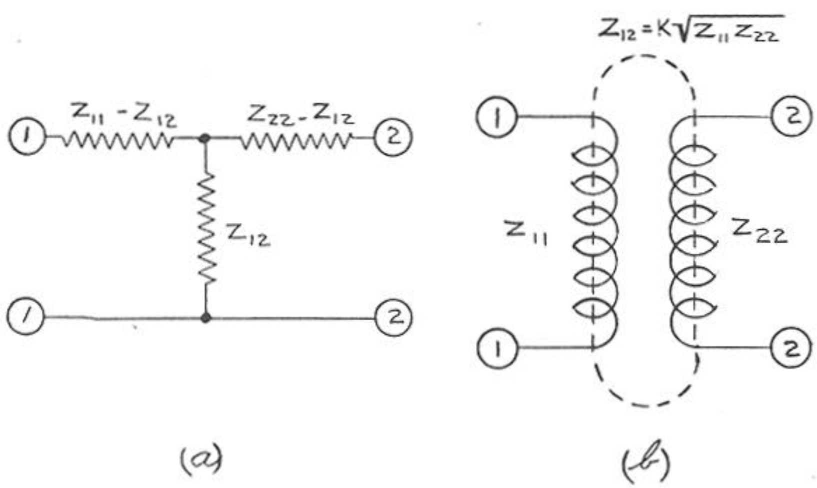
$$Z_1 = Z_{11} - \frac{Z_{12}^{\,2}}{Z_{22} + Z_2} \qquad (7)$$

This is the most general way in which one impedance can depend on another impedance, and it is well to study this relation in some detail. Mathematically, equation (7) expresses a functional relationship between two complex quantities Z_1 and Z_2, involving three arbitrary parameters. Z_1 is an analytic function of Z_2, which means that if the real and imaginary parts are given by $Z_1 = R_1 + jX_1$, $Z_2 = R_2 + jX_2$ then the following useful



CURRENT AND VOLTAGE CONVENTIONS FOR A GENERAL FOUR-TERMINAL NETWORK.

FIGURE 6



FORMS IN WHICH A FOUR-TERMINAL NETWORK MAY BE VISUALIZED.

FIGURE 7

A1386

LOCUS OF NORMALIZED LOAD IMPEDANCE FOR CONSTANT RESISTIVE OR REACTIVE COMPONENT IN INPUT IMPEDANCE IN A FOUR-TERMINAL LOSSLESS NETWORK.

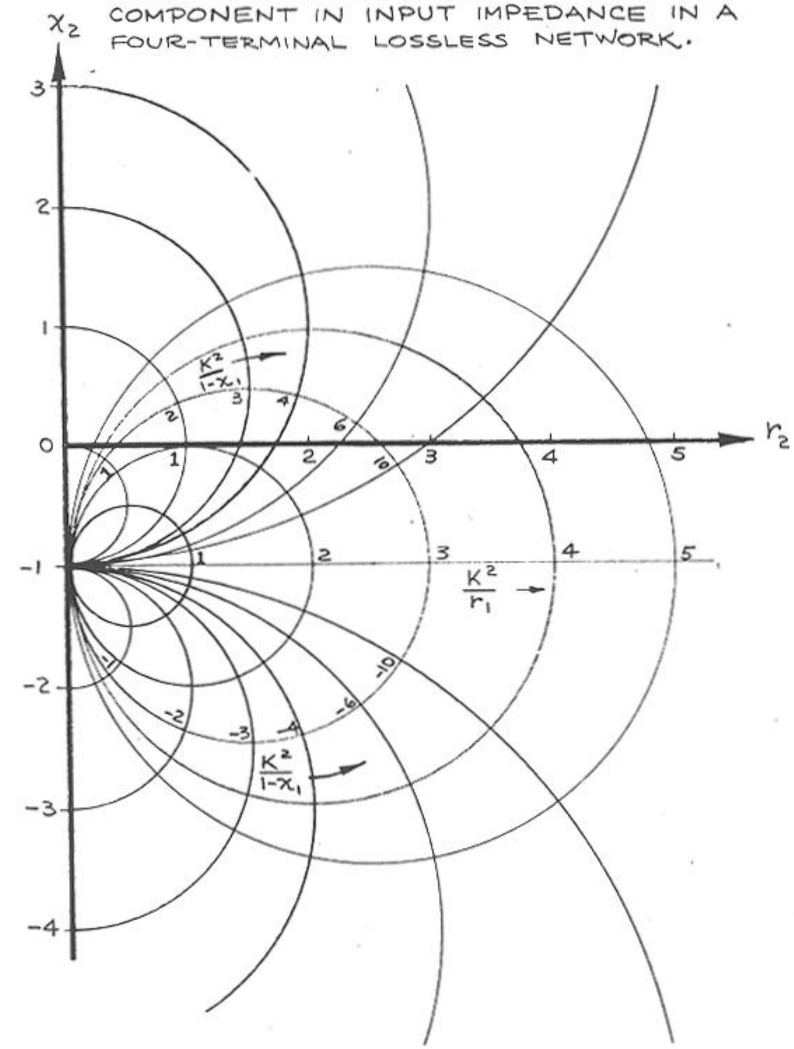


FIGURE 8

A1386

relations must hold:

$$\frac{\partial R_1}{\partial R_2} = \frac{\partial X_1}{\partial X_2} \quad , \quad \frac{\partial R_1}{\partial X_2} = -\frac{\partial X_1}{\partial R_2} \tag{8}$$

In the theory of functions of a complex variable, equations (8) are known as the Cauchy-Riemann differential equations, and they constitute a necessary and sufficient condition that a unique complex derivative $\frac{dZ_1}{dZ_2}$ exists. Geometrically, the operations performed on Z_2 in the rectangular complex plane to arrive at Z_1 are translation, inversion, change of scale with rotation, and another translation. Since each of these operations is one that preserves a circle, it is evident that if the locus of Z_2 is a circle, that of Z_1 is also a circle. If the network is lossless, its three parameters are pure reactances and the rotation cannot occur, while the translations are in the vertical direction only.

In order to visualize the relation between input impedance and load impedance of a network graphically, it is convenient to eliminate as many separate quantities from equation (7) as is possible without losing generality. The standard way of doing this is to normalize the input and load impedances with respect to some impedances defined by the network; in other words, to think of impedance ratios rather than actual impedances. This can be done in several different ways, of which two are particularly useful for microwave applications, for reasons which will appear presently.

The mathematically simplest normalization scheme is to normalize the input and load impedances with respect to the corresponding open-circuit driving-point impedances Z_{11} and Z_{22}. Rearranging equation (7) in this way, we have:

$$\frac{Z_1}{Z_{11}} = 1 - \frac{K^2}{1 + \left(\frac{Z_2}{Z_{22}}\right)} \tag{9}$$

where $K^2 = \frac{Z_{12}^2}{Z_{11} Z_{22}}$ is the complex coefficient of coupling.*

*If the network happens to be an ordinary transformer, this reduces to the familiar coefficient of coupling given by $K^2 = M^2/L_1L_2$. However, in general K is complex and is not restricted to absolute values less than unity. For example, if the network is a lossless transmission line, we have $K = \sec\theta$, where θ is the electrical length of the line, and therefore $K \geq 1$.

The transformation between the normalized impedances is seen to depend on only this one property of the network, instead of three. When we are concerned with lossless networks, it is convenient to normalize with respect to the reactances x rather than the impedances jX. In this case if we define:

$$r_1 = \frac{R_1}{X_{11}} \qquad x_1 = \frac{X_1}{X_{11}}$$
$$r_2 = \frac{R_2}{X_{22}} \qquad x_2 = \frac{X_2}{X_{22}} \tag{10}$$

we then have:

$$r_1 = \frac{K^2 r_2}{r_2^2 + (1+x_2)^2}, \quad x_1 = 1 - \frac{K^2(1+x_2)}{r_2^2+(1+x_2)^2} \tag{11}$$

Equations (11) may be re-arranged in the following form:

$$\left(r_2 - \frac{K^2}{2r_1}\right)^2 + (1 + x_2)^2 = \left(\frac{K^2}{2r_1}\right)^2 \tag{12}$$

$$\left[x_2 + 1 - \frac{K^2}{2(1-x_1)}\right]^2 + r_2^2 = \left[\frac{K^2}{2(1-x_1)}\right]^2$$

These are the equations of two families of circles in the r_2, x_2 plane which are the loci of load impedance for constant input resistance and reactance respectively. They are plotted in Figure 8. The circles of constant input resistance are tangent to the x_2 axis at the point $x_2 = -1$, while the circles of constant input reactance are the orthogonal set. A little study of these circles will enable one to visualize quite clearly how the input impedance varies with load impedance in a reactive network, in a way which does not depend on whether the coefficient of coupling is greater or less than unity.

The other method of normalization brings out an interesting connection with transmission lines, and is capable of very simple geometrical interpretation on the Smith Chart. Unfortunately, however, there is one useful class of network for which this interpretation fails, and for which the corresponding geometry is more complicated. In this method we normalize with respect to the image impedances of the network. The image impedance at terminals 1 may be defined as the geometric mean of the impedances seen looking into terminals 1 when terminals 2 are alternately open and short-circuited. In a symmetrical network the image impedances are equal to the iterative impedance, which is the impedance seen looking into an infinite cascaded chain of similar networks.

In general however, an image impedance is the impedance looking into an infinite chain of similar networks every other one of which is turned end for end so that terminals 2 of one network are always connected to terminals 2 of the next network, etc., while for the iterative impedance the networks are all oriented the same way, so that terminals 2 of one network feed terminals 1 of the next. For a transmission line the image impedances, iterative impedances, and characteristic impedance are one and the same. From equation (7) we find these image impedances to be:

$$Z_{I_1} = \sqrt{Z_{11}\left(Z_{11} - \frac{Z_{12}^2}{Z_{22}}\right)} = Z_{11}\sqrt{1-K^2}$$

$$Z_{I_2} = \sqrt{Z_{22}\left(Z_{22} - \frac{Z_{12}^2}{Z_{11}}\right)} = Z_{22}\sqrt{1-K^2} \qquad (13)$$

We then find that the formula relating the input impedance to the load impedance may be written in the form:

$$\left(\frac{Z_1}{Z_{I_1}}\right) = \frac{\left(\frac{Z_2}{Z_{I_2}}\right) + \sqrt{1-K^2}}{1 + \left(\frac{Z_2}{Z_{I_2}}\right)\sqrt{1-K^2}} \qquad (14)$$

If we compare equation (14) with the well-known expression for the input impedance to a section of transmission line of characteristic impedance Zo, propagation constant γ, length ℓ, and load impedance Z_2, which is

$$\left(\frac{Z_1}{Z_0}\right) = \frac{(Z_2/Z_0) + \tanh \gamma\ell}{1 + (Z_2/Z_0)\tanh \gamma\ell} \qquad (15)$$

we see that they are identical in form if we associate:

$$\tanh \gamma\ell = \sqrt{1-K^2} \qquad (16)$$

so that a network with coefficient of coupling K behaves with respect to its image impedances in the same way that a transmission line of total propagation constant $\gamma\ell = \tanh^{-1}\sqrt{1-K^2}$ behaves with respect to its characteristic impedance. The only difference is that the two image impedances are not in general equal, so that a general network corresponds to a transmission line which has a different "characteristic impedance" at its two ends.

It would be possible to study equation (16) in detail for the general case where K and γ are complex, in order to establish the connection between coefficient of coupling and propagation constant of the equivalent transmission line for all possible cases, but this is not necessary because nearly all networks of practical interest fall into one of the following special types, for which the analysis is greatly simplified:

(1) <u>Reactive Network Tightly Coupled</u>.

The Impedances looking into a pair of terminals with the opposite pair open and then short-circuited are pure reactances of opposite sign, so that the image impedances are pure resistances. K^2 is then real and greater than unity. From equation (16) this means that the equivalent transmission line has a pure imaginary propagation constant, and therefore has no attenuation, but has phase shift corresponding to an electrical length $\beta\ell$ where $\tan\beta\ell = \sqrt{K^2-1}$. This network then corresponds to a section of lossless transmission line, or to a filter in its passband.

(2) <u>Reactive Network Loosely Coupled</u>

The open and short-circuited input impedances are pure reactances of the same sign, so that the image impedances are also pure reactances. K^2 is real and less than unity. Therefore, the equivalent transmission line has no phase shift, but an attenuation of $\alpha\ell$ nepers, where $\tanh\alpha\ell = \sqrt{1-K^2}$. This corresponds to a waveguide below cutoff, or a filter in its cutoff region.

(3) <u>Resistive Network</u>

The open and short-circuited input impedances are pure resistances, so that the image impedances are also pure resistances. K^2 is real and less than unity, so that we again have attenuation $\alpha\ell$ as in case (2) and no phase shift. This corresponds to a resistance attenuator pad.

(4) Symmetrical Network in which Attenuation is Due to Resistances

The two image impedances are equal resistances so that the analogy with a transmission line expressed by equations (14) and (15) is exact in every detail. K2 may have any value, real or complex. This type of network is equivalent to two simpler networks in cascade, of types (1) and (3) respectively. To show that a network of type (4) may be resolved into two networks of types (1) and (3) we note that the quantity γ in equation (15) is a complex number $\gamma = \alpha + j\beta$ where α is known as the attenuation constant, and β is the phase-shift constant. If we expand the hyperbolic tangent in terms of α and β, we find that equation (15) may be split into two equations of similar form by means of which we arrive at the value of (Z_1/Z_0) in two steps

$$\left(\frac{Z'}{Z_0}\right) = \frac{(Z_2/Z_0) + j\tan\beta l}{1 + j(Z_2/Z_0)\tan\beta l}$$

$$\left(\frac{Z_1}{Z_0}\right) = \frac{(Z'/Z_0) + \tanh\alpha l}{1 + (Z'/Z_0)\tanh\alpha l} \qquad (17)$$

Note that each step involves only one of the components of γ. The physical interpretation of equations (17) is that a general transmission line with both attenuation and phase shift is equivalent to two lines of the same characteristic impedance connected in cascade, one having phase shift equal to that of the original line but no attenuation, the other having attenuation equal to that of the original line but no phase shift. The latter network evidently corresponds to a resistance attenuator pad. In addition, the order in which these idealized networks are connected makes no difference; we could equally well have connected the attenuator network to the load, and calculated the input impedance to the lossless network.

The application of the Smith Chart to networks of types (1) or (3) is very simple, and follows from the properties already described. If we locate the load impedance $(Z2/ZI_2)$ on a Smith Chart, as in Fig. 9a, the input impedance (Z_1/ZI_1) to a network of type (1) is represented by a point rotated clockwise through an angle

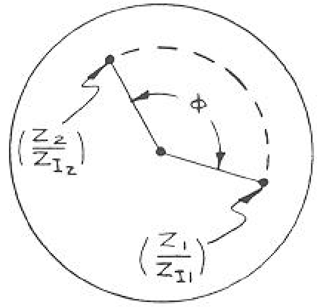
$$\phi = 2\beta l = 2\tan^{-1}\sqrt{K^2 - 1} = 2\cos^{-1}\left(\frac{1}{K}\right) \qquad (18)$$

IMPEDANCE TRANSFORMING PROPERTIES OF THREE TYPES OF NETWORKS.

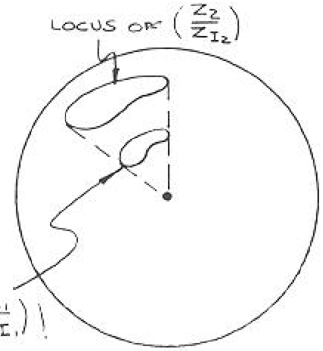
(a) TYPE (1) NETWORK - INPUT IMPEDANCE IS ROTATED THRU AN ANGLE

$$\phi = 2\cos^{-1}\left(\frac{1}{K}\right)$$

CORRESPONDS TO A LOSSLESS TRANSMISSION LINE, ELECTRICAL LENGTH $\beta l = \phi/2$.



(b) TYPE (3) NETWORK - INPUT IMPEDANCE LOCUS IS SCALED DOWN BY A CONSTANT FACTOR OF 2 FOR EACH 3 db ATTENUATION OF THE NETWORK



(c) TYPE (2) NETWORK - LOAD IMPEDANCE $\frac{Z_2}{|Z_{I_2}|}$ IS LOCATED AT 'B'. TO FIND THE INPUT IMPEDANCE $\frac{Z_1}{|Z_{I_1}|}$ COUNT ALONG THE CIRCLE ABC PAST THE NUMBER OF EQUIATTENUATION CIRCLES CORRESPONDING TO THE NETWORK ATTENUATION, LOCATING THE INPUT IMPEDANCE AT D.

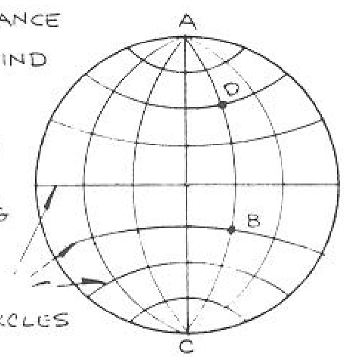
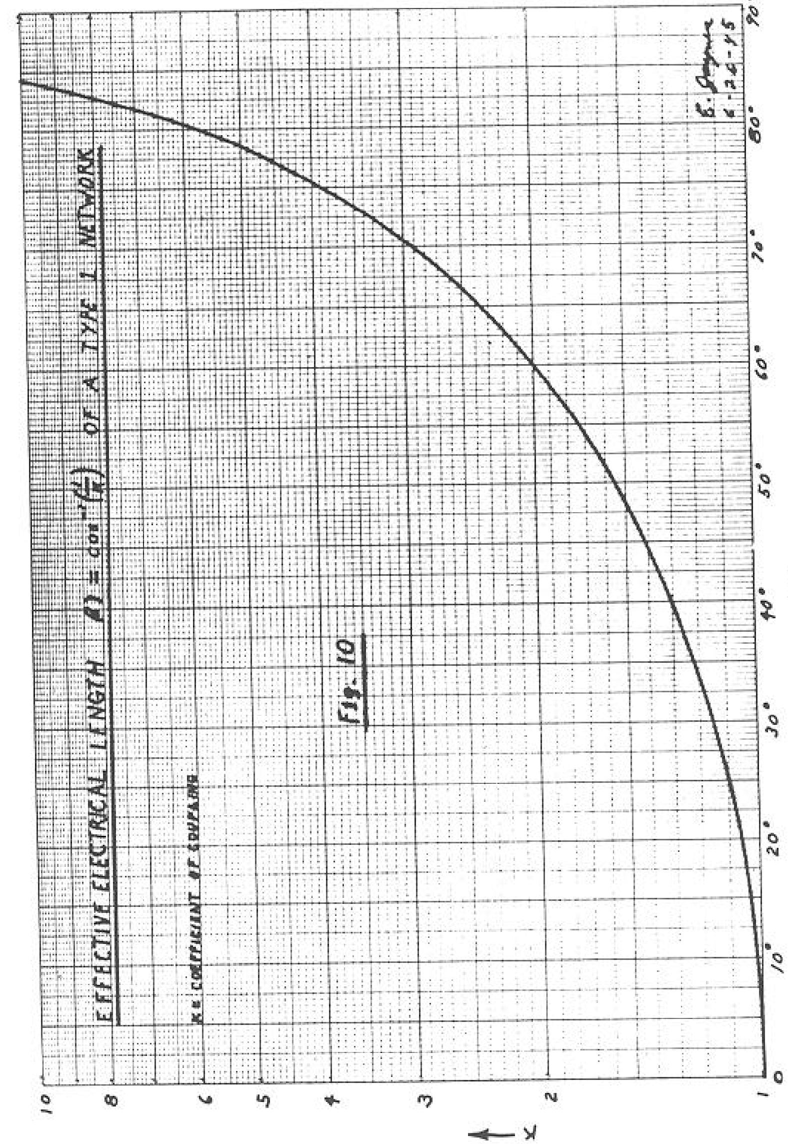


FIGURE 9

A1386



A1386

Values of the effective electrical length $\beta \ell = \cos^{-1}(1/k)$ are plotted in Fig 10.

The behavior of a network of type (3) may be found by introducing the quantities

$$\Gamma_1 = \frac{Z_1 - Z_{I_1}}{Z_1 + Z_{I_1}} \quad , \quad \Gamma_2 = \frac{Z_2 - Z_{I_2}}{Z_2 + Z_{I_2}} \qquad (19)$$

In the case of a transmission line, these quantities are called reflection coefficients and, although the name is not quite so applicable here where the two image impedances are unequal, these quantities still satisfy the same formal relations as the reflection coefficients of a transmission line, and in particular they are still the complex numbers which locate the corresponding impedances on a Smith Chart, measured from its center as an origin. If then we put $\tanh a = \sqrt{1-K^2}$, where a is the attenuation of the network in nepers, we find equation (14) may be written in the form:

$$\Gamma_1 = \Gamma_2 \, e^{-2a} \qquad (20)$$

so that the input "reflection coefficient" is equal to the output "reflection coefficient" reduced in scale by a factor e^{-2a}. This is a factor of two for each 3 db. attenuation of the network, as illustrated in Fig. 9 b.

A network of type (2) is more difficult to handle on a Smith Chart than the other special cases, because its image impedances are pure imaginary. Normalizing impedances with respect to an imaginary impedance is undersirable, not only because of the confusion that would result from the interchange of resistance and reactance, but even more important, some physically realizable impedances lead to values of (Z_2/Z_{I_2}) having a negative real component, and these lie outside the Smith Chart unit circle. However, if impedances are normalized with respect to the absolute magnitude of the image impedances, an interpretation is still possible. This relationship will merely be stated without proof, as it is not often used, the interpretation of Fig. 8 usually being simpler. If the normalized load impedance

$$\frac{Z_2}{|Z_{I_2}|}$$

is located as at B in Fig. 9c, the input impedance

$$\frac{Z_1}{|Z_{I_1}|}$$

lies on the circle ABC, and approaches either A or C, whichever represents ZI_1, as the attenuation $a = \tanh^{-1} \sqrt{1-K^2}$ is increased. In the case shown, the image impedance is inductive and represented by A, so the input impedance is located at point D. The distance DB is determined by counting past the appropriate number of equi-attenuation circles which are orthogonal to ADBE. These families of circles taken together are similar to the circles in the Hemisphere chart of Fig. 5, but rotated 90°. The law by which the equiattenuation circles are spaced is the same as the law by which constant SWR circles on a db. basis are spaced except for a factor of two; that is, calling the horizontal diameter of the Smith Chart the "circle" of zero db attenuation, the 1 db equiattenuation circle is tangent to the 2 db SWR circle , etc.

Use of Admittances

In many problems the use of admittances rather than impedances leads to simplifications in the mathematics and the physical picture, so we will consider briefly how the above results may be applied to them. The fundamental equations for the general network of Fig. 6 on an admittance basis may be written in the form:

$$I_1 = Y_{11} E_1 - Y_{12} E_2$$

$$I_2 = Y_{12} E_1 - Y_{22} E_2 \qquad (21)$$

The admittances Ymn are related to the impedances Zmn by the relations:

$$Y_{11} = \frac{Z_{22}}{Z_{11}Z_{22}-Z_{12}^2}, \quad Y_{12} = \frac{Z_{12}}{Z_{11}Z_{22}-Z_{12}^2}, \quad Y_{22} = \frac{Z_{11}}{Z_{11}Z_{22}-Z_{12}^2} \qquad (22)$$

The inverse relations giving the Zmn in terms of the Ymn are identical in form. The most straightforward way of visualizing a general network on admittance basis is the Pi-section of Fig.11. It is seen that Y_{11} is the admittance seen looking into terminals 1 when terminals 2 are short-circuited; it is therefore called a short-circuit driving-point admittance, just as Z_{11} is called an open-circuit driving-point impedance. If a load admittance Y_2 is connected to terminals 2, the input admittance to terminals 1 is given by:

$$Y_1 = Y_{11} - \frac{Y_{12}^2}{Y_{22} + Y_2} \qquad (23)$$

which is of the same form as the corresponding impedance equation (7).

From equations (22) it follows that the coefficient of coupling is given by equations of identical form in terms of either admittances or impedances:

$$K^2 = \frac{Y_{12}^2}{Y_{11}Y_{22}} = \frac{Z_{12}^2}{Z_{11}\,Z_{22}} \qquad (24)$$

Therefore if we merely substitute Y for Z in equations (9) and (14), we have the corresponding equations for the input admittances relative to the driving-point or image admittances. Since all the network equations are of exactly the same form on an admittance or an impedance basis, all of the graphical interpretations of them are equally applicable to admittances if we substitute conductance for resistance, and susceptance for reactance. Any of the equations in this section may be used as admittance equations whenever it is convenient to do so. A list of the more important equations on an admittance basis is included in Fig. 11.

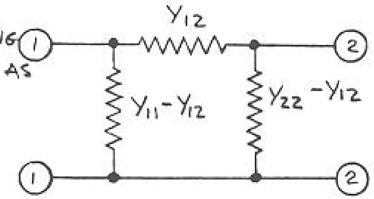
IV. Coupling Systems as Four-Terminal Networks - Choice of an Equivalent Circuit.

When we come to apply the four-terminal network theory to coupling systems, some questions regarding the equivalent circuit present themselves. There is no difficulty in deciding how to represent the equivalent circuit of the connection of a transmission line to a coupling network; we simply connect the two terminals of the line to the corresponding ones of the coupling network. In the connection to a resonant cavity, however, there is more than one way of drawing the equivalent circuit, and it is not immediately obvious whether different representations will lead to fundamental differences in the whole analysis. For example the series and shunt connections of Fig 12 and the "tapped down" connections of Fig 13 immediately suggest themselves, and although in some cases one of these schemes may correspond more closely than the others to the physical arrangement of the parts, one is likely to wonder if perhaps a single analysis would suffice for all cases and if not what he is to do in cases where he cannot decide which equivalent circuit to use.

In order to avoid getting lost in a mass of trivial academic distinctions between these circuits, it is well to recall the principle stated in an earlier section of this report, that a theory is not of much use unless it is mainly concerned with experimentally measurable quantities. Since about the only thing we can readily measure in these circuits is the input impedance to the coupling network for various frequencies and adjustments, we can make no meaningful distinction between two coupling arrangements which have identical input impedance functions. With this in mind, let us examine once more equation (7) which gives the input impedance to a general network:

NETWORK ANALYSIS IN TERMS OF ADMITTANCES

METHOD OF VISUALIZING A GENERAL NETWORK AS A PI-SECTION.



USEFUL NETWORK EQUATIONS

Y_1 = INPUT ADMITTANCE

Y_2 = LOAD ADMITTANCE

$$Y_1 = Y_{11} - \frac{Y_{12}^2}{Y_{22} + Y_2} \qquad (7a)$$

$$\left.\begin{aligned} Y_{I_1} &= Y_{11}\sqrt{1-K^2} \\ Y_{I_2} &= Y_{22}\sqrt{1-K^2} \end{aligned}\right\} \text{IMAGE ADMITTANCES} \qquad (13a)$$

$$\left(\frac{Y_1}{Y_{I_1}}\right) = \frac{\left(\frac{Y_2}{Y_{12}}\right) + \sqrt{1-K^2}}{1 + \left(\frac{Y_2}{Y_{I_2}}\right)\sqrt{1-K^2}} \qquad (14a)$$

$$\Gamma_1 = \frac{Y_1 - Y_{I_1}}{Y_1 + Y_{I_1}}, \quad \Gamma_2 = \frac{Y_2 - Y_{I_2}}{Y_2 + Y_{I_2}} \qquad (19a)$$

(REFLECTION COEFFICIENTS FOR CURRENT VECTORS)

FIGURE 11

A1386

DIRECT COUPLED EQUIVALENT CIRCUITS OF A COUPLING SYSTEM

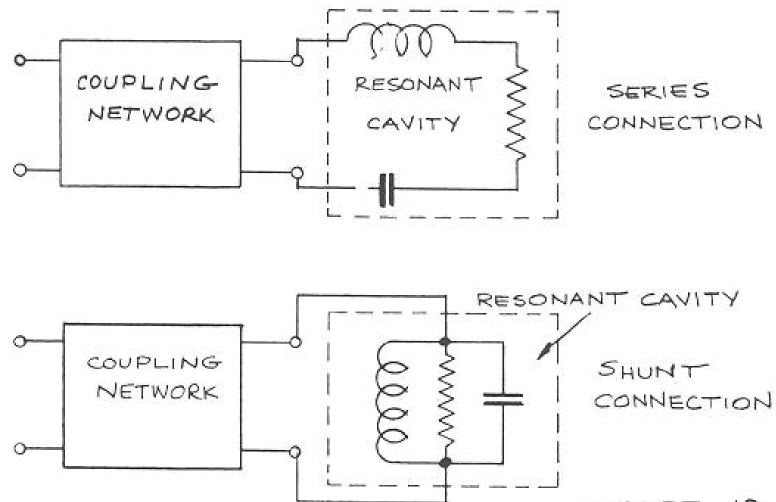


FIGURE 12

"TAPPED DOWN" EQUIVALENT CIRCUITS OF A COUPLING SYSTEM

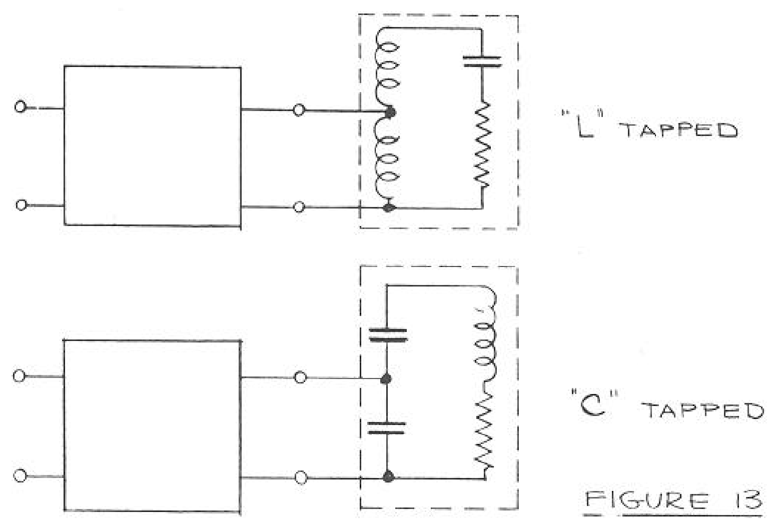


FIGURE 13

A 1386

$$Z_1 = Z_{11} - \frac{Z_{12}^2}{Z_{22} + Z_2} \qquad (7)$$

We see that, as far as the input impedance is concerned, Z22 and Z2 have no separate existence, since it is only their sum which affects the input impedance. Since Z22 has been considered a property of the coupling network, while Z2 depends on the resonant cavity, this is simply a mathematical statement of the obvious physical fact that it is never possible to draw a sharp dividing line between elements that properly belong to the coupling network and those that are part of the cavity, because every coupling system necessarily participates in the resonant action of the cavity to some extent. The mathematical statement goes further, however, and tells us that we may freely transfer elements between the coupling network and the cavity in our equivalent circuit, because this changes only the way we draw our circuits, and does not affect the equations.

With these principles in mind, we see that we may immediately eliminate the "Tapped down" cases of Fig. 13, if we consider that the part of the inductance or capacitance included between the terminals of the coupling network may just as well be considered to be part of that network, to which we have as yet assigned no specific properties. If this is done, both of the cases of Fig. 13 reduce to the series connected circuit of Fig. 12, and we have only to compare the series and shunt arrangements. These in turn may be shown to be equivalent * by the same reasoning; we may, by virtue of the generality of the coupling network, consider the inductances and then the capacitances of the two representations to be part of that network, leaving us in either case with nothing but a general network with a resistance load. One is inclined to object at this point, saying that we have generalized our cavity and problem out of existence, but each step is justified according to the above principles.

---

*Note that we are not saying that a given network may be connected in several different ways to a resonant circuit with the same behavior in each case - that is obviously untrue. What is shown is that we may regard any coupling scheme as represented by either a shunt or a series connection to the cavity if we choose the parameters of the coupling system properly for each case. When this is done, any representation leads to the same behavior in terms of measurable properties.

Perhaps it will help to offer a specific illustration of the way this equivalence works; let us compare the actual equations for the series and shunt circuits and verify directly that their behavior is the same.

In the series case, the load impedance connected to the network due to the cavity is given by

$$Z_2 = R' + j\omega L + \frac{1}{j\omega C} = R' \left[1 + jQ\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right)\right]$$

where

$$\omega_0 = \frac{1}{\sqrt{LC}}, \quad Q = \frac{\omega_0 L}{R} = \frac{1}{R'}\sqrt{\frac{L}{C}}$$

If we assume $Q >> 1$, then the frequency range in which significant impedance variations will occur is small, and we may use an approximate formula:*

$$\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega} \cong 2\,\frac{\omega - \omega_0}{\omega_0} = 2\,\delta \tag{25}$$

where δ is the fractional frequency deviation from ω_0. The load impedance may then be written as:

$$Z_2 = R'\,(1 + j2Q\delta) \tag{26}$$

Assuming that the coupling network is composed of pure reactances, we have:

$$Z_{11} = jX_{11},\ Z_{12} = jX_{12},\ Z_{22} = jX_{22}$$

so that the input impedance to the network is then:

$$Z_1 = jX_{11} + \frac{X_{12}^{\,2}}{jX_{22} + R'\,(1 + j2Q\delta)}$$

the measurements which we can make on the system are to detune the cavity and note the input impedance, then follow its variation

---

* This approximation is made here to save algebra and to introduce a notation which will be used throughout the remainder of this report. It does not affect the equivalence of different circuits, which is the main result being sought.

as we pass through resonance. When the cavity is detuned, the term $2Q\delta$ becomes extremely large compared to all other quantities in the expression

$$\frac{X_{12}^2}{jX_{22} + R' (1 + j2Q\delta)}$$

so that this term vanishes in the limit of complete detuning. The input impedance is then simply $Z_1 = jX_{11}$. As the cavity is tuned through resonance, which varies δ in equation (27), the input impedance varies along a circle as shown in Fig. 14.

In the shunt connected circuit, the load impedance of the network due to the cavity is given by:

$$Z_2 = \frac{R''}{1 + j2Q\delta} \qquad (28)$$

where R'' is the cavity shunt impedance, and the other symbols have their usual meanings. The input impedance to the network is then:

$$Z_1 = jX_{11} + \frac{X_{12}^2}{jX_{22} + \dfrac{R''}{1 + j2Q\delta}} \qquad (29)$$

this expression appears to differ considerably from equation (27) but let us study it more closely. In this case the input impedance when the cavity is detuned ($2Q\delta \rightarrow \infty$) is not jX_{11}, but instead has the value

$$j\left(X_{11} - \frac{X_{12}^2}{X_{22}}\right) \qquad (30)$$

In order to bring out similarities between equations (27) and (29) we should first rearrange (29) into a form where the expression (30) corresponds to X_{11} in equation (27). We find that the following equation results from this rearrangement:

$$Z_1 = j\left(X_{11} - \frac{X_{12}^2}{X_{22}}\right) + \frac{X_{12}^2}{-jX_{22} + \dfrac{X_{22}^2}{R''}(1 + j2Q\delta)} \qquad (31)$$

Comparing this with equation (27), we see that they are identical in their variations with respect to δ, the tuning of the cavity. The impedance locus accordingly is exactly like that of the series connected circuit of Fig. 14. There is a certain reactance present when the cavity is detuned; this is called X_{11} in one case and $X_{11} - X_{12}^2/X_{22}$ in the other, but that does not mean that the physical reactances are different, for we choose different values of the arbitrary parameters X_{11}, X_{12}, X_{22}, when we change equivalent circuits. As the cavity is tuned through resonance, the impedance in both cases varies along a circle like that in Fig. 14. The presence of (jX_{22}) in one formula and $(-jX_{22})$ in the other is not significant, as X_{22} is an arbitrary reactance which can be of either sign.

The question of the exact resonant frequency of the system will not be discussed here, as we do not yet have in our possession all of the facts necessary for the most useful definition of what we mean by a resonant frequency.

To sum up, it has been demonstrated that, given a physical cavity with its coupling system and the experimental data on the input impedance to the coupling system as a function of cavity tuning, this experimental data can be accounted for on the basis of a general four-terminal network connected to the cavity by either the series or the shunt equivalent circuit, and therefore we are at liberty to choose whichever representation is most convenient. Incidentally, if we had analyzed the shunt case on an admittance basis instead of with impedances, we would have arrived at the desired result with far less algebra, and therefore the shunt representation would be preferred in problems worked out on an admittance basis, while the series arrangement is simpler when working with impedances.

There is one way, however, in which the relation between coupling network and cavity may be represented with a greater economy of elements in the complete equivalent circuit, and therefore in which the analysis requires fewer quantities than for other arrangements. The cavity has been represented by two reactances $\pm X$ and a resistance, and the coupling network contains a parameter X_{22} which we have shown to be experimentally indistinguishable from the cavity impedance. We can see from equation (27) that if we combine X_{22} and the cavity impedance into a single impedance of the form $R(1 + j2Q\delta)$ the effect is merely to shift slightly the zero-point of δ, but this "detuning" changes only our equations, and is not experimentally observable. We now have to reapportion this combined impedance between the coupling network and the cavity in whatever way leads to the simplest equations in what follows. It will be found that the most convenient method is simply to place the new $X_{22} = X$, the reactance of one of the resonant elements of the cavity. In

VARIATION OF INPUT IMPEDANCE TO A NETWORK COUPLED TO A RESONANT CAVITY

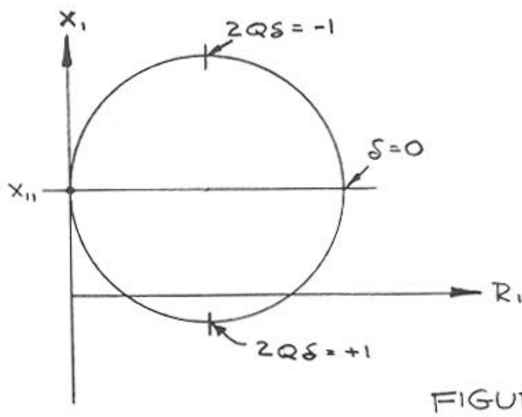
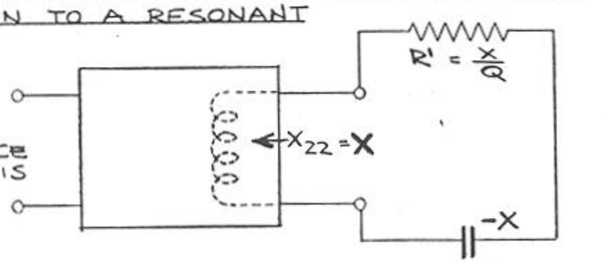


FIGURE 14

SIMPLIFIED EQUIVALENT CIRCUITS FOR THE CONNECTION TO A RESONANT CAVITY

(a) IMPEDANCE ANALYSIS



(b) ADMITTANCE ANALYSIS

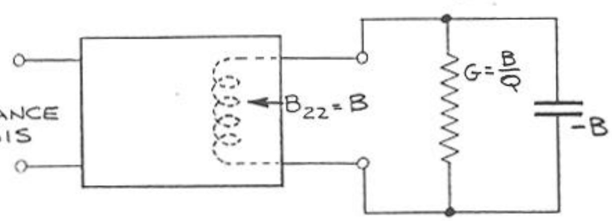
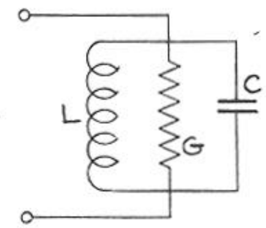


FIGURE 15

A1386

(a) PARALLEL RESONANT CIRCUIT



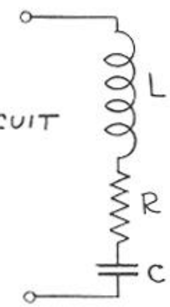
IMPEDANCE ACROSS TERMINALS IS:

$$Z = \frac{1}{G + j\omega C + \frac{1}{j\omega L}} = \frac{(1/G)}{1 + jQ\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right)} \cong \frac{(1/G)}{1 + j2Q\delta}$$

WHERE:

$$\omega_0 = \frac{1}{\sqrt{LC}}, \quad Q = \frac{\omega_0 C}{G} = \frac{1}{G}\sqrt{\frac{C}{L}}, \quad \delta = \frac{\omega - \omega_0}{\omega_0}$$

(b) SERIES RESONANT CIRCUIT



IMPEDANCE ACROSS TERMINALS IS:

$$Z = R + j\omega L + \frac{1}{j\omega C} = R\left[1 + jQ\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right)\right] \cong R(1 + j2Q\delta)$$

WHERE:

$$\omega_0 = \frac{1}{\sqrt{LC}}, \quad Q = \frac{\omega_0 L}{R} = \frac{1}{R}\sqrt{\frac{L}{C}}, \quad \delta = \frac{\omega - \omega_0}{\omega_0}$$

FIGURE 16

A1386

other words, <u>the driving-point impedance seen looking from a cavity into any coupling network may always be put equal to one of the reactances X of the cavity equivalent circuit</u>. The suggested ways of representing the equivalent circuit of the connection for problems on an impedance or admittance basis are shown in figures 15a and 15b respectively. Thus, we have eliminated one of the quantities from our analysis without losing generality, by the choice of an equivalent circuit which involves only as many elements as the system has independent physical properties. The advantages of this equivalent circuit will become quite evident in later developments.

V. Impedance Relations in a Transmission Line Coupled to a Cavity-Definition of Resonance in the Overall System.

In this section we will consider some simplifications in the analysis of coupling systems which become possible when energy is fed into a system through a distributed-constant transmission line. The term "transmission line" is used here in the broadest sense, including all transmission systems in which an impedance may be meaningfully defined, such as two-wire lines, coaxial lines, and waveguides in regions where only one wave type exists. In all such systems there is more or less of an ambiguity in deciding just where the boundary separating the transmission line from the coupling system should be located. Because of the generality of the coupling system, it may be considered to include an arbitrary length of transmission line, and the parameters of this overall coupling system will depend on how much line is included. Since our choice of a dividing line can obviously make no difference in the physical behavior of the system, it is apparent that we should choose it where it will be most convenient mathematically, without regard to the physical construction of the system. This may be done as follows. When the cavity is detuned, we have seen that the input impedance to a lossless coupling system is a pure reactance. Now whatever the value of this reactance, it will be transformed by the transmission line so that at certain points separated by a quarter wavelength it will appear as alternately a short circuit and an infinite impedance. At these points, as we might expect, the form of the impedance variation as the cavity is tuned through resonance is particularly simple, and it will be convenient to consider one of them as the boundary between the transmission line and the coupling network. Which of these points is chosen will depend on the individual problem.

It is assumed in this section that the coupling network is not resonant; that is, that its parameters vary only slowly with frequency, so that they may be considered constant over the narrow frequency range in which the impedance variations of a high-Q cavity occur. The impedance developed along the transmission line will then depend only on the difference between the frequency of measurement and the resonant frequency of the system, so that the same resonance curve is traversed by varying the frequency as by tuning the cavity. Experimentally it makes little difference whether the frequency or the cavity tuning is varied if the cavity Q is greater than about 50 and the network is simple. However, in some situations, such as when a cavity is very heavily loaded or when several wavelengths of line are between the cavity and the point at which impedances are measured, this assumption becomes doubtful. In such cases corrections may be applied to the data, in ways to be considered later.

If we use the series circuit of Fig. 15a for the connection between coupling network and cavity, as explained in the preceding section, the input impedance to the coupling system reduces to:

$$Z_1 = jX_{11} + \frac{X_{12}^{2}}{R(1 + j2Q\delta)} \qquad (32)$$

At a point where the impedance looking toward the detuned cavity is a short circuit, we evidently have $X_{11} = 0$, so that the input impedance is simply:

$$Z_1 = \frac{X_{12}^{2}}{R(1 + j2Q\delta)} \qquad (33)$$

this is exactly the impedance of the parallel resonant circuit of Fig 16a, with resonant shunt impedance equal to X_{12}^2 R, at a "frequency"

$$\varepsilon = \frac{\omega - \omega_0}{\omega_0}.$$

Therefore all impedance measurements made on the transmission line are the same as if a lumped constant parallel resonant circuit were connected as a termination on the line at a point where a short circuit occurs looking toward the detuned cavity. This lumped-constant circuit has, of course, the same Q as the cavity, since Q for both is determined in terms of energy storage and dissipation in the same system. It will be called a virtual resonant circuit, by analogy with the virtual images used in optics.

It is not quite correct to say that the virtual circuit has the same resonant frequency as the cavity, since the latter has not yet been defined; in fact the question of the resonant frequency of a cavity has been carefully avoided thus far. At first sight it might seem that there is no difficulty in finding an intrinsic resonant frequency for a cavity. If it is of simple shape, then a resonant frequency may be calculated from Maxwell's equations, assuming smooth, unbroken walls and therefore no coupling system. If the shape is not simple, the calculation would be impractical, but the principle would be unchanged. If then a coupling system is added, one might be able to calculate a new resonant frequency according to some criterion of resonance, and say that the difference represents the amount that the cavity is "detuned" by the coupling system. This detuning would, of course, depend on the impedance connected to the other side of the coupling network. This interpretation may be useful at low frequencies, or when the coupling is very loose, but when a microwave resonant cavity is tightly coupled to some other circuit, the necessary physical modifications in the cavity and the change in internal field distributions are so great that it can no longer be called the same cavity. In that case, the concept of an intrinsic resonant frequency of the cavity becomes meaningless, and we can only talk about the resonant frequency of the system as a whole.

However, we soon find that even this overall resonant frequency is not very easy to define, unless we know precisely what we mean by the term "resonance". Accordingly, let us try to find some condition upon which to bestow this name that is theoretically reasonable, experimentally useful, and mathematically unique. Our first guess at this condition might be based on equation (32). According to this equation, the impedance at any point along the transmission line starts out at some reactance X_{11} when the cavity is detuned, and moves along a circle as in Fig. 14 when the tuning is varied. When the term δ vanishes, the impedance has traversed exactly half of this circle, and the resistance component of the input impedance is a maximum. Therefore, it would appear that a good definition of resonance is simply the condition $\delta = 0$. The only difficulty with this idea is that the frequency so defined depends on what point along the transmission line the impedance locus is observed, although this is not apparent from equation (32). To show this, consider a coupling system which is so adjusted that the input impedance locus passes through Z_0 at a particular frequency, as shown in Fig. 17. Evidently if the line is matched at one point it is also matched at all other points, so that all impedance loci measured at various positions along the line must pass through Z_0 at the same frequency. Fig. 17 shows two of these circles, one of which is the impedance locus seen at the position of a short circuit when the cavity is detuned. The condition $\delta = 0$ must obviously define two different frequencies for the two impedance circles. This should serve to emphasize the fact that δ as used in equation (32) is not a physically invariant quantity, but merely a convenient linear frequency scale that involves one of the parameters of the general coupling network, and whose zero-point shifts with the length of

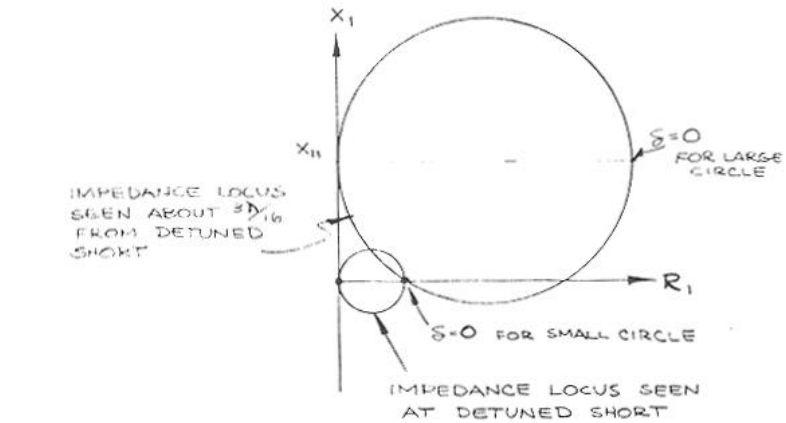


FIGURE 17

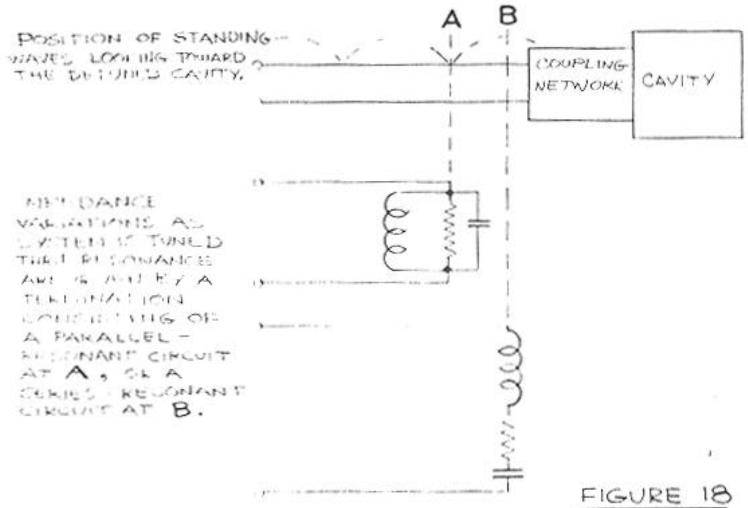


FIGURE 18

A1386

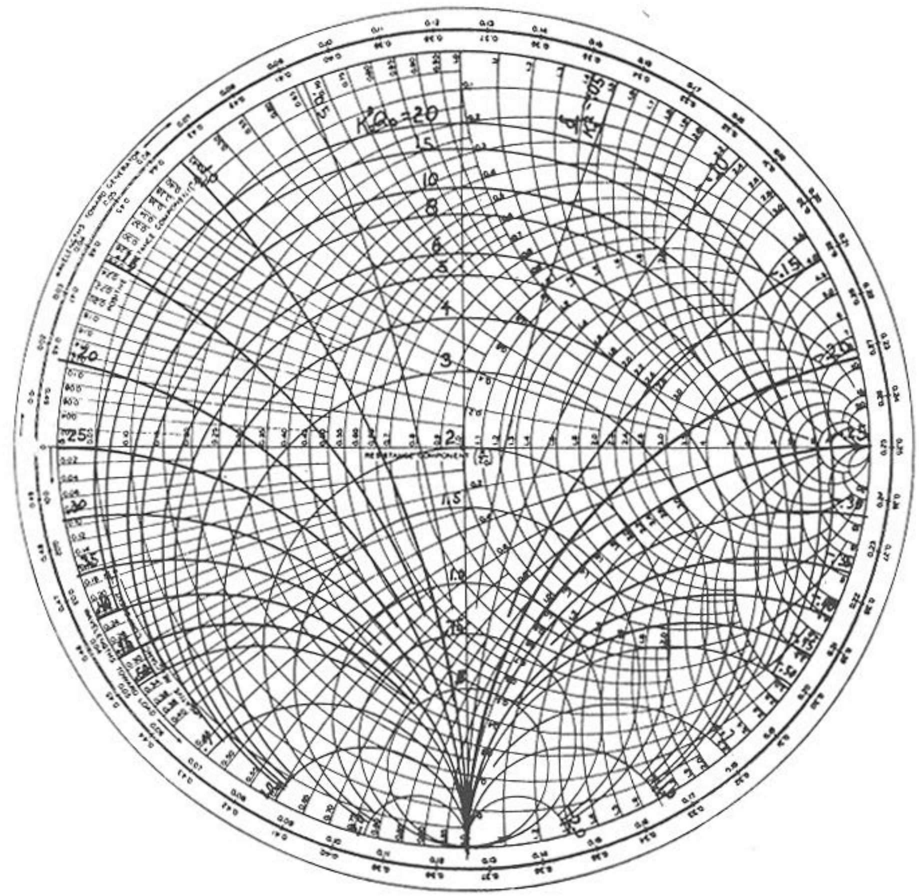


Fig. 19

A1386

E. Jaynes
2-5-45

line that is considered to be included in the coupling network.

There is, however, one unique property of the frequency at which the line is matched in the above example; the power fed into the cavity is a maximum at that frequency, and this condition is true for the impedance locus at any other point on the line, at the frequency where it passes through Z_0. Moreover, this property is universal even when the impedance does not pass through Z_0. The frequency at which the SWR looking toward the coupling system is a minimum is a physical property of the system as a whole, and can not depend on where we consider the boundary between transmission line and coupling system to be located. In addition this frequency has the following properties:

(1) It is the resonant frequency of the virtual parallel resonant circuit which we have imagined to be located at the position of a voltage minimum looking toward the detuned cavity.

(2) It is the frequency at which a charge released on the inside walls of the cavity will oscillate until its energy is dissipated, if the transmission line is matched looking backwards from the cavity. It is thus the "Natural frequency" of the system as seen from the cavity. (This will be proven in the next section.)

(3) It is the frequency at which the position of the voltage minimum on the transmission line looking toward the cavity has either returned to the position of the voltage minimum with the cavity detuned, or is exactly a quarter wavelength away from this position depending on whether the tightness of coupling is less or greater than the critical value which matches the line.

Note that this frequency is truly a property of the entire system; its value depends on the parameters of the coupling system and on the characteristic impedance of the transmission line, as well as on the adjustment of the cavity. It satisfies the requirements of reasonableness, usefulness, and uniqueness, whereas it has been shown that any attempt to define a resonant frequency of the cavity alone leads to serious logical difficulties. Therefore, we shall never speak of the resonant frequency of a cavity, but only of the resonant frequency of the entire system. It may seem that we have devoted a disproportionate amount of discussion to this matter which really amounts to only a few megacycles difference at the most, but it is a very subtle point which has caused much confusion in the past, and has resulted in many unintelligible statements and arguments.

We have seen that the variation of impedance along the transmission line as the system is tuned through resonance is the same as if the line were terminated in a parallel resonant lumped-constant circuit at the position of a voltage minimum looking toward the detuned cavity. In order to avoid repetition we shall hereafter call this point the detuned short. At a point a quarter-wavelength away from this position, the impedance variation is again simple, and may be found quite simply from the well-known transforming property of a quarter-wavelength line. This line transforms an impedance Z into a new impedance (Z_o^2/Z), so that the parallel resonant circuit with an impedance

$$\frac{\beta Z_o}{1 + j2Q\delta} \qquad (34)$$

where β is the ratio of shunt impedance of the parallel resonant circuit to z_o, is transformed into an impedance

$$\frac{Z_o}{\beta}(1 + j2Q\delta) \qquad (35)$$

which is recognized as the impedance of a series resonant circuit. These relationships are illustrated in Fig. 18. The resonant frequencies of these virtual circuits are the same and equal to the resonant frequency of the system as defined above.

VI. Behavior of Coupling Systems as seen from the Cavity; Unloaded, Loaded, and Coupled Q

In the preceding section we have discussed the impedance relations seen in an external transmission line looking toward a resonant cavity; we now reverse our viewpoint and consider the conditions as seen looking out from the cavity into the coupling network. In so doing we are turning our attention to quantities most of which are not directly measurable, so that experimental verification of the results must be done by devious means, and therefore some further attention to the validity and rigor of the equivalent circuits is in order.

The introduction of a coupling system into a cavity involves a certain amount of perturbation of the internal field distributions, which has two significant effects; the resonant frequency must be considered as belonging to the entire system and is in general different from the resonant frequency of the cavity without coupling, and the losses in the cavity are increased by the amount of power fed out through the coupling system. As in the preceding section, if we were to write down the exact formulas for several equivalent circuits we should soon become lost in a maze of intricate equations which owe their intricacy to the inclusion of many trivial higher order effects the detection of which is beyond our experimental techniques. An example of such a trivial effect is the difference between the impedance functions of a parallel resonant circuit when the loss is represented first by a large resistance in parallel with the reactances, then by a small resistance in series with the inductance. If we work out the impedance functions, we find that the impedance of the latter is equal to that of the former multiplied by a factor $(1 - j\frac{1}{Q}\frac{\omega_o}{\omega})$, where the quantities have their usual meanings. This factor differs from unity by less than 1% in almost all practical cases, and the difference usually amounts to about 0.1%. However, even this distinction is artificial, as the resistance actually present in microwave cavities is due to skin effect and varies as the square root of the frequency, a fact which is not apparent from the form of the impedance equations, and which leads to another small uncertainty of comparable size. A standing-wave detector of extremely good mechanical design might enable one to make measurements of standing-wave ratio to an absolute accuracy of about 0.2 db, which means roughly 2% error in the value of an impedance near Z_0 but due to imperfect connectors and mechanical errors in probe position the probable error in measurements is usually several times this amount. Thus we are not concerned here with effects of the order of 1% or less, but instead choose for each type of circuit the impedance function that is mathematically simplest for this degree of accuracy in practical cases. These simplest functions are of the symmetrical forms $(1 + j2Q\mathcal{E})$ and $(1 + j2Q\mathcal{E})^{-1}$ for series and parallel resonant circuits respectively, corresponding to the circuit connections of Fig. 16.

However, it remains to be demonstrated that the various equivalent circuits which were shown to lead to identical results as far as measurements of impedance in the transmission line looking toward the cavity are concerned, also lead to identical results as seen from the cavity. "Identical" as used here means that the resonant frequency as seen from the cavity is identical with the resonant frequency defined in the preceding section, and that the loading seen by the cavity is the same as would be predicted by connecting the line to the virtual resonant circuits of Figure 18. Fortunately, both of these results may be established by very general reasoning that does not involve any approximations of the type discussed above. We have defined the resonant frequency of the system as the frequency at which the SWR in the transmission line looking toward the cavity is a minimum; in other words, the frequency at which the maximum power is transferred between a matched signal generator and the lossy element in the cavity. By the reciprocity theorem, this must also be the frequency at which maximum power transfer takes place in the other direction, from a generator inserted in series with the lossy element to a matched load in the transmission line. Now for either of the circuits of Fig. 15, the resonant frequency would ordinarily be defined as the frequency at which the impedance seen looking out from the terminals of the resistance of the cavity is a pure resistance. But this is just another way of stating the condition just mentioned for maximum power transfer from cavity to load, which takes place when the reactances cancel. Therefore the resonant frequency as seen from either cavity or transmission line is the same.

This result may also be gotten from still more general reasoning involving the second law of thermodynamics instead of the reciprocity theorem. It is well known that all resistive elements generate noise due to thermal motion of the atoms and electrons within them. The amount of this thermal agitation is such that the available noise power from any resistance in a bandwidth ΔF is equal to $kT \Delta F$, where k is Boltzmann's constant* and T is the absolute temperature. When the transmission line is connected to an impedance containing a resistive component and its other end coupled to a cavity at the same temperature, it is necessary in order to preserve thermal equilibrium that the fraction of the available noise power from the cavity that is delivered to the transmission line termination be equal to the fraction of the available noise from this terminating impedance that is delivered to the cavity. Furthermore, by the principle of detailed balancing, this must hold true for each individual frequency component, otherwise the insertion of a filter would upset thermal equilibrium. The equality of these fractions means that the amount of loading of the cavity by the transmission line is exactly equal to the amount of loading obtained by connecting whatever impedance is seen looking back into the transmission line from the virtual resonant circuits of Fig. 18 across their terminals, since the power transfer in the direction from transmission line to these virtual circuits was shown in the preceding section to be identical with the power transfer to the cavity, by virtue of their identical impedance functions.

The above discussion has furnished us with a powerful method of determining loading of a cavity; now that the validity of the virtual resonant circuits of Fig. 18 has been established for power flow in either direction, we may calculate loading or detuning due to the terminating impedance of the transmission line by considering that the virtual circuit <u>is</u> the cavity, to whose terminals the transmission line is connected. The virtual circuit at the detuned short is parallel resonant, and builds up a shunt impedance $R = \rho Z_0$ at resonance. The quantity ρ may be called the SWR at resonance, if it is kept in mind that when it is less than unity, this SWR must be taken as (Emin/Emax) rather than its reciprocal. We now define three different values of Q which are useful in describing the system.

---

* In practical units, $k = 1.37 \times 10^{-23}$ watts per degree per cycle.

In general the Q of any system may be defined by:

$$Q = 2\pi \frac{\text{Energy stored in system}}{\text{Energy dissipated per cycle}} = \frac{\omega \times \text{Energy stored}}{\text{power dissipated}} \qquad (36)$$

The three different values of Q are concerned with power dissipation in different elements. They are:

$$\text{"Unloaded Q"} = Qu = \frac{\omega \times \text{Energy stored}}{\text{Power dissipated in cavity walls}} \qquad (37)$$

$$\text{"Coupled Q"} = Qc = \frac{\omega \times \text{Energy stored}}{\text{Power dissipated in load}} \qquad ($$

$$\text{"Loaded Q"} = Q_L = \frac{\omega \times \text{Energy stored}}{\text{Power dissipated in both cavity and load}}$$

In terms of admittances at the detuned short, we have, when the line is matched:

$$QU = \frac{\omega C}{G}, \quad Qc = \frac{\omega C}{Y_o}, \quad QL = \frac{\omega C}{G + Y_o} \qquad (38)$$

where C and $G = 1/\beta Z_o$ are the capacitance and conductance of the virtual resonant circuit and $Y_o = 1/Z_o$ is the characteristic admittance of the line. The following relations are then seen to exist between the different Q's:

$$\frac{Qu}{Q_L} = 1 + \beta, \quad \frac{Qu}{QC} = \beta, \quad \frac{Qc}{QL} = 1 + 1/\beta \qquad (39)$$

These relations are extremely useful, as the quantity β is directly measurable with a standing-wave detector.

If the transmission line is not matched as seen looking back from the coupling system, the loading of the cavity and the resonant frequency of the system as seen from the cavity will in general be altered. If the transmission line places an admittance $G_L + jB_I$ at the detuned short, this admittance is in parallel with the virtual resonant circuit at that point, and we have for the values of Qc and cavity detuning:

$$Qc = \frac{\omega C}{G_L}, \quad \frac{\Delta F}{F_o} = -1/2 \; \frac{B_L}{\omega C} \qquad (40)$$

Where ΔF is the detuning in megacycles and F_o is the resonant frequency when the line is matched. Note the particularly simple form of equations (40). It is only at the position of this virtual resonant circuit that the loading and detuning are so separated that one depends only on the load conductance and the other only on the load susceptance. The general case where the load impedance is measured at an arbitrary point on the transmission line is needed in the design of certain components and will be considered later, but we will first obtain one extremely useful result from equations (40) as an illustration of their convenience when it is possible to use them.

Let us assume that the transmission line is not perfectly matched, so that there is a SWR equal to S (voltage ratio), and furthermore that the exact length of the line is unknown, so that we do not know the actual values of G_L and B_L but only that they lie on a circle of constant SWR = S. We wish to know the maximum amount by which the resonant frequency of the system can differ from its value when the line is matched. These conditions are recognized as the case of an oscillator which is connected to an imperfectly terminated feeder line, with resulting frequency pulling. Since according to equation (40) the detuning depends only on the susceptance placed across the virtual resonant circuit, we need to know the maximum value of susceptance that is reached on the constant SWR circle. This is readily found to be:

$$B_L = \pm \frac{Y_o}{2} (S - 1/S)$$

so that the maximum detuning is given by:

$$\frac{\Delta f}{F_o} = \pm \frac{Y_o}{4\omega C} (S - 1/S) = \pm \frac{(S - 1/S)}{4\, Q_c} \qquad (41)$$

where $Q_c = \omega C/Y_o$ is the coupled Q when the line is matched. This is a well-known formula for the frequency pulling in an oscillator which conforms to certain conditions which make the frequency of oscillation equal to the resonant frequency of the system.

The laws governing loading and detuning in the general case where the point of measurement of load impedance does not coincide with a virtual resonant circuit could be found from equations (4) and the laws of impedance transformation along the line, but it will be more instructive to go back to the general network equivalent circuits of Fig. 15 and work out these relations from first principles. One feature of the equivalent circuits of Fig. 15 is that the coupled Q is given simply by the ratio of reactance to resistance in the impedance seen looking into the coupling network from terminals 2. This is a consequence of our placing X22 = X in the equivalent circuit. If we use the series arrangement of Fig. 15a, the impedance seen looking back into the coupling network may be found from equation (11) by reversing the subscripts, since our notation is symmetrical with respect to the two sets of terminals. The resistive component of this impedance is given by:

$$\frac{R}{X22} = \frac{K^2 \, R_L/X_{11}}{(R_L/X_{11})^2 = (1 + X_L/X_{11})^2}$$

the coupled Q is then equal to X22/R, so that we have the following relation:

$$K^2 \, Qc = \frac{RL}{X11} + \frac{\left(1 + \frac{XL}{X11}\right)^2}{\left(\frac{RL}{X11}\right)} \qquad (42)$$

This formula neglects the reactance coupled into the cavity by the load in comparison to X22, an approximation which is justified if this coupled reactance is such that the detuning relative to the resonant frequency when $Z_L = \infty$ is small. If the coupled reactance is such that this detuning is 1%, which is several times greater than is ordinarily encountered in practice, then the error in equation (42) is 2%, etc.

To find the amount of detuning, we use equation (11) to find the reactance seen looking into the network:

$$X = X_{22} - \frac{K^2 \left(1 + \frac{XL}{X11}\right) X_{22}}{\left(\frac{RL}{X11}\right)^2 + \left(1 + \frac{XL}{X11}\right)^2} \qquad (43)$$

The second term represents the coupled reactance Xc, while X22 is the "normal" reactance when terminals 1 of the network are open-circuited. Now detuning is evidently a relative quantity; it is meaningless to ask how far the resonant frequency of the system is altered by detuning without reference to some frequency which is considered its nominal value. The non-existence of any unique value of this standard frequency in the present case is due to the absence of any transmission line on which values of standing-wave ratio might be measured. To put it differently, one more quantity of the dimensions of an impedance must be specified before this standard frequency can be established. In the case of a transmission line connection this is the characteristic impedance of the line; in general it may be as any arbitrary impedance, and the standard frequency is then the resonant frequency when this impedance is connected to terminals 1. The mathematically simplest standard is obtained by setting this impedance equal to infinity, in which case the impedance seen looking into terminals 2 of the coupling network is simply jX22. The detuning relative to this standard is then:

$$\frac{\Delta f}{f_o} = -1/2 \frac{Xc}{X_{22}} = \frac{K^2}{2} \frac{(1 + X_L/X_{11})}{(R_L/X_{11})^2 + (1 + X_L/X_{11})^2} \qquad (44)$$

This equation bears a certain symmetrical relation to equation (42), and is capable of a simple interpretation on a rectangular impedance chart, but from a practical standpoint it is not the most useful way of setting up a detuning standard. The practical use of any theory consists of substituting certain experimental data into its formulas and thereby calculating other quantities which are less easy to measure directly. The quantities which are most easily measurable at microwave frequencies are impedances, and these are determined with a standing-wave detector which has a certain characteristic impedance. If we take the standard impedance for any coupling system to be the characteristic impedance of the device which we contemplate using experimentally for measurements on that system, we will have a detuning standard from which we can get useful results with the minimum amount of manipulation of the data.

Let us first consider a special coupling system which has X11 equal to Z_o of the standing wave detector. For this network, load impedances normalized with respect to X11 as in equation (44) are identical with the same impedances normalized with respect to Z_o, which is the way the experimental data is taken. In this case, it is found from equation (43) that the detuning relative to the standard resonant frequency defined by Z_o is given by:

$$\frac{\Delta f}{f_1} = \frac{K^2}{2} \left[\frac{(1 + X_L/Z_o)}{(R_L/Z_o)^2 + (1 + X_L/Z_o)^2} - \frac{1}{2} \right] \qquad (45)$$

The graphical interpretation of equations (42) and (45) can be made most simply in terms of a Smith Chart normalized with respect to Z_o. If values of load impedance are plotted on this chart, lines of constant Qc and constant detuning can be drawn, from which these quantities can be read directly. These lines are shown in Fig. 19.* The lines of constant K^2Qc constitute a family of circles tangent to the bottom of the chart, while lines of constant δ/K^2 where $\delta = \Delta F/F_1$ are the orthogonal set. The effect of equation (44) rather than (45) to define detuning would have made the circle labeled ($\delta/K^2 = -0.25$) which extends to the point ($ZL = \infty$) the center from which detuning is measured, with resultant loss of symmetry in the chart.

* Following Page 25

We must now find a graphical method of interpreting our equations for the general case where X11 is not equal to Z_o. The rules for this case may be reasoned out as follows: Connect the standing-wave detector to the cavity through terminals 1 of the coupling system. When the cavity is detuned the impedance looking into the coupling system is a pure reactance jX11, which is transformed by the line into all possible reactances at various other points. In particular there will be a point A, midway between a voltage minimum and a voltage maximum, at which this reactance is equal to Z_o in magnitude; let us assume in order to be definite that we have chosen a point where it is inductive($X11' = Z_o$) If we consider point A to be the boundary between the transmission line and a new coupling system, it is evident that a plot of load impedances seen at this point normalized with respect to Z_o is identical with a plot of impedances normalized with respect to the driving-point reactance X_{11}' of the new network. Therefore, for this particular choice of a boundary the chart of Fig. 19 is again applicable. Let us call the coefficient of coupling of this new network Ko.

Now the values of Qc and detuning relative to Zo for any load impedance seen at point A are physical properties of the system which can not depend on where we consider the boundary between the transmission line and the coupling network to be located. Therefore, if we move this boundary back to its original position B where the detuned reactance of the coupling network is X11 without the prime, the values of Qc and δ for the load impedances seen at this point must still be equal to their values obtained at the point A. The load impedances seen at point B are simply the impedances seen at point A, rotated on the Smith Chart in a clockwise direction (since we are here going away from the load) through an angle equal to twice the electrical length of line between A and B. If the coupling and detuning are unchanged, then the lines of constant coupling and constant detuning for the load impedances at point B must have rotated along with the points representing the impedances. In moving the boundary from A to B, we have gone toward the cavity, and therefore the point on the Smith chart representing jX11 is located by rotating the point $jX_{11}' = jZ_o$ counter-clockwise through the same angle. Since in Figure 19 the point of convergence of the constant coupling and detuning lines was over the point on the Smith Chart representing the impedance ($-jX_{11}'$), it is seen that after the rotations, this point of convergence is over the impedance (-jX11), which gives us our law for the orientation of the grid of constant coupling and detuning lines.

The value of coefficient of coupling associated with these lines after the rotation is still K_o, however, so we need to know the relation between K_o and the coefficient of coupling K of the network whose boundary is point B. This may be found from equation (42) by setting $ZL = Z_o$. The equation then reduces to:

$$K^2\,Qc = \frac{X11}{Z_o} + \frac{Z_o}{X11} \tag{46}$$

at the point A where $X11' = Z_o$, we have

$$Ko^2\,Qc = 2$$

and as we move along the line Qc remains constant, since it is a physical property of the system that does not depend on our choice of a boundary. We therefore find that:

$$K^2 = \frac{Ko^2}{2}\left(\frac{X11}{Zo} + \frac{Zo}{X11}\right) \tag{47}$$

which is the desired law of transformation. A method of measuring K^2 for a network coupled to a cavity may be found from equation (44) which gives the detuning relative to the standard resonant frequency defined by $ZL = \infty$. If we observe the resonant frequency of the cavity through a second coupling system when the one to be measured is alternately open and short-circuited at terminals 1 (XL is ∞ and then zero) the coefficient of coupling is seen to be:

$$K^2 = 2\,\frac{\Delta F}{Fo} \tag{48}$$

where ΔF is the difference of the two resonant frequencies.

The rules for predicting cavity loading and detuning relative to Z_o may then be summed up as follows:

1. Measure X11 of the coupling system with a standing-wave detector, and K^2 as described above.

2. Calculate $Ko^2 = \frac{2\alpha K^2}{1+\alpha^2}$ where $\alpha = \frac{X11}{Zo}$

3. Place a transparent chart of the constant coupling and detuning circles of figure 19 over the Smith chart on which load impedances are plotted, with the centers aligned and the point of convergence of these lines over the impedance (-jX11).

4. For any value of load impedance read off the corresponding values of Ko^2Qc and δ/Ko^2.

This procedure and its explanation may sound quite difficult, for it is very hard to describe in words, but in practice it is found to be very simple, and it is of great importance in designing coupling systems such as crystal mixers where the crystal RF impedances may vary over a known range and it is desired to ascertain what this will mean in terms of receiver performance.

VII. Resonant Cavity Filters

In the preceding sections we have developed the fundamental laws which govern coupling systems, and have studied the behavior of a coupling system between a line and a cavity in great detail. We now make use of these fundamentals in studying a cavity which contains two coupling systems, such as a TR cavity in a receiver, whose function is to protect the receiver against strong signals and to provide a certain amount of RF selectivity. The properties of such a system that are of interest to us here are the bandwidth and insertion loss as functions of the adjustment of the coupling systems, and the method of determining these quantities experimentally.

It will be necessary to adopt a new notation for some of the relevant quantities in order to avoid duplication due to the second coupling system. In the preceding sections, we have used the convention that the terminals of the coupling network that connect to the transmission line or the load impedance are denoted by the subscript 1, while terminals 2 were always on the cavity side. In this section we shall not have occasion to use quantities such as X_{11}, but will work with the more useful derived quantities such as K^2 and Q_c. Therefore we shall use the subscript 1 to denote the input coupling system and 2 to represent the output coupling system. The following quantities may then be used:

Q_u = Unloaded Q of cavity

Q_{c1} = Coupled Q of input circuit

Q_{c2} = Coupled Q of output circuit

Q_{L1} = Loaded Q of input circuit = $Q_u Q_{c1}/(Q_u + Q_{c1})$

Q_{L2} = Loaded Q of output circuit = $Q_u Q_{c2}/(Q_u + Q_{c2})$

Q_{L12} = Total loaded Q due to losses in cavity, input circuit, and output circuit.

$\beta_1 = Q_u/Q_{c1}$ = SWR at resonance looking into the input coupling system with the load disconnected from the output circuit.

$\beta_2 = Q_u/Q_{c2}$

γ_1 = SWR at resonance looking into the input coupling system with the load connected to the output circuit.

The conditions of the problem that are met in practice are usually the following: given a cavity with a certain unloaded Q, a filter is to be made having a bandwidth of Δf megacycles between half-power points. How should the coupling systems be adjusted to attain this bandwidth with a minimum amount of insertion loss, and what is this minimum possible insertion loss? The bandwidth of the system depends on the overall loaded Q, which is denoted here by Q_{L12} while the insertion loss depends also on the particular combination of Q_{c1} and Q_{c2} that is used to achieve this value of Q_{L12}.

The determination of γ_1 in terms of the other parameters depends on the physical reasoning that when the load is connected to the output circuit of the cavity, the effect as seen by the input circuit is the same as if the unloaded Q of the cavity were decreased to Q_{L_2}.* Since the SWR at resonance looking into a coupling system is proportional to the Q of the cavity as seen from the input circuit, we have for γ_1:

$$\gamma_1 = \frac{Q_{L2}}{Q_{c1}} \doteq \frac{1}{Q_{c1}} \left(\frac{Q_u Q_{c2}}{Q_u + Q_{c2}} \right) = \frac{\beta_1}{1 + \beta_2} \qquad (49)$$

The quantity β_2 has been defined simply as the ratio Q_u/Q_{c2}. If the output coupling system feeds another matched transmission line, then β_2 is the SWR at resonance seen looking back from this transmission line into the output coupling system when the input circuit is disconnected, and it is exactly analagous to β_1. However, the load is usually a crystal mixer and there is no point at which such a SWR could exist. In that case β_1 and β_2 are corresponding quantities only in the equations, since β_2 still satisfies the same relations as this hypothetical SWR. In the similar way a quantity

$$\gamma_2 = \frac{\beta_2}{1 + \beta_1}$$

corresponding to γ_1 but not necessarily representing a SWR physically present, may be defined. Ordinarily β_1 and γ_1 are the only quantities which can be directly measured.

The total loaded Q may be found by combining Q_u, Q_{c1}, and Q_{c2}. Since the Q for any part of the system is inversely proportional to the losses in that part and the various losses add directly to give the total loss, the law of combination is found to be:

---

*This assumes that direct coupling between the input and output circuits may be neglected as far as its effect on input impedance is concerned.

$$\frac{1}{Q_{L12}} = \frac{1}{Q_{c1}} + \frac{1}{Q_u} + \frac{1}{Q_{c2}} \qquad (51)$$

or,

$$\frac{Q_u}{Q_{L12}} = 1 + \beta_1 + \beta_2 = \beta_1\left(1 + \frac{1}{\gamma_1}\right) \qquad (52)$$

The bandwidth is then equal to:

$$\Delta F = \frac{F_o}{Q_{L12}} \qquad (53)$$

If we compare this with equation (51) we see that the overall bandwidth is simply the sum of the separate "partial bandwidths" due to the individual sources of loss.

The insertion loss of a cavity filter may be considered due to two separate effects: First, the input impedance to the device at resonance does not match the input transmission line if $\gamma_1 \neq 1$. so that a portion of the incident power is reflected there. Second, not all of the power that enters the cavity is delivered to the load since there is some loss in the cavity walls. From standard transmission line theory we find that the fraction of the incident power that enters the cavity at resonance is:

$$\frac{P_1}{P_o} = \frac{4\gamma_1}{(1 + \gamma_1)^2} \qquad (54)$$

The cavity efficiency, defined as the ratio of power delivered to the load to total power entering the cavity is, since the Q of an element is inversely proportional to the corresponding power loss:

$$\frac{P_L}{P_1} = \frac{1/Q_{c2}}{1/Q_u + 1/Q_{c2}} = \frac{Q_{L2}}{Q_{c2}} = \frac{\beta_2}{1 + \beta_2} = \frac{\beta_1 - \gamma_1}{\beta_1}$$

The overall efficiency is the product of (54) and (55).

$$E_{ff} = \frac{P_L}{P_o} = \frac{4\gamma_1}{(1 + \gamma_1)^2} \cdot \left(\frac{\beta_1 - \gamma_1}{\beta_1}\right) \qquad (56)$$

Since all of the quantities in equation (56) are easily measurable this affords us a convenient way of determining insertion loss of a cavity filter. Equations (54) and (55) are plotted separately in Figs. 20 and 21. It should be noted that although the quantities β and γ always refer to voltage ratios in equations, they

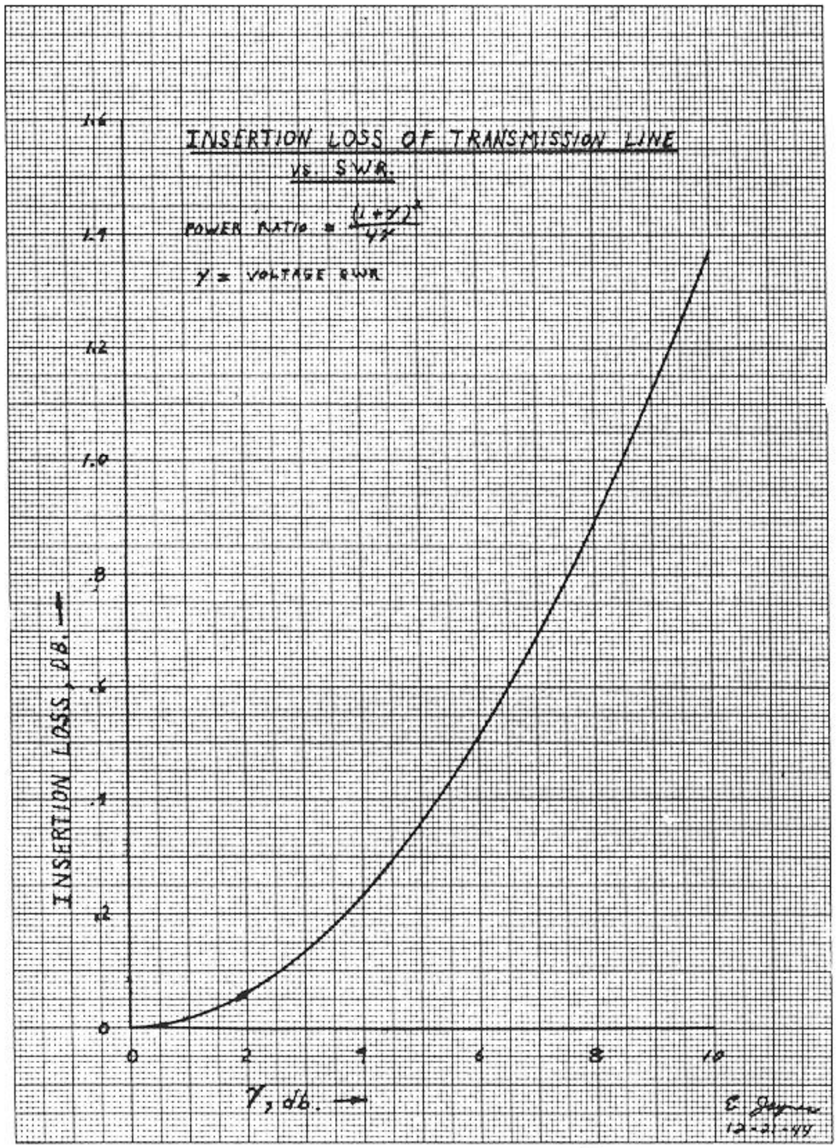


Fig. 20 A1386

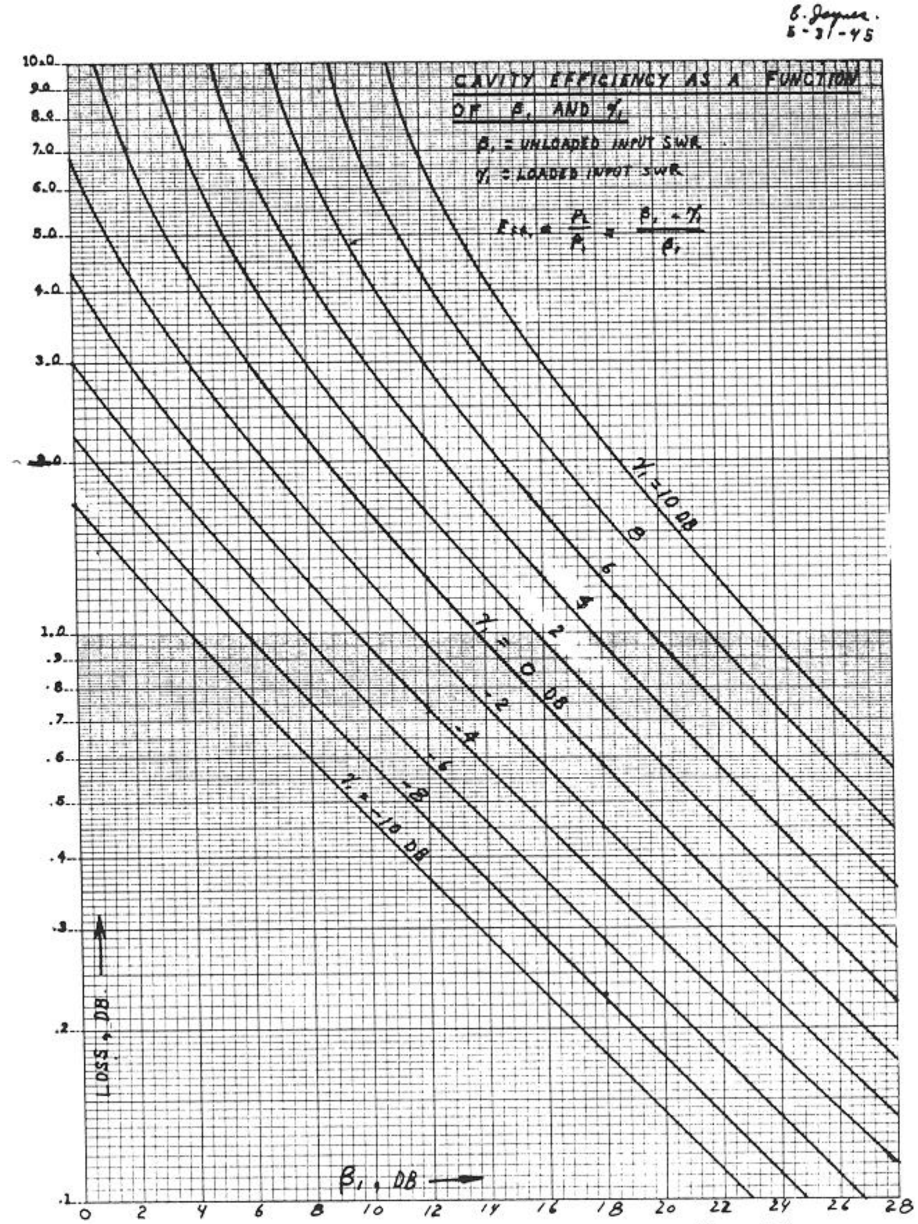


Fig. 21 A1386

are usually measured experimentally in db, and therefore they are expressed in db on these graphs. When the SWR in db comes out negative, this means that $\gamma_1 < 1$ in voltage, or $Q_{L2} < Q_{c1}$. This condition is detected experimentally by the fact that the position of the voltage minimum at resonance is the same as its position when the cavity is detuned, while for positive SWR in db ($\gamma_1 > 1$) the voltage minimum at resonance is a quarter wavelength away from this position. A SWR of zero db means that the input is matched at resonance.

We must now study the equations for insertion loss and bandwidth to find the relation between input and output loading which gives the minimum insertion loss for a given bandwidth. It is seen from equation (52) that the condition for constant bandwidth is

$$\beta_1 \left(1 + \frac{1}{\gamma_1}\right) = \text{Const.} = \frac{Q_u}{Q_{L12}}$$

Solving this relation for γ_1 and substituting the value obtained into equation (56), we have:

$$E_{ff} = 4\beta_1 \left[1 - (1 + \beta_1) \frac{Q_{L12}}{Q_u}\right] \frac{Q_{L12}}{Q_u} \qquad (57)$$

It is found by differentiation that the value of β_1 leading to the greatest efficiency for a given bandwidth is:

$$\beta_1 = \frac{1}{2} \left(Q_u/Q_{L12} - 1\right) \qquad (58)$$

This is the value to which β_1 should be experimentally adjusted, The corresponding value of γ_1 is:

$$\gamma_1 = \frac{Q_u - Q_{L12}}{Q_u + Q_{L12}} = \frac{\beta_1}{1 + \beta_1} \qquad (59)$$

It is seen that this is always less than unity, corresponding to a negative value when expressed in db. When the output coupling system has been adjusted to this value of γ_1, it is found from equation (49) that the value of β_2 is:

$$\beta_2 = \frac{\beta_1}{\gamma_1} - 1 = \beta_1 \qquad (60)$$

The coupling is therefore symmetrical between input and output when this optimum adjustment has been reached.

The efficiency at the optimum adjustment may be found from equations (57) and (58) to be:

$$E_{ff.} = \frac{Q_{L12}}{Q_u}\left[\frac{Q_u}{Q_{L12}} + \frac{Q_{L12}}{Q_u} - 2\right] \qquad (61)$$

This is the maximum possible efficiency for a cavity filter with unloaded $Q = Q_u$ and a bandwidth $\Delta f = f_o/Q_{L12}$. In terms of β_1, this maximum efficiency is:

$$E_{ff} = \left(\frac{2\beta_1}{1 + 2\beta_1}\right)^2 \qquad (62)$$

The following problem often arises in practice: Given a cavity with one of the coupling systems fixed, how should the second coupling system be adjusted in order to minimize the insertion loss, regardless of bandwidth? In case the output coupling is fixed, the distribution of power between cavity losses and load is fixed, so the minimum insertion loss is obtained when we have removed the reflection loss at the input, which means $\gamma_1' = 1$. For this case of matched input, we have the following relations:

$$\beta_1 = 1 + \beta_2$$

$$\gamma_1 = 1 \qquad (63)$$

$$E_{ff} = (1 - 1/\beta_1)$$

$$\text{Bandwidth} = \Delta f = 2\beta_1 \frac{f_o}{Qu}$$

When the input coupling is fixed, we may find the output coupling adjustment which leads to maximum efficiency by differentiating equation (56) with respect to γ_1, keeping β_1 constant. The optimum value of γ_1 is thus found to be:

$$\gamma_1 = \frac{\beta_1}{2 + \beta_1} \qquad (64)$$

When this adjustment has been reached we find the following relations to hold:

$$\beta_2 = 1 + \beta_1$$

$$\gamma_2 = 1$$

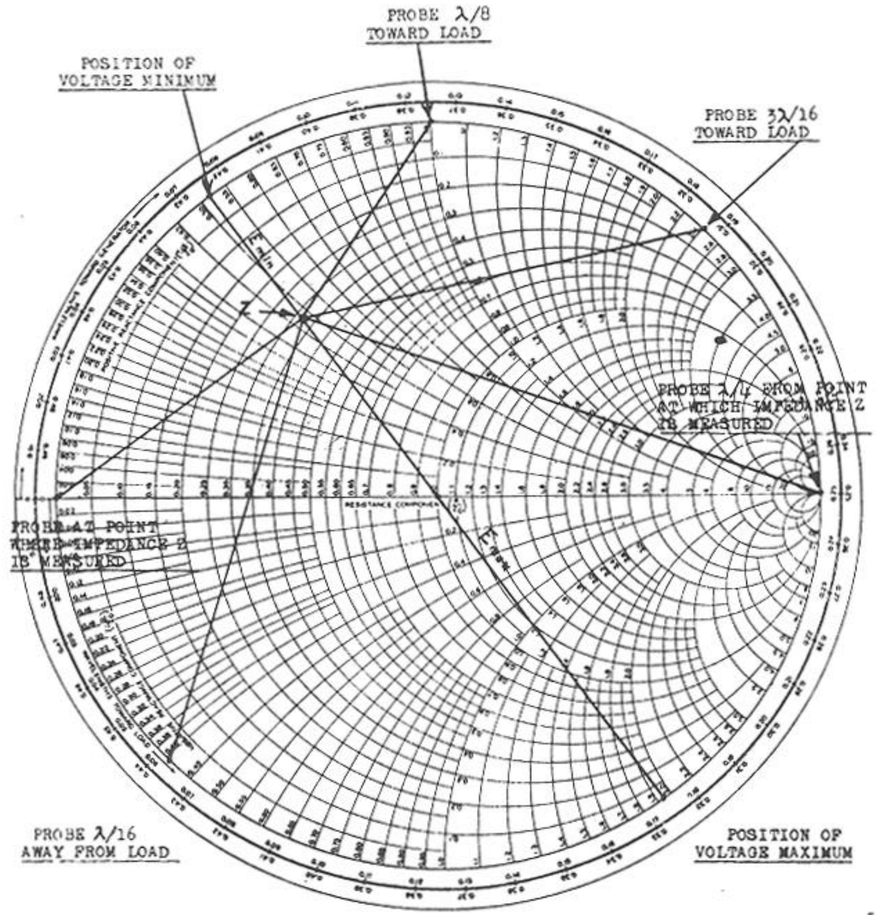
$$E_{ff} = \frac{\beta_1}{1 + \beta_1} = \frac{2\gamma_1}{1 + \gamma_1} \qquad (65)$$

$$\text{Bandwidth} = \Delta f = 2\beta_2 \frac{f_o}{Qu} = 2(1 + \beta_1)\frac{f_o}{Qu}$$

VIII. Experimental Technique for Q Measurements

In this chapter we shall apply the theory developed above to the problem of measuring the values of Qu, Qc, and Q_L when a cavity is coupled to a line. Before studying the available methods, however, we need to establish some conventions for representing the behavior of a standing-wave detector on a Smith Chart.

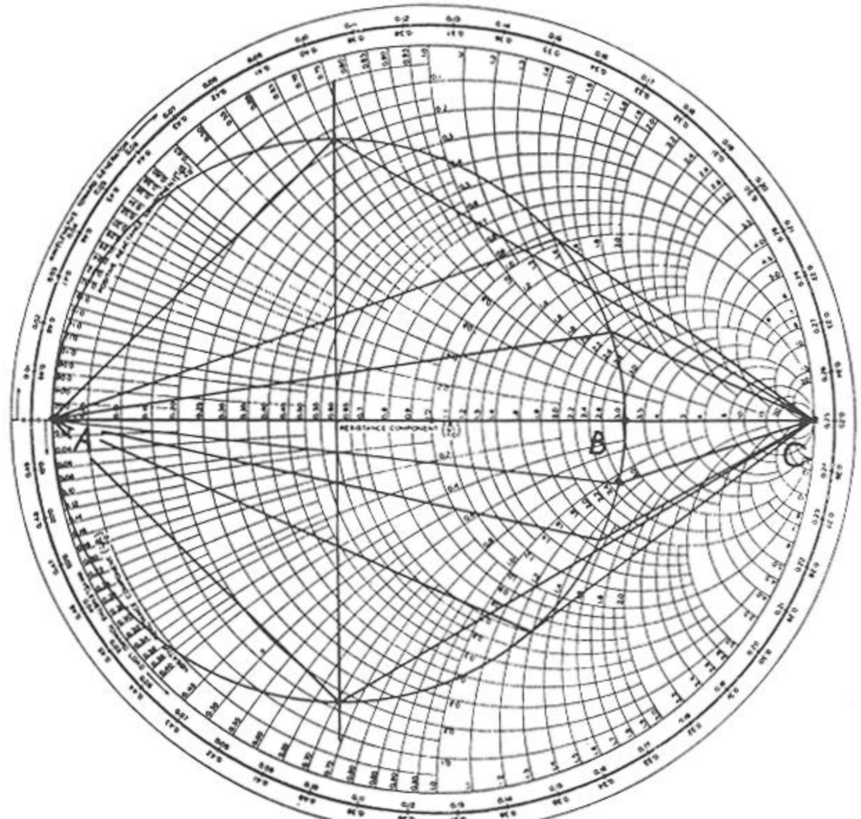
The voltage induced on the probe is, of course, proportional to the total voltage across the line or waveguide at the point where the probe is located. This total voltage is shown in the vector diagram of Fig. 4. For a given incident wave, the voltage at any position along the line is proportional to the vector from the zero impedance point on the Smith Chart to the point representing the load impedance seen at that position. As the probe is moved along the line, this impedance moves along a constant SWR circle, in a clockwise direction as the probe moves away from the load. Since the phase of the voltage induced on the probe is not ordinarily measured or desired, the length of the vector representing the total voltage on the line in Figure 4 is the only thing in which we are interested. The manner in which this length varies with probe position may then be visualized by allowing the point representing the load impedance to rotate as the probe is moved. However, it is evident that as far as the length of this line is concerned, it is immaterial whether the point representing the load impedance rotates in a clockwise direction or the point on the periphery of the Smith Chart to which the total voltage vector is drawn rotates an equal amount counter-clockwise. If we visualize the movement of the probe as a movement of this point, we can keep the impedance plot fixed, so that the impedances are always represented as their values measured at the same position, such as the position of a virtual resonant circuit. This is illustrated in Figure 22, in which the lines whose lengths are proportional to the probe voltage are shown for an arbitrary load impedance Z, for various probe positions. The point on the periphery of the Smith Chart to which this line is drawn will be called simply the probe position in what follows; this should cause no confusion, as it will be apparent from the context whether we are referring to probe position on the Smith Chart or its physical position on the line. The rule for locating the probe position on the Smith Chart is seen to be the following:



ILLUSTRATING CONVENTIONS USED IN INTERPRETING ACTION OF A STANDING-WAVE DETECTOR. The length of each line gives the voltage induced on the probe at the corresponding position, measured from the point on the line at which the load impedance Z is measured.

A1386 Fig. 22

NAVSHIPS (NRL) 131



Illustrating how a system may be tuned to resonance (point B) by tuning for maximum probe pickup with probe at A, the position of a detuned short, or for minimum pickup with the probe a quarter wavelength away, at C. It is seen that the minimum in the latter case is much sharper than the maximum at point A.

A1386 Fig. 23

NAVSHIPS (NRL) 131

The angular position on the Smith Chart measured from the zero impedance point at the extreme left, is equal to twice the electrical length of line between the probe and the reference point at which the load impedance is measured, and the sense is clockwise when the probe is on the load side of this reference point.

The probe position on the Smith Chart is the same whether the point has rotated clockwise through an angle 2θ, or counter-closckwise through an angle (360°-2θ), corresponding to the fact that the voltage magnitude on the line repeats at intervals of a half-wavelength. From diagrams like Fig. 22 one can visualize immediately how the probe voltage varies with the probe position or how the voltage induced on the probe at any position varies with the load impedance as measured at any other position, if the amplitude of the incident wave in the line is not changed by this change in load impedance. (This requires that the generator be matched to the standing wave detector when looking in the other direction.) This configuration is therefore a most powerful mental tool for rapidly predicting and interpreting nearly every type of data which can be obtained by means of a standing-wave detector. Its ability to correlate a large number of factors is a good example of the amazing versatility of the Smith Chart, which makes it invaluable for almost any kind of RF measurements.

Returning to the subject of Q measurement techniques, there are three different principles by means of which this may be done in terms of impedance measurements looking toward the cavity. In ordinary cases these yield the same results with varying degrees of accuracy and ease, but there are some cases, such as when the cavity is loaded on the other side by a resonant element or a non-linear element such as a crystal mixer, when the results by these techniques are not in agreement because they are based on slightly different definitions of Q, which are not necessarily equivalent in the case of non-linear elements. Which technique should be used in any particular case is a matter of individual judgment, and depends on the use which is to be made of the data, the accuracy required, the degree of previous knowledge about the coupling system, and the exact type of equipment available.

The first step in each of these methods is to tune the system to resonance and determine the relative values of Qu, Qc and Q_L by measuring σ and applying equations (39). Since this part of the procedure is standard, we shall now describe it separately, and take up the different methods at the point where they deviate.

In Fig. 23 is shown a Smith Chart on which has been drawn the locus of the impedance seen looking toward the cavity at the position of the detuned short, as the system is tuned through resonance. The virtual circuit at this point is parallel resonant, so that it has zero impedance when it is detuned, and builds up a maximum shunt impedance $Z = \beta Z_0$ at resonance. As the tuning is varied, the impedance locus is a circle, which is shown in Fig. 23 as enclosing the center of the Smith Chart, corresponding to $\beta > 1$. This condition, however, is not necessary for the method to apply. It is seen that there are two ways of tuning the system to resonance. We may set the probe at the position of the virtual parallel resonant circuit, indicated by point A on Fig. 23 and, with the signal generator matched to the standing-wave detector, tune the system for maximum probe voltage. It is apparent from the diagram that the impedance when this condition obtains is represented by point B. Unless the cavity is very loosely coupled to the line ($\beta << 1$) however, it is more accurate to move the probe a quarter-wavelength to position C, and tune for minimum probe pick-up. The minimum thus obtained is quite sharp, as is apparent from Fig. 23 and the principle that the probe voltage is proportional to the length of the line on the Smith Chart between the point C representing the probe position and the point which moves along the impedance locus as the tuning of the system varies. When the system has in this way been brought to resonance, the standing-wave ratio in the slotted line is measured, noting whether the voltage minimum is at the position of the detuned short, or a quarter wavelength away from this position. In the former case, the impedance locus does not encircle the center of the Smith Chart, and we have

$$\beta = \frac{E_{min}}{E_{max}}$$

while in the latter case the coupling is tighter so that the impedance locus encircles the center of the chart, and we have:

$$\beta = \frac{E_{max}}{E_{min}}$$

In any case, β is given by the ratio of probe voltage at the detuned short to the voltage a quarter wavelength away. The relative values of Qu, Qc, and Q_L are then determined by equations (39), which are repeated here:

$$\frac{Q_u}{Q_L} = 1 + \beta \, , \; \frac{Q_u}{Q_c} = \beta \; , \; \frac{Q_c}{Q_L} = 1 + 1/\beta \qquad (66)$$

It is to be emphasized that the signal generator must be matched to the standing wave detector in order that this procedure should be valid. The reason for this is that the lines on the Smith Chart from the probe position to the impedance plot represent the probe voltage relative to the amplitude of the incident wave in the line. It is necessary that the wave reflected from the load be completely absorbed at the generator end of the standing wave detector in order to avoid multiple reflections which would result in a series of waves of progressively decreasing amplitudes all traveling toward the load, whose vector sum would form a new incident wave of different amplitude from the original one. In an extreme case where there is a large mismatch at both load and generator, the standing-wave detector becomes self-resonant for certain frequencies, and very large values of probe voltage may result.

When we speak of tuning the system through resonance, it is understood that the quantity $\delta = \frac{\omega - \omega_o}{\omega_o}$ is being varied. It is usually immaterial whether this is done by changing the signal generator frequency, or the resonant frequency ω_o of the system. If the parameters of the coupling system vary rapidly with frequency, however, (due to partial resonance or because the transmission line between cavity and probe is several wavelengths long), it is preferable to vary ω_o by tuning the cavity, as errors due to the varying coupling system are avoided.

The above procedure will determine only the relative values of Qu, Qc, and Q_L; one more piece of information is necessary in order to find their actual values. The measurement of β determines the size of the circular impedance locus of Fig. 23, and we must now find how rapidly the impedance moves along this circle as the tuning is varied. The general law by which points differing by a constant frequency are spaced along this locus may be found very simply from the basic property of the Smith Chart with which it was introduced in Section II. As illustrated there in Figs. 2 and 3 it was shown that a plot of an impedance Z on a Smith Chart normalized with respect to a resistance R_o is identical with a rectangular plot of the impedance

$$Z_p = \frac{ZR_o}{Z + R_o} \qquad (67)$$

obtained by connecting R_o in parallel with Z, the origin of the rectangular coordinate system being the zero impedance point at the extreme left of the Smith Chart. In the present case we have $R_o = Z_o$ the characteristic impedance of the transmission line. We may express the impedance Z of the virtual parallel resonant circuit as in equation (34) by:

$$Z = \frac{\beta Z_o}{1 + j2Qu\delta} \qquad (68)$$

the impedance Zp of the loaded circuit formed by connecting Z_o in parallel with this virtual circuit is then found to be, using equation (66):

$$Z_p = \frac{ZZ_o}{Z + Z_o} = \frac{Z_o \beta/(1 + \beta)}{1 + j2Q_L\delta} \quad (69)$$

This is, of course, simply the impedance function of a parallel resonant circuit with a Q equal to Q_L. Therefore a plot on the Smith Chart of the load impedance of the virtual resonant circuit involving Qu, is identical with a rectangular plot of the impedance of a resonant circuit having the loaded Q, and the law of spacing of the points on this locus is given by:

$$\tan \phi = - 2Q_L\delta \quad (70)$$

where ϕ is the phase angle of the impedance Zp, as shown in Fig. 24. In particular, the half power points of the loaded circuit are given by $\phi = \pm 45^o$, or:

$$2Q_L\delta = \pm 1 \quad (71)$$

these points are shown in Fig. 24, and are directly above and below the center of the impedance locus circle. Now that we have located the half-power points of Q_L on the Smith Chart, we see that if we can find experimentally the difference ΔF between the signal frequencies which bring the impedance to these positions, we have very simply:

$$Q_L = f_o/\Delta f \quad (72)$$

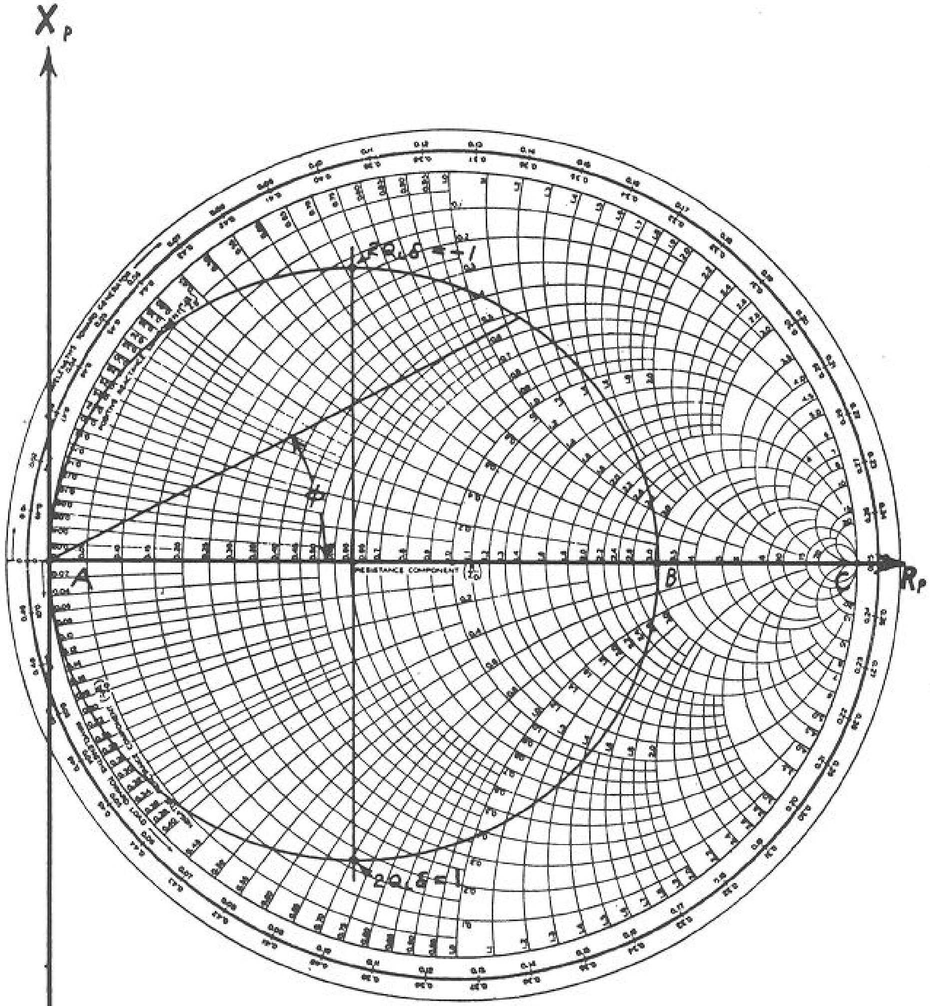
and the values of Qu and Qc may then be found from equation (66) and the measured value of β.

It is at this point that the three methods of Q measurement separate; there are three principles by which one can set the frequency to the half-power points of Q_L and a number of variations and refinements. They depend on observing the change in the position of the voltage minimum, the change in the SWR, and the change in the magnitude of the impedance Zp respectively, as a means of determining position along the impedance locus. The procedures are considered separately below.

1. The "$\Delta\lambda$" Method. - The experimental procedure for this method was described by Lawson* in an early Radiation Laboratory report, but unfortunately a table of values to which the experimenter must refer in the course of the measurement was calculated

---

*Rad Lab Report #64-3, dated May 18, 1942



SUPERPOSITION OF RECTANGULAR COORDINATES AND THE SMITH CHART

in such a way that the same circle represents the impedance locus of both the unloaded and loaded resonant circuit. The half-power points are then located for Q_L as shown.

A1386 Fig. 24

NAVSHIPS (NRL) 131

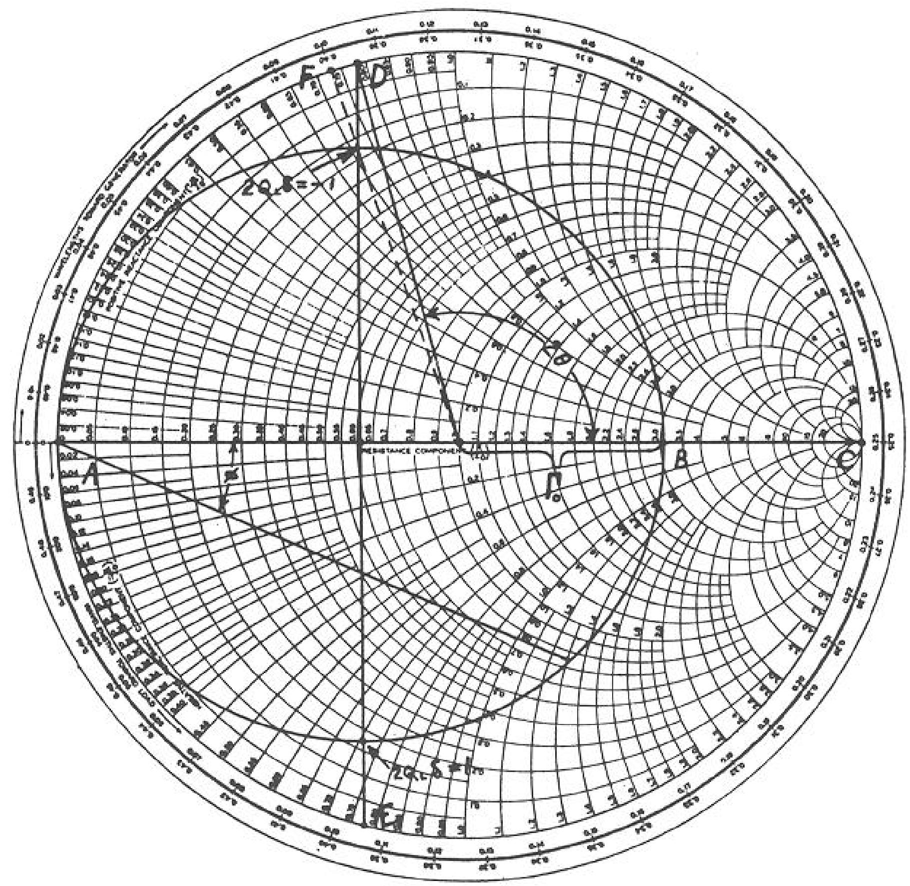


Fig. 25. Illustrating the Q measurement technique of Lawson and Wheeler.

A1386

NAVSHIPS (NRL) 131

by Lawson from an incorrect formula which leads to large errors unless the cavity is tightly coupled to the standing-wave detector. As far as the writer is aware, the exact formula for this method was first given by H. A. Wheeler.

In this method one first measures β as explained above, and from this value determines by means of a previously prepared graph a distance $\Delta\lambda$ through which the probe is to be moved away from position C of Figure 24. The probe is moved this amount in one direction from point C, and the signal generator frequency changed until the voltage "picked up" at the probe is a minimum. (It is again necessary that the signal generator be matched to the line.) This frequency is noted, the probe moved the same distance $\Delta\lambda$ the other side of point C, and the frequency again adjusted so as to minimize the probe pickup. The difference between these frequencies is the bandwidth Δf from which Q_L is calculated using equation (72).

The graphical interpretation of this procedure is shown in Fig. 25. The points where the probe must be located in order that the probe pickup is a minimum at the half-power points of Q_L as the frequency is varied, are evidently points D and E, directly above and below the center of the impedance locus circle. If the number of electrical degrees between these points and point C is θ, the angle on the Smith Chart between them is 2θ, as shown. We may find from Fig. 25 the formula relating $\Delta\lambda$ to β as follows. If the reflection coefficient at resonance (point B) is called Γ_o, we have from the geometry of the system, since the radius of the Smith Chart is unity:

$$\cos 2\theta = \frac{\Gamma_o - 1}{2} \tag{73}$$

From equation (5) we have the relation:

$$\Gamma_o = \frac{\beta - 1}{\beta + 1}$$

substituting this value of Γ_o into equation (73) and reducing by means of various trigonometric identities, the result is:

$$\tan\theta = \sqrt{1 + 2/\beta} \tag{74}$$

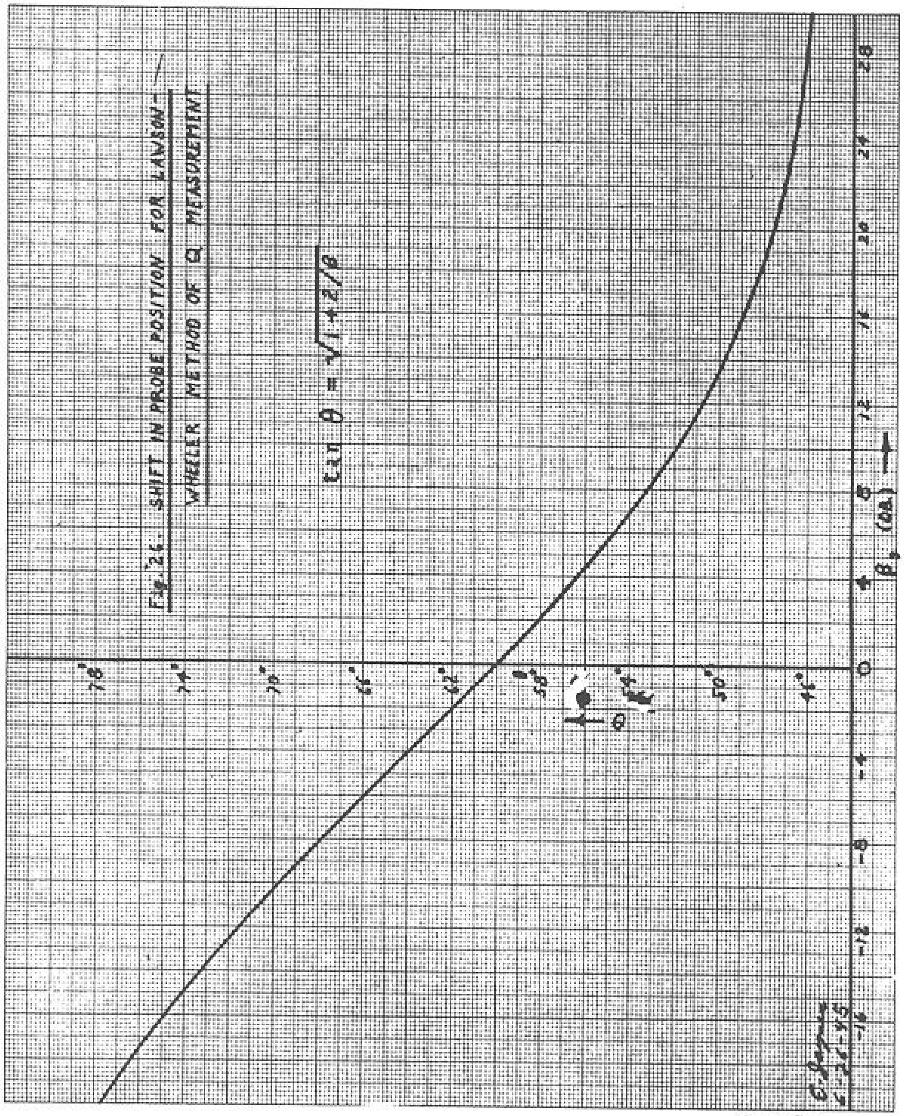
This is the exact formula given by Wheeler. Values of θ vs β are plotted in Fig. 26. The angle θ is related to the physical distance $\Delta\lambda$ through which the probe is moved by the equation

$$= \frac{\lambda\theta}{2\pi} \tag{75}$$

Note that equation (74) gives the probe position for which the probe pickup is a minimum at the half-power points of Q_L as the tuning parameter δ is varied; this is not the same as the probe position at which the pickup is a minimum as the probe position is varied when the system is kept tuned to the half-power points of Q_L. Confusion on this point is quite common, even with people who have used this method almost daily over long periods of time. The difference between these positions is shown in Fig. 25. The correct probe position according to the experimental procedure described is at D, while the position of the voltage minimum on the line is slightly to the left, at point F. The table of values of $\Delta\lambda$ given by Lawson in the above reference locates the probe at position F, resulting in an error that becomes serious when β is less than about 6 db.

Since the half-power points of Q_L are located in this method by tuning to rather sharp minima, the errors introduced by a small amount of generator mismatch, by variation of generator output with frequency, and by resonance in the probe circuit are quite small and can usually be neglected in practical cases. However, the error due to the fact that the electrical length of line between cavity and probe varies with frequency is often not negligible. At the higher microwave frequencies it is difficult to avoid having several wavelengths of line present, so that the position of a detuned short shifts rapidly as the frequency is varied. Now the probe position given by equation (74) is relative to this detuned short (although for convenience the electrical angle is measured from the "detuned open", a quarter-wavelength from the detuned short) so that when this point shifts, the probe should be shifted the same amount in order that the experimental procedure described should locate the true half-power points of Q_L. But a knowledge of the amount to shift the probe to take this into account would imply that Q_L was already known, so that this can not in general be done. However, a formula for the first-order correction to the data obtained neglecting this shift can be derived as follows:

Referring to figure 25, the probe position is no longer fixed at D or E as one tunes to the half-power points of Q_L by changing the frequency (since the impedance chart always refers to impedances seen at the shifting detuned short), but instead moves in a direction opposite to the direction of motion of the point on the impedance locus, so that the probe pickup reaches a minimum before the half-power points of Q_L are reached, and the Q values arrived at are too high. If the number of wavelengths of line between cavity and probe is N, the electrical length is $\psi = 2\pi N$, and the shift of probe position for a fractional frequency change δ is $d\psi = 2\pi N\delta$. At half-power points of Q_L we have $2Q_L\delta = \pm 1$, $d\psi = \pm \pi N/Q_L$. Assuming that, to a first approximation, the probe pickup will be a minimum when the impedance point is the same distance from the point ($2Q_L\delta = -1$) that the probe position is from point D, we have, since the radius of the Smith Chart is unity and



A1386

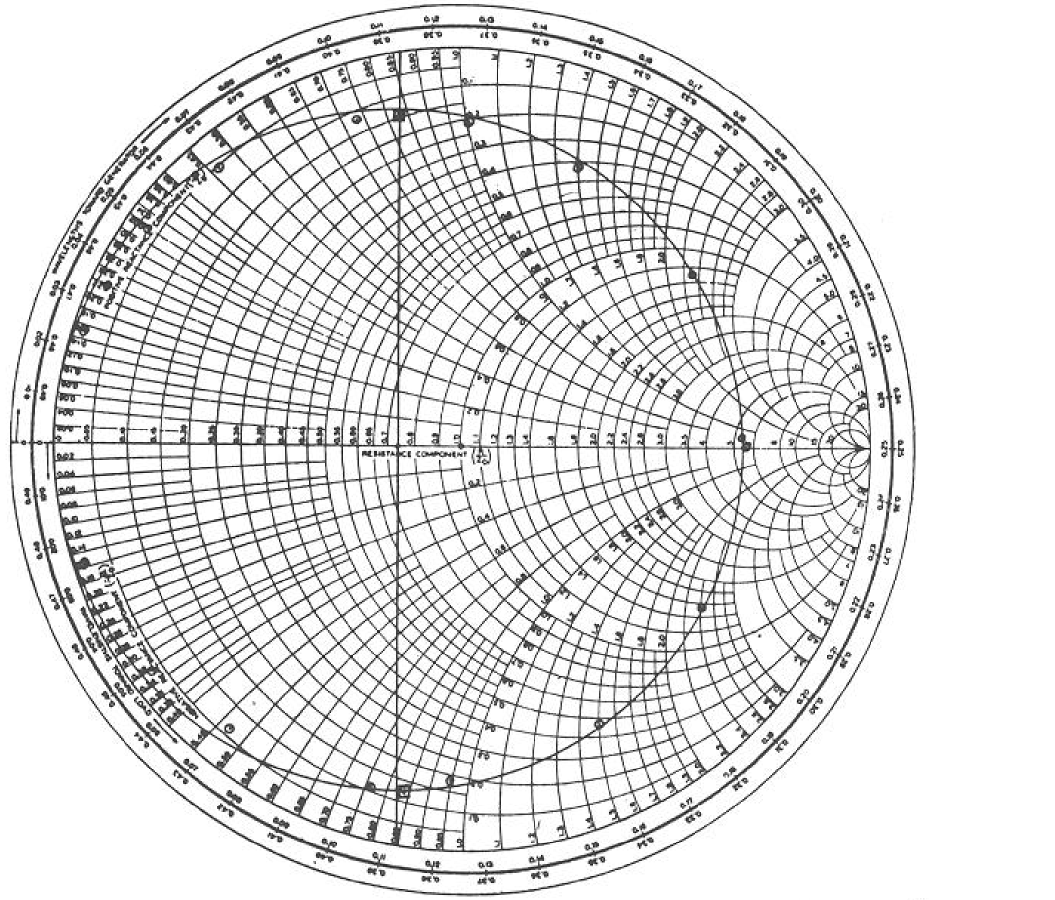


Fig. 27. Experimental values of input impedance to a coupling system at several frequencies. Half-power points of Q_L as determined by the "$\Delta\lambda$" method are indicated by the small squares.

A 1386

$d\psi$ is expressed in radians:

$$|d\Gamma| = d\psi = \pi N/Q_L \qquad (76)$$

where $|d\Gamma|$ is the absolute magnitude of the change in reflection coefficient from the half-power point ($2Q_L\delta = -1$). In order to find the frequency error represented by this displacement, we need to know the value of the derivative:

$$\left| \frac{d\Gamma}{d(2Q_L\delta)} \right|$$

evaluated at the half-power point. We may find Γ as a function of $(2Q_L\delta)$ by substituting Z_p from equation (69) into equation (4). The resulting expression is differentiated, the absolute value taken, and the condition $(2Q_L\delta) = -1$ imposed, giving the result:

$$\left| \frac{d\Gamma}{d(2Q_L\delta)} \right| = \beta/(1+\beta) \qquad (77)$$

We then find the frequency error to be given by:

$$d(2Q_L\delta) = \frac{|d\Gamma|}{\left| \frac{d\Gamma}{d(2Q_L\delta)} \right|} = \frac{\pi n(1+\beta)}{Q_L\beta} \qquad (78)$$

this is equal to the percentage error in Q_L, so we have:

$$(\text{True } Q_L) = (\text{Apparent } Q_L) \cdot \left[1 - \pi N(1+\beta)/Q_L\beta \right]$$

or

$$(\text{True } Q_L) = (\text{Apparent } Q_L) - \pi N(1 + 1/\beta) \qquad (79)$$

this formula may be used to correct for the effects of a long line. It is seen that the per cent error due to a line of given length increases rapidly with tightness of coupling (lower Q_L), so that when one couples more tightly to a cavity in the belief that the resulting sharper minima of probe voltage mean greater accuracy, he may be defeating his own purpose if this correction is not taken into account. Experience indicates that with this method the Q_u of a cavity may be measured with greatest accuracy and reproducibility when the coupling to the standing-wave detector is such that β lies between 10 and 20 db. For tighter coupling the bandwidth over which the signal generator must be tuned is so wide that errors due to variations in the coupling system as well as the line length become appreciable, and for loose coupling the probe voltage minima are too broad to locate accurately.

2. <u>The Standing-Wave Ratio Method and Variations</u>. -

In this method, described by Lawson and McCreary*, position along the circular impedance locus is determined by measuring standing-wave ratio. As a function of frequency, the SWR has a minimum value equal to β at resonance, and the rate at which the SWR rises each side of resonance constitutes the additional piece of information necessary to determine the Q's. The theoretical way in which this SWR rises with detuning may be found by substituting the value of Z_p from equation (69) into equation (4), which transforms Z_p to the reflection coefficient Γ. Taking the absolute value of this reflection coefficient, we have:

$$\left|\Gamma\right| = \frac{1}{1+\beta}\sqrt{1 + 2\beta(\alpha^2 - 1)/(\alpha^2 + 1) + \beta^2} \qquad (76)$$

where $\alpha = 2Q_L\delta$ is the detuning parameter.

the standing-wave ratio corresponding to this reflection coefficient is then:

$$\eta = \frac{1 + |\Gamma|}{1 - |\Gamma|} \qquad (77)$$

If we expand this expression by substituting equation (76) and solve for α, we have:

$$(1 + \beta)\,\alpha = 2Q_u\delta = \sqrt{(\eta - \beta)(\beta\eta - 1)/\eta} \qquad (78)$$

This equation gives the detuning parameter as a function of β and η, the measured values of SWR. If δ is measured with a wavemeter, we can then calculate Q_u from equation (78).

Although a measurement of Q based on equation (78) is a relatively simple procedure, it should be noted that the accuracy so attained is limited by the accuracy with which it is possible to set the system to resonance, since knowledge of δ involves a knowledge of the resonant frequency. In order to keep the total error small, the detuning should be such that $\alpha > 1$. A more accurate procedure is to measure several values of δ and the corresponding values of η, taken on both sides of resonance. It is found from equations (76) and (77) that at the half-power points of Q_L ($\alpha = \pm 1$), the SWR reduces to:

---

*Rad. Lab Report #64-6, "Measurement of the Q-Value of a T-R Box", dated July 13, 1942.

$$\eta_L = \frac{1 + \beta + \sqrt{1 + \beta^2}}{1 + \beta - \sqrt{1 + \beta^2}} \qquad (79)$$

If the measured values of η are plotted versus frequency and a smooth curve drawn through the points, the frequencies where this curve reaches the SWR given by equation (79) are the half-power points of Q_L. This eliminates the error due to uncertainty as to the exact resonant frequency.

The above method has been described several times in the literature, but is quite tedious, and therefore is not very practical when a large number of determinations must be made. It has been found that some increase in the convenience of the method can be achieved by working in terms of Q_c rather than Q_L. If equation (76) is written in terms of the detuning parameter relative to Q_c,

$$\varepsilon = 2Q_c\delta = \alpha(1 + \beta)/\beta \qquad (80)$$

then the magnitude of the reflection coefficient is given by:

$$|\Gamma|^2 = \frac{\beta^2\varepsilon^2 + (\beta - 1)^2}{\beta^2\varepsilon^2 + (\beta + 1)^2} \qquad (81)$$

at the half-power points of Q_c, we have $\varepsilon = \pm 1$, which reduces (81) to a very simple form. The resulting SWR is then found as before, from equation (77), to be:

$$\eta_c = \frac{\sqrt{1 + 2\beta + 2\beta^2} + \sqrt{1 - 2\beta + 2\beta^2}}{\sqrt{1 + 2\beta + 2\beta^2} - \sqrt{1 - 2\beta + 2\beta^2}} \qquad (82)$$

which at first glance seems to be far more complicated than equation (79) and more cumbersome to use. Its advantage, however, lies in the fact that in practical cases β is usually of the order of 3 to 10 (voltage ratio) and the series expansion of (82) in powers of $1/\beta$ converges very rapidly. The first few terms of this series are found to be:

$$\eta_c = 2\beta + \frac{1}{2\beta} + \frac{1}{8\beta^3} + \frac{3}{512\beta^5} + \ldots\ldots\ldots \qquad (83)$$

It is seen that even when $\beta = 5$, the second term is only 1% of the first one, so that we need use only the first term for $\beta \geq 5$. The resulting simplification is evident; we do not need to compute values of η_c from the exact equation, and do not need any previously prepared graph or table of values. We simply find the amount of detuning on either side of resonance which doubles the SWR*, and these are the half-power points of Q_c .

For purposes of comparison, the first few terms of the series expansion of equation (79) are found to be

$$'/_L = 2\beta + 2 + \frac{3}{2\beta} + \frac{1}{2\beta^2} - \frac{1}{8\beta^3} + \ldots\ldots\ldots\ldots \qquad (84)$$

the asymptote is here $(2\beta + 2)$, which is more complicated than the 2β of (83), especially when dealing with values of SWR measured in db, and in addition, the approach to this asymptote is less rapid than for the other series, since the term in $1/\beta$ is here three times as large. Therefore the error incurred by using the first two terms of (84) is about three times the error resulting from using only the first term of (83). The first two terms of (83) are accurate to 1% down to about $\beta = 1.7$ (4.6 db).

Incidentally, if we work out the formula for the SWR at the half-power points of Q_u, we get exactly equation (82) with β replaced by $1/\beta$. The series (83) in powers of $1/\beta$ then becomes a series with the same coefficients, in powers of β . Therefore, if the cavity is very loosely coupled to the line ($\beta \ll 1$) the frequencies where the SWR is 6 db greater than at resonance are the half-power points of Q_u, so that this method may be applied with equal ease for very loose or very tight coupling.

3. <u>The "3 db Down" Method</u>. - This method, due to Jaynes, is a modification of the "brute force" method sometimes used, which consists of installing two coupling loops in the cavity and measuring the bandwidth between half-power points of the resulting filter. Aside from the obviously undesirable mechanical features of this "brute-force" method, the correct apportioning of the measured bandwidth between Q_u and the values of Q_c for the two loops would involve a number of additional measurements which are almost invariably neglected.** Since the input impedance to a single coupling system contains all the necessary information, lack of a standing-wave detector would seem to be the only good reason for using the brute-force method.

---

* Since SWR is usually measured directly in db, this means that one would find the two frequencies when the SWR is 6 db greater than

** It is, of course, generally realized that in the limit of very loose coupling the measured bandwidth approaches the value due to to Q_u alone, but it is rarely that anyone using the brute force method has any idea of the extent to which this is true in any specific case. In addition it is not generally realized that the two couplings must approach zero independently; it is not enough that the insertion loss be increased merely by decoupling at one end.

The concept of the virtual parallel-resonant circuit seen looking into a coupling system enables us to apply essentially the same principle to a cavity without cutting two holes in it. The virtual resonant circuit is coupled to the matched signal generator by a direct connection, and the probe when located at the detuned short (which is the position of this virtual circuit) serves as the second coupling system. The amount of loading due to the signal generator connection is already known from the measurement of β, and the effect of the probe can easily be made negligible by having 30 to 40 db of probe isolation. Because of the loading at the input to the virtual resonant circuit, its bandwidth will be that corresponding to Q_L. The way the probe voltage varies with frequency can be seen from Fig. 23. With the probe at point A, it follows from the geometry of a circle that the voltage induced on the probe at the half-power points of Q_L is less than the maximum value at resonance by a factor $\sqrt{2}$, which is 3 db. Since the probe usually feeds a square-law detector, the procedure for finding Q_L is simply to set the probe at the detuned short, tune to resonance, and find the frequencies at which the output of the square-law detector is reduced to half of the maximum value. This method, when it can be used, is far simpler than the first two; it requires no computation from special formulas, no previously prepared graphs, no plotting of experimental points, and its operation is exactly the same regardless of the value of β, whereas the convenience and accuracy of the other methods vary considerably with β. However, it requires very high quality measurement equipment in order to equal the accuracy attainable with rather crude equipment by the other methods, as the effects of generator mismatch, variation of generator output with frequency, and probe resonance are not masked by sharp minima. In general, this method will save a considerable amount of time if the accuracy required is not greater than about 15%.

Experimental Confirmation

We may put the above theory on solid ground by studying some actual experimental data taken on the input impedance to the coupling system of a resonant cavity. This data will verify the theoretical result that the impedance locus is a circle, and will indicate directly how accurately the half-power points of Q_L may be located. In Fig. 27* are shown the values of input impedance measured at the detuned short with a carefully adjusted standing-wave detector, at several different frequencies. Actual values of frequency are omitted for security reasons. The type of coupling system was a rather indefinable combination of loop and iris coupling, and the cavity was from a receiver preselector. The SWR looking back toward the generator was 0.8 db, so that the amplitude of the second wave traveling toward the load was less than 5% of the amplitude of the backward-traveling wave.

* Following Page 47

It is seen that the experimental points lie very nearly on the circle determined by the origin and the load impedance at resonance, but they tend to fall inside this circle at high standing wave ratios, due to losses in the transmission line and to a superimposed rotation as the detuned short shifts with frequency. The half-power points of Q_L are located by geometrical construction, and the points found by method (1) are indicated. It is seen that the preceding theory is an accurate description of what takes place in a coupling system, and that method (1) is capable of high accuracy. The values of Q_u obtained by the three different methods were as follows:

"$\Delta\lambda$" method	822
SWR method	791
"3 db down" method	768

The data for the "$\Delta\lambda$" method were corrected for the length of line (1.5 wavelengths) between cavity and probe. The true value of Q_u was probably between 791 and 822, making the error in the first two methods about 2 to 4%, while the error in the "3 db down" method was probably 4 to 7%.

Edwin T. Jaynes

EDWIN T. JAYNES
Ensign, USNR