SYLVANIA ELECTRIC PRODUCTS INC.

ELECTRONICS DIVISION

70 FORSYTH ST. BOSTON, MASS.

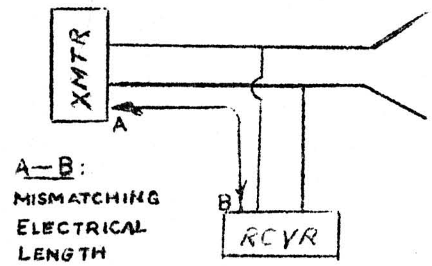
OPERATION AND DESIGN
OF TR AND ATR TUBES

Lecture No. 1

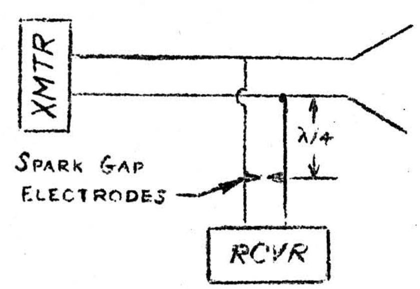
History

Early Radars:

Since even in the early radar systems a single antenna was used both for transmission and reception there was always the problem of effectively isolating the transmitter from the receiver so that a minimum of signal would be wasted. In the earliest systems, where the transmitter power was very low by modern standards, and the receiver input was to a rugged triode amplifier or detector, it was not necessary to protect the receiver. The purpose of the duplexer was simply to avoid undue loss of signal. This could be accomplished by accurately tuning the various line lengths.



Since this adjustment was critical, it was soon replaced by a fixed position spark gap, the forerunner of the modern TR. In its earliest form the TR switch was quite literally a spark gap. Tungsten electrodes with a narrow gap between them were placed across the line. This rudimentary arrangement worked quite well, until the development of improved radar systems.

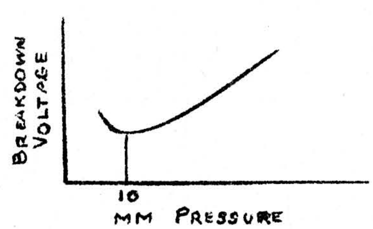


Recent Radars:

With the advent of very high frequencies for radar systems, it was necessary to increase the sensitivity of the detectors. Unfortunately, this made the detectors highly susceptible to damage, and at the same time, transmitter powers were going up, compounding the difficulties. The higher the frequencies got, the worse the situation became. The simple spark gap no longer sufficed. When crystal detectors were introduced an improved system was imperative. A crystal is an unusually sensitive device. Modern radar crystals detect signals of the order of 10 micro-microwatts. On the other hand, the power of some of the modern transmitters extends to the megawatt values. Thus the range between transmitter power and minimum detectable signal power is of the order of 10^{17}. The duplexing system must be efficient!

Gas:

The first step toward improving duplexers was made when the electrodes were placed in an evacuated envelope so that the spark gap was under reduced pressure. A plot of the breakdown voltage vs. the pressure will show that minimum breakdown voltage is at about 10 mm pressure. Between the minimum firing voltage at this point and at atmospheric pressure is a difference of a good many orders of magnitude. The gap, then in the evacuated envelope will fire that much sooner.



Along with the investigation of pressures within evacuated envelopes, studies were made of varying gas fills. As a result of experiments with a wide range of gases, it was discovered that best performance is obtainable with the noble gases (such as argon) and a few of the other simple gases (such as hydrogen). The tube design eventually centered on hydrogen and argon. Most of the modern tubes have argon fills.

This brought into prominence the problem of recovery time. At reduced pressures, the life time of electrons in active space charge is long. Furthermore, with the increased sensitivity of radar systems, it was desirable to reduce still more the minimum observable distance.

As explained in the first lecture, reductions in recovery time are best effected by mixing water vapor with the argon or hydrogen gas fill. The water vapor is introduced in order to give the electrons a place to go after the end of the arc. The probability that an electron will adhere if it hits either an argon or hydrogen ion is small, so small as to be practically negligible. Classically, you cannot have two particles colliding and adhering, and at the same time conserve both energy and momentum. Two-body gas-phase reactions are practically an impossibility unless one of those bodies has available to it various vibrational modes to accept the extra energy from the collision. That is what water does. The molecule of water is sufficiently complicated to make an electron tend to stick when it hits the molecule. A negatively charged water ion is, to all effects and purposes, out of the field as far as the r-f is concerned. It is too heavy to move with the field. It was found then, that a mixture of argon or hydrogen and water is satisfactory. The argon (or hydrogen) ionizes sufficiently readily so that spike leakages are small, while water absorbs the electrons rapidly after the end of the r-f, permitting rapid recovery.

Resonant Modes:

Once the gas problem was more or less solved, attention was focussed on the geometrical design of the tube. How could its effectiveness be increased? For low frequency circuits the solution is easily attained. Consider a spark gap that is to be fired as fast as possible. There are two things that can be done. One, place a step-up transformer on the input. A small voltage at the input shows up as a large voltage across the gap, making it more apt to break down. A corresponding step-down transformer is put on the output so that there is no net effect on small signals. Secondly, we can insert across the gap sufficient capacity to resonate with the inductance in the circuit (the transformer windings). The resonant build-up quickly increases the voltage until there is a breakdown of the gap.

The problem is how to do this with microwaves. Achieving resonance is easy enough. Any box or any enclosed structure of any shape is resonant at some frequency. One of the best ways of visualizing it is to consider an organ pipe, or the way sound will resonate in a room.

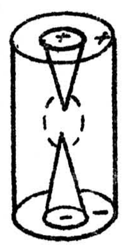
That the cavity can operate in several modes simply means that there are that many different ways in which oscillations can occur. An organ pipe is excited in such a way that all those modes occur simultaneously. But, if a tuning fork is placed on an organ pipe, the pipe will not generate any of the harmonics. If the tuning fork is tuned to the fundamental or any harmonic, resonance at that frequency occurs. These different resonances are called the "modes" of the pipe.

Now let us consider a transmission line. Short circuit the two ends. Then excite the line in some way. The boundary conditions (so-called by mathematicians) are clear enough. Inside we have the oscillations as determined by the transmission line equations. At the ends we must have zero voltage. The line, then, will be excited appreciably only if the frequency is such that the line is effectively a half wavelength, one wavelength, three halves or in general, any number of half wavelengths long. Those are all different modes.

A cavity can be analyzed in the same manner. Considering a rectangular cavity we can look at it along one axis and observe that it constitutes a waveguide with short-circuited ends. The same conditions as before exist. The boundary conditions are that across the two end plates there must be zero voltage, so resonance occurs at those frequencies for which the waveguide is a half wavelength or any multiple of a half wavelength. Similarly, we can look along another axis. Again we have a waveguide but with different walls, and, again, the ends are shorted. In general the dimensions of this waveguide are different from the first one considered. Thus the frequencies of operation are different, but the modes are the same type. Then we could view the cavity along the third axis and obtain still a third set of modes. In a more complicated cavity, the modes will appear more complicated, but the same type of operation occurs. Any cavity has an infinite number of modes in which it can operate. They may or may not be evenly spaced. They may be much more complicated. But the basic situation is exactly the same.

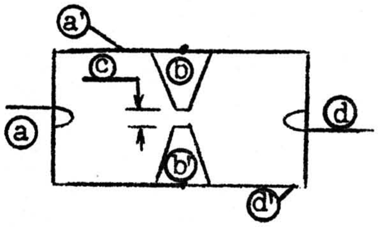
Narrow Band TR Tubes:

This, roughly, is what is done with the TR tubes. In a cylindrical cavity is placed a gap as illustrated. This structure can oscillate. In the particular method of operation in which we are interested,



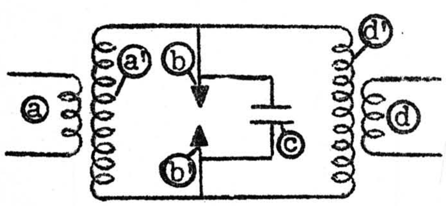
the top and upper cone act more or less uniformly in phase. The bottom and lower cone operate in phase opposite to the upper cone. At a given moment the upper cone has a positive charge, and the lower a negative. Half a cycle later the upper cone is negative, the lower, positive. A high potential is developed across the gap. Now a quarter cycle after the start, say, the electrons which were concentrated on the lower cone will be running up the sides to the top, and if you like to think of it that way, they will keep on coasting after they have neutralized the positive charge on the top until they build a negative charge there. Such a device will store energy, and that is what we want. If we store energy, we will build up the voltage, and when we build up the voltage, the breakdown will occur more quickly.

We must get energy in and out of the cavity. One way of doing this is to put loops through the sides. This was the principal method used before waveguides became as common as they are now. These



loops couple into the cavity's magnetic field, which hasn't been mentioned so far, but which should be fairly plain. It will be remembered from the previous analysis that the path traced by electrons was from the bottom (negative charge) to the top (positive charge) along the sides, in effect forming a loop of current. That means that there is established an r-f magnetic field going round and round the cones. The loops, then, are so positioned that they will couple into this alternating magnetic field.

Effectively there is the same performance as occurs with the use of a step-up transformer at the input and a step-down transformer at the output. The combination of loop and cavity walls behave



as a step-up (or step-down, depending upon position with respect to signal) transformer, and provides the voltage build-up necessary to break down the gap at the maximum possible rate. Corresponding parts are labeled with the same symbols in the above figures.

At this point we might consider the spark gap circuit from the point of view of optimum design. If the posts of which the gap is composed have a relatively large area, a very small gap spacing will give too much capacity for the amount of inductance that is present, so that the wrong frequency will be obtained. If, however, narrow rods are used for the posts, the current traveling down the rod will set up a magnetic field around it which will effectively slow it down. In other words, add inductance. What is wanted, then, for each post, is that the area at the gap location be as small as possible, and the area at the base location be as large as possible. Hence, we have our present cone shaped construction. The cone shaped construction can actually cause trouble again, because

of the shunt capacity. But there does exist an optimum cone angle, found by experiment, which has a minimum inductance for minimum shunt capacity.

In modern TR's fed from and into waveguides, the loops are not required, and a hole, or "iris" is normally used. Their behavior is identical with that of the loops although the analogy to a transformer may not be so obvious.

Broad-Band TR Tubes:

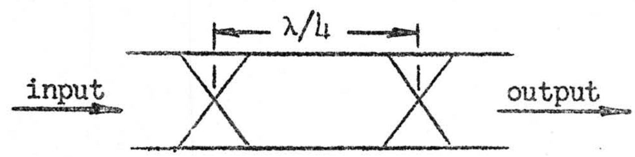
The broadband TR came into the field toward the end of the war. Broadbanding the TR tube was the logical outcome of experience with, and improvements in radar systems both by our Services and by the enemy. Accurate alignment of radar systems with narrow band tubes was exceedingly difficult to perform out in the field. Also, there is a good deal of advantage to being able to shift the frequency of operation. For one thing, the nature of a target can often be guessed by watching the behavior of its echo on the scope under varying frequency. For another, counter measure devices were being used more frequently and effectively by the enemy. A rapid 5% or 10% shift in frequency may enable the radar operator to see the target before the jamming measures could locate the new frequency and block information. A further reason for the broadband TR tubes is the development of some of the more complicated duplexer systems. The balanced duplexers (to be discussed later) depend upon a very precise match between two arms of a magic tee. The adjustments necessary to make such a system work properly are much too critical, or require too much auxiliary equipment to be done in the field. Such a system must be based on a broadband tube. As a matter of fact, the system calls for more than broadbanding. It calls for knowing what is called "the position of the effective short circuit." That is, when the tube fires and appears as a short circuit to the line, exactly where does the short circuit appear? The balanced duplexers call for careful matching of the effective short-circuits in their branches, so that accurate multiples of quarter wavelengths of line may be used.

Broad-banding, then, was the next logical step in the developing design of TR's.

Principles of Broadbanding:

If you were given the problem of broadbanding a TR tube, starting from the narrow-band design, the first thing you might try is simply reduce the Q by the necessary amount. However, we want a tube which will cover, say, a 10% band with an SWR(V) of less than, say, 1.4. The Q would then have to be about 0.3. And this is much too low for the tube to serve its protective purposes.

We can, however, use the principle of destructive interference. If we put two identical elements along a line a quarter-wave apart:



then some of the input power is reflected at the first element. If the fraction reflected is small, the second element will reflect almost as much. But the reflection from the second element, by the time it gets back to the first, will be dead out of phase with the reflection from the first. Hence they will cancel, and if almost no power is reflected, almost all power must be transmitted.

This is the principle by which broad-banding is accomplished. In detail, of course, it gets much more complicated. We shall consider it in detail later. But this is the basic technique. Using it, we are able to increase the Q's of the elements from the 0.3 mentioned before, to about 4.5 or 5. Theoretically we could go higher - to about 10. But 4.5 - 5, in a multiple element tube, has been found to be sufficiently high to give adequate protection.

Summary:

To summarize, then, the basic problem is to use the same antenna for both transmitting and receiving. In the earliest radars this was not difficult. Since the power was low and the receiver rugged, it was only necessary to avoid undue loss of power. This could be done by adjusting the line lengths.

This adjustment was critical so spark-gaps were substituted—the prototype of modern duplexers.

When transmitter power became large, and particularly when easily damaged crystal detectors were introduced, a better system was needed. First, the spark gap was enclosed in an envelope and the pressure reduced to the 10 mm. or so which gives minimum breakdown voltage. Then, other gases were tried. Argon and hydrogen were found most effective. But these gases did not recover. So water vapor was added to absorb the electrons after the r-f pulse was finished. This combination is still used.

To further increase the tube's effectiveness as a duplexer, the gap was designed to operate across a resonant cavity. The resultant voltage build-up increases the rate at which it will break down, making it more effective. This, in its basic form is the modern narrow band TR tube.

Modern applications place emphasis on a broad-band structure. This is desirable for many reasons, e.g.:

(a) To eliminate field tune-ups.

(b) To permit frequency changes to aid target evaluation.

(c) To permit frequency changes to avoid enemy "jamming".

(d) To permit the use of more efficient duplexers which could not practically be tuned up in the field.

etc.

To do this a multiple element tube is used. By properly arranging these, their reflections can be made to cancel (destructive interference) over the desired band, even with fairly high reflections from the individual elements.

OPERATION AND DESIGN

OF TR AND ATR TUBES

Lecture No. 2

Survey of TR & ATR Behavior

Introduction:

This series of lectures is being given as an aid to engineering personnel directly concerned with development and production problems in TR and ATR tubes. By presenting the background and theory, it is hoped to provide not only the means for improvement in our products, but also, and in particular, bring to the foreground the difficulties confronting engineering personnel, so that we as a Service Group can be of assistance.

The lectures on the TR type tubes will cover the general history of the tubes, the various types and usages, measurement and tuning considerations, and filter theory algebra. Lectures on TR tube properties which depend upon the plasma and ionization are scheduled to be given by Dr. Varnerin.

This second talk will be a general survey of TR tubes, showing how we propose to break down the general field and giving some idea of the overall picture. The third one will cover uses. We shall briefly review the various types of duplexers, branched and balanced and perhaps some of the more unusual designs. You are not apt to be directly concerned with these, but should know something about them because the system for which the tube is to be used governs the specifications of the tube.

In the fourth lecture, that on measurements and tuning, we plan to discuss the measurements that are made on TR tubes, what may be learned from these measurements, and some indications of how they tie in with design. We shall also describe some of the different methods available for tuning broadband TR tubes. Tuning is a very complicated subject, because a complete discussion of it must include mistuning. Most of the latter I shall neglect entirely, but may give a little of it in the lectures on the low-level operation of TR tubes.

The next three lectures will consider the TR tube as a broadband filter. We shall discuss its equivalent circuit, and investigate a little of matrix algebra. Matrices will be introduced to inform you what a matrix is and what it tries to do. Matrices are simple. They are a shorthand notation which permits you to carry out easily some very tedious operations of ordinary algebra. By using matrices you can do a good many arithmetic operations at once and have them automatically come out right -- you hope. Since matrices are used in any filter problem, I believe you ought to have a nodding acquaintance with them and their uses.

The lectures about the properties of TR's which are dependent on the plasma and ionization will be given by Dr. Varnerin. Presumably, one lecture will analyze TR tubes while firing. The other phase is the process of recovery.

Returning to the present lecture, we will make a general survey of TR and ATR behavior.

Four Phases of the TR Cycle:

I believe you all know what a TR tube is. Basically, it is a simple tube — you might call it a monode. It gets a little complicated, of course, when you start considering the details. There are four phases to its operation.

(1) We may start off with the low level phase. With a small signal impressed, how does the tube behave?

(2) The second phase is that of the tube during firing. When a high power pulse of r-f energy reaches the TR from the magnetron, how fast does it fire and, in fact, how does it fire? (It might be mentioned at this point that among Dr. Varnerin, myself and a few other people, there is at this time considerable discussion as to just how the things fire. There are in existence two completely opposing theories, one of which is my own pet one, the other being more or less standard. These two theories are entirely contradictory, so much so that they disagree on such a basic point as to which element fires first. I mention this now to illustrate the gaps that exist in our knowledge of what goes on in the TR tube. About something as fundamental as this there is still confusion, and room for conflicting theories.)

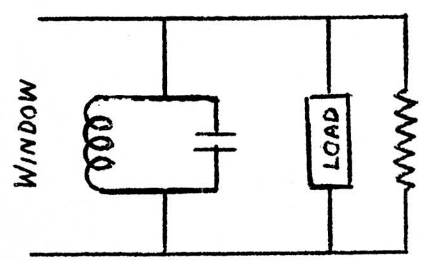
(3) To continue, the third phase is the operation during the high level signal. Basically, the picture here is uncomplicated.

What you want, and almost have, is a dead short-circuit across the tube. You do not quite get that, but you do get close to it in a good tube. The high level operation is of considerable practical importance for design. It is during this part of the cycle that destruction takes place in the tube. The tube is taking the beating during this period which will ultimately lead to the end of its life.

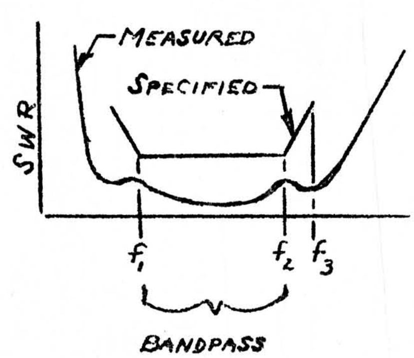
(4) The fourth phase is the recovery time. Recovery is what limits the practical application of the TR tube. Radar systems which use TR tubes cannot be required to pick up an echo in an interval less than the recovery time of the TR tube. The absolute altimeter and a few other devices must be able to detect signals from as close as a few feet. Hence, the absolute altimeter does not use pulse systems; and cannot until a new type of TR is invented. If anyone present can design a TR with no recovery time, there will be an ample market for it.

<u>Low-Level Phase:</u>

Let us go back to the first phase in which we consider the TR tube as a broadband filter. The primary application, here, is to broadband TR tubes. With narrow band TR's and ATR's, we are, to be sure, interested in bandwidth. However, the narrow band TR is a very elementary kind of filter -- simply a shunt resonant circuit. And there is little point in applying filter theory to ATR's because little can be done about them. There is a theorem known as "Foster's Reactance Theorem" which states briefly, that, if you have an element (in this case the window of the ATR), you cannot broadband it by the addition of a reactance network in parallel. That is, the bandwidth cannot be made any wider than that of the window itself. To increase the bandwidth it is necessary to introduce loss into the circuit. Foster's Theorem is not applicable to TR tubes, since the complete network does have loss. The TR tube is loaded by the output, which, ideally, is a pure resistance. Hence, when we speak of the filter properties of TR tubes, we shall be speaking primarily about broadband TR's.

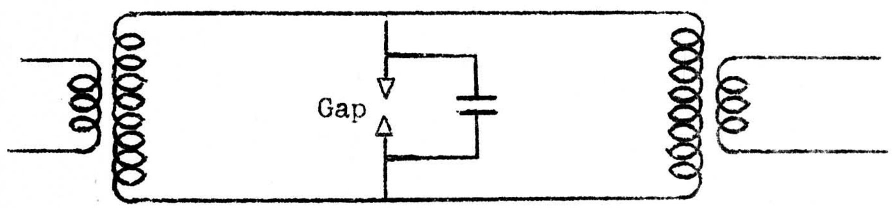


It is somewhat deceptive to speak of a tube's filter properties. There are several fundamental differences from the point of view of conventional filter theory. For example, let us consider an i-f bandpass filter. For such a filter we specify that the characteristic must be kept sufficiently flat across a required band. At the same time, for i-f applications, it is usually essential that there be as complete as possible rejection outside the band. In TR tubes, on the other hand, the specifications normally only require that the standing wave ratio be kept below a designated level over a specified band. Another point may be specified at which a somewhat higher standing wave ratio may be tolerated. For the value at points beyond, there is no concern. There is no interest at all in the rejection of signals outside of the band.



A straight piece of waveguide would fit the low level specifications of any TR tube. It is perfectly flat at all frequencies. Hence, we must apply to the TR tube some other criterion. This has to do with its high level behavior; it must adequately protect the crystal. And we must ask, then, what it is about the tube that determines how good protection it gives. Clearly this is determined by how fast it fires and when it is fired, how completely.

Let us consider a narrow band TR. We may write an equivalent circuit for it as follows:



Equivalent Circuit of a Narrow Band Tube

The input is a voltage step-up transformer. This is what we want because the more the voltage is stepped up, the quicker the gap is going to break down. Furthermore, once the gap is broken down so that the impedance across it is reduced to a small value, then the higher the input step-up, the less the effect of the residual impedance - the more effective the approximate short circuit. The measure of the effectiveness of the step-up is expressed by the doubly loaded Q of the circuit, which we will discuss in more detail later.

To put the matter briefly, then, the other criterion in TR tube design is that the Q of the firing element should be as high as possible. Now that leaves open the question, in multiple element tubes, as to which element should be high Q. Are we willing to sacrifice Q in one element to get a higher Q in another? This is a question to which, as far as I know, nobody knows the answer. The general feeling on the subject is that Q is good no matter where you have it, but you would like at least some element in the tube to be as high Q as possible.

When we compare the broadband design practice to conventional bandpass filter theory, then, there is this very fundamental difference in what constitutes optimum design. In conventional filter theory we define optimum as that which, giving us the required bandpass characteristic, also gives us as sharp a cutoff as possible. In the TR tube the behavior in the pass band is specified as before, but our definition of optimum is that at least one element of the tube shall have as high a Q as possible. These two different definitions of optimum may be equivalent, but under some circumstances will lead to quite different results.

Transition to the Fired State:

The second phase of the TR tube is its transition to the firing stage. This phase is an extremely difficult one to handle since not enough is known about gaseous discharges.

Furthermore, until very recently most of the basic work on gaseous discharges was done with d-c or low frequency a-c. Data obtained at low frequencies is, on the other hand, practically entirely inapplicable in any respect to r-f discharges. Low frequency and d-c discharges depend on the migration of electrons in one direction, and positive ions in the other. An elementary calculation will indicate that positive ions can gain practically no energy from an r-f field -- they are just plain too heavy. The r-f field is going too fast. Any theory of the u-h-f breakdown of gases must be based on the electrons and only on the electrons. That is not to say that the ions can be neglected. They are important since they put on a more or less fixed space charge through which the electrons must move. Recently, Dr. S. Brown at M.I.T. and others, have been working at r-f. The subject is being clarified and a good deal about the problem is understood now. Dr. Varnerin will tell you more about this.

Unfortunately, the present state of knowledge is not sufficient to help us much. This new understanding applies to discharges that are isolated. In a TR tube they are far from being isolated. The gap is too short. The cones come in so that the minimum distance between electrodes is of the order of .055". That is not very many mean free path lengths long at the pressures at which a TR tube is usually filled. It is of the order of 5 to 20 mean electron free path lengths. When electrons start oscillating in that discharge region the chances are excellent that they will hit the cones. The problem of what happens then is complicated.

They may bounce off, induce secondaries, or be absorbed either with or without emitting photons — light — which can cause ionization elsewhere. The problem gets immensely involved, and as far as I know no one has even tried to approach it. That leaves us back about where we started. Qualitatively, we have some knowledge of microwave discharges, which we can apply to the TR tube in, again, a qualitative way.

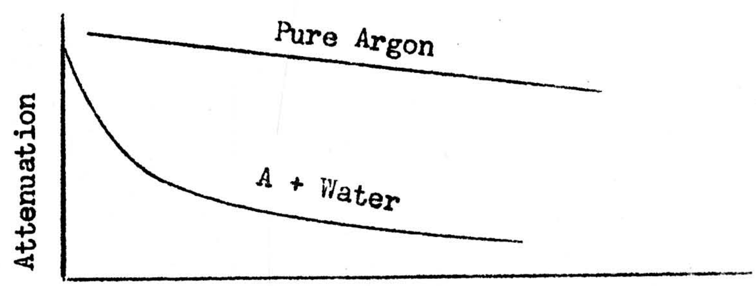
An additional drawback to a clearer understanding of the TR tube is, of course, the keep-alive. This, by the way, is the point around which the discussion I mentioned before revolves. Just what is the function of the keep-alive? I dare say some of you here, thought you knew. According to current standard theory the keep-alive enables the gap in which it is placed to fire at the earliest possible time. To explore at this meeting the other theory about the keep-alive would take too much time. It can be pointed out, however, that the theory opposing the standard one hinges on the point that while it may be true that the keep-alive enables the gap in which it is placed to fire at the earliest possible time, this fact may not be important.

<u>High Level:</u>

The next phase is the high-level phase. This is, in some ways, the simplest part of the whole picture — primarily because there is not too much that could be done about it if it were understood, and it is not understood, anyway. It is dependent upon the gas discharge phenomenon. While we understand the latter in a very rough qualitative way, we are not yet able to calculate precisely what is happening. However, we do know that during this period most of the damage is done to the TR tube, and it is during this period that the main limitations are placed on the TR tube behavior.

Recovery:

The recovery time generally determines the usefulness of a tube. As I indicated before, the major problem in connection with this fourth phase of the tube cycle is concerned with minimizing recovery. Argon gas (most often used) ionizes easily, but will not recapture electrons rapidly, tending toward a lengthy recovery time. Water will recapture electrons easily. However, water (or other gases suitable for reducing recovery time) cannot be substituted for argon entirely. Its ability to absorb electrons would inhibit ionization, giving high spike leakage. Best results are obtained by adding to the argon fill sufficient water vapor to reduce recovery time to a useful value.



As an indication of the magnitude of this effect, in well baked-out tubes filled with pure argon, recovery times of the order of 1000 μs. have been measured. Minute traces of water will reduce this significantly. With, say, 5 mm partial pressure of water, the recovery time is reduced to the order of 5 μs.

Summary:

To summarize, then, the four phases of a TR tube's operation are:

(a) Low-level--when we may treat it as a passive linear network.

(b) Firing--when we are concerned with the speed of breakdown and the function of the keep-alive.

(c) Fired--when we are concerned with the completeness of breakdown and effectiveness of the short circuit, and when the principal damage to the tube is done.

(d) Recovery--how fast does the tube recover? This often limits the applicability of a TR tube.

We shall consider these in more detail later.

OPERATION AND DESIGN

OF TR AND ATR TUBES

Lecture No. 3

The Uses of TR Tubes

Today we shall consider the various ways in which TR tubes may be used. Although most of us here can hold ourselves immune from such considerations, concentrating our attention on the specifications as received, yet the circuitry into which they are to be placed does, after all, determine the specifications.

The Basic Functions of TR Duplexers:

The fundamental purpose of any duplexing system whether it uses TR's or not, is to properly interrelate the various components of the system during the two phases of transmission and reception.

1. During transmission the transmitter is to be connected to the antenna but not to the receiver. The disconnection between transmitter and receiver during this period is a very critical one and must be done exactly. The maximum allowable power that may leak to the receiver is of the order of one part out of 10^8.

2. During the receiving period, the antenna is to be connected to the receiver but not to the transmitter.

This behavior defines what is known as the duplexer, or duplexing system. Normally, the basic function is performed by the TR and ATR tubes. It is, however, possible to accomplish it without using the TR tube. The same effects can be achieved by straight plumbing, but at the expense of some of the other factors. There are two minor functions, or rather desirable characteristics that the duplexer should also fulfill.

3. The duplexer must act fast; the switching operation should be done as rapidly as possible. To avoid losing a near target, switching from transmission to reception must be done swiftly, somewhere between 3 and 4 microseconds.

4. The loss in the duplexer must be small. A 3 db loss in the duplexer, that is, 3 db during transmission and 3 db during reception means a total loss of 6 db or 3/4 of the power.

There are some other factors of interest to circuit engineers. Stability of the tube is one of them, particularly as regards temperature. How well does the system stay in tune as the temperature is changed? For with a change in temperature there is going to be a change in operating frequency, if only a few megacycles. If a narrow band TR tube is used, that slight shift in frequency may put the transmitter entirely outside the frequency of operation of the TR tube. The disadvantage of the narrow bandwidth in such a case places an upper limit on the Q of the tube. (Q will be discussed in more detail in the measurements lecture.) Q is inversely proportional to bandwidth. Or, viewing the problem from another angle, the variation in frequency with temperature, or more exactly, the allowable tolerance of this variation governs the lower limit of the bandwidth that may be used.

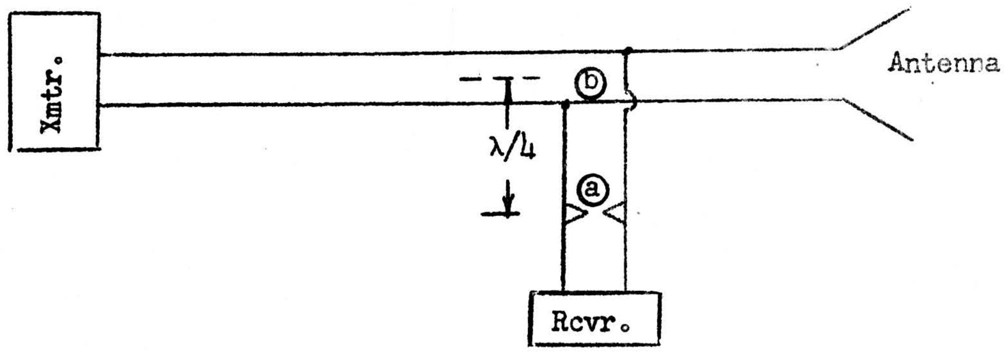
Another factor to be taken into account is the line power. Fluctuations in line power will change the voltage level delivered to the magnetron and the local oscillator, thus changing the frequency output. In general, though, this problem is less serious than variations in TR performance caused by changes in temperature, or simply by age.

There is to be considered, too, the matter of shock resistance. Will the tube go out of tune when subjected to excessive vibration? That is a particularly serious problem in some of the modern applications of TR tubes. It is clearly very critical indeed for a TR in a guided missile. It is still an important problem for the older applications. A battleship, when it fires its guns, has a momentary acceleration of the order of 100 gravities. The TR must be built to withstand, or else shock-mounted to withstand this acceleration.

Of all the requirements the chief of them perhaps is whether or not the TR tube can be tuned in the field and how accurately. It can easily happen that the precision demanded is so high that the necessary equipment is not available in the field. But even if it can be done, it is still very inconvenient. This, as we saw last time is one of the primary reasons for producing broadband tubes. There are other reasons also, as we mentioned, such as the ability to tune the radar set either to avoid enemy jamming or to get a better evaluation of the target.

Branched Duplexers:

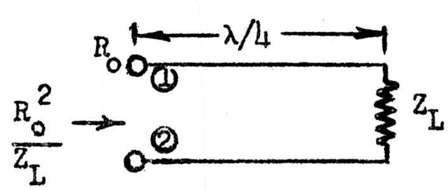
The original duplexing system began with a transmission line (either parallel wires, coax, or waveguide) at one end of which was placed the transmitter, and at the other end the antenna. Somewhere along this line is to be inserted a means for conducting signal to the receiver.



To satisfy the requirement of the duplexer during the transmission time that the transmitter be decoupled from the receiver, a spark gap is placed at (a). When signal travels down the line from the transmitter, the spark gap fires. If the spark gap provides a good enough short, no power will get to the receiver.

In general, this would seriously interfere with the transmission from the transmitter to the antenna. We can arrange things, however, so that it does not.

If we look into a quarter-wavelength of line with an impedance Z_L at the end, we see the reciprocal impedance. That is, if we measure the impedance between the terminals 1 and 2, we will get a value R_o^2/Z_L. R_o is called the "characteristic impedance" of the line and is a constant, at a given frequency, with the dimension of ohms, determined solely by the dimensions of the line. (It is usually written R_c instead of the Z_c that might be expected from the name "Characteristic Impedance" since it is real if the line propogates without loss - i.e.



it is always resistive for lossless transmission lines, and for wave-guides at frequencies above cut-off.)

If, now, we put the spark-gap a quarter wave back from the junction, when it fires we have a short circuit ($Z_L = 0$) as (a) but the transmitter sees an open circuit ($R_c^2/Z_L = \infty$) at (b) looking toward the receiver. Since this open circuit is parallel to that seen looking toward the antenna, the branch does not interfere with the transmission of power to the antenna. The first requirement then, is satisfied.

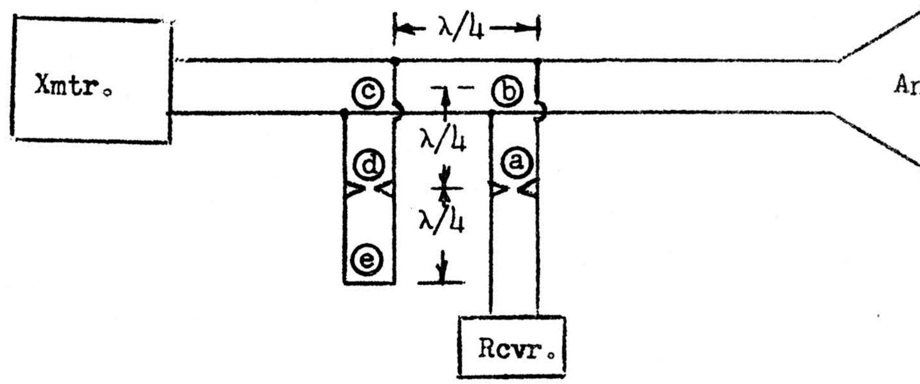
The first part of the second requirement, which calls for the coupling of the antenna to receiver during the receive-cycle, is satisfied if the spark gap, (a), is arranged in such a manner that when it is not fired it does not interfere with the energy going through it. If we build the spark-gap into a cavity (as was mentioned in Lec. 1) this can be satisfied. With the proper cavity design, and providing there is no ionization, it is possible to eliminate reflection by the TR.

The other part of the second requirement, which states that the antenna be decoupled from the transmitter is not so critical. Sometimes no special equipment is used. The transmitter when it is off usually presents an impedance which is very different from the characteristic impedance of the line. This mismatch may be used to direct the received energy into the receiver. For example, suppose that looking back into the output of the transmitter, we see an open circuit at the resonant frequency.

If the line length from the transmitter to the junction, (b), is a multiple of a half wave long, at (b), looking toward the transmitter, we will see an open circuit. Since this parallels the presumably flat line to the receiver, all received power goes to the receiver. The line length from the transmitter to the junction, (c), can be used to satisfy the second part of the second requirement - to "decouple" the transmitter from the antenna during "receive".

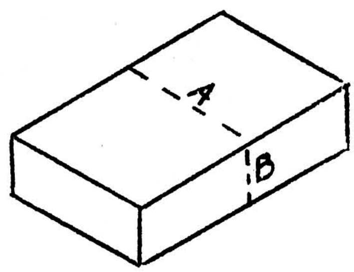
This is not always a good arrangement, however. If, because of a frequency change or what not, the line length from transmitter to junction becomes an odd multiple of a greater wavelength, the open circuit at the transformer output shows at the junction as a short circuit. Since this shunts the receiver line, it will reflect all received power back out the antenna. So, this system can lead to trouble.

So, although the method of adjusting the line length from transmitter to junction is feasible (it was used in the early radars and is still used occasionally), it leads to difficulty. It is difficult to make the length adjustable since the line must also carry high power. It is a fairly critical adjustment anyway. And if you "pre-plumb" the line then you must impose upon the manufacturers of magnetrons very tight tolerances on frequency, particularly if the transmission line is appreciably long. The answer to this problem was the ATR or anti TR. A quarter wavelength down the main line from the TR, another junction is made, and another line shunted across the transmission line. The spark gap is spaced a quarter wavelength from the junction down the new line, and is backed up by a short placed a quarter wavelength from it.

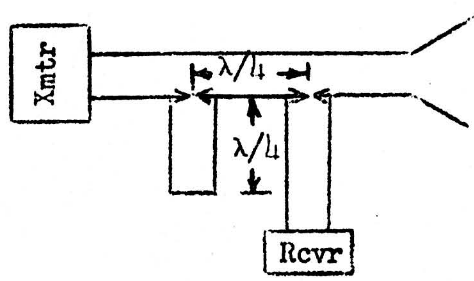


During the transmit-cycle energy coming from the transmitter fires the ATR. This puts a short at (d) and the power sees an open circuit at (c) looking into (d). Except for the small amount of power absorbed in firing the ATR, the system then operates as previously explained. During the receive-cycle, that is the cycle when energy is traveling from the antenna into the line, the short at (e), which is a halfwave from (c), reflects a short at (c). The zero impedance at junction (c) reflects infinite impedance at junction (b), and no signal coming from the antenna travels to the transmitter. The signal travels down the line to the receiver. It will be noted that during the transmit-cycle, the ATR and TR are fired; during the receive-cycle, the ATR and TR are unfired.

The ATR-TR arrangement shown is the basic duplexer system. Any number of modifications may be made, principally for convenience. In a waveguide a component inserted in the wall at A is in series with the guide. A circuit inserted as at B acts as a shunt to the guide.



So, although it is difficult to use series connections with coaxial transmission lines, they are quite feasible with waveguides. This allows us to simplify the arrangement somewhat. The ATR can be placed in series with the line, the short a quarter-wavelength behind it. During the receive-cycle the short circuit behind the ATR reflects an open circuit at the



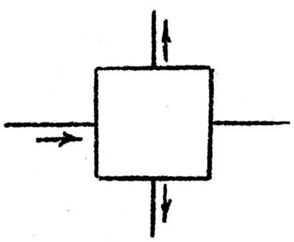
junction of the ATR and the waveguide, which effectively provides a break in the line to the transmitter insofar as incoming signals are concerned. Similarly, the TR arm is moved up a quarter-wavelength, by placing the TR in series with the waveguide. A mixture of series-shunt connections can be used for a duplexer. One branch may be connected in series (as are both branches of the above simplified arrangement); and the other branch may be connected to the line in shunt (as is done in the ATR-TR arrangement first illustrated).

Mention should also be made of the possibility of using similar ATR-TR arrangements for multi-channel applications. For instance where a beacon is involved, it may be desirable to receive several frequencies simultaneously and duplex against all of them.

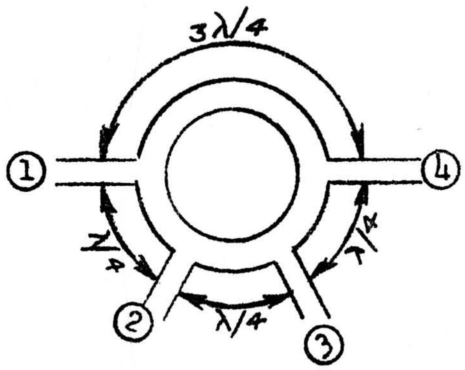
Balanced Duplexers:

Balanced duplexers depend essentially on a group of components which can all be called hybrid circuits. There is a group of them and quite a number of possible variations among them. The commonest of these components are the "rat race" and the "magic tee". A third component, not usually considered hybrid, solely because its applications are different, is the directional coupler.

All the devices accomplish one job. They may be represented by a box with four input-output lines. Signal fed into any



one of them divides between two of the others, but does not go to the fourth. The reason for this division of signal is probably most easily seen with the "rat race."



On a ring of waveguide, four arms, which we shall label (1), (2), (3), and (4), are spaced such that the distance between, say, (1) and (4) is $3\lambda/4$. The distances between (1) and (2), (2) and (3), and between (3) and (4) are made $\lambda/4$. A wave entering the ring at arm (1) divides in two, one part traveling clockwise, the other counterclockwise. What happens at arm (2)? The wave going counterclockwise from (1) to (2) has traveled one quarter-wavelength. The wave traveling from (1) to (4) to (3) to (2) has traveled a full wavelength plus one quarter-wavelength. It arrives at (2) delayed by an extra cycle but in-phase with the wave coming in the counter clockwise direction; signal feeds into arm (2). Similarly, the wave traveling clockwise that reaches arm (4) will be in phase with the wave traveling counter-clockwise from (1) to (2) to (3) to (4), each having traversed $3\lambda/4$. But at (3), we have a different situation. The signal that comes from (1) to (3) via the clockwise route travels a full wavelength, whereas the signal from (1) to (2) to (3) travels but a half wavelength. At (3) these signals will be equal and of opposite phase, thus cancelling each other. Hence, no signal goes to arm (3).

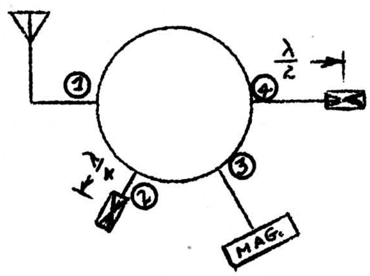
A further fact may be noted. If a signal is fed in (1), it divides between (2) and (4) as noted. But the relations are easily seen to be such that the resultant signals in (2) and (4) are 180° out of phase with each other. If we had fed the signal in (3), then we would have found that no signal goes to (1) but the power divides, as before, between (2) and (4). In this case, however, the resultant signals in (2) and (4) are <u>in</u> phase.

This is important because we can reverse the direction of our signals (since network is linear and passive). Hence, if we put equal signals (of identical frequency) in (2) and (4), then power will divide between (1) and (3) in a way depending on the phase difference of the two inputs. If they are in phase, all power will go to (3). If they are out of phase, it will all go to (1).

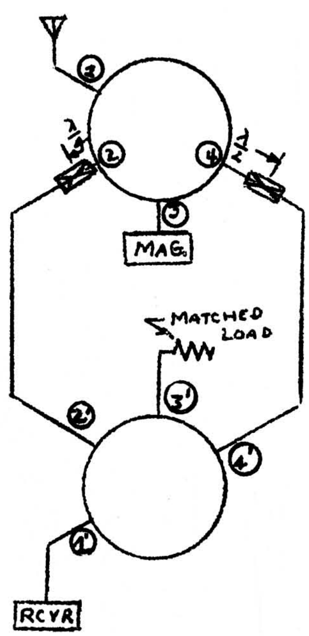
To complete the picture, we can, of course, interchange the roles of (1) and (4) and of (2) and (3) (simultaneously) in the above.

The other hybrid circuits--the magic tee, the directional coupler, the circular magic tee, etc.--behave the same way, (except that the directional coupler does not usually divide the power equally).

The duplexing function can be achieved efficiently by using two hybrid circuits. Consider the connections of one part of the duplexer first. To the (1) arm attach the antenna. Put the magnetron on the opposite arm (3). The magnetron then does not couple directly



with the antenna. The power from the magnetron divides instead between arms (2) and (4), and resulting signals are in phase. A TR tube is put in arm (2) and another in arm (4) but one of the TR's is spaced $\lambda/4$ further from the ring than the other. The power from the magnetron fires the TR's and is reflected back to the ring.



Because of the difference in the distances from the ring to the two TR's, the energy from the (2) arm comes back out of phase with the energy from the (4) arm and hence, all the power goes out to the antenna. To dispose of any leakage power, we hook on to the first rat race the other half of the duplexer. In doing so we make up for the $\lambda/4$ difference in arms (2) and (4) so that the distance to arms (2') and (4') are equal. Power from the magnetron was originally coming down the arms in phase, so that it arrives at the corresponding arms of the second rat race still in phase. Hence, the leakage power will travel out to arm (3'), where we can absorb it in a matched load. Theoretically this disposes of all the leakage.

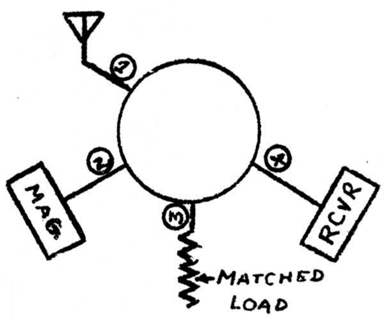
At the end of the transmit cycle, the antenna receives a signal which is fed through arm (1) to the ring. There it splits into two equal signals traveling in counter directions. These two signals feed through arms (2) and (4), out of phase and arrive at arms (2') and (4'), still out of phase. Since power fed through arms (2') and (4') 180° out-of-phase will all go through arm (1'), the receiver is connected to the last named arm.

The major advantage of the balanced duplexer is the elimination of leakage. It is theoretically possible to balance out any effects of leakage. There is also considerable advantage from the standpoint of the signal-to-noise ratio, particularly with respect to the noise generated by TR tubes.

One of the disadvantages to the balanced duplexer is that the TR tubes in the arms must be carefully matched and meticulously spaced to effect the correct phase difference. This means accurate knowledge of the position of the short circuits when the TR's are fired, and, hence, close tolerances on them. Otherwise, several misfortunes can occur. For example, when the magnetron fires, incorrect positions of the shorts can send a high standing wave ratio back into the magnetron, damaging it. The leakage may have its phases somewhat messed up and enter the receiver. Also, if the duplexer is used with high power circuits, the rat race tends to magnify any reflections. It does not take too much of mismatch to get things arcing merrily.

There are a number of different versions of the form of balanced duplexer discussed — various types such as the circular polarization and turnstile duplexers — but they need not enter the discussion at this time. What is, perhaps, of interest at this point is the fact that a TR tube is not essential for duplexing.

The linear balanced duplexer uses no TR. Either a rat race or magic tee may be used. Let us consider the rat race. If the magnetron is put, say in arm (2), power is going to divide between



arms (1) and (3). Hence, the receiver is put at arm (4), so that no power will be coupled directly from the magnetron to the receiver. The antenna may then be put at arm (1), and a matched load at arm (3). It will be noted that half of the magnetron power is wasted; half of it goes to the antenna and half to the load where it is absorbed. This waste of power is true also in the case of the received signal, which divides between the magnetron and receiver. That energy that enters the magnetron is either absorbed or reflected. If reflected, it is split between the matched load, which absorbs it, and the antenna, which radiates it. Thus, another half of the power is lost during reception, or a total of 3/4 ths. In other words there is a 6 db loss, which has to be tolerated if no TR is to be used. It can be shown that this will necessarily be true of any linear duplexer.

Summary:

To review, then, there are two main types of duplexing circuits called "branched" and "balanced". The former is quite straightforward. A TR is used to block the signal to the receiver, and an ATR, usually, but not always, to prevent undue loss of received signal because of the transmitter. The ATR can be avoided if the line length from transmitter to the junction is properly adjusted.

Connections to the line can be either series or parallel although only the former is mechanically feasible for coax. Impedances are adjusted by making use of the fact that a quarter wave line effectively "inverts" its terminating impedance, changing a short-circuit to an open and vice versa.

Balanced duplexers depend on the directional properties of "hybrid circuits", of which the rat race, magic tee and directional coupler are common examples. These are devices with four input or output circuits. Their central property is that a signal into any one divides between two others but does not go to the fourth. By combining two such circuits, with two TR's between, a duplexer can be designed which is theoretically considerably more efficient than a branched duplexer. This, however, makes TR design more difficult since the two tubes used must be carefully matched - particularly as regards the "position of the effective short-circuit". It also demands broad-band TR's, since the adjustments that would be needed, were narrow band TR's used, could not practically be made in the field. Given suitable TR tubes, however, the balanced duplexer is considerably better than the branched type.

OPERATION AND DESIGN

OF TR AND ATR TUBES

Lecture No. 4

Measurements

This lecture will be concerned with the measurements done on TR and ATR tubes. We shall attempt to show what the significance of these are, and outline how they are done. Since most of you are already familiar with much of the procedure, we shall be very brief on all but Q measurements and leakage. The former we will consider in detail because the concept of Q has profound implications in the design of TR and ATR tubes. Also, it is tricky. It is easy to be confused about it.

Leakage we shall consider in detail to emphasize the gross inaccuracies inherent in the present technique of spike measurements in broad band TR's. Spike measurements are a useful control - nothing more. We shall emphasize this because it is a point that has, on occasion caused considerable confusion.

Definitions: Q:

We shall first consider Q. Quite a number of different Q's are used. The letter Q originally stood for "quality". It dates back to the early days of ham radio when the higher the Q of a coil, the more expensive it was.

In the present, particularly in regard to TR and microwave structures, we often work very hard indeed to get low-Q structures, so that the connection of Q with "quality" has become somewhat tenuous, but the designation is still used.

For the definition of Q there are a number of relations that could be used, but the basic definition is:

$$Q = \frac{2\pi \text{ x power stored}}{\text{energy per cycle}}$$

We have not specified the particular energy being considered. The different Q's are defined in terms of the different kinds of energy involved.

The Q that is common to all elements is the unloaded Q, usually written Q_u.

$$Q_u = \frac{2\pi \text{ x power stored}}{\text{energy dissipated internally}}$$

The energy dissipated internally includes I^2R, dielectric, and, occasionally, radiative losses. Its significance is that it sets an upper limit to the loaded Q's.

The Q with which we are primarily concerned in TR tubes is the doubly loaded Q, usually designated with the subscript 2L.

$$Q_{2L} = \frac{2\pi \text{ x power stored}}{\text{energy delivered to the load + energy to the generator + energy dissipated}}$$

A certain amount of energy is being stored in the cavity, and three things can happen to it. (1) It can be dissipated internally; (2) it can go out through the output to the load; (3) or it can return through the input to the generator.

Q_{2L} is a measure of the combined effectiveness of these dissipations. For a constant drive - constant power in - it is a measure of the stored energy and, hence, of the resonant build-up that occurs in the structure. Since, as we have seen, the build-up determines the speed and effectiveness of a TR as a protective switch, its Q_{2L} is very important in its design.

For cavities with a single coupling to the line, such as ATR's, the significant Q is singly loaded, Q_L. The stored energy can either be dissipated internally or delivered through the single output. This is also true of such other devices as magnetrons.

$$Q_L = \frac{2\pi \times \text{power stored}}{\text{energy delivered to the output line + energy dissipated}}$$

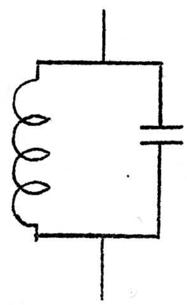
The fourth Q, which is more important to magnetrons than to TR tubes, is the external Q, Q_E. For this the definition is

$$Q_E = \frac{2\pi \times \text{power stored}}{\text{energy delivered}}$$

Note that this Q does not include the energy dissipated internally.

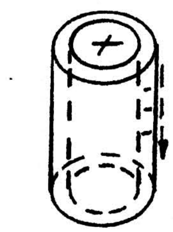
The above four are the main Q's. There have been other Q's defined at times for special usages, but they are of little consequence for the purpose of this lecture. The ratios of power stored to energy delivered are important because the more power that is being stored, the harder it is to match into the device. There is a law known as Bode's Law which puts an absolute limit on the band over which a given element can be matched into a given generator. The limit can be expressed in terms of the energy stored in that element.

The question of how the energy is stored in an element is not primarily important to design work, but it is of some interest. Beginning with an ordinary tank circuit, it will be remembered that the



energy can be stored either by charging the condenser, or by current through the coil. At the resonant frequency, the energy stored in such a circuit is oscillating from one method of storage to the other.

The same considerations apply in a general sort of way to a microwave

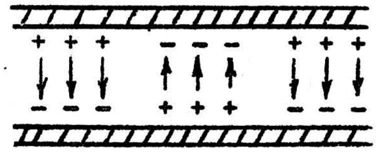


cavity. At a given time we will have maximum voltage developed across the cavity. A quarter cycle later electrons run around the edge of the cavity; and half a cycle later maximum voltage across the cavity is again developed, but this time in an opposite direction. The flow of the electrons through the metal, leaving a positive charge behind increases the potential energy of the cavity. It takes work to do it and represents stored energy.

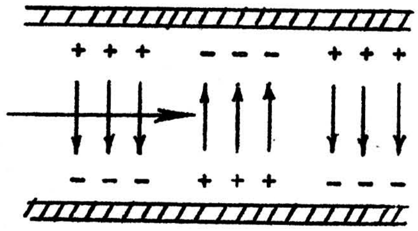
On the other hand, when the electrons are moving fastest, and there is no net separation of charge, they create a moving field. In effect we may consider that the energy is stored as momentum, although the actual effect is to create a varying magnetic field.

The energy stored in the cavity oscillates between these two forms - one an electric or potential energy field, the other a magnetic field, analogous to kinetic energy.

It is still not obvious how a resonant window for example in a waveguide stores energy. Energy can be propagated down a waveguide in

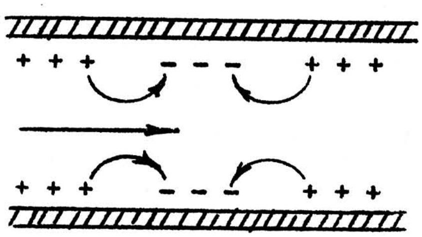


a number of different ways. All of these can be described in terms of the motions of the electrons in the walls. Supposing we were able to photograph the net charge on the walls at a given instant. In the normal mode we would see the charge as illustrated. If we took movies, these



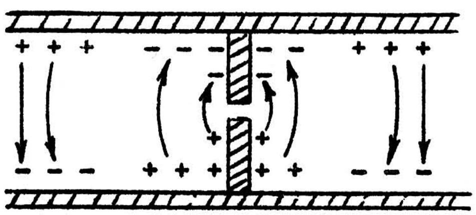
concentrations of + and - charge would appear to be moving in the direction of the propagation. This does not mean that the electrons are so moving, any more than that the molecules of water need do more than oscillate around a point to produce moving waves on the surface.

The other modes correspond to different distribution of charges. In one, for example, the charges on the top and bottom have the same polarity. Again these charges move in the direction of propagation. These various ways in which energy can be propagated are called the "modes" of the waveguide.



For a given waveguide and a given mode, there is a minimum frequency for which each type of propagation is allowed. It is called the "cut-off frequency" of the mode in that size waveguide. If you try to drive the waveguide in any mode at a frequency lower than its cut-off, you find you cannot do it. Or, rather, you can, but it does not propagate. The amplitude, the magnitude of the resulting charge concentrations, die off exponentially as they go away from the driven point. The energy that you try to drive into the mode is reflected back to the source.

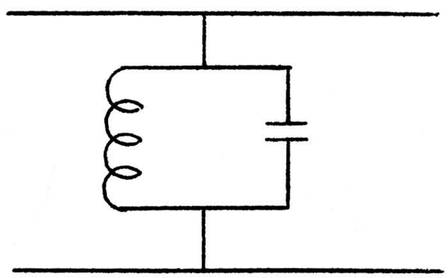
These modes can, however, store energy even though below cut-off. Charges are separated in the neighborhood of the driving source and this represents energy which was absorbed from the generator at the time it was turned on. This energy does not propagate--hence it is stored.



When a wave in the normal mode hits a resonant window in the waveguide, there are certain special boundary conditions that have to be met. If the metal part of the window were a perfect conductor, it would not be possible to establish a voltage difference along the surface of the window and in its immediate neighborhood. If such a gradient were established, current would immediately flow through it, and would flow to that extent necessary to reduce the gradient to zero. The last statement is another way of saying that the metal will short-circuit the field. However, the normal mode necessarily has a uniform voltage field which extends all the way across the guide. Consequently, when such a wave hits the window, it cannot in itself satisfy the boundary conditions. Then the metal part of the window does short circuit part of the field, and in so doing sets up some of the higher modes. They will not propagate down the line. Nevertheless, energy does go into them, energy that stays in the immediate neighborhood of the window, representing stored energy.

Measurements, Q:

Q's are measured by the effect of the element on the line. There are various methods of evaluating the effect. Most of them depend, either directly or indirectly, on finding the half-power point. We can approximate most of the microwave networks with a simply shunt



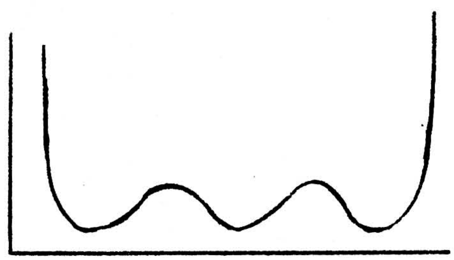
resonant circuit, at least over a narrow frequency band. Then, if we denote by f' and f" the two frequencies at which the power transmitted is one-half the power put in the circuit, we can find the relation between these two frequencies and the center frequency, f_o. By working through the network it will be found that f' and f" are related to the center frequency in such a way that the ratio of the center frequency to their difference is equal to the loaded Q, either singly or doubly loaded, depending upon how the circuit is loaded.

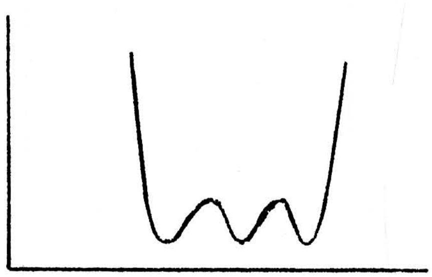
$$Q_{(2)L} = \frac{f_o}{f' - f''}$$

There are various other indirect ways of measuring loaded Q, such as measuring the rate of change of the susceptance in the neighborhood of the resonance. Also the "Pease" method of measuring standing wave ratios in the neighborhood of resonance. They may all be referred back to the half-power point method.

These Q's are important. In a narrow-band tube the Q of the tube (doubly loaded) determines the band over which the tube may be used without retuning. Since narrow band tubes are mostly tunable, there is no special concern about the ultimate bandwidth usability. It is important, though, that the tube shall pass a wide enough band so that any slight variation in the frequency of the magnetron will not throw the latter completely out of the passband of the TR.

In a broadband tube it is the Q of each element which determines the design. For example, a five-element tube might have windows with Q's of 2.1. The other elements would have Q's of 4.5 (These figures are approximately correct.) A tube designed to give these Q's will have a particular bandpass characteristic. If we put a flat load on the output, and plot the SWR against the frequency, the bandpass obtained will be as pictured. If the value of each Q is doubled that is, if a tube is constructed in which the windows have Q's of 4.2, other elements, Q's of 9.0, then the bandwidth will be halved. The height of the bumps would be the same in each case. The whole curve is simply reduced in width by the factor by which the Q's were multiplied. The shape of the curve depends only upon the relation to each other of the element Q's, assuming all elements are tuned to the same frequency. Thus, the Q's of the elements are the factors upon which the whole design of the TR hinges.





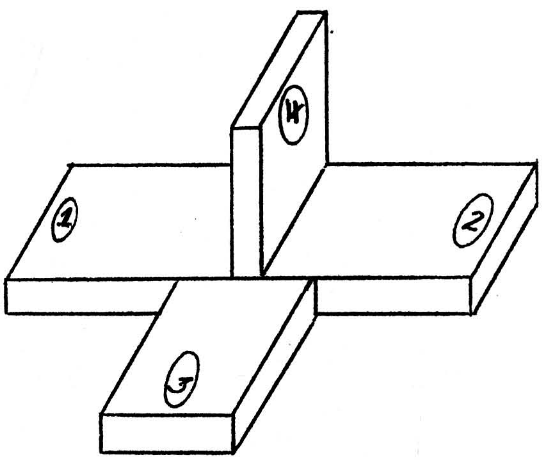
<u>Other Low-Level Measurements</u>: (Insertion Loss, Interaction Loss)

Insertion loss is defined as the loss in signal that results from replacing a matched load by the tube in question. A detector in the line measures the amount of power going into the load. The circuit is opened and the tube inserted so that the detector then reads the power out of the tube. The difference between the two readings represents the amount of power lost in the tube. It generally includes the power reflected by the tube. However, as far as TR tubes are concerned, the definition is more restricted in meaning. Insertion loss measurements are confined to center band, where presumably, no power is reflected. Furthermore, the keep-alive is not turned on, so that the insertion loss, then, measures only the I^2R and dielectric losses internal to the tube and in its windows.

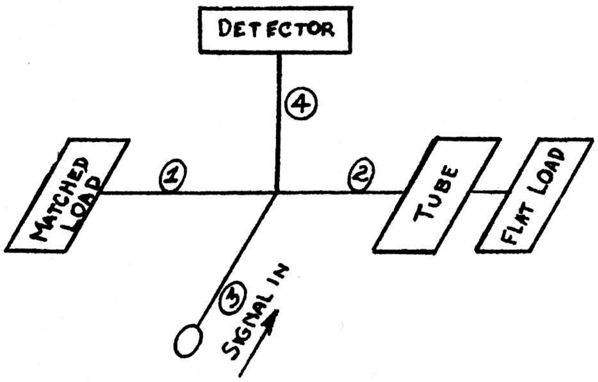
The interaction loss is measured in the same manner as the insertion loss, except that in the case of the former the keep-alive is turned on. The interaction loss is the difference in insertion loss with the keep-alive on and with the keep-alive off. It is, presumably, a measure of the power coupled out of the keep-alive. If the interaction loss is too high it means that the keep-alive is too closely coupled. Normal procedure in such circumstances consists either of pulling the keep-alive back of the gap, or changing the gas fill, or in some other fashion decoupling the keep-alive.

Impedance Bridge (Triple Pipper):

The SWR, insertion and interaction loss as discussed represent the principal low level measurements made on TR tubes. In connection with the getting of these values there is a handy device worth noting at this point, the Triple Pipper. The basic element of the Triple Pipper is the Magic Tee, a type of hybrid circuit described in the last lecture. The Magic Tee has the same properties as the rat race and the directional coupler. It consists of a piece of waveguide coupled into another section of waveguide on the narrow side, and a third section of waveguide set with its axis perpendicular to that of the other two sections. If we label the sides (1) and (2), and the two arms (3) and (4), we can look into the properties of the unit. As in the rat race (3) couples into (1) and (2), but not into (4). Signals of equal amplitude introduced in-phase to (1) and (2) will all go to (3). If the signals are introduced out of phase they go to (4).



The Triple Pipper is used as an impedance bridge in the following fashion. Signal is sent through arm (3). A matched load is placed in (1). The tube to be measured is set in arm (2) with a flat load behind it. Since all power from (3) divides between (1) and (2), a detector is placed in (4). Assuming that the circuit is



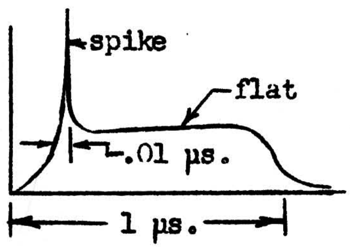
well balanced, no signal will reach arm (4) directly from arm (3). All the power reaching (1) is absorbed by the matched load. Signal reaching arm (2) is either absorbed by the load on the output of the tube under test, or is reflected. If reflected, the energy splits between (3) and (4) and half of it is detected at (4). Thus we have a way of comparing the impedance of the matched load with that seen looking into the tube.

In actual testing use with TR's, three signals are sent in through (3). The signals are either pulsed or superimposed on each other. This permits us to compare the impedance of the TR tube at three different frequencies simultaneously. The frequencies selected usually are those for which SWR's have been specified. Assuming that the SWR of the tube does not jump out of bounds at points between those three, the Triple Pipper gives a rapid approximate bandpass characteristic.

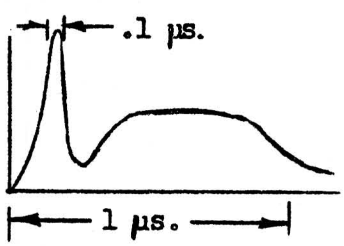
We might mention here, that we have built a kind of generalized triple pipper for the 1B58. Instead of inserting only three frequencies, a sweep oscillator was built to sweep over the entire band. Using a magic tee as before, and putting the detected signal on a synchronized scope, the entire bandpass is presented continuously.

High Level Measurements — Leakage Power:

Of the high level measurements made on the TR, probably the most important is leakage. The type of leakage through a TR tube depends very much on the tube. In the narrow band tube, the high Q, single element TR, the input pulse rises rapidly. The tube breaks down and the arc develops. Somewhere later on there is the end of the high powered pulse and the end of the leakage pulse. The sharp rise and fall at the beginning of the leakage pulse is known as the "spike". The remainder is called the "flat". The spike is actually a spike with a time duration in the neighborhood of .01 μsec. or less. The flat normally is quite flat.

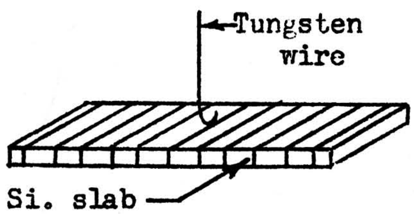


With the broad-band tubes, the rise occurs at the beginning of the leakage pulse in the same way as it does with the narrow-band TR's. But it breaks down with somewhat different results, giving a marked dip in energy level between the spike and the part that corresponds to the "flat" of narrow-band tubes. The spike here has a time duration of the order of .1 to .2 μsec. Also, the spike of the broad-band tube does not have the clean outline of the narrow-band tube. In the broad-band case, there is definite evidence of structure on the spike, that is, oscillations of some kind. As yet, there has not been built a synchroscope with a sufficiently wide band amplifier to analyze the fine structure in the spike. (It would need about 1 Kc to 100 Mcs.) The "flat" is far from being flat. The difference in energy level from

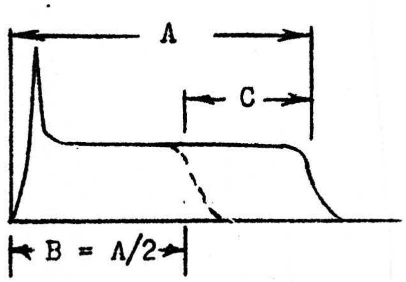


the dip to the top of the flat may be as much as 20 db. That makes things difficult in obtaining the values of the leakage pulse.

At this point we must digress for a brief interval to consider the burnout properties of crystals. A crystal can be burned out either by steady application of too much power, or by the quick application of too much energy. In a silicon crystal, for example, there is a silicon slab and a tungsten wire making contact with the slab. Most of the heat is generated in the neighborhood of the contact. If the heat can be dissipated fast enough compared to the duration of its generation so that thermal equilibrium is essentially reached, then we are interested in power. If, however, the energy is applied very fast so that none to speak of is dissipated, then the energy itself will burn out the crystals. Hence, it is necessary, if the TR tube is going to protect the crystals that not only should the flat be less than a given power level, but also that the spike should be less than a given energy level. The difficulty comes in separating for measurement the two components of the leakage pulse.



In the narrow-band case the components of the leakage pulse can be evaluated with some degree of precision. If we measure the length of time elapsed from start to finish of the pulse (A), and then apply an input pulse one-half the duration of the full pulse, presumably the portion of the value representing the spike will be the same in both measurements of the one pulse, while the flat will extend only half as long.

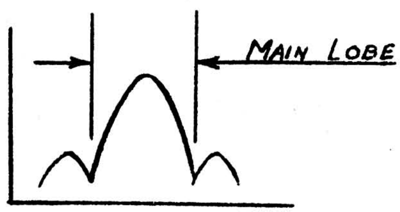


Then, if we take the difference, obtaining the amount of leakage during C, and subtract that from the second measurement, the leakage during B, all we are going to have left is spike. This appears to be a quite satisfactory way of measuring spikes for narrow-band TR's. The results obtained by it are consistent. They agree reasonably well with results obtained by other methods.

Unfortunately, such an argument cannot be applied to the type of leakage usual to broad-band tubes. The same measurements, the same calculations are generally made for broad-band TR's. There is no harm in making the measurements. In fact, they serve as an extremely useful control, indicating when manufacturing processes are getting out of hand. The numerical answer, however, has little relation to the actual spike energy. Actually, the calculations will often reveal "negative" spike energies. This is somewhat disconcerting, not because it really means anything bad, but because it is hard to convince non-technical people that it doesn't. The difficulty, of course, is that the flat is not flat. The dip after the spike means that from B you should subtract something significantly less than C. When you subtract C, you are overdoing it and cancelling some or all of the spike. Hence, the negative spikes.

There are other methods for measuring the spike of broad-band TR's. In one of them a small fraction of the power from the magnetron is used to cancel out the flat. The main advantage of applying this method is that it assures a positive spike energy and eliminates explanatory notes. It still does not help the dip.

Another method utilizes the spectrum of the leakage pulse. Regardless of its shape any pulse has a spectrum more or less as



depicted here. The width of what we might call the "main" lobe is dependent upon the width of the pulse. It is twice the reciprocal of the duration of the pulse. That is, if for example, we have a one microsecond pulse, the width between the vertical bars in the diagram is 2 Mc. A half microsecond pulse would give a spectrum width of 4 Mc. The entire leakage pulse is usually in the order of one microsecond, so that its spectrum will be confined principally within about 2 Mc. of the center frequency. On the other hand, the spike of the broad-band TR will spread over about 20 - 40 Mc. It is plausible, theoretically, to filter out the middle two megacycles, leaving only most of the spike energy. This method has not as yet reached a practicable working state.

In connection with the leakage power, we might note the direct coupled leakage, which is of some use in design work. Direct coupled leakage measures that leakage through the tube when all the gaps are short circuited. Frequently the cones are literally soldered together and the leakage going through measured directly. This measurement is made principally on narrow-band TR's. It is useful in indicating the ultimate that can be obtained with a particular design. This technique might very well be employed more frequently than it is. Particularly in attenuator tubes.

Other High-Level Measurements:

The following definitions will complete the survey of high-level measurements.

Harmonic Attenuation is the attenuation that is obtained through the tube at a higher frequency, normally a harmonic. This measurement is important because magnetrons, being non-linear devices, generate harmonics, which can burn out the crystal just as easily as the fundamental frequency. In some instances it has been necessary to use a pre-TR to protect the crystal from harmonics.

Arc Loss represents the amount of high level power lost in the arc. It is employed mainly for ATR's.

Minimum Firing Power is the lowest power required to fire the tube. It is not normally of much interest because in present day applications we are more concerned with high powers than we are with the lowest power. This measurement, though, does come up once in a while.

Recovery Time is the time that it takes a tube to recover from being fired by a high-level signal. The magnetron pulse kicks the TR tube into an ionized state. At the end of the magnetron pulse the ions have not yet disappeared. The ions are of no particular importance, but the electrons are. So long as the electron concentration is high in the tube, little r-f signal will pass through. Recovery time, then, is the length of time elapsed from the end of the magnetron pulse to such time that the tube will pass the r-f signal with less than a given loss, usually 3 or 6 db.

In the Keep-Alive various things are measured. Interaction Loss was defined earlier in this lecture. The Voltage Drop is that voltage necessary to keep the Keep-Alive operating with a given Keep-Alive Current. The Break-Down Voltage is the voltage required to start the Keep-Alive. This may be much higher than the Voltage Drop - particularly if no radio-active material has been put into the tube to supply initiating ions. Note that the time within which the tube is expected to fire must be specified for the Break-Down Voltage since the process is a statistical one.

For ATR's certain other measurements are made. The "Equivalent Conductance" is the conductance (reciprocal of resistance) that the ATR presents when it is fired. The "Tuning Susceptance", which is a measure of the accuracy of its tuning, is the imaginary component of the equivalent admittance of the tube when fired. The "High Level VSWR" is a direct measure of how well the ATR passes the transmitter pulse.

OPERATION AND DESIGN

OF TR AND ATR TUBES

Lecture No. 5

Solution of TR Equivalent Network By Matrix Algebra

TR Equivalent Network Redefined:

The material to be presented today will be new, not so much in subject matter as in form. That is, part of the lecture will review information in a manner already known to those of you with electrical engineering degrees. The other part of the lecture will cover the information from a point of view probably unfamiliar.

The difference in viewpoint is not entirely a trivial one. The trend in mathematical development is towards finding a more efficient form for stating relationships. Generally, in mathematics a simpler expression aids in the discovery of fundamental properties. New relations became apparent and problems of added complexity can be handled.

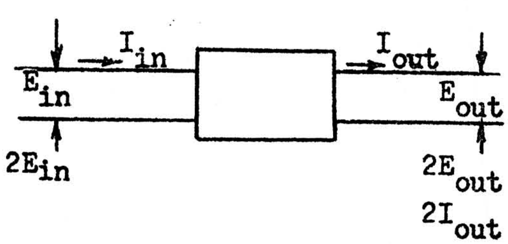
We are interested in analyzing the behavior of the TR tube. A common method of analysis is to consider the equivalent circuit of the TR as a bandpass filter. As explained once or twice before, this terminology is somewhat anamalous because in the performance of that filter there are some very distinct, significant departures from conventional filter theory. A better designation would be to call the TR tube a passive linear network.

It would perhaps be well to go through the title, "passive linear network", and discuss the meaning of each term. Starting with "network" the work implies a black box from which various terminals—in this case two pairs—lead. To any input there corresponds a definite output. The box may or may not be active. In other words, there may or may not be a signal in the box independent of input, but the signal that comes out of the box is modified by that put in.

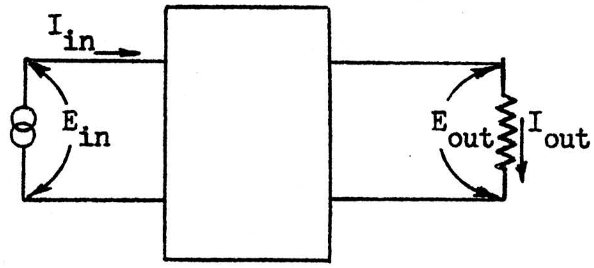
The term "passive" implies that no signal will go out of the box unless a signal is put in. The term connotes more than that, though. It means also that the energy out does not exceed the energy in, regardless of what the input may be or how the input is loaded. It is possible to design a Class A Amplifier, with or without feedback, from which, neglecting d-c, the signal out is both greater than and proportional to the signal fed in. That circuit is linear, but not passive.

The linearity qualification means that the signal put out by the box is proportional to the signal put in.

Assume we have designated one pair of terminals of the network the input, and the other the output. At the input we have a voltage, E_{in} and a current I_{in}. At the output, we have E_{out} and I_{out}.



If we double the input, make it $2E_{in}$, we double the output readings, giving $2E_{out}$ and $2I_{out}$ so that, with a given load, all quantities are proportional. The proportionality constants between them are the various impedances and admittances of the network.



E_{out}/I_{out}, which is determined only by the load, and not at all by the network, is the load impedance. The reciprocal, I_{out}/E_{out}, is the load admittance. Similarly, the input or "driving point" impedance is E_{in}/I_{in}, and its reciprocal, the admittance. Note that this is a function not only of the network, but also of the load impedance.

Two other ratios are also used. E_{in}/I_{out} or E_{out}/I_{in} (which are equal by the so-called "reciprocity theorem") is called the mutual" or "transfer" impedance. $I_{in}/E_{out} = I_{out}/E_{in}$ is similarly the mutual or transfer admittance.

Basic Elements of Networks:

In ordinary electrical work the networks with which we are concerned involve three elements or combinations thereof, inductance, capacitance and resistance. There are also interactions possible among the individual elements. Two inductances which directly interact we call a transformer. Theoretically, the same condition can be brought about with capacitors but such devices are rarely used. What, then, is the difference among the three elements?

For those who have not been introduced to the exponential way of describing the behavior of the three components, we will take a quick review at this point. If we put into a circuit an alternating voltage, we can define that voltage in terms of the maximum voltage and a sinusoidal time dependency.

$$E_{in} = E_{max} \cos \omega t$$

where ω is the usual $2\pi f$, f being the frequency and t, the time. At the output there will be a current which also has a maximum value, and which can be defined in terms of its maximum value and the same time dependency, except that the information at the output may be ahead or behind that of the input. Thus we can say

$$I_{out} = I_{max} \cos (\omega t + \phi)$$

When the input E_{in} has reached its peak, the output I_{out} may have either gone by its peak value, or not quite reached it. This difference in timing is represented by the ϕ of the second definition.

A convenient way of expressing $E \cos \omega t$ is to write it as the real part of $\left\{ E(e^{j\omega t}) \right\}$. $e^{j\omega t}$ is defined as

$$e^{j\omega t} = \cos \omega t + j \sin \omega t$$

where $j = \sqrt{-1}$.

When we take the real part of $e^{j\omega t}$, therefore, we simply remove the $j \sin \omega t$. The current $I \cos (\omega t + \phi)$ can be abbreviated in similar fashion to the real part of $\left\{ I\left[e^{j(\omega t+\phi)}\right]\right\}$. The convenience of this terminology is illustrated when it comes to the definition of impedance, Z, (input, output, or transfer).

$$Z = \frac{\text{voltage}}{\text{current}} = \frac{E\, e^{j\omega t}}{I\, e^{j\omega t+\phi}} = \frac{E}{I} e^{-j\phi}$$

where the $e^{-j\phi} = \cos \phi - j \sin \phi$.

The impedance, then, has two parts. One, $\left\{\frac{E}{I}\cos\phi\right\}$ indicates the proportionality between those parts of the E and the I which do not have a difference in timing. It is called the "resistive component" and is usually designated by R.

The other part $\left\{-j\frac{E}{I}\cos\phi\right\}$, is the ratio between the parts of E and I which have a difference in timing of a quarter cycle. Leaving off the j, the terms $\left\{-\frac{E}{I}\cos\phi\right\}$ is called the "reactive component" and usually designated by X. By combining resistive and reactive components, we can express the relation between any voltage and current (assuming linearity) regardless of the ratio of peak values and the time difference involved.

Likewise, we define the admittance, Y,

$$Y = \frac{\text{current}}{\text{voltage}} = \frac{I\, e^{j(\omega t+\phi)}}{E\, e^{j(\omega t)}} = \frac{I}{E}\, e^{j\phi}$$

$$= \frac{I}{E}\,(\cos\phi + j\sin\phi)$$

The two components, real and imaginary, of admittance are called the conductance $(G = \frac{I}{E}\cos\phi)$, and the susceptance $(B = \frac{I}{E}\sin\phi)$.

If we examine the frequency dependency of the effects of the various elementary circuit components, we find another basic difference. Resistances are independent of frequency, while the reactance or susceptance of an inductance or capacitance varies either directly or inversely with the frequency. We may summarize these properties of the basic network elements by the following table.

Element	Component of Complex Expression	Time Relation	Frequency Dependence	
			As Imped.	As Admit.
R	Real	No change	Independent	Independent
L	Imaginary	$\pm \tau/4$	ωL	$-1/\omega L$
C	Imaginary	$\pm \tau/4$	$-1/\omega C$	ωC

(τ = period of driving force)

Elements of TR Tubes:

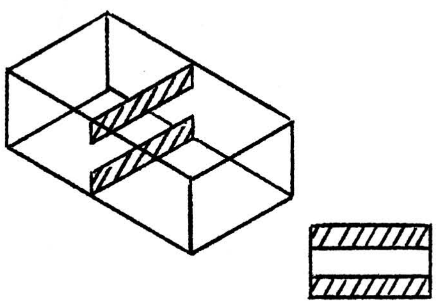
The elements we work with in TR tubes are a good deal more complicated than these elementary components. In TR's we use irises, windows, cones and, in general, elements which can be put in a wave-guide. We need to find relationships for these more complicated elements. How much of our table can be generalized to apply to them? The answer is that a good deal can be. By considering various subtle implications of passivity and physical realizability, it can be shown that the real part of any impedance or admittance can always be expressed as a function of only the even powers of the frequency. The function may be independent of frequency ($n = 0$), may be squared, to the fourth power, inverse squared, etc. Reactance and susceptance, on the other hand, are functions of odd powers of the frequency.

$$\text{real components} \begin{pmatrix} R \\ G \end{pmatrix} = F(\omega^{2n})$$

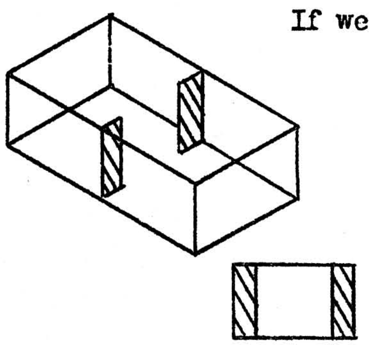
$$\text{imaginary components} \begin{pmatrix} X \\ B \end{pmatrix} = F(\omega^{2n+1})$$

Both of these relations are perfectly general. They apply not only to any component, but to any linear network. Even if non-passive, and at all frequencies.

If we have any kind of an iris in a waveguide which does not dissipate power, it is necessarily expressible as a function of an odd power of the frequency. The exponent 2n+1 can vary from $-\infty$ to $+\infty$, which does not tell us too much about it. However, by suitably



choosing the geometry of the element, much of the complexity can be eliminated. If we study an iris like the one pictured, we find, within a wide frequency range, that its susceptance is positive and that its variation is approximately linear with frequency. Hence, it may be said to be capacitative. It is not correct to say that it is a shunt capacity, for its behavior is not exactly the same. But it may be considered as one with reasonable accuracy over reasonable frequency ranges.



If we use an iris such as this one, we find in a similar manner that, over a wide band around a given operating frequency, the only significant term in the susceptance is

$$B \doteq -a/\omega$$

where $\underline{a}$ is a positive constant. The performance of the iris, then, approximately satisifes our definition of inductance.

Likewise, in writing equations for r-f loads, we can neglect the frequency dependence entirely, obtaining resistance.

Thus the complicated components of the TR may, over restricted frequency ranges, be used as capacities, inductances and resistances. The question now is how to calculate the results of the combinations of these elements.

General Expression for Linear Networks:

We have been considering passive systems. Another way of stating linearity is that the input voltage may be written as a linear combination of the output voltage and output current. As,

$$E_{in} = a(\omega)E_{out} + jb(\omega)I_{out}$$

And the input current may be written as a linear combination of voltage and current outputs.

$$I_{in} = jc(\omega)E_{out} + d(\omega)I_{cut}$$

The factors a, b, c, d are functions of frequency, but not of current or voltage output.

The two equations just given define the network as linear. Should the network, for example, contain a crystal, the above equations would not hold, for the input voltage would not be a linear function of the output.

Let us examine more closely the make-up of the equations. The a expresses the dependence of the input voltage on the output voltage. Therefore, it is a pure number and can have any frequency dependance. On the other hand, the b, stating the relationship between input voltage and output current, is a mutual impedance and is bound by the odd and even functions of ω discussed before. The j-factor precedes the b and c because the systems that we are interested in TR tubes, are essentially lossless, and lossless components (pure inductances and pure capacitances) are represented by the imaginary expressions. The relationship $b(\omega)$ has the dimensions of resistance, whereas $c(\omega)$ has the dimensions of admittance.

Definition of Matrix:

Such equations completely and exactly (exactly, if set up correctly) describe the behavior of the system. It is pertinent to ask then, what is the significant information contained in the equations. We may deal with a variety of outputs. In such circumstances there will be a corresponding variety of inputs. We wish to know what in the equations

$$E_{in} = a(\omega)E_{out} + jb(\omega)I_{out}$$

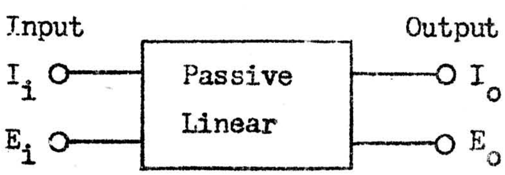
$$I_{in} = jc(\omega)E_{out} + d(\omega)I_{out}$$

is fundamental to the network and independent of everything but the network. The answer is, of course, — the constants. Let us put them in brackets for more convenient examination.

$$\begin{Vmatrix} a(\omega) & jb(\omega) \\ jc(\omega) & d(\omega) \end{Vmatrix}$$

The symbology used to write down the constants is known as a matrix. It is simply a short-hand notation and does nothing but summarize the equations. The form permits us to see relations which we might otherwise miss. The form also enables us to perform easily operations that would be complicated if the original equations were handled directly.

The matrix we have just defined is called the "transmission" matrix. It is so called because it tells how the conditions at the input are transmitted to the output, or vice versa. The order taken is a matter of which is considered the input. This matrix



is by no means the only one in use. We have four quantities to relate — input voltage and input current, output voltage and output current. We may express any two in terms of the other two as linear equations, abstract the constants and write down the matrix. For example, if we take the two voltages and express them in terms of the two currents,

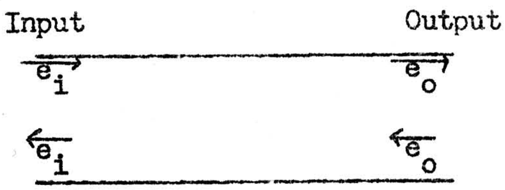
$$\begin{Vmatrix} \dfrac{\partial E_i}{\partial I_i} & \dfrac{\partial E_i}{\partial I_o} \\ \\ \dfrac{\partial E_o}{\partial I_i} & \dfrac{\partial E_o}{\partial I_o} \end{Vmatrix} \qquad \text{(voltages expressed in terms of the currents)}$$

we have the "impedance" matrix. If we reverse the order of dependency, i.e., relate the currents in terms of the voltages,

$$\begin{Vmatrix} \dfrac{\partial I_i}{\partial E_i} & \dfrac{\partial I_i}{\partial E_o} \\ \\ \dfrac{\partial I_o}{\partial E_i} & \dfrac{\partial I_o}{\partial E_o} \end{Vmatrix} \qquad \text{(currents expressed in terms of the voltages)}$$

we have the "admittance" matrix.

A matrix of some importance is the so-called "wave" matrix. It will be remembered that any condition in the transmission line can be resolved into a forward and a backward wave. The wave moving forward from the input, $\overrightarrow{e}$, will at the output set up a reflected wave moving backward, $\overleftarrow{e}$, down the line from the output. At both output and input there will be in general a combination of forward and backward moving waves. It is possible to derive the relationship between the waves, and set up the matrix which defines the dependency of the forward and backward waves at the input on those at the output.



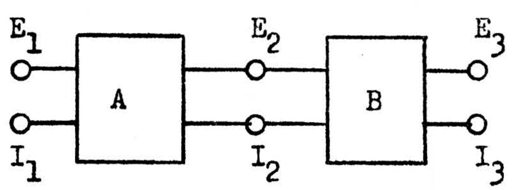
$$\begin{Vmatrix} \dfrac{\partial \overrightarrow{e_i}}{\partial \overrightarrow{e_o}} & \dfrac{\partial \overrightarrow{e_i}}{\partial \overleftarrow{e_o}} \\[2ex] \dfrac{\partial \overleftarrow{e_i}}{\partial \overrightarrow{e_o}} & \dfrac{\partial \overleftarrow{e_i}}{\partial \overleftarrow{e_o}} \end{Vmatrix}$$

There are, as a matter of fact, a three-fold infinity of matrices that can be set up, but most of them appear to be fairly useless. Those principally used are the ones we have indicated above. The impedance matrix is convenient in handling the combination of networks in series; the admittance matrix, of networks in parallel. Of major interest from the standpoint of TR tubes are the transmission and wave matrices since these are adapted to the combination of networks into a "ladder" structure. The two are equivalent, containing exactly the same information, so that we will confine ourselves to the transmission matrix.

In passing, though, it is worth noting that the wave matrix is basically related to another matrix used for different purposes, viz., the "scattering" matrix. The scattering matrix is of assistance where there are a number of branches off a wave-guide, and it is necessary to find out how the power divides among the various branches. The case of prime interest is that in which the branches each are connected to a matched load, so that there is no reflection with which to contend. Under such conditions the scattering matrix is used. It tells us how the power put into the waveguide scatters among the various outputs. It is written in terms of the waves, and is a simplified version of the wave matrix, with the backward wave eliminated.

Matrix Multiplication:

We are ready to examine in further detail the convenience of the transmission matrix. What happens when two or more networks are combined together? Let us say that we have one network, A to which we connect another network B.



At the input terminals of A we have input voltage and current which we will label E_1 and I_1. At the output terminals of A we will have output voltage and current E_2 and I_2, which also act as input to network B. At the output of network B we have voltage E_3 and current I_3.

The equations and corresponding matrix for network A are, considering E_2 and I_2 as the output voltage and current,

$$E_1 = a\,E_2 + jb\,I_2 \qquad\qquad \begin{Vmatrix} a & jb \\ jc & d \end{Vmatrix}$$
$$I_1 = jc\,E_2 + d\,I_2$$

where it is understood that a, b, c, d are functions of ω.

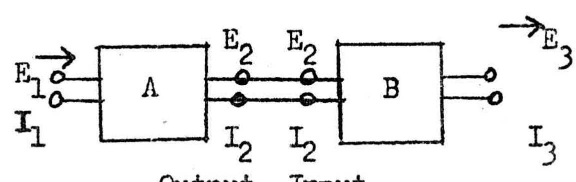
Considering E_2 and I_2 as the input voltage and current of network B, the equations and corresponding matrix are

$$E_2 = A\,E_3 + jB\,I_3 \qquad\qquad \begin{Vmatrix} A & jB \\ jC & D \end{Vmatrix}$$
$$I_2 = jC\,E_3 + D\,I_3$$

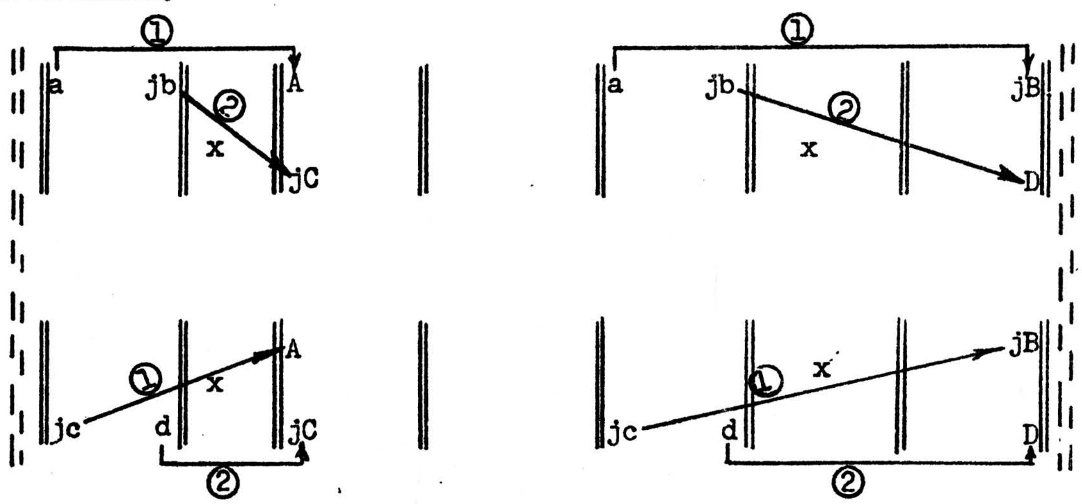
where A, B, C, D are functions of ω.

It is evident from inspection of the two sets of equations that the relationship between E_1, I_1 and E_3, I_3 is obtainable by direct substitution and algebraic manipulation. A speedier, less tedious solution is possible by the use of matrices. The expression of E_1, I_1 in terms of E_3, I_3 is the result of multiplying together the matrix for network B.

$$\begin{Vmatrix} a & jb \\ jc & d \end{Vmatrix} \times \begin{Vmatrix} A & jB \\ jC & D \end{Vmatrix}$$



The rules for obtaining a matrix product are defined not from the standpoint of what seems reasonable to intuition, but solely from the standpoint of what gives the correct result for the operations it summarizes. The product of the two matrices will give a third matrix whose terms are found by the proper combination of the terms of the first two matrices.



$$\begin{Vmatrix} a & jb \\ jc & d \end{Vmatrix} \times \begin{Vmatrix} A & jB \\ jC & D \end{Vmatrix} = \begin{Vmatrix} aA - bC & j(aB - bB) \\ j(cA + dC) & -cB + dD \end{Vmatrix}$$

(network A) (network B) (combination of networks)

The operations sketched above the matrix product are intended merely as an aid in following the rules.

The term in the upper left hand corner of the resulting matrix is obtained by multiplying the term in the upper left hand corner of the A-matrix by the term in the upper left hand corner of the B-matrix, (a x A), then adding to this product that obtained by multiplying together the term in the upper right hand corner of the A-matrix and the term in the lower left hand corner of the B-matrix, (jb x jC)

$$(a \times A) + (jb \times jC) = aA - bC$$

Generalizing, each term in the first row of the A — matrix, going across from left to right, is multiplied by the corresponding term in the first column of the B—matrix, going from top to bottom. The products are then added.

For the second term, the upper right hand of the resulting matrix, to the product of the upper left hand term in the A — matrix and the upper right hand term in the B—matrix, (a x jB), is added the product of the upper right hand term in the A—matrix and the lower right hand term of the B—matrix, (jb x D).

$$(a \times jB) + (jb \times D) = j(aB + bB)$$

In other words, each term in the first row of the A—matrix, going from left to right, is multiplied by the corresponding term in the second column of the B—matrix, going from top to bottom, and the products are added.

In like fashion, to get the lower left hand term of the combination matrix

$$(jc \times A) + (d \times jC) = j(cA + dC)$$

Or, each term in the second row of the first matrix, traveling from left to right, is multiplied by the corresponding term in the first column of the second matrix, going from top to bottom, and the products added.

And for the final term

$$(jc \times jB) + (d \times D) = -cB + dD$$

It is stressed that the rules for combining elements of the matrices were fitted to the correct answers. It might be observed that the same is true of the multiplication of ordinary numbers. However, the procedure was learned by most of us at an early enough age so that it now seems intuitive.

We have emphasized this point because matrix multiplication has some peculiar properties. For example, as we can see immediately, the answer depends upon the order in which the terms are multiplied. We have multiplied the matrix for network A by that for network B—i.e., A-matrix x B-matrix. If we reverse the order—i.e., B-matrix x A-matrix

$$\begin{Vmatrix} A & jB \\ jC & D \end{Vmatrix} \times \begin{Vmatrix} a & jb \\ jc & d \end{Vmatrix}$$

and apply the rule for getting the upper left hand element of the resulting matrix, we have

$$(A \times a) + (jB \times jc) = aA - Bc$$

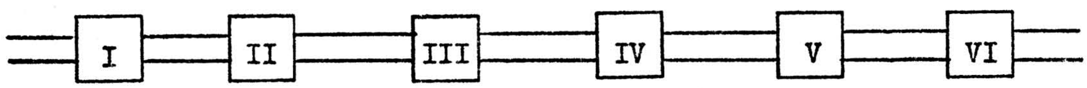
By inspection this answer differs from that obtained for the corresponding element in the first multiplication. And so on with the other terms. Mathematicians label this property by saying that matrix multiplication is "non-commutative." In other words, a change in the order in which the factors of a product are multiplied results in a change in the answer.

A-matrix x B-matrix $\neq$ B-matrix x A-matrix

This property is, of course, necessary for our purposes. For, if the order of the elements in a network are interchanged, the response of the network is certainly changed. Therefore, we could not use for the solution of networks a symbology which did not take into account the order in which elements appear.

<u>Application of Matrix Multiplications:</u>

If we have a series of network elements connected together, and we know the matrix of each element,



$$\|I\| \times \|II\| \times \|III\| \times \|IV\| \times \|V\| \times \|VI\|$$

$$I \times II, \quad (I \times II)III, \quad \left[(I \times II)III\right] IV, \quad \text{etc.}$$

we can by successive multiplications arrive at the matrix for the total network. The principal question then is: What can we do with the answer, now that we've got it? The key formula relates the insertion loss to the matrix elements. In db, in its most general form, it is:

$$L_{db} = 10 \log \left| \frac{aZ_L + dZ_g + j(b + cZ_gZ_L)}{Z_g + Z_L} \right|^2$$

where a, b, c, d are the elements of the transmission matrix with which we have been working $\begin{Vmatrix} a & jb \\ jc & d \end{Vmatrix}$

Z_g is the generator impedance, which generally is resistive, It may, however, be complex.

Z_L is the load impedance, which for our purposes is purely resistive. That is, we are assuming a purely dissipative load.

The insertion loss is the ratio, in db, of the power delivered to the load through the network to that that would be delivered were the network removed and the load connected directly to the generator output. If the insertion loss is known, the input standing wave ratio can be calculated. Hence, this formula gives us what we need for comparing a proposed TR structure to the specifications.

We can simplify the formula by utilizing the characteristics peculiar to our network, the TR tube.

In the first place, the TR being lossless, does not absorb power, so that the input and output short-circuit and open-circuit impedances must be purely reactive. If we short-circuit the output, and refer to the two equations summarized by the transmission matrix, we get, by dividing the expression for E_1 by that for I_1,

$$Z_{in} = a/jc$$

Since this must be a pure imaginary, if a is pure real, c must also be pure real. For open circuit at the output ($E_2 = 0$)

$$Z_{in} = jb/d$$

so that again, if b is pure real, d is pure real. From the condition of short circuit at the input we find that d/jc is pure imaginary. The condition of open circuit at the input gives us the fact that jb/a is pure imaginary. Thus, if a is real, b, c, and d are also real. We can, as a matter of fact, choose one element as we like. So we can say that, if a network is lossless, its matrix elements are real.

Then, since the driving impedance, in addition to the load, may be assumed purely resistive in our problem, the formula for insertion loss may be simplified to

$$L_{db} = 10 \log \left[\frac{(aZ_L + dZ_g)^2 + (b + cZ_gZ_L)^2}{(Z_g + Z_L)^2}\right]$$

We can simplify the formula further. For TR applications we can assume that the generator and load impedances are both equal to the characteristic impedance, R_c, of the line or waveguide. We may then "normalize" the components. To do this, we divide all impedances, including b of the matrix, and multiply all admittances, including d of the matrix, by R_c. The normalized values of Z_L and Z_g, in our case, are then unity. Hence the matrix becomes

$$\begin{Vmatrix} a & j\frac{b}{R_c} \\ jcR_c & d \end{Vmatrix} = \begin{Vmatrix} a & jb' \\ jc' & d \end{Vmatrix}$$

The insertion loss formula then becomes

$$L_{db} = 10 \log \left[(a + d)^2 + (b' + c')^2\right]$$

This gives us the insertion loss of the TR tube. From L_{db} we can calculate the standing wave ratio, which according to the specifications, must be kept below specified limits over a certain band.

<u>Summary:</u>

We have, then, developed the elements of network theory that are needed for the consideration of the low-level behavior of TR tubes. So far, we have been fairly abstract, dealing with generalized circuit elements and the mathematical tools useful in manipulating their interactions.

Next time we shall begin to discuss some of the conclusions we can draw about TR tubes. Our main objective will be first to consider what we want in a TR tube, and the specific elements available to us for getting what we want. Later we shall discuss what we actually get and why.

OPERATION AND DESIGN

OF TR AND ATR TUBES

Lecture No. 6

Applications of Filter Theory to TR Tube Design

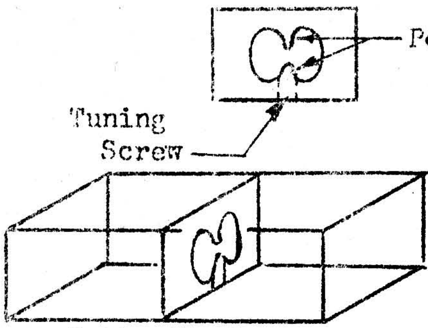
Part I

In this lecture we shall begin to relate some of the theory we have been discussing with the actual design practice. We cannot tie the two completely together for the theory as yet is insufficient. For the broad-band tube the theory will tell us readily what the Q of each individual element should be to obtain a given passband, but not the physical design required to yield a specific Q. Finding the Q in terms of tube geometry is not an impossibility. By applying electromagnetic theory to a particular design, with about two years' calculations, we would arrive at a value of the Q. The answer most likely would not be what we want, so we would have to begin again with a new design.

As a matter of fact, even the simpler problems of specifying design in terms of Q are not feasible in the practical sense. Such a specification would have to include consideration of the effect of tolerance and this is difficult. In particular, the mathematical expressions become quite complicated when we consider mistuned tubes. Much remains to be done along these lines. However, we can set objectives and define what we would like to have in a TR tube.

Basic Types of Resonant Gap Structures:

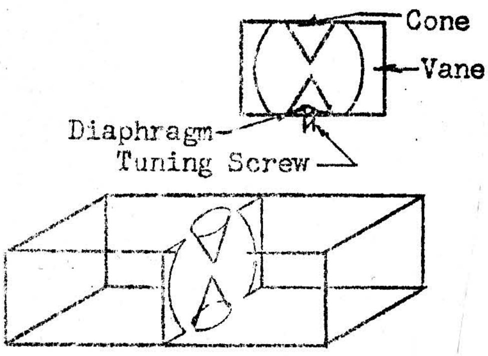
Basically, there are two types of structures that are used for the resonant gap elements: the "butterfly" design and the cone-and-iris design. The "butterfly" is a single piece of metal with cutouts in the central area



Butterfly Design

suggesting in outline its descriptive name. One post is replaced by a screw, so that the structure may be tuned by varying the width of the gap. This unit is inserted across the waveguide to form the resonant gap. The "butterfly" type is used mostly for larger sizes of waveguides, as in such tubes as the 1B58 and TR-368.

In the cone-and-iris type, the bases of two cones are mounted, respectively, at the top and bottom of the waveguide, so that the axes of the cones are perpendicular to the waveguide top and bottom. Through slots in the walls of the waveguide, vanes are inserted in the plane of the cone axes and perpendicular to the walls of the waveguide. One of the cones is mounted on a diaphragm so that again the gap is tunable. This type is used mostly for small sizes of waveguide, or higher frequency tubes, such as the 1B63A.

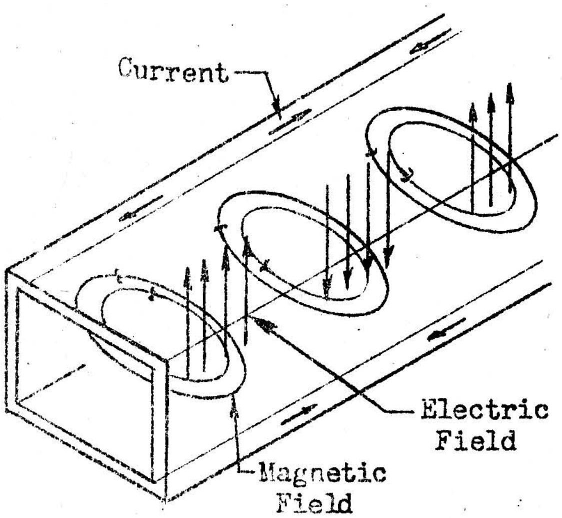


Cone & Iris Design

Both designs are essentially equivalent insofar as electrical behavior is concerned. Each has an iris which narrows the long dimension of the guide, and elements (posts or cones) which form a shunt capacity across the guide. From the electrical design standpoint, the cone-and-iris structure is the better of the two. Although both the cone and the post have self-inductance, the shape of the cone minimizes it. The butterfly type, though, is easier to put together, and it is easier to short circuit the gap as required for one of the best tuning methods.

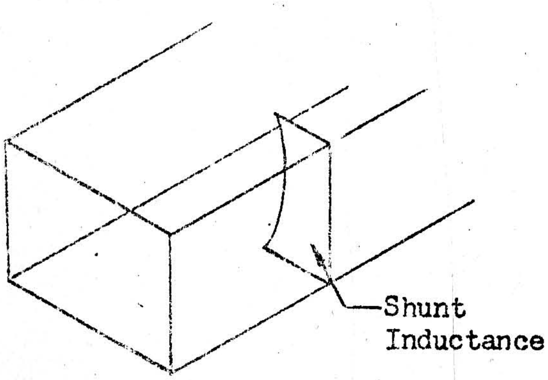
Equivalent Circuit of Resonant Gap:

Let us go back a bit in fundamental theory to get a better understanding of the gap performance. In the normal mode of propagation down a waveguide,



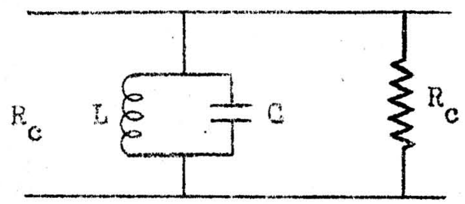
the electric field is across the guide and the magnetic field goes around in loops, perpendicular to the electric field. To establish these loops there must be current both on the side walls and top and bottom of the waveguide. The fact that the electric field goes only across the guide means that it is difficult to simulate a series capacity. A shunt capacity, however, is simple. Anything which blocks the wave so that the electric field has a shorter path will act like a shunt capacity across the guide. Cones, posts, or simply a vane across the guide will provide shunt capacity.

The fact that the currents run on both the top and bottom of the guide means that we can set up either a shunt or a series inductance. The series inductance element is not in general usage for TR tubes, though it has been



used on occasion with windows. Shunt inductance is readily simulated by a vane inserted from the side wall. The vane effectively inserts inductance because it compresses the magnetic lines. The loops we mentioned before must "squeeze" to get through. Hence the main effect is inductive. In the extreme case, the vane short-circuits the waveguide. Hence it is in shunt.

The element may then be represented by a parallel resonant network shunted across the line, where the cones or posts are the shunt capacity,



and the vanes are the shunt inductance. The components of this circuit may be described in terms of two parameters; resonant frequency and the Q which in this case is doubly loaded. As before, we write the characteristic impedance of the line as R_c, and place a matched load at the end of the line.

If the line is lossless and propagating (above "cut-off") the characteristic impedance is real or resistive. This is why we write the impedance as R_c rather than as Z_c. The doubly loaded Q, in terms of the network elements may be calculated to be:

$$Q_{2L} = \frac{1}{2} R_c \sqrt{\frac{C}{L}}$$

The factor $\sqrt{C/L}$ has the dimension of reciprocal ohms. This multiplied by the ohms of R_c will give us a pure number of Q_{2L}. It will be remembered from discussions on loaded Q's, that the loaded Q is the ratio of the power stored to the power dissipated or delivered. Where our network is in the middle of the line, the power that is stored can be taken out of the circuit either by reflection back on the input or from the output. Consequently, the network can be loaded by both R_c's. Since they are in parallel, the combination of the two R_c's is equivalent to $\frac{1}{2} R_c$.

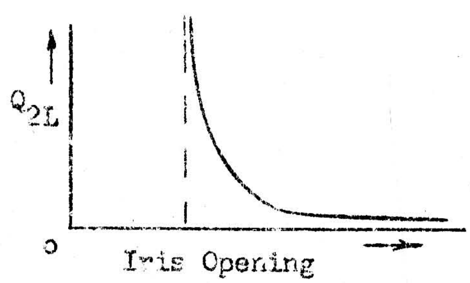
This says that the doubly loaded Q can be raised in value either by (1) increasing the capacitance, or (2) decreasing the inductance. Capacitance is increased by bringing the cones closer together. Inductance is decreased by

If the Q is changed by modifying one element of the resonant network and not the other, the frequency, too, will be changed, as is the case for ordinary resonant networks. The resonant frequency of our cone-gap circuit is

$$f_o = \frac{I}{\sqrt{CL}}$$

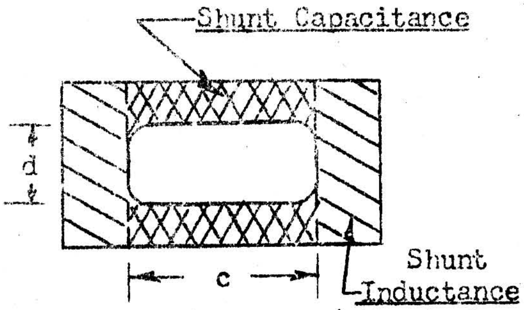
Raising capacity decreases the frequency, and lowering the inductance increases it. Consequently, when a TR is at the proper resonant frequency and it is desired to increase the Q_{2L} of one or more gap elements, alterations are made both to the capacity and to the inductance. First the iris is made smaller, i.e., the vanes are moved in closer to the cones, and then the gap is tuned by diminishing the distance between the cones until the tube is brought back to the desired resonant frequency.

The dependency of Q on C and L is not linear. If it is known that a change by a given amount in L produces a definite change in Q, it does not follow that another change in L by the same amount will produce another change in Q equal to the first. If we plotted the doubly loaded Q vs the iris opening, we would find that for wide separations in the iris, the Q is low. As we decreased the amount of iris opening, the Q would rise slowly at first, then rapidly as a critical value is approached.



Windows:

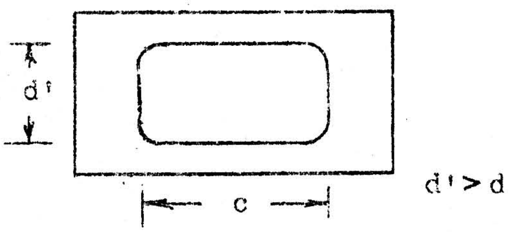
The behavior of windows as far as Q changes are concerned is similar to that described for iris openings in the resonant gap circuit. To a first approximation the window structure may be considered as two superimposed irises, one contributing shunt capacitance and the other shunt inductance.



Resonant at ω_o, Q_{2L} (without glass)

That the window has glass in it means simply that the capacity is partially filled with a dielectric. The glass has no appreciable effect on the inductance.

Suppose we consider first the window with no glass. Say that the opening has a length c and a width d. Without glass the window has a certain resonant frequency and a certain Q_{2L}. Now if we wish to make the window with glass and duplicate the resonant frequency and Q_{2L}, then, theoretically the length c of the window remains unchanged, but the width d must be increased by some amount to d'.



Resonant at ω_o, Q_{2L} (with glass)

Just how much greater d' must be than d is something that unfortunately has to be left to experiment. We do not know how effective the glass is in the window. The effectiveness of the glass depends on how many of the electric lines of force thread through it and how many go alongside it.

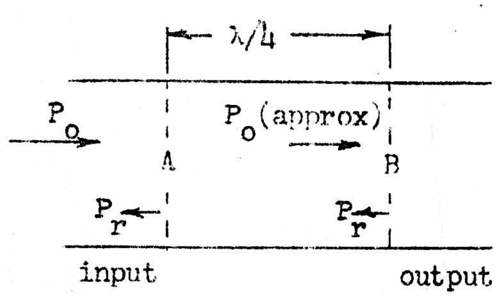
Multiple Element TR Tube

The resonant gap structures and tuned windows are the elements from which broadband TR tubes are constructed. We have previously gone into the reasons for broad-banding TR's. Recapitulating, briefly they are to give more flexibility to the system in which the tubes are used, and to eliminate field tuning operations. Specifications for broadband TR's usually call for something in the neighborhood of a 10 - 12% (sometimes more) passband over which the tube shall introduce a maximum standing wave ratio of, for example 1.4.

If we attempted to produce such a bandpass with a single resonant gap, the Q_{2L} of the gap would have to be somewhere in the neighborhood of 1/4.

Translated in terms of width spacing between cones or posts, it would mean that we could not get the capacitive elements close enough. Even if it were possible to do so, the leakage would undoubtedly be far too high since the direct coupling (cf measurements lecture) would be excessive. Since the Q is inversely proportional to the bandwidth, if we were willing to tolerate a Q_{2L} as low as 1, the single element construction would give us a bandpass of 2-1/2 to 3% only. For the 1B58 and TR-368 the doubly loaded Q's of the elements are probably in the neighborhood of 5 - 6. Even this figure is not high enough for a single resonant gap tube which will protect the system. Just what the minimum allowable figure for a single element is, we have not yet found out. We have experimented with single gap tubes which had a doubly loaded Q of approximately 6. They did not give adequate protection.

We must then increase the bandpass somehow without sacrificing protection to the system. Suppose we have a line into which we have put some element at A such that if we feed into the line a signal whose power is P_o, the element at

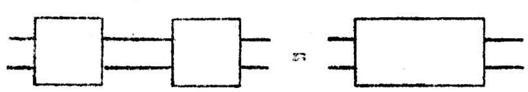


A will reflect P_r watts. Suppose a quarter-wave down the line from A we insert at B another element identical to that at A. If we can assume that P_r is a negligible fraction of P_o, then the input power to the element at B is approximately P_o, too. Then the reflected power from B is also P_r. But the power reflected by B, when it arrives back at A, will have traveled exactly one-half wavelength further than if it had been reflected at A. It will be 180° out of phase with the reflected power at element A, and, ideally, will exactly cancel it, leaving no reflection in the line. The power has to go someplace, and assuming that both elements in the line are lossless, the input power can only go to the output. This cancellation is known as destructive interference, destructive as far as reflection is concerned. This is the principle upon which multiple element TR's operate.

Filter Theory of Multiple Element TR's.

Unfortunately, the principle of cancellation does not lend itself easily to tube design calculations. With the wide bandwidths compared to the Q's used, the power input to the second element at frequencies near the edge of the pass-band is no longer even reasonably approximate to the input of the first element. Furthermore, the Q's may be different so that reflections are not the same, and waves bounce back and forth between elements. The situation may be still further confused if the distance between elements is not effectively quarter-wave. This, then, is the reason for some of the more refined analytical methods - specifically matrix algebra.

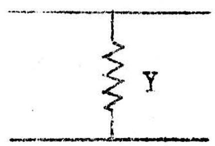
If we consider the multiple element TR in terms of bandpass filter theory,



$$\begin{Vmatrix} A & B \\ C & D \end{Vmatrix} \times \begin{Vmatrix} A' & B' \\ C' & D' \end{Vmatrix} = \begin{Vmatrix} AA'+BC' & AB'+BD' \\ CA'+DC' & CB'+DD' \end{Vmatrix}$$

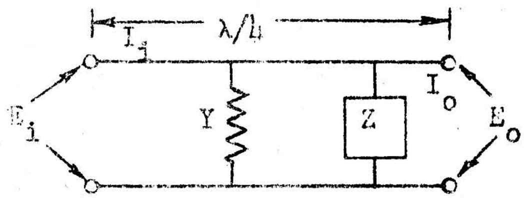
we can handle the calculations neatly with matrix algebra. It will be remembered that any network with input and output terminals can be described in terms of a matrix. The matrix, an array of four functions each of which depends on frequency alone, relates the input to the output. It will also be recalled that if one network is connected to a second which has another matrix, then the combination resulting may be considered as one big network whose matrix is the product of the matrices of the two components.

For simple elements, matrices are uncomplicated. For example, an admittance in shunt to the line is represented



$$\begin{Vmatrix} 1 & 0 \\ Y & 1 \end{Vmatrix}$$

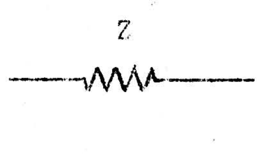
It might be worthwhile to derive that matrix. We go back to our network relations. The input voltage and current are respectively E_i and I_i; and the output voltage and current E_o and I_o. The input voltage must equal the output voltage independent of the current. Hence, the first equation.



$$\left.\begin{aligned} E_i &= E_o + 0 \cdot I_o \\ I_i &= Y E_o + I_o \end{aligned}\right\} \quad \begin{Vmatrix} 1 & 0 \\ Y & 1 \end{Vmatrix}$$

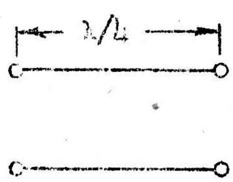
The input current divides between the shunt admittance and the load, and is, of course, equal to the sum of those two currents, giving us the second equation. The matrix, then, expressing the relation between input and output, summarizes the relation by noting the coefficients.

To continue, a series impedance in the line becomes



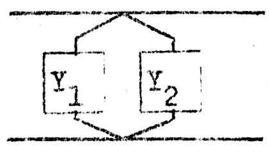
$$\begin{Vmatrix} 1 & Z \\ 0 & 1 \end{Vmatrix}$$

and a line, a quarter wavelength long, if everything is normalized,



$$\begin{Vmatrix} 0 & j \\ j & 0 \end{Vmatrix}$$

Getting back to the network of the resonant gap, we note that we have

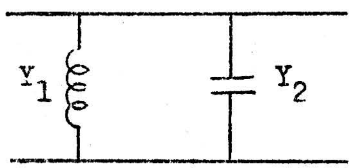


two admittances in shunt, say Y_1 and Y_2, at the same place in the line. As usual, we get the matrix representation for the combination of the two admittances, by obtaining the product of the matrices for the elements.

$$\begin{Vmatrix} 1 & 0 \\ Y_1 & 1 \end{Vmatrix} \times \begin{Vmatrix} 1 & 0 \\ Y_2 & 1 \end{Vmatrix} = \begin{Vmatrix} 1 & 0 \\ Y_1+Y_2 & 1 \end{Vmatrix}$$

We find, as we might expect, that the admittances add.

For the resonant gap circuit we are interested in, the first admittance



element is the inductance of the iris, the second, the capacitance of the cones. We may then replace Y_1 with the admittance of the inductance, or $-j/\omega L$; and Y_2 with $j\omega C$. The parallel resonant network in shunt to the line is then conveniently expressed by

$$\begin{Vmatrix} 1 & 0 \\ j(\omega C - 1/\omega L) & 1 \end{Vmatrix}$$

Utilizing the fact that $\omega_o = 1/\sqrt{LC}$, or that $\omega_o\sqrt{LC} = 1$, we can effectively divide ωC by 1, and multiply ωL by 1, so that the term

$$j(\omega C - 1/\omega L) = j\left[\frac{\omega C}{\omega_o\sqrt{LC}} - \frac{\omega_o\sqrt{LC}}{\omega L}\right]$$

$$= j\sqrt{\frac{C}{L}}\left[\frac{\omega}{\omega_o} - \frac{\omega_o}{\omega}\right]$$

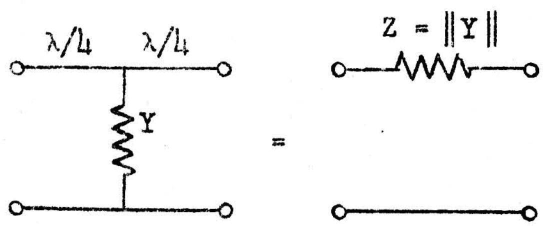
If we let $a = \sqrt{C/L}$ and $x = (\omega/\omega_o - \omega_o/\omega)$, then the matrix for the resonant gap circuit becomes

$$\begin{Vmatrix} 1 & 0 \\ jax & 1 \end{Vmatrix}$$

It will be noted that when the network is normalized $a = 2\,Q_{?L}$.

The factor x is known as the "reduced frequency parameter." It is approximately proportional to the amount in megacycles the network is off resonance. The advantage of using x is that with it all bandpass filters may be converted into low pass or high pass filters, permitting a single theory to cover the different types of filters.

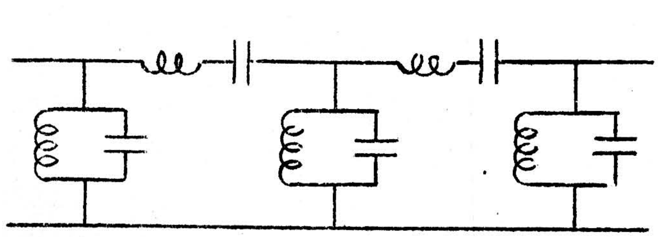
We have examined the network response of the resonant gap by itself in the line. Suppose now we consider the overall effect of the gap and the line. If we take a quarter wavelength of line, follow it by a shunt admittance of some kind, and then by another quarter wavelength of line, we have a product of the three matrices equal to



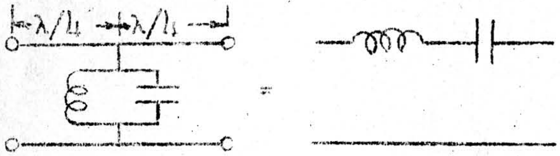
$$\begin{Vmatrix} 0 & j \\ j & 0 \end{Vmatrix} \cdot \begin{Vmatrix} 1 & 0 \\ Y & 1 \end{Vmatrix} \cdot \begin{Vmatrix} 0 & j \\ j & 0 \end{Vmatrix} = -\begin{Vmatrix} 1 & Y \\ 0 & 1 \end{Vmatrix}$$

It is observed that the result is equivalent to the matrix for a series element. In other words, a shunt element of admittance Y, preceded and followed by a quarter wavelength of line, acts as a series element whose normalized impedance is numerically equal to the normalized admittance of the original element.

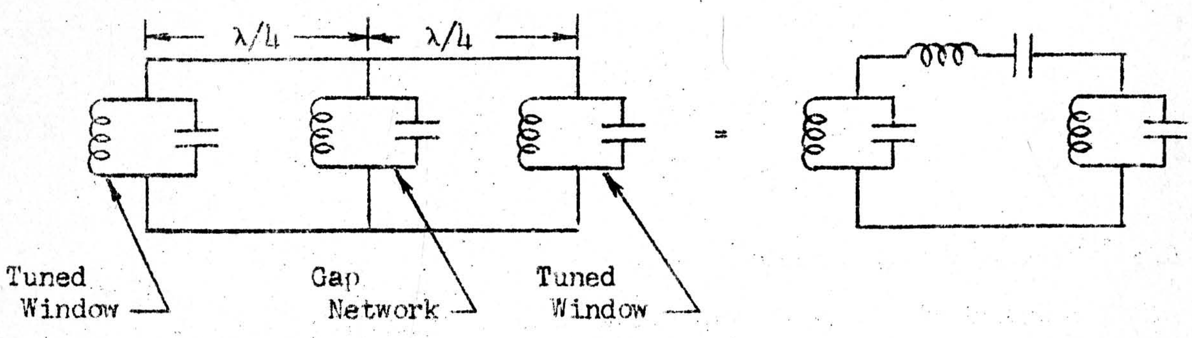
The ordinary bandpass filter, which we wish to simulate in the TR tube might have as its input a parallel resonant circuit, followed by a series resonant network, then a parallel resonant, etc. It was pointed out previously that it is difficult to set up a series capacity in a waveguide operating in the



normal mode. Therefore, we have to find some means of producing an effective series capacity in the waveguide. To do so we utilize the effect we have uncovered above, that of a shunt element preceded and followed by a quarter wavelength of line. For our purposes, a shunt parallel resonant network so placed is equivalent to a series resonant circuit inserted in series with the line. Hence, if we have a single gap tube with tuned windows, its circuit is equivalent to the ordinary band-pass filter.



$$\begin{Vmatrix} 0 & j \\ j & 0 \end{Vmatrix} \cdot \begin{Vmatrix} 1 & 0 \\ jax & 1 \end{Vmatrix} \cdot \begin{Vmatrix} 0 & j \\ j & 0 \end{Vmatrix} = \begin{Vmatrix} 1 & jax \\ 0 & 1 \end{Vmatrix}$$

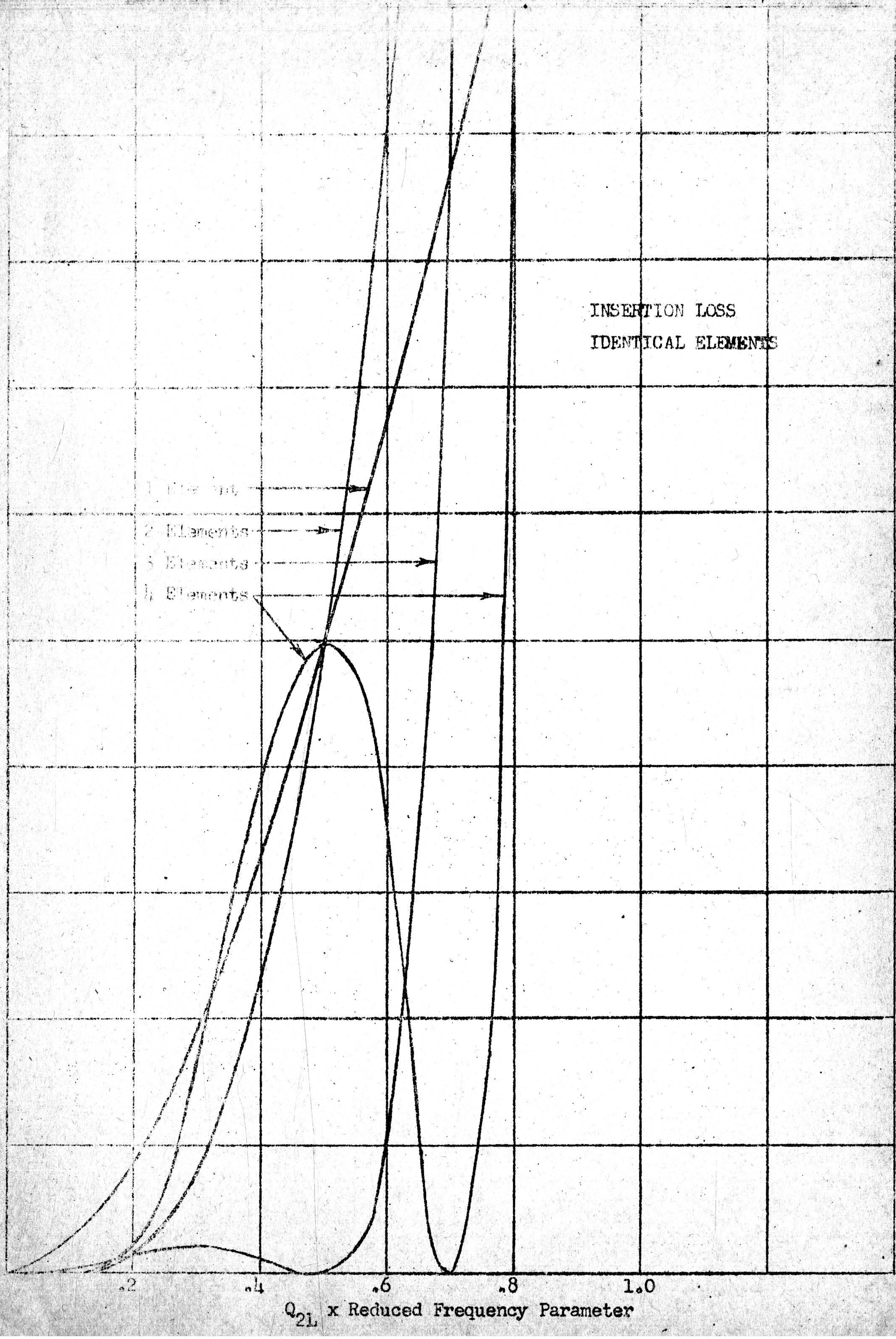


We see, then, that a bandpass filter can be obtained by inserting elements in the line spaced a quarter of a wavelength apart. What, then, should the elements be for best results? Do we, for example, want to make all the elements the same? Without going into the details of the calculation, the transmission curves for one, two, three and four identical elements are plotted on the attached graph. These show the insertion losses in db vs Q_{2L} times the reduced frequency parameter.

It will be seen that two elements in the line give almost twice the bandpass of one element. Adding a third element increases still further the working band. The humps that are getting more marked in the curves, represent points of high reflection, or high standing wave ratios. Addition of a fourth element raises the hump so high that the tube as it stands is unusable.

These curves are true for identical elements only. They state that if enough identical elements are used in a tube, cancellation of reflection by destructive interference over the desired bandpass becomes less likely. There still is cancellation, but occurring at such high frequencies that the humps have become too high for practical use.

These curves are known as Tschebbyscheff functions of the second kind. They do not represent the optimum solution. We shall investigate next time some of the other solutions that should be considered.


INSERTION LOSS
IDENTICAL ELEMENTS
2 Elements
3 Elements
4 Elements
.2
.4
.6
.8
1.0
Q_{2L} x Reduced Frequency Parameter

OPERATION AND DESIGN
OF TR AND ATR TUBES

Lecture No. 7

Applications of Filter Theory to TR Tube Design, Part II

In the last two sessions we have been treating the TR tube as a bandpass filter. It differs from conventional bandpass filter theory, particularly in regard to the criterion of optimum design chosen. Ordinary filter theory concerns itself with the rejection band. In the TR tube, the rejection band is of little interest.

The criterion for optimum design in the TR tube that appears best is that the Q of the elements be as high as possible. This is reasonable since the stored energy represented by the Q when added to the incoming energy builds up the voltage across the gap, causing more rapid breakdown.

We come then to the design of the tube where the design hinges upon the Q's. Unfortunately, we cannot anticipate by calculations what the Q of an element will be. We know just that to raise the Q of a tuned gap we must narrow the gap and, to hold the resonant frequency constant, narrow the iris. At this stage of the theory, then, actual determination of the design parameters must be experimental.

Filter Solution By Tschebbyscheff Polynomials:

At the end of the last lecture we showed the bandpass characteristics for tubes composed of what are known as "recurrent structures." That is, all the elements of a tube have the same value. The curves for these structures fall into the general category of Tschebbyscheff Polynomials or Functions, a name that is referred to frequently in filter theory.

There are two kinds of Tschebbyscheff Functions. The curves plotted belong to the second kind, which we will look into now. They are polynomials in y which can be written

$$U(y) = \frac{\sin n \cos^{-1} y}{\sin \cos^{-1} y}$$

By way of interest we might note the solution of the polynomials for particular values of n. When $n = 1$

$$U_1 = 1$$

When $n = 2$

$$U_2 = 2y$$

When $n = 3$

$$U_3 = 4y^2 - 1$$

And the general recursion **formula** is

$$U_{n+1} = 2y\, U_n - U_{n-1}$$

To obtain bandpass characteristics of TR tubes with identical elements, we express the relationship between the theoretical performance of the tube and the number of elements in terms of the voltage transmission function, t,

$$\left|\frac{1}{t}\right|^2 = 1 + (2\, Q_{2L}\, x)^2\, U_n^{\,2}\, (2\, Q_{2L}\, x)$$

where x is the reduced frequency parameter, $(\omega/\omega_o - \omega_o/\omega)$ and U_n are the Tschebbyscheff Functions, in this instance expressible in $(2\, Q_{2L}\, x)$, instead of the y written above.

We presented the plots of the polynomials for the above expression of the voltage transmission function at the end of the lecture previous to this. Referring again to the family of curves, it is seen that when three elements are used, even though the bandpass has widened considerably, it has developed a slight bump. Additional elements rapidly make the bump more pronounced. In fact, just with four elements, the SWR for that particular section of the band has exceeded specification limits generally set. These bumps are much too serious for the resultant tubes to be of practical use. We have to seek, then, other solutions for the TR tube as a bandpass filter.

On a straight optimization basis, over a wide range of application, the ideal filter circuit can be written in terms of the Tschebbyscheff Functions of the first kind, viz

$$T_n(y) = \cos\,(n \cos^{-1} y)$$

where the first three polynomials are, respectively,

$$T_1 = y$$

$$T_2 = 2y^2 - 1$$

$$T_3 = 4y^3 - 3y$$

and the general recursion formula is

$$T_{n+1} = 2y\,T_n - T_{n-1}$$

In the application of the U_n functions we started off with the "recurrent structure" design and expressed its performance in terms of the functions. For the T_n functions, however, we do not know what the design is for the ideal filter circuit. We state that we would like to have a design for which the transmission function is expressible as

$$\left| \frac{1}{t} \right|^2 = 1 + \varepsilon^2 \, T_n^{\,2}(2 \, Q_{2L} \, x)$$

where ε^2 is some quantity, frequently called the "tolerance."

A plot of the T_n functions will indicate their possible application to filter design. When $y = 1$, the polynomials all go through either the point +1 or -1. For values of y between +1 and -1, the polynomials of T_n become either +1 or -1 a given number of times (depending on the order of n.) That is, in between the limits of $y = +1$ and $y = -1$, the function oscillates in value between $T = +1$ and $T = -1$. This is illustrated in Fig. I by a plot of a T_n function of the third order. When we square the function T, its plot will oscillate between 0 and +1 for values of y between +1 and -1. (Fig. II). We set the ε such that the performance is within the power desired. (Fig. III) Finally, by using Q_{2L} as a horizontal scale factor, we can adjust the bandwidth. So, if we can devise a filter which will have the transmission function as determined by the above relationship with T_n, the filter performance will be within specifications. This can be done, but we will not go into more details, mainly because at the present stage of the tube art, the solution is not practical.

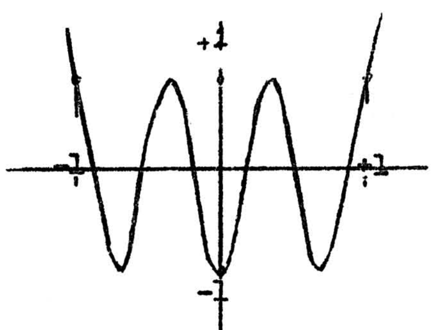


Fig. I

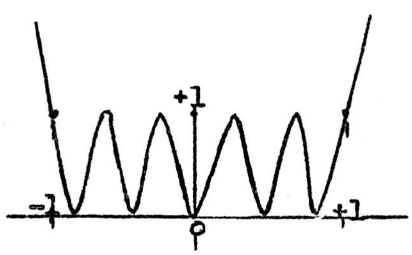


Fig. II

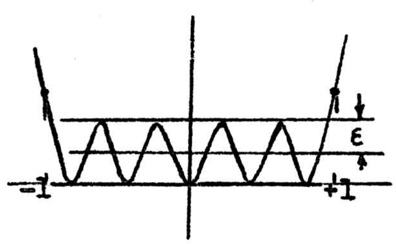


Fig. III

Mathematically, the solution employing Tschebbyscheff polynomials of the first kind appears to give, over a wide range, the theoretical optimum design. However, when, from the application of these polynomials, we specify that an element have a Q of 6.3, or be tuned to a resonant frequency of 0, we cannot, at the present time construct to such specifications. Practical design has to take into account machining and tuning tolerances. Small errors will throw this solution completely off.

Butterworth Solution:

From the standpoint of the unavoidable tolerances, there is another solution which appears more desirable. It has various designations - "Butterworth", "maximally flat", and "semi-infinite slope." It involves solutions of the transmission function as expressions of the form:

$$\left| \frac{1}{t} \right|^2 = 1 + c^2 x^{2n}$$

where the c is some positive constant. The plot differs from the previous solutions in that it has no bumps. The curve rises very slowly from the origin (or center band), and is quite flat for some length before its slope increases at a rapid rate. (Fig. IV) It is the behavior of this curve which gives rise to the designations, "maximally flat", and "semi-infinite slope."

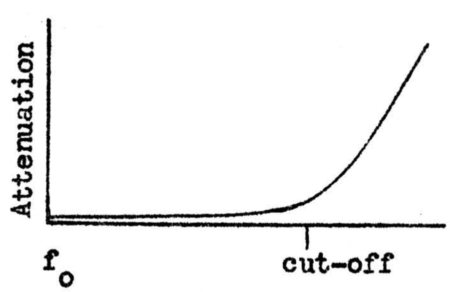


Fig. IV

The description "maximally flat" is appropriate with reference to the minimum curvature of the function at the origin. The expression for the transmission function in terms of the Butterworth Function is a polynomial of degree 2n. Every derivative except the $2n^{th}$ vanishes at the origin. The curve is "maximally flat."

It gets the name "semi-infinite slope" from another characteristic of its plot. If the value of the function is plotted in db vs the reduced frequency parameter, it will be observed that for points on the reduced frequency scale below that corresponding to cutoff, the function has a nearly constant value. For points above the cutoff, the insertion loss increases in a way that quickly approaches a linear rise (Fig. V). It approximates to a curve which has a constant value up to a point, and constant rate of change beyond this point. Such a curve is called a "semi-infinite slope."

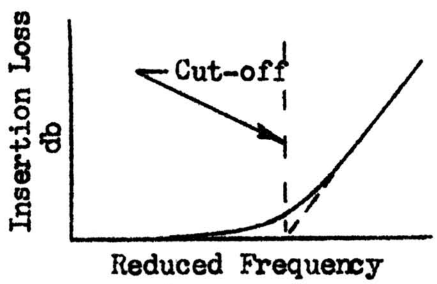


Fig. V

It is beyond the scope of this lecture to analyze the passband of TR's as approximated by Tschebbyscheff and Butterworth solutions. Some of the findings of such analyses are summarized below. However, it might be of interest to point out that the Butterworth solution is the limit of the T_n function, as the tolerance is reduced to zero.

The type of design criteria chosen for a tube depends upon such factors as tolerance and errors in tuning. For a one-element tube the design criterion is necessarily Butterworth. Likewise, for a two-element tube, providing the elements are the same, and tuned to the same frequency.

To obtain the Butterworth design with three elements, the windows should each have a Q whose value is about one-half that for the center element.

In a four-element tube, to give a Butterworth response, the Q's of the windows should be ($\sqrt{2}$ - 1) that of the two center elements, or approximately .4 of the value of the center elements which should be identical.

For the five-element tube, the value of the Q's may be referred to the value of the Q of the center element. Ideally, for Butterworth, the first and third gaps should have Q's about .8 that of the middle gap; and the windows, approximately .3 of that for the second gap. (Fig. VI.) However, most 5-element (3 gap) tubes are made with identical internal elements. With this design, the best that can be done is to make the window Q's about 3/8 the gap Q. This design is not bad. Bumps will occur but they are not high and the bandwidth is good. Most of our recent designs have aimed at this structure.

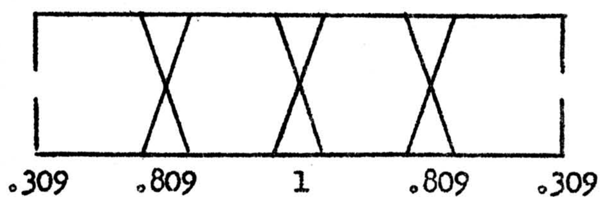


Fig. VI

<u>Line Effect on Bandpass:</u>

Thus far we have neglected line effects. We are connecting the tube elements together by lengths of line that we say are a quarter-wave long. But, at the edge of the 10-12% band, with which we are concerned, the line is no longer exactly a quarter-wavelength long. The problem is not trivial; neither, on the other hand, is it too serious.

For example, in Fig. VII there is plotted the Butterworth solution for a five-element tube with a bandpass of approximately 12%. The dotted curve represents theoretical performance of the tube when line effects are neglected. The solid curve includes line effects. Oddly enough, the second has a slightly wider bandwidth -- one of the few cases where line effects actually widen the band.

The lengths of line connecting tube elements has introduced fairly large sized humps into the bandpass curve. However, these can be compensated for to a great extent. The lengths of line have a Q of approximately $\frac{\pi}{2}$, which may be considered as centered in the element. If the Q of each element is reduced in value by subtracting $\frac{\pi}{2}$, the resulting bandpass will be very close to that anticipated.

Generally, the introduction of line effects is not serious enough to cause much trouble. For this particular tube the humps have a standing wave ratio of about 1.15, compared with the 1.4, or more, limit usually specified. The line effect problem, then, is not of too much significance providing the absolute ultimate is not expected from the design.

Current Design of TR Tubes:

Our present designs of five element TR's are very close to being Butterworth. However, as mentioned, we make all three gaps identical and then find the most suitable windows. Fig. VIII shows the plots for such a tube without windows, and with windows of various Q's. The Q's of the three center elements are kept constant at 4.1. The Q's of the windows are varied from 0 up.

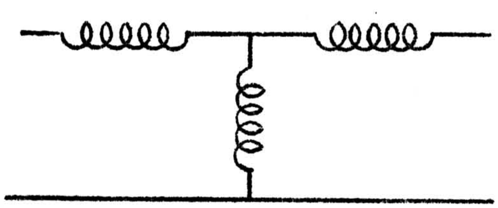
With no windows it is seen that in the passband there is one hump that peaks at about 1.45 SWR. As the window Q's are increased in value, the hump gets smaller and slides over somewhat, widening the usable bandpass. A Q of 1.5 for the windows appears to give optimum passband. Values of window Q above 1.5 make for rapid increase of bandwidth, but also for a greatly increased hump within the passband.

In this construction, therefore, it would seem that window Q's should be roughly 3/8 the Q of the three center elements. Comparison with the figures for the ideal five-element tube will show that the present design is not too far off from the Butterworth design.

<u>Determination of Effective Spacing Between Elements:</u>

We have already touched upon the fact that a length of line in the tube may be a quarter-wavelength physically, but not electrically. At this point we shall take up a method of determining the electrical value of the lengths between elements.

In general the equivalent circuit of any element in the wave-guide is rarely a simple shunt network. Usually it is a complete T-section. Hence, it may introduce a positive or negative length of line, so that what begins as a physical quarter-wavelength is electrically not exactly a quarter-wavelength. This condition can lead to mistuning and other complications.



A method of measuring the spacing between elements, which gave satisfactory results for the 1B58 can be applied generally.

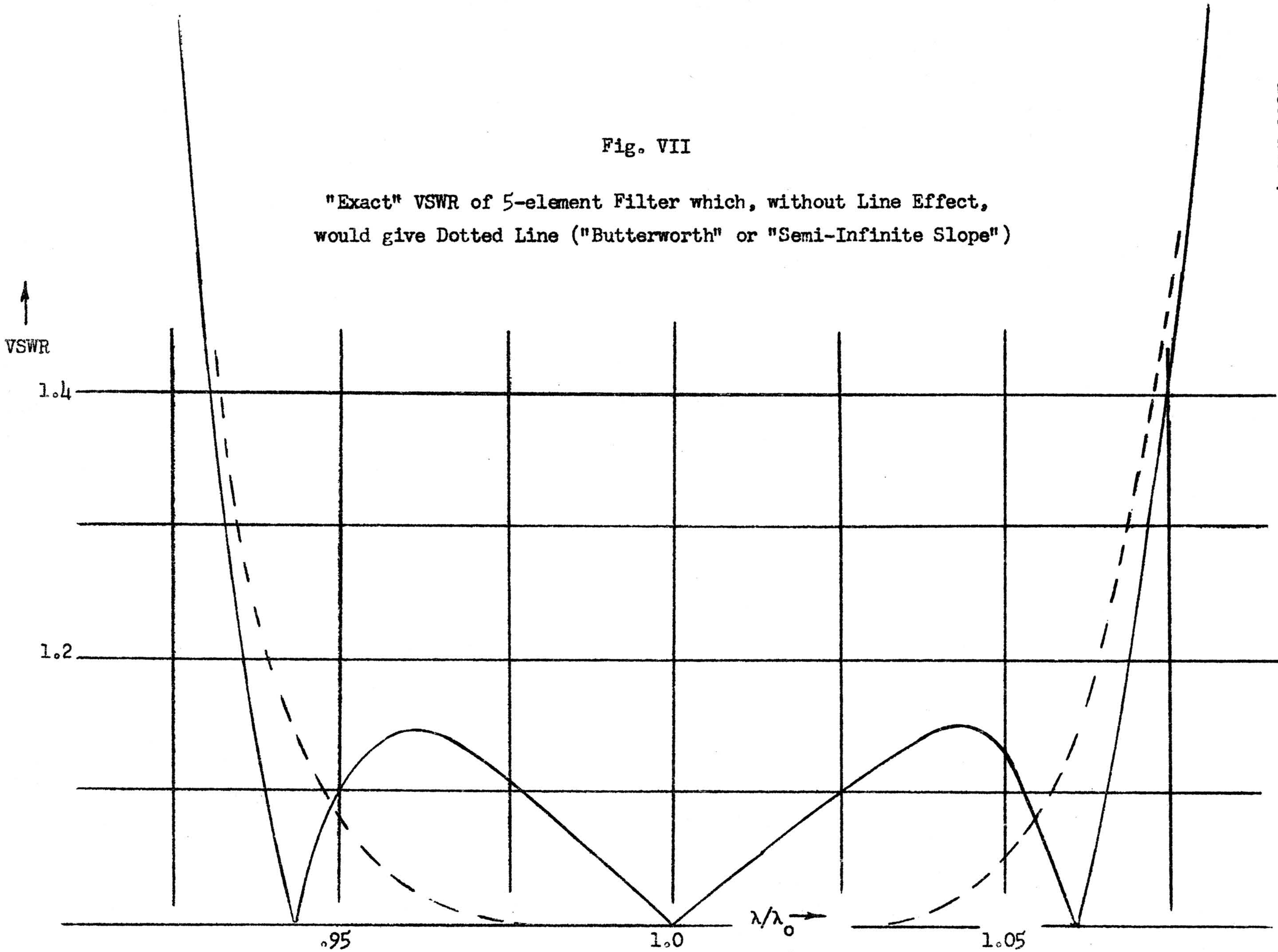
Three elements are accurately constructed to be as nearly identical as possible. Their Q's are measured with great accuracy. The elements are built so that they can be bolted together to form a TR tube. This is obviously not a pumpable tube, but it is one on which cold measurements can be made. The advantage of the tube being demountable is that each element can be tuned separately before the three sections are bolted together to obtain the overall passband.

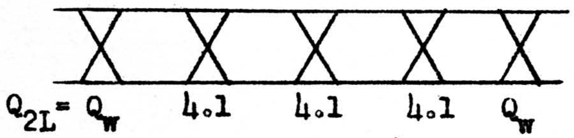
The theory behind this calculation of line lengths goes back to the Tschebbyscheff Functions of the first kind. In Fig. IX, by means of these functions, there is plotted a set of theoretical passbands for a 3-element tube with line spacings between elements of 85°, 88° and 90°. It will be noted that each curve has two bumps, and a position of minimum at three different frequencies. It is perfectly feasible in obtaining the actual passband to measure the positions of minimum SWR quite accurately. By means of the minimum points and the Q it is possible to calculate the effective spacing between elements, within 1% accuracy. The answer can be applied as a correction factor to yield the physical lengths necessary to give quarter-wave spacing.

With this method, it appears quite feasible to determine the effective length of line separating the elements to an accuracy of the order of 1°. As we shall see later, this is sufficiently precise for our purposes.

Fig. VII

"Exact" VSWR of 5-element Filter which, without Line Effect, would give Dotted Line ("Butterworth" or "Semi-Infinite Slope")

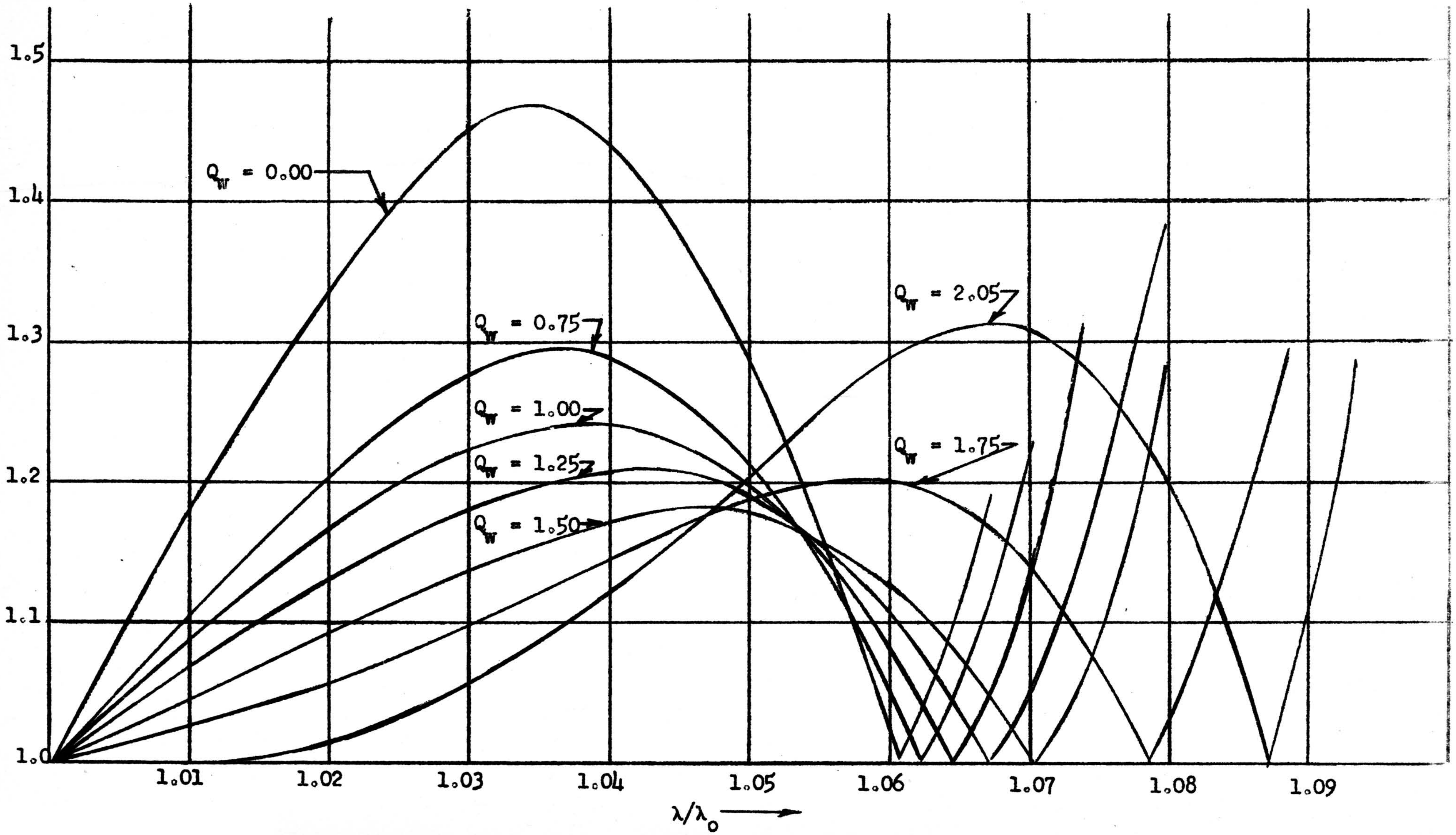




Identically Tuned

Fig. VIII

Approximate Band Plot
(Neglecting Line Effect)
Five Element Tube, Varying Windows



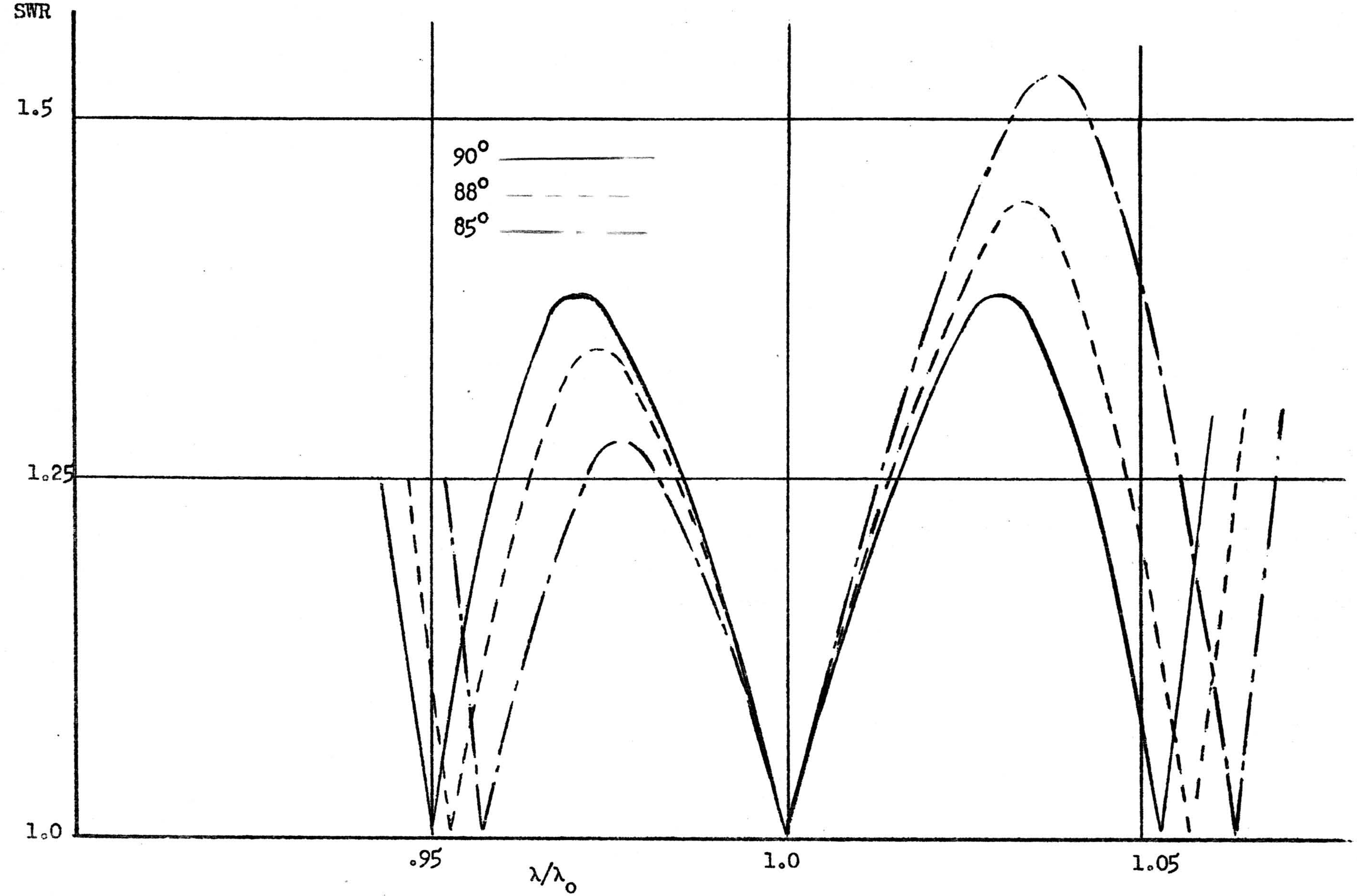


Fig. IX

Theoretical SWR(V) of a filter consisting of three identical elements with $Q_{2L} = 4.1$, spaced with center-band electrical lengths noted.

OPERATION AND DESIGN OF TR AND ATR TUBES

Lecture No. 8

Gas Discharges, Part I

Previous lectures on TR and ATR tubes have highlighted mechanical construction of components and network equivalents as they affected tube performance and characteristics. In this and the following talk, the emphasis will be on the fill of tubes -- the breakdown properties of gas and its interaction with the r-f field -- the little that is known and "guessed." Preliminary to a discussion of the actual discharge process we will consider the diffusion of electrons in a gas, since an understanding of the latter process is important in determining what the breakdown characteristics will be.

Definition of Diffusion:

Diffusion of electrons or molecules, or of any other such particles results in a medium from random currents causing a net flow of the particles out of the volume of the medium. Let us assume an ideal case of gas consisting of molecules or electrons that do not collide, and consider one square centimeter of area within the gas. Particles will be flowing in all directions, some passing through the area and some not. (See Fig. I) We want to calculate how many actually do go through this segment of unit area. Suppose we consider an even more specialized case in which all the particles are traveling in one direction only. Assuming that the particles have an average velocity

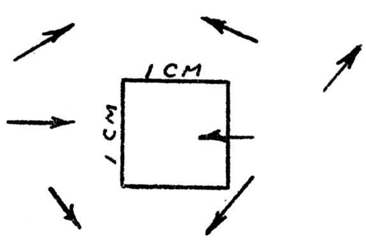


Fig. 1

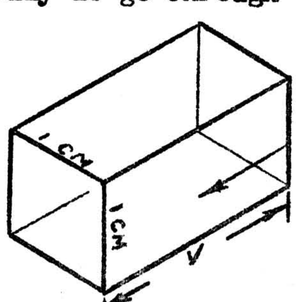


Fig. 2

$\bar{v}$, the number that pass through the segment in one second are contained in a volume of length $\bar{v}$ and cross-section equal to the unit area. (See Fig. 2). If there are n electrons per cubic centimeter, the number of electrons that go through the segment, or the current Γ is equal to the number of electrons in one cubic centimeter times the length $\bar{v}$ or $\Gamma =$ $n\,\bar{v}$.

However, the electrons do not travel in one direction only; they go in all directions. If we carried out the above process to take into account non-directional flow, we would find that the average current passing in one direction is

$$\Gamma = \frac{n\,\bar{v}}{4}.$$

If the density of electrons is constant in space, then equal numbers of electrons will flow in both directions and there will be no net flow.

We shall now look at the case where density is not constant, permitting a net current flow, and where collisions are taken into account. If the electrons collide with each other, then the number of collisions made determine how many electrons are able to pass through our unit area. To evaluate the quantity, we consider a density plot at some point in the gas. (See Fig. 3.)

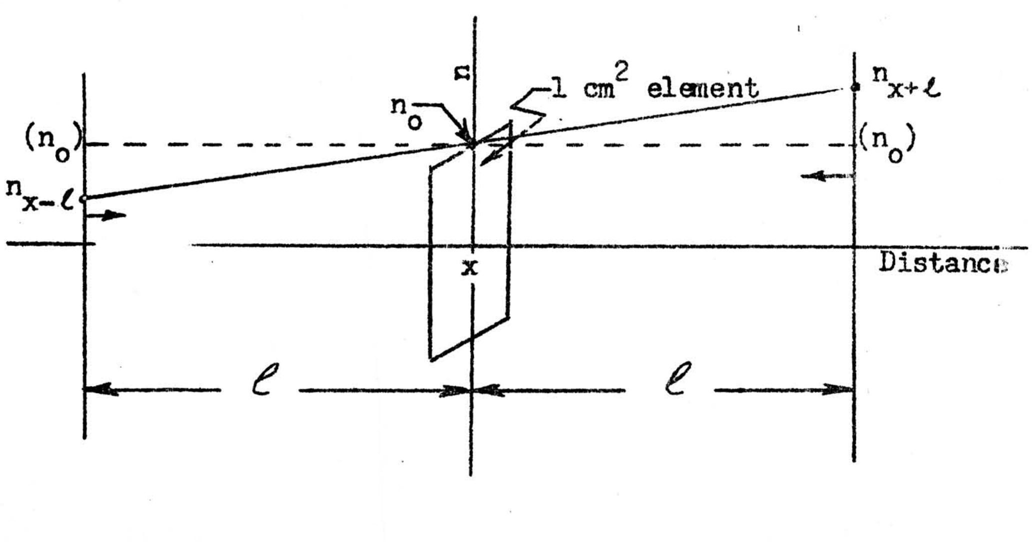


Fig. 3

We let the density of electrons be given on the vertical axis. Distance, x, traversed by the electrons is indicated on the horizontal axis. We set up two imaginary boundaries, each a distance ℓ from the origin. This quantity ℓ we define as the mean free path, or that distance which an electron travels on the average before it collides with a gas molecule. Within the region included between the two boundaries, then, is a space that we can say approximates a collision-free area. We establish the 1 cm^2 element on the vertical axis and count the number of electrons going through it.

The density of electrons on the boundary drawn to the right of the unit element is $n_{x+\ell}$ and on the boundary to the left, $n_{x-\ell}$. The number of electrons which go through the unit area element, coming from the left boundary will be

$$\Gamma_1 = n_{x-\ell}\,\bar{v}/4 \tag{1}$$

and the number coming from the right boundary

$$\Gamma_2 = n_{x+\ell}\,\bar{v}/4 \tag{2}$$

The difference between these two quantities will be the net number going through the unit area element,

$$\Gamma = \Gamma_1 - \Gamma_2 \tag{3}$$

We can approximate the density plot by a straight line, labeling as n_o the intersection of the line with the unit area element. Then we can write $n_{x+\ell}$ and $n_{x-\ell}$ in terms of n_o and the slope of the density plot.

$$n_{x-\ell} = n_o - \ell\frac{dn}{dx} \tag{4}$$

$$n_{x+\ell} = n_o + \ell\frac{dn}{dx} \tag{5}$$

Substituting (4) and (5) in (1) and (2), respectively, and solving (3), we get

$$\Gamma = -\frac{\ell\,\bar{v}}{2}\frac{dn}{dx} \qquad (6)$$

The net flow of current depends upon the gradient or the slope of the density curve.

Diffusion Coefficient

The factor preceding the slope of the density curve in (6) is known as the diffusion coefficient. Its actual value differs from that arrived at in the current formula because our derivation has been simplified, by assuming that there were no collisions in the region between the boundaries, and that no electrons came from the regions outside the boundaries. Had we taken these factors into account more carefully, then we would have arrived at a value for the diffusion coefficient

$$D = \ell\,\bar{v}/3 \qquad (7)$$

By examining (7) we can relate the dependency of the diffusion coefficient upon gas pressure. Since ℓ is the mean free path for collisions of electrons with molecules, we can evaluate it qualitatively in terms of the gas composition. The gas is composed of a number of randomly distributed particles. An electron will travel through the gas for a distance ℓ before it makes a collision and is deflected. If the pressure is increased, there will be a greater density of molecules, and consequently, the mean free path will be shorter. In other words, the diffusion coefficient varies inversely with the gas pressure

$$D \propto \frac{1}{p}$$

Ambi-polar Diffusion Coefficient

Thus far we have been talking about an idealized diffusion, free diffusion. The only force on the individual diffusing particles considered was that of collisions. Since there are positive ions as well as electrons in a discharge, there are mutual electric fields -- the electrons move in the field of the positive ions, and the positive ions move in the field of the electrons. That alters considerably our simplified picture. Suppose, initially (at zero time), the distribution of positive ions and negative electrons is that pictured in Fig. 4. We will assume that the number or density of positive ions is that same as that of the electrons, so that the n stands for either positive ions or electrons.

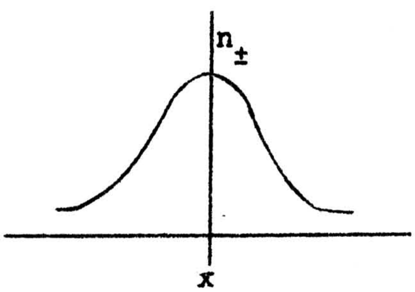


Fig. 4

Because their densities are equal there is no net space charge field. At some time later, the electrons and the ions will have flowed away from the high concentration region, since they will flow in a direction of lower density. The electrons will flow much faster than the positive ions. From (7) we see that the diffusion coefficient depends directly upon velocity. If the electrons and positive ions have equal energies, the electrons, being much lighter, must have much greater velocity. Thus, the electrons will tend to flow out first. In fact, they will tend very quickly to form the distribution pictured in Fig. 5, while the positive ions have roughly the same distribution as initially.

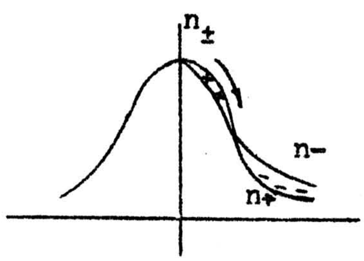


Fig. 5

In one portion of the region there is a net positive charge; in another, a net negative charge. Thus we have two different charges which tend to attract each other. The electrons try to pull the positive ions out of the high concentration area. The positive ions try to pull the electrons back into the area. As a result, when densities are high enough, the positive ions will not allow the electrons to move out of the region which they occupy together. This phenomenon occurs when the density of positive ions becomes greater than 10^6 or 10^7 per cubic centimeter. The densities of positive ions and electrons thus must remain very nearly equal.

This type of diffusion where electrostatic interaction is very strong is known as ambi-polar diffusion, ambi-polar, because it has two poles, or two charges with opposite signs. We can define a diffusion coefficient which takes this phenomenon into account. The ambi-polar diffusion coefficient, D_a, is

$$D_a \doteq D_+ \left[1 + \frac{\bar{u}_-}{\bar{u}_+} \right] \tag{8}$$

where D_+ is the diffusion coefficient of the positive ions

$\bar{u}_-$ is the average energy of the electrons

$\bar{u}_+$ is the average energy of the positive ions.

In an active discharge, where the average energy of electrons is considerably larger than the average energy of the ions, D_a may become relatively large. By relatively large is meant that the ambi-polar diffusion coefficient may be of the order of one-tenth the diffusion coefficient of the electrons.

$$D_a \doteq \frac{D_-}{10}$$

Although diffusion has slowed down to a great extent over free diffusion, it may still be fairly large.

In the case of a decaying plasma, such as occurs in TR tubes, the situation immediately after the r-f ceases, is that pictured in Fig. 5. One portion of the gas has a net positive charge, another portion, a net negative charge. A very short time later, of the order of microseconds, the electrons have cooled down to thermal energies. The average energy of electrons as well as that of ions is thermal energy (about 1/30 electron-volt), and the ratio $\frac{\bar{u}_-}{\bar{u}_+}$ becomes 1. Then the value of the ambipolar diffusion coefficient is of the order of magnitude of a thousandth that of the electron diffusion coefficient,

$$D_a \doteq \frac{D_-}{1000},$$

considerably slower than the diffusion during an active discharge.

We have spent quite some time in this lecture on diffusion because it is important to the analysis of a-c discharge. But, before we go directly into the latter topic, we will discuss the d-c breakdown process.

D-C Discharges

Let us consider a discharge tube with an anode and a cathode.

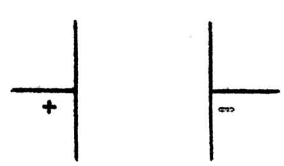


Fig. 6

There is a gas pressure between the two plates. (See Fig. 6.) In the absence of an electron in the region between the plates there will be no discharge, regardless of the potential applied, short of the enormous potential required to literally pull electrons free from the metal (field emission.) Normally, the discharge will not occur unless there is one electron initially present some place in the gas.

As a result of natural radioactivity, an electron will enter the region between the electrodes. The electron will move toward the anode. On its way it will gain enough energy so that it will disrupt a neutral molecule on collision, and set free a new electron. This is an ionization process. electrons will thus be born. The new electrons moving toward the anode will produce still more free electrons, and so on. An avalanche is formed in which there may be a very large number of electrons.

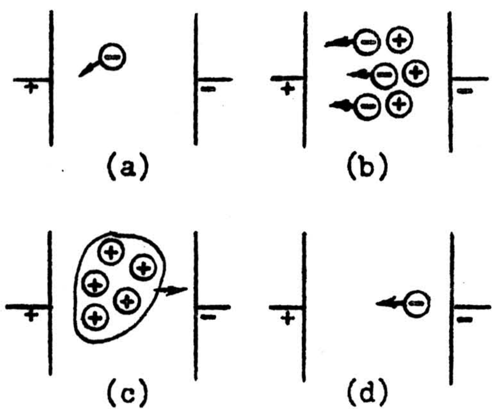


Fig. 7

All the new electrons are collected on the anode, leaving behind in the region between plates a cloud of positive charges. These positive charges move toward the cathode. When they reach the cathode there is a possibility that in drawing off electrons to neutralize themselves and become neutral molecules again, they will also pull out an extra electron. If just one new electron is freed by this process, the entire process repeats itself. We have, then, the condition for a sustained discharge.

This type of breakdown depends not only upon the gas, but also upon the cathode material, since the cathode material will determine how readily extra electrons will be pulled out by positive ions to maintain the discharge. The interaction of the positive ions with the cathode is a necessary and important condition in determining what the breakdown field will be.

A-C Discharges

The situation in the a-c breakdown differs somewhat from that just described. Since there is no d-c field there will be no net transport of electrons under the action of a field. Actually the mechanism of breakdown of a gas under an a-c field is quite similar to an atomic bomb explosion, though less spectacular! We could talk about an atomic bomb and by substituting in the description the appropriate terminology we would have the explanation of an a-c discharge.

For example, Fig. 8 might represent a lump of uranium, or it might be an r-f cavity. Let's investigate its properties first in terms of a lump of uranium. Let us assume that inside there are ten neutrons bouncing around with a certain energy, colliding with uranium atoms, and eventually getting out. Assume that all ten escape, but in the process of escaping have become involved in nuclear reactions so that nine new neutrons have been created inside the lump. By the same procedure, the nine neutrons get out, leaving behind about eight. The process is repeated until there are no neutrons remaining within the lump or bomb. This lump of uranium would consitute what is known as less than the "critical mass."

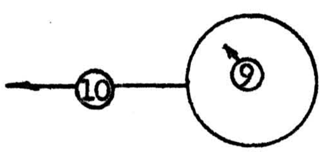
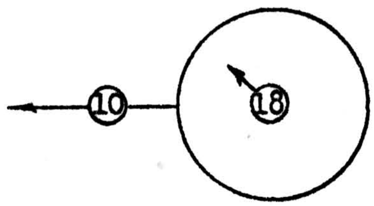


Fig. 8

Suppose, however, we make one change -- make our mass of uranium larger. If we increase the diameter by 2, the area is increased by 4, and the volume by 8. There is now four times the chance of a neutron escaping. But, since we have 8 times the volume, there is a relatively greater chance of creating a larger number of neutrons inside. We start out again with 10 initial neutrons which find their way out. In the process the ten have been involved in relatively more nuclear reactions and leave behind not 9, but, say, 18 neutrons. The 18 make their way out of the bomb and leave behind something like 36 neutrons. This process keeps building up until the rate of nuclear reactions becomes so large that a vast amount of energy is liberated. When the size of the bomb is such that the density of neutrons increases rather than decreases, we have what is known as the "critical mass."



Diameter = 2 x Dia. of Fig. 8

Fig. 9

The explanation would hold for an a-c breakdown in a cavity if instead of neutrons we talk about electrons. Instead of nuclear reactions creating new neutrons we would talk about an electric field creating new electrons. For example, if Fig. 8 represented an r-f cavity in which was fed r-f power resulting in but one field strength within the cavity, no other r-f field could be produced. By a process exactly analogous to nuclear reaction in the cavity of Fig. 8, there would be more electrons escaping than being created, and it would not be possible to break down the gas. If in a larger cavity, such as might be represented by Fig. 9, the same electric field is established, then the gas in the cavity will break down.

The process by which both electrons and neutrons escape from a given volume is, of course, diffusion. The condition of electrons being released can be complicated by other mechanisms, such as attachment of electrons, but generally diffusion is the prime consideration.

Ionization, Excitation

In discussing diffusion earlier we referred sketchily to ionization. Let us examine the ionization process more closely. What happens in an ionization? For one thing, we know that when an electron with sufficient energy hits a molecule, it liberates a molecular or atomic electron. We can plot an electron energy scale, as in Fig. 10.

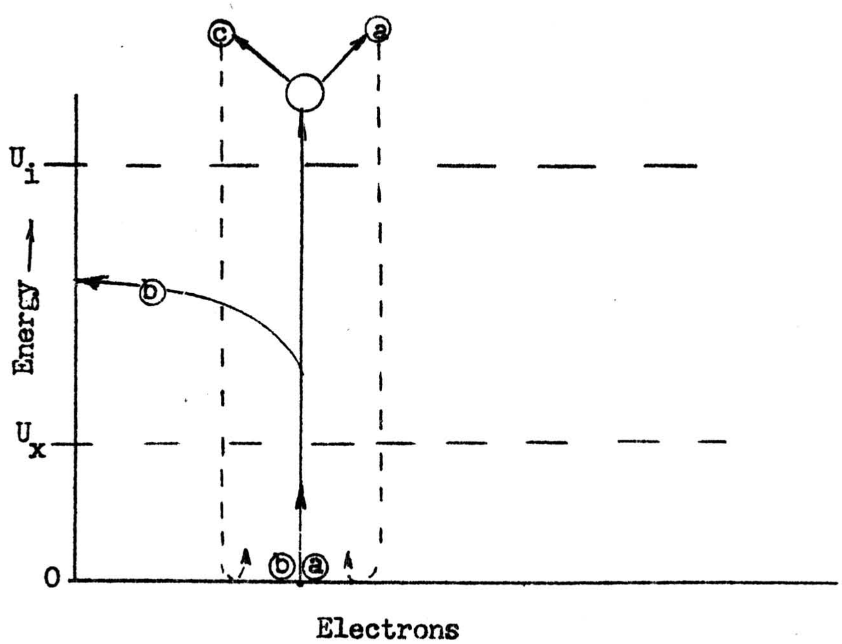


Fig. 10

At one point along the scale we establish what we call the excitation energy, U_x; at another point, the ionization energy, U_i. An electron that has an energy of U_i is capable of exciting, but not ionizing, a molecule; that is, it merely disturbs the arrangement of electrons. In the region bounded by energy levels U_x and U_i an electron can seriously distort the atom or molecule without ionizing it.

In exciting a molecule the electron expends energy. In the entire region bounded by U_x and U_i the electron has to fight against losing energy while it is in the process of gaining energy from the field so that it can reach the level necessary for ionizing.

Assume we have two electrons born with zero energy as a result of some action, such as radioactivity or the action of the field. To keep track of our two electrons we will name them (a) and (b). The two electrons begin a journey up the energy scale. On the way up they are going to lose energy. They will collide with neutral atoms and molecules, and in so doing lose energy either by recoil or excitation. The loss by recoil is brought about simply by the electron hitting an object of finite mass, and bouncing off.

If the electric field is strong enough to compensate for the energy losses, the electrons can keep on traveling. Electron (b) may at some point in the excitation region diffuse out of the volume. Electron (a) can receive enough energy from the field to keep going to the ionization region, where it can hit a molecule and be deflected. In hitting the molecule, though, it knocks out another electron, which we can call (c). Here, again, are two electrons, (a) and (c). (c) now takes the place of (b) which diffused out. Generally, these electrons have very low energies, so that they are at the bottom of the energy scale from where they begin climbing up. Thus there are two new electrons to repeat the process, and the ionization is sustained. This explanation is the same as previously given in terms of an r-f cavity, but more attention has been given to energy requirements.

It will be noted that the excitation region is a critical stage through which an electron has to pass before it can ionize a gas. For each gas there are several types of excitation possible. These excitation states, found in the excitation range are largely responsible for the size of the electric field required to break down a gas. We can depict excitation regions for various gases in an energy level diagram such as Fig. 11.

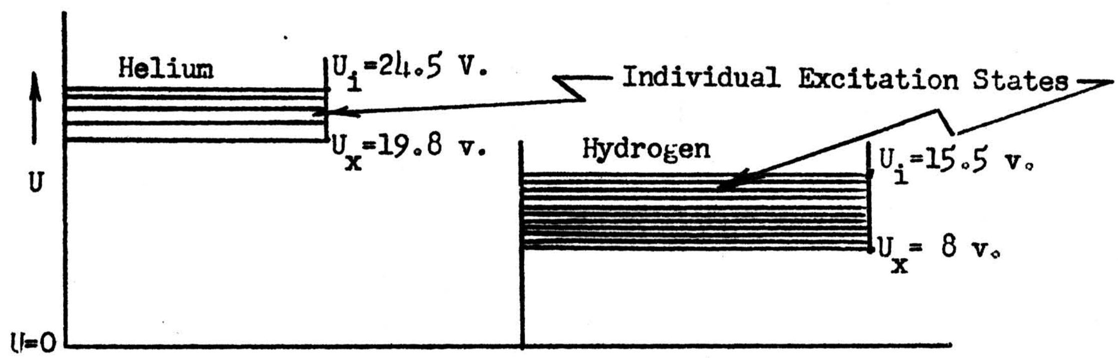


Figure 11

If we compare the excitation region of helium with that of hydrogen, we observe, first, that though it takes more energy initially to excite a helium atom, it requires less additional energy to ionize the gas. Also for the case of hydrogen, the region is more densely populated with excitation states. For the monatomic gases, such as helium, argon, xenon and krypton, the number of available excitation states is relatively small, and in a confined region. For a gas such as hydrogen, there is a comparative galaxy of excitation states in a relatively spacious region. In other words, an electron bent on ionizing helium travels through a fairly short obstacle course; whereas an electron with the same goal in hydrogen has a longer, more arduous obstacle course. That explains why hydrogen is more difficult to breakdown than helium or argon.

Summary of Gas Breakdown Phenomenon

To review, a new electron must be brought up the energy scale for every electron which diffuses out. That is, the field must be sufficiently large to make up for excitation, and the compensation must be supplied before one electron diffuses out. Then we can get the breakdown as we illustrated in the small and large containers (Figs. 8 and 9.) With an effectively large container, electron density builds up at a nearly exponential rate until the complete breakdown is achieved.

At this point we can define more clearly what we mean by complete breakdown. We recall that, when ambi-polar diffusion is established, the field is such that fewer electrons can diffuse out of the region. Suppose we plot the field against time in an r-f cavity of some sort, as in Fig. 12. If we start off with a few electrons and increase the quantity slowly, there is no particular reaction on the field. When we reach very large densities, ambi-polar diffusion sets in and fewer electrons are lost, and a smaller field is required to maintain the discharge in the region from A to B. The field drops simply because of the internal impedance of the generator supplying energy. The high

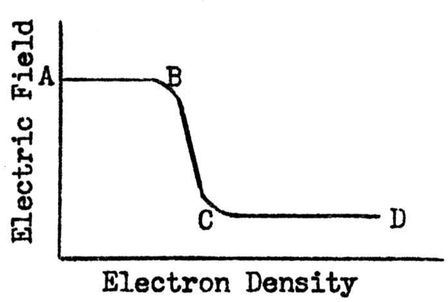


Fig. 12

field cannot be maintained at the level B - C because of the drain on the generator resulting from the discharge conditions. It drops down to the level labeled C - D, the "maintaining field." The drop from B to C signifies what we normally call "breakdown."

The breakdown occurs when the electron density is of the order of 10^6 to $10^7/\text{cm.}^3$. With such a density, the r-f field strength will decrease to the maintaining field value, but we must build up a larger density to get the reflection of the r-f that is required in TR tubes. The density must build up until the number of electrons per cubic centimeter is about 10^{11} for S-band, and 10^{12} for X-band, before there is any appreciable reflection of the r-f by the electrons.

The next lecture will discuss at greater length the interaction of the r-f field with the plasma.

OPERATION AND DESIGN

OF TR AND ATR TUBES

Lecture No. 9

Gas Discharges, Part II

In the previous lecture were presented the breakdown properties of gas for both d-c and a-c conditions, stressing those a-c properties important in the operation of TR tubes. This discussion will cover the type of interaction between an r-f field and plasma, and also the recovery properties of the plasma.

Wave Propagation in Plasma:

We have in the preceding lecture considered the conditions for ionization of a gas to occur under the action of an r-f field. When the ionization is sufficiently intense so that the density of positive ions and electrons is greater than about $10^5/cm^3$ the region is considered to contain a plasma. We shall examine some of the properties of this plasma by determining the characteristics of the current resulting from the r-f field.

Our first step is to find an expression for the current resulting from an r-f field propagated in space. With a time varying electric field we associate a displacement current whose magnitude is proportional to the time rate of change of the electric field, and which may exist along with real conduction current.

Let the electric field strength be expressed by

$$E = E_0 \cos \omega t \tag{1}$$

where E_0 is the magnitude of the electric field,

ω the radiant frequency of the field,

and t time.

We define electric displacement, D, in free space as equal to the field strength times the dielectric constant of free space, or permittivity, $\mathcal{E}_o$.

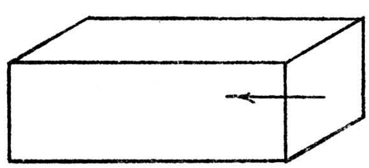
$$D = \mathcal{E}_o E \qquad (2)$$

The displacement current I_D, is the rate of change of this displacement with respect to time. I_D is now given by

$$I_D = \frac{\partial D}{\partial t} = \mathcal{E}_o \frac{\partial E}{\partial t}$$

$$= - \omega \, \mathcal{E}_o \, E_o \sin \omega t \qquad (3)$$

If electrons are present, they will move in the field E and constitute an electric current which we will call I_e. To obtain the formula for I_e we can visualize electrons moving in a rectangular box whose cross-section is one square centimeter. Let us assume, to begin with, that the electrons are moving in one direction under a d-c field. Let us say that the electrons travel along the long axis of the box of length v, with a velocity of v. The current density equals the number of electrons N in the box times the charge of the electrons $-e$. (The negative sign is used to take into account the negative charge of electrons.) Since the number of electrons in the box is equal to the number of electrons per cubic centimeter, n, times the length of the box, the current passing through the box is



v cm.

$$I = - Ne = - nve \qquad (4)$$

Suppose, now, that the electrons are in the r-f field defined by (1). By making use of the fundamental relations of force, mass and velocity, we can derive the formula for v

$$v = -\frac{e}{\omega m} E_o \sin \omega t \tag{5}$$

where m is the mass of the electron.

Substituting (5) in (4), we find that the current density due to the electrons in an r-f field is

$$I_e = \frac{n e^2}{m \omega} E_o \sin \omega t \tag{6}$$

The total current occurring in the field is the sum of the displacement current and the electron current

$$I_T = I_D + I_e$$

$$= \left[-\omega \varepsilon_o + \frac{n e^2}{m \omega} \right] E_o \sin \omega t \tag{7}$$

Now suppose that the r-f had been propagated in some dielectric material such as polystyrene. In that case we would have to account for another factor--the dielectric constant of the material K_e. The total current in the dielectric would be

$$I_T = -\omega \varepsilon_o K_e E_o \sin \omega t \tag{8}$$

If we re-arrange (7)

$$I_T = -\omega \varepsilon_o \left[1 - \frac{n e^2}{m \omega^2 \varepsilon_o} \right] E_o \sin \omega t \tag{7a}$$

we see the similarity between (7a) and (8). In (7a) the expression in brackets is analogous to K_e in (8), and in fact we can call the expression the dielectric coefficient of the plasma K_p.

$$K_p = 1 - \frac{n e^2}{m \omega^2 \varepsilon_o} \tag{9}$$

The plasma acts as a dielectric whose dielectric coefficient K_p is less than 1.

Dielectric Effect on R-F Field:

From what we know about the interaction between dielectrics and wave propagation we can infer the most important effects of the plasma on r-f. For example, a rectangular waveguide of dimensions 1" x 2" has a cut-off wavelength of 10 cm. That is, the dimensions of this waveguide are too small to propagate 10 cm waves. But, if we fill the waveguide with a dielectric material such as polystyrene, for which $K_e = 6$, the waveguide appears to be larger electrically, and the 10 cm wavelength is readily propagated down the guide.

If, however, we have a plasma in the waveguide, the effect will be opposite to that of polystyrene. Since K_p is less than 1, the dimensions of the waveguide instead of becoming effectively larger, will become effectively smaller. A waveguide which ordinarily transmits 10 cm wavelengths will, as a result of the presence of plasma, attenuate 10 cm waves.

Electron Density and R-F Reflection:

The similarity between the effect of a plasma and that of a dielectric material provides one means of measuring electron density. For if we can measure the effective dielectric constant of a plasma, we can determine the density of electrons by formula (7a).

The formula for the total current given by (7a) also tells us at what densities to expect a strong interaction between the plasma and the r-f field. In particular, when K_p becomes zero, there is a very pronounced effect between the r-f and the plasma, and the plasma will in general become reflecting. For, when $K_p = 0$, a plasma infinite in extent will cause total reflection of the r-f. If the density is larger ($K_p < 0$), then nearly complete reflection will occur for a plasma less than infinite in extent.

This particular property of the plasma accounts for the reflection of the power incident to the TR tube. In the range of densities with which we are principally concerned, it is quite different from ordinary metal-like reflection. However, when n becomes larger by several orders of magnitude, the plasma reflection takes on the aspects of metal-like reflection.

Taking as a convenient value that density for which $K_p = 0$, we can plot density with respect to frequency. The plot demonstrates the fact that the density at which the strong interaction between the r-f field and the plasma occurs, increases as the square of the frequency.

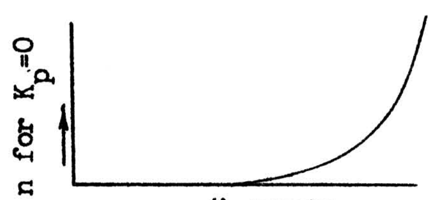


Fig. 1

The phenomenon of r-f reflection from a plasma is evident in the reflection of radio waves in the ionosphere. If we beam a radio wave directly to the ionosphere from the surface of the earth, we will at some time later receive an echo. If we plot the time elapsed between the sending and receiving of signal against the frequency of the signal transmitted, we will find that for a certain range of comparatively low frequencies the elapsed time corresponds to a distance from the earth's surface of about 100 kilometers. Beyond the first band of frequencies we will find that the time elapsed between transmitted signal and received echo increases with the frequency, until we reach a second band of frequencies for which the time elapsed is constant, corresponding to a distance of about 200 Km. Beyond

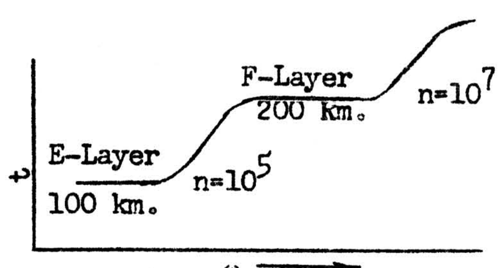


Fig. 2

this group of frequencies, the time increases with the frequency until a third plateau is reached, etc.

To interpret this curve we assume, first, that up to a certain height from the surface of the earth there is no reflection from electrons. Then at some particular height we find an electron concentration which causes the reflection of the r-f at relatively low frequencies. If we refer to a plot of expected reflection of the r-f, such as Fig. 1, we infer that the density of electrons in the region at 100 Km. is of the order of 10^5 electrons per cubic centimeter. The region at 100 Km. corresponds to the so-called E-layer of the ionosphere. If we increase the frequency so that the density of 10^5 electrons per cubic centimeter is not effective in reflecting the radio wave, the r-f penetrates the E-layer and reflects at a somewhat higher altitude where larger electron densities are encountered. The second region of reflection corresponds to the F-layer. This is the method used for getting plots of the ionospheric layers.

The problem of interaction in a TR is due partly to this kind of reflection. When a keep-alive current is so large that the electron density is great enough to cause an appreciable change in the dielectric constant from that of free space, then some reflection of the r-f will occur. However, the desired attenuation observed in TR's is caused by much larger densities than are present in the keep-alive gap before firing.

Plasma Decay:

The build-up of electron densities in an ionized gas results at some point of the process in establishing a plasma region characterized by large quantities of free positive and negative particles. This condition does not stay fixed, normally, because electrons are removed from the region by some means or other, bringing about the decay of the plasma. We can account for the loss of electrons by the following removal mechanisms:

1. loss by diffusion,
2. loss by attachment,
3. loss by recombination.

By isolating each mechanism for convenient study we can arrive at three distinct partial rates of loss which, combined, will give the total rate of loss.

In the absence of all other loss mechanisms, the rate of loss by diffusion of electrons between parallel plates would be given in terms of the separation between the plates, the diffusion coefficient and the electron density. In particular we are interested in the ambi-polar diffusion coefficient, so that

$$\text{Rate of loss by diffusion alone} = (\pi/L)^2\, D_a\, n$$

where L is the parallel plate separation, and as before,

D_a is the ambi-polar diffusion coefficient, while

n is the electron density.

The recombination loss is given by the recombination coefficient α times the density squared,

$$\text{Rate of loss by recombination alone} = \alpha n^2$$

Note that the density is squared. We can see this in the following way. The rate of loss must be proportional to the number of electrons per cubic centimeter, n_-, and also to the density of particles (positive ions, n_+) with which the electrons must collide to effect a recombination. Hence this rate is proportional to the product $n_+ n_-$. Since $n_+ \doteq n_-$ in most plasmas $(n > 10^5/\text{cm}^3)$, the rate of loss is proportional to the square of electron density

Attachment loss is relatively easy to visualize. In each collision between an electron and a molecule there is a probability that the electron will be physically caught by the molecule. This probability may be something in the order of 1 chance in 10,000; in which case an electron making 10,000 collisions will, on the average, be attached by a molecule. If the probability of attachment is called h, and the frequency of collisions ν_c, the loss due to attachment can be written

Rate of loss by attachment alone = $h \nu_c n$.

The rate of removal of electrons by either diffusion or attachment is linearly dependent upon the density. Thus, if we know the loss rate by either mechanism, we can calculate the time required for the plasma region to have decayed a given moment. As a measure of the decay time of a plasma we can define a time required for the density to decrease to 1/e, or about 1/3 its value. The time that it takes for the gas to reach this state of recovery we can call τ. This decay time is really only a figure of merit and will be only a fraction of the recovery time of the TR tube. The actual recovery time may be from 5 to 10 times as long. We shall assume that the recovery time T is five times as great as τ, our figure of merit.

The τ corresponding to the plasma decay as a result of recombination is not as simple to calculate since, as was pointed out previously, the recovery rate involves the square of the density. However, figures will be given for this recovery time as a function of density, so that the effect of all three mechanisms can be examined.

It is cautioned that the values appearing for the various τ's are given merely for illustrative purposes. They do not take into account the difference between gases or the more refined features of the theory, and are not intended for use in design consideration.

Diffusion & Recovery Time:

The decay time of a gas which is brought about solely by diffusion is

$$\tau_D = L^2/\pi^2 D_a \qquad (10)$$

where the D_a indicates that ambi-polar diffusion is the mechanism operating on the electrons.

It will be recalled that the ambi-polar diffusion coefficient D_a is much greater for fast electrons than for slow electrons. This will give us two possibilities. If we assume a 10 mm pressure in a container of about one cubic centimeter, we may calculate values of τ_D for both fast and slow electrons.

Immediately after the r-f is turned off the conditions existing in the plasma correspond to fast electrons. For this case we will find that

$$\tau_D \doteq 10 \ \mu\text{sec}$$

To designate the recovery time of the tube we may use T with a subscript appropriate to the mechanism in operation. Then, if the electrons remained at high energy, we would have

$$T_D \doteq 50 \ \mu\text{sec}$$

Within a few microseconds after the r-f is turned off, the electron energy has cooled down to that typical of slow electrons. For slow electrons we find

$$\tau_D \doteq 500 \ \mu\text{sec}$$

and the recovery time of the tube

$$T_D \doteq 2500 \ \mu\text{sec}$$

If diffusion alone were operative on the gas, we would expect the recovery time of the tube to be about 2500 μsec.

Attachment & Recovery Time:

If we consider attachment alone as the mechanism responsible for the plasma decay, we may write

$$\tau_A = \frac{1}{h\,\nu_c} \tag{11}$$

Again we have a choice of fast electrons or slow electrons. Assuming water vapor and 10 mm pressure, for fast electrons

$$\tau_A \doteq .25\ \mu\text{sec}$$

For slow electrons

$$\tau_A \doteq 2.5\ \mu\text{sec}$$

However, the situation is somewhat complicated by the fact that the electrons can be described neither as fast nor as slow. At one time many electrons of many energies exist. The distribution of electrons in energy is a curve humped at some point. In the case of an active discharge the maximum might occur at about 5 volts. The attachment probability

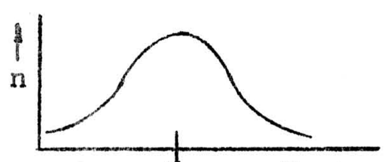


Fig. 3
Energy Distribution of Electrons

h is dependent on energy as well. For example, if we consider water and plot the attachment probability as a function of energy, the curve would have a minimum of attachment probability at about 5 volts, where most of the electrons are found.

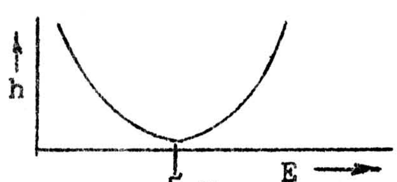


Fig. 4

During a discharge a relatively large number of fast electrons making collisions will become attached. This is true also of the electrons in the low energy levels. But not many electrons located in the middle energy range of the distribution are going to make attaching collisions. The recovery of the plasma, then, is dependent not only on the attachment probability, but also on the particular manner in which the electrons are spread in energy.

If the electrons maintained the spread of energy depicted above, it would take a long time before the major portion of electrons were attached to molecules. Consequently, the tube would have a long recovery time. Fortunately, electrons can lose energy by elastic collisions with neutral molecules. The high energy peak is characteristic of a hot discharge, that is, a discharge with the r-f field imposed upon the gas. When the field is removed, electrons, losing energy by collisions, tend to migrate towards low energy, where the probability is high, and where they will be attached and removed from the plasma. This change to the relatively low energy range is effected in a few microseconds. A few more microseconds elapses before the attaching process is completed. Recovery time can be expected to range from 5 to 10 μsec.

The mechanism of attachment is somewhat interesting in its own right. Let us look a little more closely at it. Normally, when an electron collides with a neutral atom or molecule there will be no attachment, since electronically the neutral atom or molecule is in a balanced state--does not require further electrons to fill its energy complement. Yet there is a rather wide prevalence of attachment, and it might be interesting to find out under what conditions the mechanism occurs.

Since we are concerned primarily with water, we will look into the conditions involved there for attachment. The attachment of higher energy electrons is probably due to a dissociation of the water in the following type of reaction:

$$e + 5.4 \text{ volts} + H_2O \longrightarrow (HO)^- + H$$

The electron on the high energy side of the energy distribution carries, say, +5.4 volts into the reaction with a water molecule. The result will be $(HO)^-$, a negative ion. A positively charged atom of hydrogen can later combine with the negative ion.

The above explanation will not account for attachment at low electron energies. The latter phenomenon does not seem to be too well understood. It is believed, though, that the type of reaction occurring at low energy levels is mainly one of condensation of water around a charge, in particular, a condensation involving two molecules

$$2\ H_2O + e = (H_2O)_2^-$$

<u>Recombination and Recovery Time:</u>

Since the rate of removal of electrons by recombination is proportional to the square of the electron density, it is difficult to define a τ in the manner applicable to the diffusion and attachment mechanisms. When densities are very high, recombination is an efficient remover of electrons. For example, if the density is of the order of 10^{12} we can define a pseudo τ_R which for hydrogen will be something like 0.5 µsec. As soon as the density drops to, say, 10^{11}, however, τ_R of the gas increases to about 5 µsec.

The recombination mechanism would seem to be most valuable in the higher frequency tubes. For a comparatively low frequency tube, such as the S-band TR, to attain rapid recovery it is preferable to reduce the concentration of electrons to well below 10^{10}. But for a 3 cm TR, and tubes of much shorter wavelengths, the method of recombination might easily be the most useful in effecting reduced recovery time. For the higher frequency tubes we would want to reduce n to a low value after firing, but not as low as that for S-band. For we have shown in our qualitative analysis of the dielectric coefficient of the plasma that attenuation for a high frequency field occurs only at densities much higher than necessary for a low frequency field. Hence, recombination will be expremely rapid in the range of high densities.

Removal Mechanisms & Pressure:

We can summarize the several mechanisms operating to remove electrons from a plasma region by the following table of comparisons:

Relative Differences in Electron Removal Mechanisms

$p \doteq 10$ mm in a volume of 1 cm^3

	Diffusion	Attachment	Recombination
Rate of Loss	$\left(\frac{\pi}{L}\right)^2 D_a\, n$	$h\, \nu_c\, n$	$\alpha\, n^2$
Recovery time of plasma	$\tau_D = \left(\frac{L}{\pi}\right)^2 \frac{1}{D_a}$ $\doteq 500$ µsec	$A = \frac{1}{h\, \nu_c}$ $\doteq 5$–10 µsec	for $n = 10^{12}$ $\tau_R \doteq 0.5$ µsec for $n = 10^{11}$ $\tau_R \doteq 5.0$ µsec
Recovery time of tube	$T_D \doteq 2500$ µsec	$T_A \doteq 25$–50 µsec	for $n = 10^{12}$ $T_R \doteq 2.5$ µsec for $n = 10^{11}$ $T_R \doteq 25$ µsec

As a general review of these removal mechanisms, the pressure dependence serves to give somewhat of an insight to the problem. In the case of diffusion, the time involved increases with the pressure. The higher the pressure, the longer the time.

For attachment, the greater the pressure, the shorter the recovery time. In addition, the increased pressure permits more collisions and, consequently, an accelerated loss of energies so that larger numbers of electrons attain the range of relatively low energies at which the attachment probability is high.

In the case of recombination, there is no direct correlation between the rate of electron loss and pressure. In general, though, the higher the pressure, the faster the recombination.

OPERATION AND DESIGN OF TR AND ATR TUBES

Lecture No. 10

Tuning Methods and Tuning Errors of TR's

Today, we will consider the problems associated with tuning broad-band TR tubes. These tubes are made, as we have seen, as a series of four or five tuned elements properly spaced. The outer two elements are the windows. These are pretuned and need not concern us. The internal elements, however, must be tuned after assembly. The problem is to obtain a sufficiently accurate tuning to resonance of these two or three elements after the tube has been constructed.

The problem is important and no completely satisfactory solution has been found. For example, we have made a few demountable tubes for experimental purposes, in which each element could be removed for individual tuning and then replaced. The resultant bandpass was in the neighborhood of 17% or 18%, half again as large as production models.

There are two methods of tuning in general use:

1. the multiple-frequency method, usually utilizing a triple pipper, and

2. the single frequency, employing short-circuiting.

Multiple-Frequency Tuning:

The first method normally is dependent on the triple pipper. By this device the reflection coefficients at three frequencies are examined simultaneously. The tunable elements are adjusted until all three reflected waves are brought down below a given limit.

An advantage of this method is that if the three frequencies selected are the specified check points, then the tube can be made automatically to pass those particular points. This expedient is to be used with caution, however, since the SWR in between the check points examined might quite conceivably reach very high values.

This latter difficulty would be avoided if we had sweep oscillators to view the response at all frequencies of the band at one time. This has been done for the 1B58, at 10 cm. For other frequency ranges, however, we do not have such a device, and reliance must be placed on the triple pipper.

Although these methods are commonly used, we shall not consider them further. Unfortunately, no analysis of the possible errors of such methods has been made. The reason is not that the problems are unimportant, but simply that they present fundamental difficulties of analysis. We hope eventually to find some satisfactory approach to the problem. But, until we do, there is little we could say.

Single-Frequency Tuning:

The second method of tuning tubes involves short-circuiting the elements and plotting position of minimum at center frequency. If we symbolize the elements of a tube by crosses (Fig. 1a), the outside two represent the windows. These are already tuned to resonance, so we place circles around the midpoints to indicate the elements are transparent at the frequency chosen. The internal elements are numbered 1, 2 and 3.

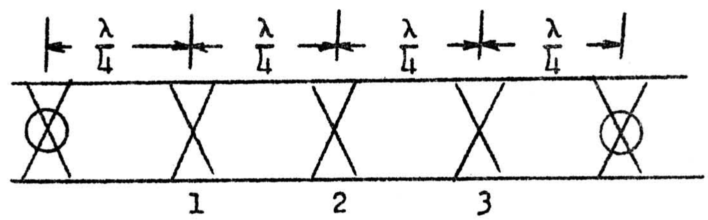


Fig. 1 a

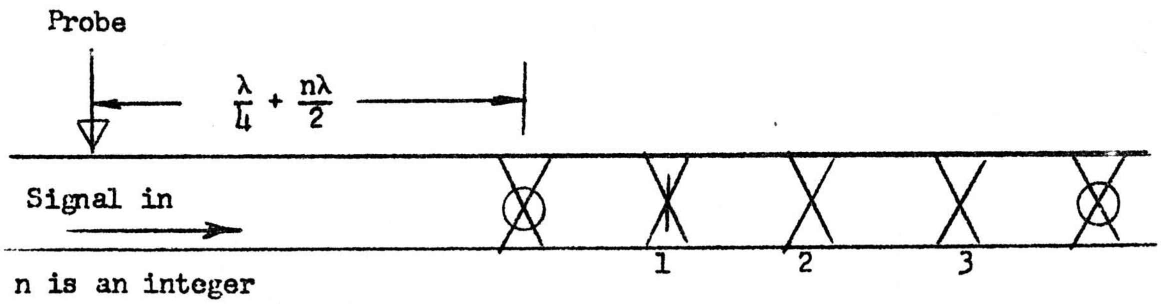


Fig. 1 b

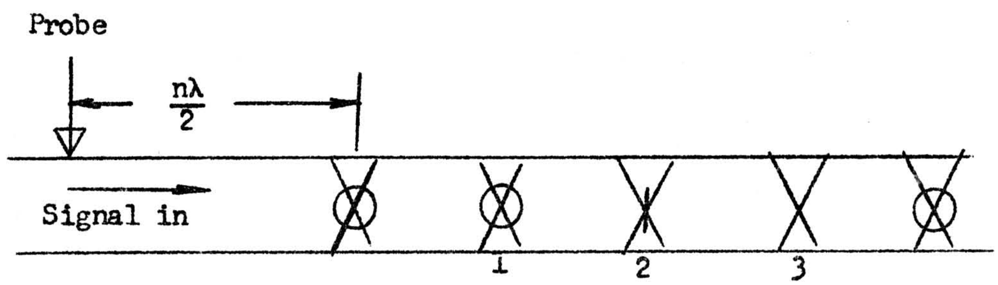


Fig. 1 c

The No. 1 element is short-circuited as indicated by the bar through it (Fig. 1 b). At some point a quarter wavelength plus an integral number of half wavelengths from the window, the probe measures a position of minimum, since the elements are supposed to be a quarter wavelength apart.

The probe is moved a quarter wavelength down the line towards the tube. The short is removed from No. 1 element and placed in No. 2 element. The No. 1 element is adjusted to give a position of minimum reading for the probe. No. 1 element now is presumably tuned to transparency.

The No. 2 element may then be tuned in similar fashion by short-circuiting No. 3. Alternatively, the tube may be reversed and the No. 3 element tuned. Finally, the remaining element is tuned to minimum SWR. If the tube is properly designed, and if the tuning procedure is performed with exact accuracy, the two short-circuiting procedures, called the 1-2-3 and the 1-3-2 methods, should give identical results.

The problem of either of these methods is to obtain a sufficiently good short circuit. In the butterfly gap design, there is not too much trouble in getting a good contact with a short. The cone and iris design presents a trickier situation since great care must be taken to avoid damaging the cones.

Tuning Errors:

The errors in tuning possible with the use of the short circuit method may be classified as follows:

1. Reactive short circuit,
2. Residual phase shift,
3. Incorrect spacing,
4. Random

The fourth type is a "catch-all." The effect of these errors on tuning of the tube will be discussed for each type in the order given.

Reactive Short-Circuit:

If in the short circuiting method of tuning, the position of the effective electrical short does not coincide with the position of mechanical short, the displacement from the physical center of the gap may be described as a reactive short circuit. The short may be effectively near, instead of at the physical shorting device, but still may be sufficient to block practically all the power. At least that part which does leak through may be neglected. Thus we can say that there is no resistive loss in the element, or that it is a purely reactive circuit.

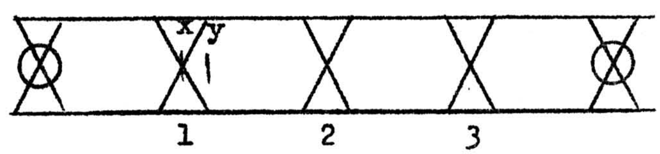


Fig. 2 a

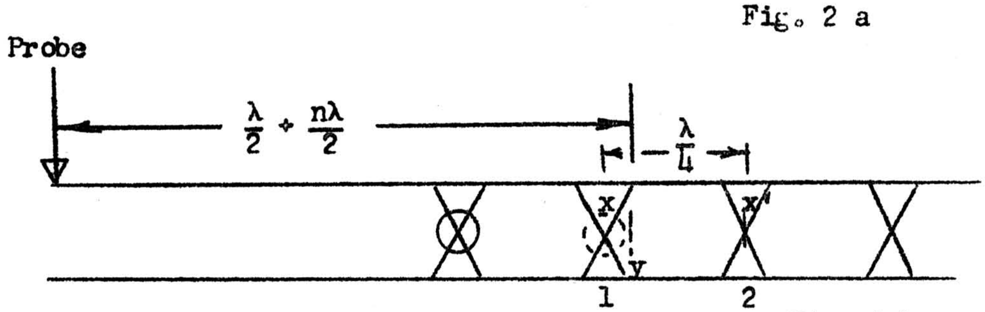


Fig. 2 b

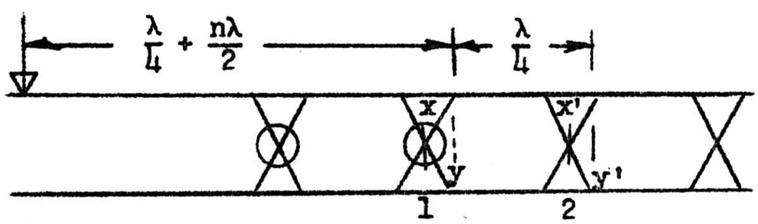


Fig. 2 c

The short may be effectively displaced either in front of or behind the physical shorting device. Assume, for this discussion that it appears behind as in Fig. 2a (at y instead of at x).

The first consideration is how does this apparent shift affect the tuning of the tube. At first glance there would seem to be a considerable effect since the basis of the method is the quarter-wave separation of the elements so that the short circuit on one appears as an open circuit - infinite shunting impedances - on the other. However, on closer inspection, we find that a reactive short-circuit has no effect on the designs in use. The original reference point was set by the position of minimum when No. 1 was shorted. The probe then will be on odd number quarter wavelengths from y (and not x). If the No. 2 element behaves exactly as the No. 1 element, its short will appear not at x' but at y' (Fig. 2c), and the distance between the two positions of effective short will be a quarter wave. So that after the probe has been shifted a quarter wave, and the No. 1 element is adjusted until the probe again indicates a minimum, this latter reading will occur at the point where the element being adjusted is exactly transparent—in other words, when it is in theoretical tune.

The above explanation would hold as well for the position of the short appearing electrically on the other side of the physical short.

We see, then, that the reactive short circuit has no effect on the tuning of the tube.

Residual Phase Shift:

If upon retuning the shorted element, it effectively expands or contracts some amount, the error in tuning that results from this change we can call "residual phase shift."

Suppose we consider a black box containing a passband network. (Fig. 3)

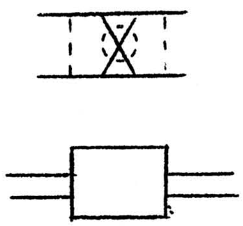


Fig. 3

There are two characteristics of interest with respect to the bandpass performance of that box. One is the attenuation, and the other, the phase shift. The phase shift is related principally to what happens to the attenuation far from the passband and, in particular, as to how symmetrical it is on both sides of the passband. Since waveguide is quite unsymmetrical, having a cutoff at one end, any element in the waveguide almost necessarily has a phase shift at any frequency at which it does not have attenuation.

We are interested in what effect the phase shift has on the tuning. Assume that when all of the internal elements of a tube are shorted they appear to be equally spaced a distance ℓ apart. (Fig. 4 a) Then we might find when the tube is being tuned that the first element seems to have expanded (or contracted). We can say that effectively a piece has been inserted in the line between the first two elements. Suppose in the procedure we find that the next element behaves as though it had expanded, too. An additional piece is then effectively inserted between the first and second elements. In this manner, as we successively tune down the line of the tube, we find that the elements become spaced an electrical distance apart that is greater than the physical spacing.

Fig. 4a

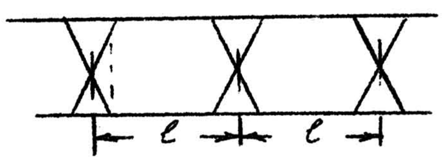


Fig. 4b

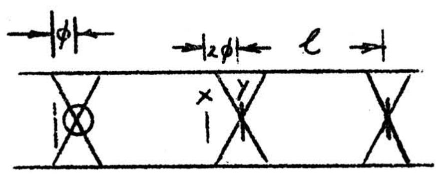


Fig. 4c

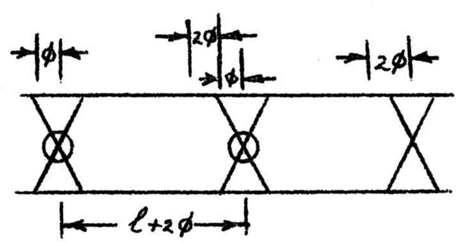
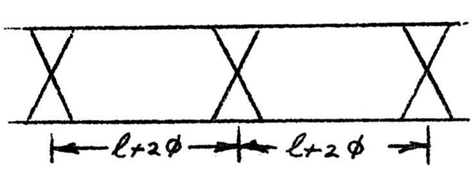


Fig. 4d



Let us say that the short circuit of the first element appears physically at x. In the process of tuning the tube it appears to have shifted to y.

(Fig. 4b) If we call the distance between the two positions 2ϕ, then all the other elements effectively have been moved 2ϕ from the first, and the first element itself appears to have been displaced ϕ.

In tuning the second element we note that it behaves in the same fashion as the first. It, too, then is effectively displaced a single ϕ, while there is another section 2ϕ long apparently added to the electrical length between it and the succeeding elements. (Fig. 4c)

Then, if we check the spacing between the electrical short for the first element and that for the second, we find that it is no longer ℓ but $\ell + 2\phi$. (Fig. 4c) This will hold for the other spacings. We find that all the elements are displaced by the same amount 2ϕ. (Fig. 4d)

Had we considered the condition of the elements contracting 2ϕ we would find that the spacing between elements changed electrically from ℓ to $\ell - 2\phi$.

The residual phase shift then would throw the tube out of tune since the electrical spacing between elements changed with retuning. The error, however, is readily compensated by changing the length of the line between elements. All that is necessary is to make the electrical distance between the <u>tuned</u> elements equal to the guided quarter wavelength

$$\ell \pm 2\phi = \lambda_g/4$$

Thus, in the case of expansion ($\ell + 2\phi$), the initial distance between elements is shortened physically by 2ϕ. In the case of contraction ($\ell - 2\phi$), the initial distance is increased by adding a section equal to 2ϕ. If the design is modified in this way, ideal tuning can - theoretically - be achieved.

<u>Incorrect Spacing</u>:

Suppose we have a tube for which any residual phase shift has been compensated. And in tuning this tube we have so chosen the point of reference (Fig. 2) that any reactive element of the short circuit has been automatically

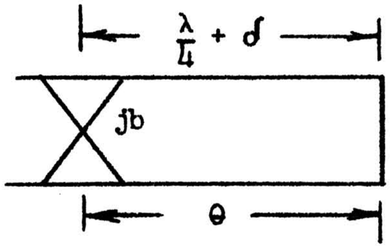


Fig. 5

cancelled. But, in the mechanical construction an error was made in the spacing between elements, so that the short circuit appearing behind the element is a distance not equal to a quarter wavelength. (Fig. 5) We can evaluate the consequences of this error by calculating the admittance to which each element will be tuned. The admittance of the element at center frequency is a measure of the degree to which it is mistuned.

The matrix of the shunt admittance (normalized) of the element is $\begin{Vmatrix} 1 & 0 \\ jb & 1 \end{Vmatrix}$. If the electrical distance between elements is θ, the matrix of the section of the line between the elements is $\begin{Vmatrix} \cos\theta & j\sin\theta \\ j\sin\theta & \cos\theta \end{Vmatrix}$.

The short circuit has no voltage and can have any finite current. Its column matrix is $\begin{Vmatrix} 0 \\ 1 \end{Vmatrix}$.

The admittance of the circuit in Fig. 5 is the product of these three matrices

$$\begin{Vmatrix} 1 & 0 \\ jb & 1 \end{Vmatrix} \times \begin{Vmatrix} \cos\theta & j\sin\theta \\ j\sin\theta & \cos\theta \end{Vmatrix} \times \begin{Vmatrix} 0 \\ 1 \end{Vmatrix} = \begin{Vmatrix} j\sin\theta \\ \cos\theta - b\sin\theta \end{Vmatrix}$$

The input impedance, then, of the circuit is

$$Z_{in} = \frac{j\sin\theta}{\cos\theta - b\sin\theta}$$

We tune the element to an open circuit, or make $Z_{in} = \infty$. Then,

$$\frac{j\sin\theta}{\cos\theta - b\sin\theta} = \infty$$

or

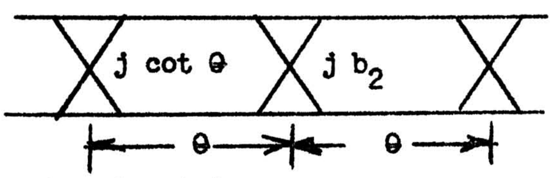
$$\cos \theta - b \sin \theta = 0$$

so that

$$b = \cot \theta$$

In other words, at center band the first element we have tuned has an admittance equal to $\cot \theta$. If the spacing is correct, and $\theta = 90^\circ$, then b is tuned to zero admittance - ideal tuning.

By this method we can find the admittance values of the other elements. Assume that we use the 1, 2, 3, method of tuning a five-element tube. We will find that the admittance of the second element (b_2) is equal to $2 \cot \theta$. The third element is tuned to unity SWR. To obtain this, it must be tuned to $b_3 = \cot \theta$, the same as the first.



$b_2 = 2 \cot \theta$ Fig. 6

We have then expressed the susceptance of each internal element in terms of the electrical spacing between elements. We can consider the effect on the bandpass when this spacing is not 90°.

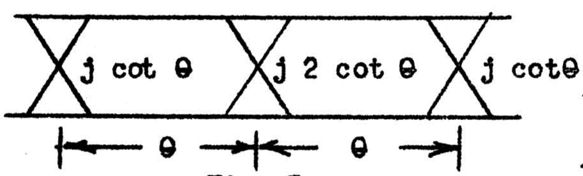


Fig. 7

When the bandpass of the three internal elements alone is plotted, there seems to be little effect resulting from incorrect spacing. In Fig. 8, curve A is the bandpass for three elements spaced 90° apart. Curve B is the bandpass of the tube with no windows and spacing between elements of 80° or about 10% short. The difference is small. But when windows tuned exactly to center frequency are placed on the tube, the passband is definitely affected adversely, as shown by curve C, which should be compared with the dotted curve for 90° spacing. The effect of an error in spacing is seen to be quite serious.

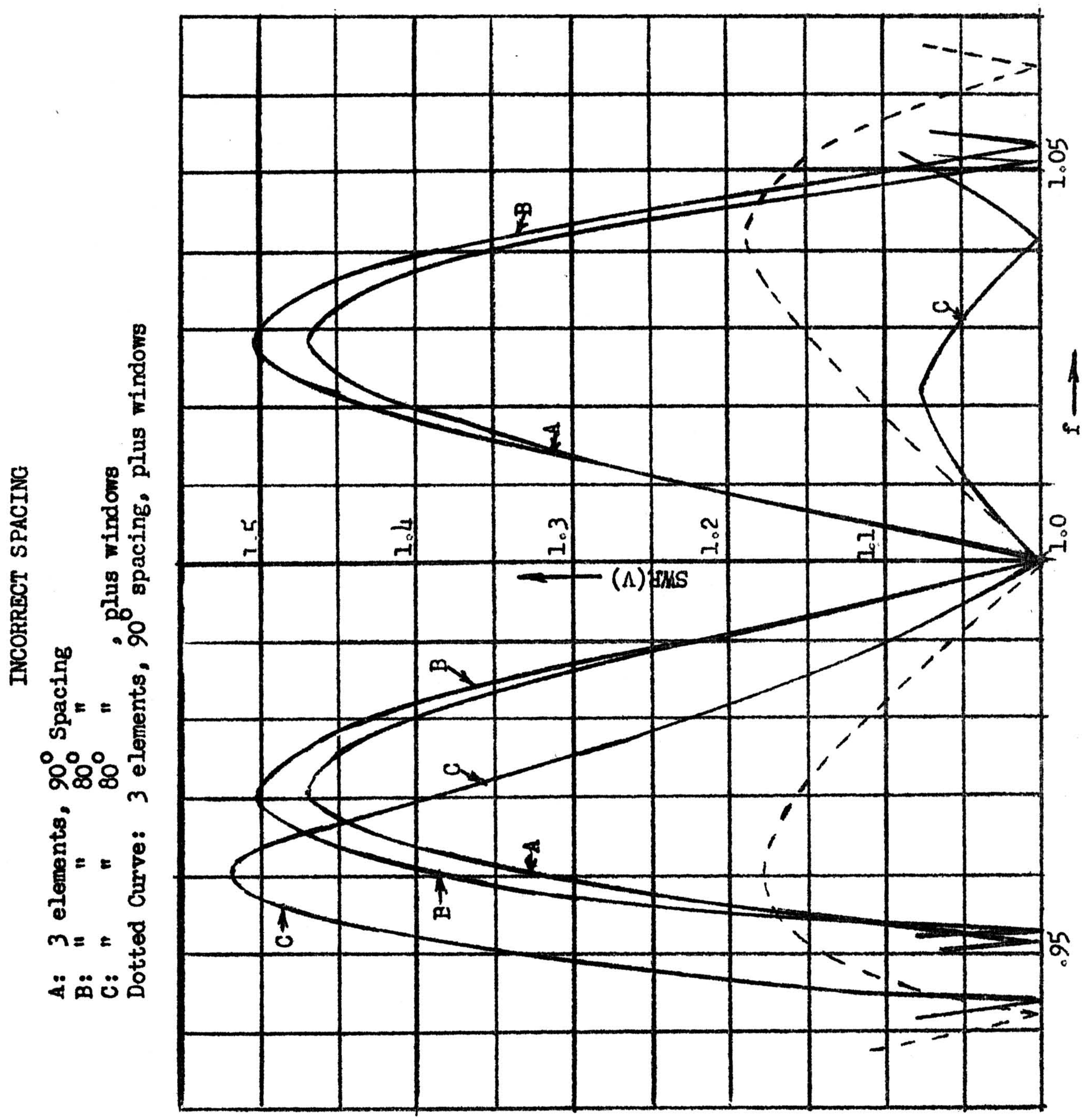

INCORRECT SPACING
A: 3 elements, 90° Spacing
B: " " 80° "
C: " " 80° " , plus windows
Dotted Curve: 3 elements, 90° spacing, plus windows
1.5
1.4
1.3
1.2
1.1
SWR(V)
.95
1.0
1.05
f
A
B
C

Fig. 8

Random Errors:

In this category we may place all errors that cannot be classified under any of the three types treated above. Since, however, this has too broad a scope to be handled, we shall inquire into the most probable types of such errors. Let us, then, assume a chance error is made in tuning the first element. Another small random error is perhaps made in the second. After the final stage is reached the first two elements can be retuned, if necessary, to give a flat SWR around the center of the band. Since the most accurate adjustment in the procedure is that of making the passband flat around the center frequency, we will restrict the random error classification to cases which do involve a low SWR at center band.

Then, looking into the tube at the center frequency, the normalized conductance is unity, and the normalized susceptance is zero. We can label the admittances at center band of the three elements respectively b_1, b_2 and b_3. (Fig. 9) The condition that the normalized conductance, the real part of the admittance, (normalized resistance) equal 1, can be found to be that

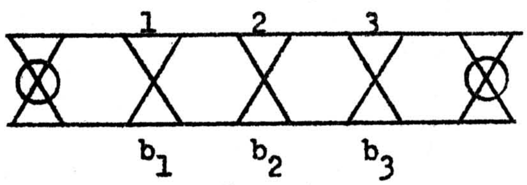


Fig. 9

$$b_2 b_3 = b_1 b_2$$

The condition that the reactance is zero can also be found:

$$b_1 + b_3 - b_2 = b_1\, b_2\, b_3$$

The first condition is satisfied either when $b_2 = 0$, or when $b_1 = b_3$. $b_2 = 0$ corresponds to the No. 2 element being on tune. If that is the case, from the condition for zero reactance

$$b_1 = - b_3$$

This relationship of the admittances corresponds to the No. 2 element being tuned to center frequency, the No. 1 element tuned above (or below) center frequency a certain amount, and the No. 3 element tuned below (or above) center frequency by this same amount. It amounts in effect to a stagger-tuned condition.

On the other hand, if

$$b_2 \neq 0$$

then, for matched conductance we must have

$$b_1 = b_3$$

Putting this into the reactive condition,

$$b_2 = \frac{2\,b_1}{1 + b_1^{\,2}}$$

If b_1 is small, b_2 may be approximated

$$b_2 \doteq 2\,b_1$$

The last expression can be translated in terms of the comparative tuning of the elements. If the No. 1 element is tuned high, the No. 3 element will be tuned equally high, and the center element will be tuned twice as high. This tuning error is quite similar to the effect obtained with incorrect spacing between elements.

The first random error (stagger-tuned effect) rapidly becomes disastrous to the passband of the tube. In Fig. 10 the theoretical passband of a tube without windows, internal elements tuned accurately, is plotted as curve A. If the first and third elements of the tube are mistuned by 1/2%, the bandpass of the three elements deteriorates to curve B, where the outside edges are rising rapidly. When windows are placed on the tube (curve C), deterioration

STAGGER TUNED

A: 3 elements, accurately tuned

B: 3 elements, 1^{st} & 3^{rd} mistuned by $\pm \frac{1}{2}$ %

C: Tube B with windows added

Dotted Curve: Tube A with windows added

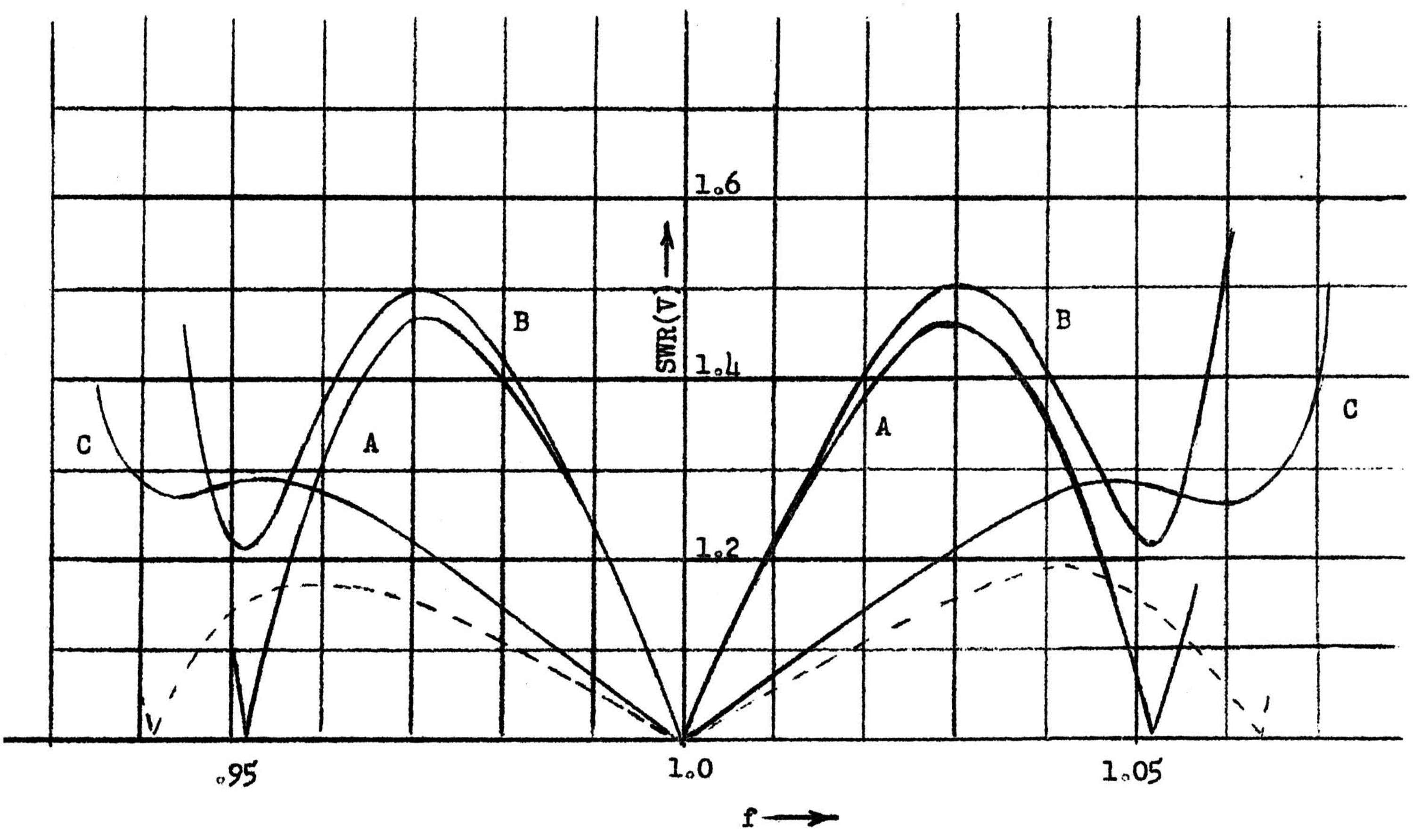


Fig. 10

of the bandpass is still pronounced. (Cf. the dotted curve showing the bandpass including windows if accurately tuned.)

From this analysis of passband characteristics it is seen that 1/2% is about the maximum that can be tolerated by way of stagger mistuning.

The other type of mistuning due to random errors affected the individual elements in a fashion resembling the effect obtained by incorrect spacing. In the random error case, however, the spacing has not changed electrically. In Fig. 11 the dotted curve shows the theoretical passband obtainable from a tube without windows, and with the first and third elements mistuned about 1%, the second 2%. From the point of view of the amount of attenuation within the bandpass, the curve is passable. But it is badly asymmetrical. When windows are placed on this tube, the asymmetry of attenuation characteristic becomes worse as shown in the solid curve. The amount of mistuning represented in this plot, then, is about the amount that can be tolerated.

With either type of random mistuning 1/2% is approximately the maximum that can be tolerated. Not only is this a small tolerance, but also it is very nearly the probable minimum experimental error. These tubes are being tuned by shifts of the position of minimum, and so are dependent upon the accuracy with which the minimum is set. It is doubtful that the position of minimum can be set more accurately than 1/2%.

<u>Comparative Summary of Tuning Errors:</u>

<u>Reactive Short Circuit:</u> The electrical short of each element appears either in front of, or behind the shorting device, but any effect on the bandpass characteristics is automatically eliminated by the fact that the initial position of minimum is referred to the electrical short of the first

RANDOM TUNING--3-ELEMENT TUBE

1^{st} & 3^{rd} elements mistuned about 1%,

2^{nd} element mistuned about 2%

Dotted Curve--no windows

Solid Curve--with windows

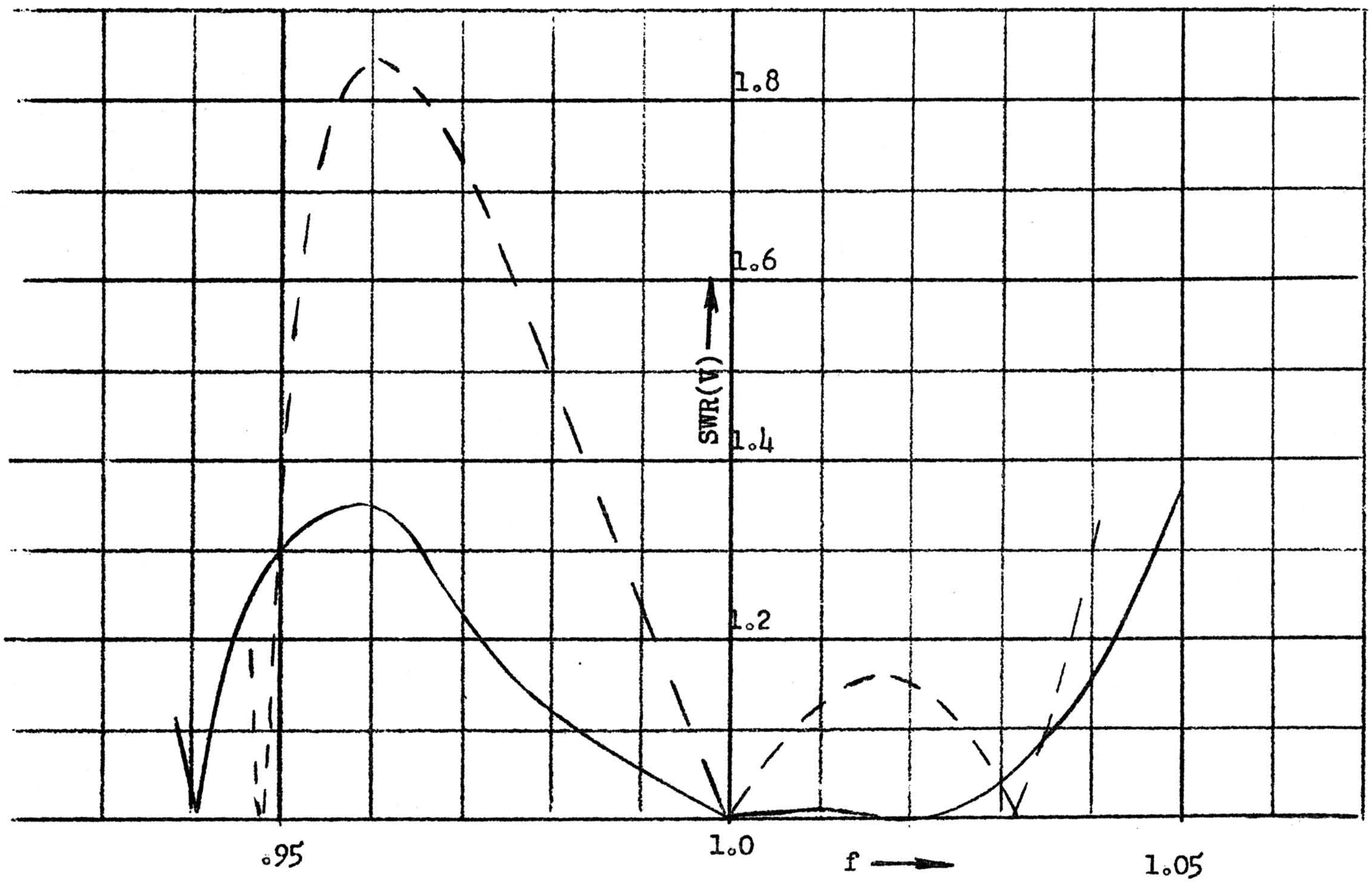


Fig. 11

element. Elements remain electrically spaced a quarter wavelength apart. If the shift in short location does not effectively disappear with tuning of the tube, the reactive short circuit is classifiable as residual phase shift.

Residual Phase Shift: The position of short in the element appears to charge electrically as a result of phase shift. The performance of the element resembles an expansion or contraction. If the elements appear to be expanding (the electrical position of short displaced further from the probe) it is equivalent to adding to the line between elements a section electrically equal to the phase shift. If the elements appear to be contracting, the end result is that a section equal to the phase shift is subtracted from the line. The error is easily compensated for by shortening or lengthening the spacing between elements.

Incorrect Spacing: If an error has been made in electrical spacing which is not compensated for by either of the above errors, as much as 10% off does not disturb the bandpass characteristics until the windows are placed on the tube. Spacing between elements not equal to a quarter wavelength results in the susceptance of the middle element being twice that of the outer elements.

Random Errors: Difficulties, such as foreign matter on the contacts which prevent the physical short from being a good electrical short, introduce random errors. The only random errors considered are those which occur in tubes with passband flat at center frequency. Two forms are possible: a stagger-tuned effect; and a condition in which the admittance of the center element is twice that of the outer two. The maximum random error that can be tolerated before the tube is thrown badly off tune is in the order of magnitude of 1/2%.

OPERATION AND DESIGN

OF TR AND ATR TUBES

Lecture No. 11

Test Equipment and Techniques, Pt. I

In this and the following meeting there will be discussed the testing techniques and test equipment used on TR and ATR tubes. They are given to present an idea of the purpose behind the various tests, the kind of equipment employed and the limits of the accuracy that can be expected with the equipment.

It was believed that, in the short period allotted, the most information could be conveyed if the first session were conducted on a lecture basis, and the second session were a round-table discussion in which questions on test procedure could be answered at length.

This lecture will be split into two topics: ATR tubes, and TR tubes. Before going into these it would be well to inquire into the general purpose of testing.

Functions of Tests:

As far as the Test Group is concerned, there are basically three functions that it has to fulfill:

1. Obtain operating characteristics of a tube.

2. Act as a control on production.

3. Obtain information to aid design and development.

The first function gives the necessary information to the buyer of the tube. The test results tell the equipment and design engineer what the performance of a particular TR or ATR is going to be in the system for which it is intended.

The rate of production of the tube is dependent on test results. If the test results show a good tube, production can continue at a rapid rate. If they indicate tube performance is questionable, then production is slowed down and processing is performed with caution. If tests show poor performance, then the time has come to get into the processing and locate the trouble.

The third function of testing is important from the standpoint of the design and development section. From the infancy of the tube, through its adolescence, until the pilot run and production, tests indicate to the design engineer the status of the design and electrical performance of the tube.

ATR Functions & Tuning:

Prior to describing the various tests performed on ATR tubes we will review the definition of the ATR. It will be recalled that an ATR is a device employed to decouple effectively the transmitter from the receiver during the receive cycle of the radar system, and present a good match to the magnetron during the transmit period. It offers zero impedance to high level and infinite impedance to low level signals, theoretically.

The ATR tube consists of a cavity, at one end of which is a low-Q window, resonant at center frequency. Inside the tube is a flexible baffle. (Fig. 1)

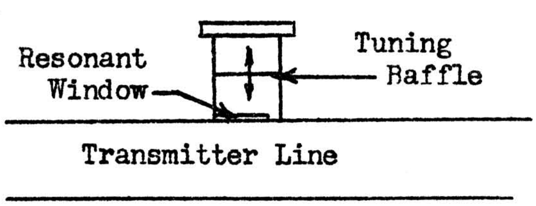


Fig. 1

The baffle is moved in an up or down direction to tune the cavity to resonance. After the tube is tuned, it is filled with gas at low pressure. When high power hits the window, the gas ionizes and the tube effectively presents a low impedance to the signal in the transmitter line.

When low power of the proper frequency reaches the tube, the ATR cavity resonates and sets up a large reflection towards the source, so that very little signal can pass by the tube down the transmitter line. Actually, the tuning of the tube is a matter of adjusting the baffle until the minimum signal is transmitted past the tube at low power, at which time the SWR is a maximum. (Fig. 2)

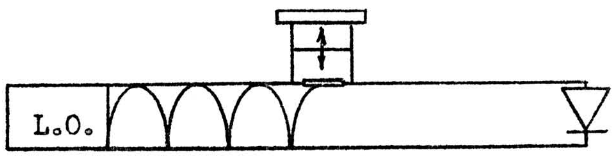


Fig. 2

ATR Tuning Bench Schematic

The low-level source signal is supplied by a local oscillator in the line. A crystal load indicates the amount of signal traveling down the line past the tube.

ATR Production Tests:

Before the tubes are given the various electrical production tests after filling, they are subjected to temperature cycles from -40° to 100° C. to check their ability to withstand severe weather and operating conditions.

The essential production tests are the measurement of susceptance, arc loss and firing time.

The equivalent normalized susceptance test is one of the most important tests given an ATR. It essentially tells how well a tube is tuned by measuring the phase of the standing-wave. Normally, each ATR is tested at least twice, once just after tuning before filling, and once after filling, plating and polishing. The tests after filling check whether or not the baffle has been detuned (moved) during the filling process, or whether the tube has otherwise been detuned by plating, polishing, or careless treatment.

The setup (Fig. 3) for measuring the phase of the standing wave consists of a local oscillator to supply the signal, a roving tuner (standing-wave detector) with an Ames or other type of dial gauge to read position of minimum, a reversible seat for the ATR and a matched load.

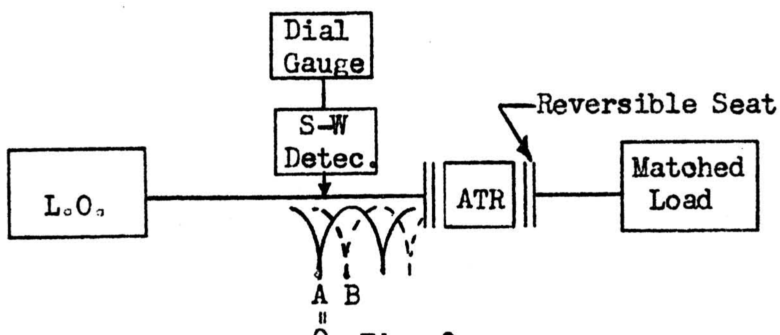


Fig. 3

A "standard" ATR is put in the tube seat. The probe of the standing-wave detector is moved along the line to locate a position of minimum in the standing wave set up by the tube. The tube seat is reversed, and the position of minimum reading is repeated. The average of the two readings is taken as the position of minimum for the standard tube (at A in Fig. 3), and the dial gauge is set to zero for this position.

The tube which is to be tested is then inserted in the reversible seat and the probe moved along the line to locate the minimum point of standing wave with the seat in both forward and reverse positions. The deflection from zero on the dial in each case is noted, and the average of the two readings is used to determine B for comparison with the standard tube. The travel from A to B in thousandths is translated into susceptance by means of a prepared chart on which the phase difference in thousandths has been calibrated into equivalent normalized susceptance.

Arc Loss & Firing Time are tested in a high power setup such as diagrammed in Fig. 4.

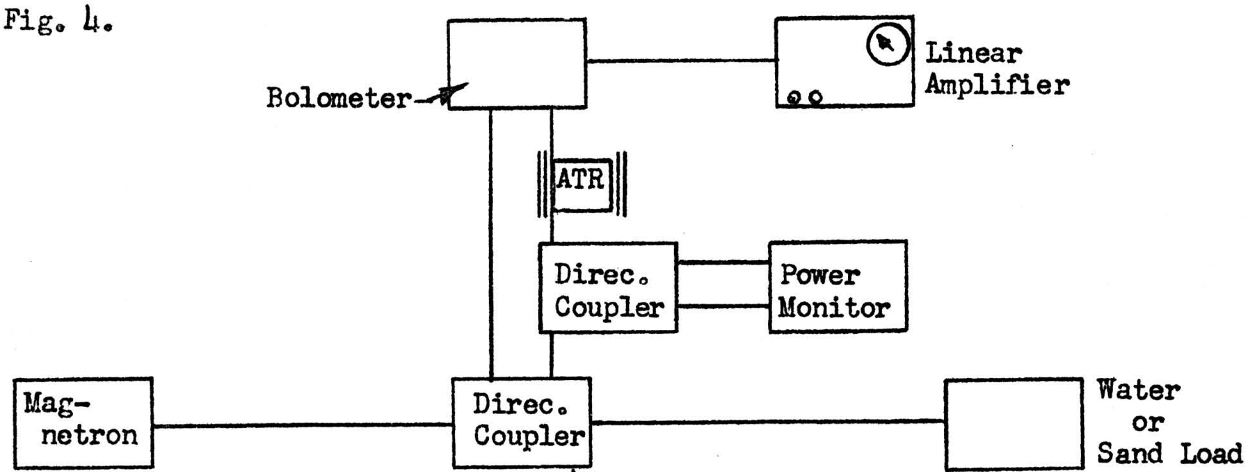


Fig. 4

The magnetron power output is measured by means of the calibrated water load, or thermistor bead mounted on a directional coupler. A dummy tube is placed in the ATR seat. Then the power from the magnetron is adjusted until the amount of power in the section of line holding the ATR seat is equal to the power specified for this test of the tube. The power going past the dummy ATR reaches the bolometer in which change of energy is indicated by the linear amplifier. When the tube under test is placed in the seat, the drop in indicated power is the arc loss of the tube.

The check on firing time is a matter of noting that the tube ionizes within 10 seconds after the power is turned on, or after the tube is inserted under power. Generally, when tubes are a long time in firing, they later become leakers.

ATR Design & Type Approval Tests:

It is not always possible to draw a sharp distinction between design and type approval tests. Depending upon the individual specifications written for a particular tube, the same type tests may be "design" for one tube and "type approval" for another.

The equivalent normalized conductance is one of the important design tests. It is measured in a low power setup such as Fig. 5.

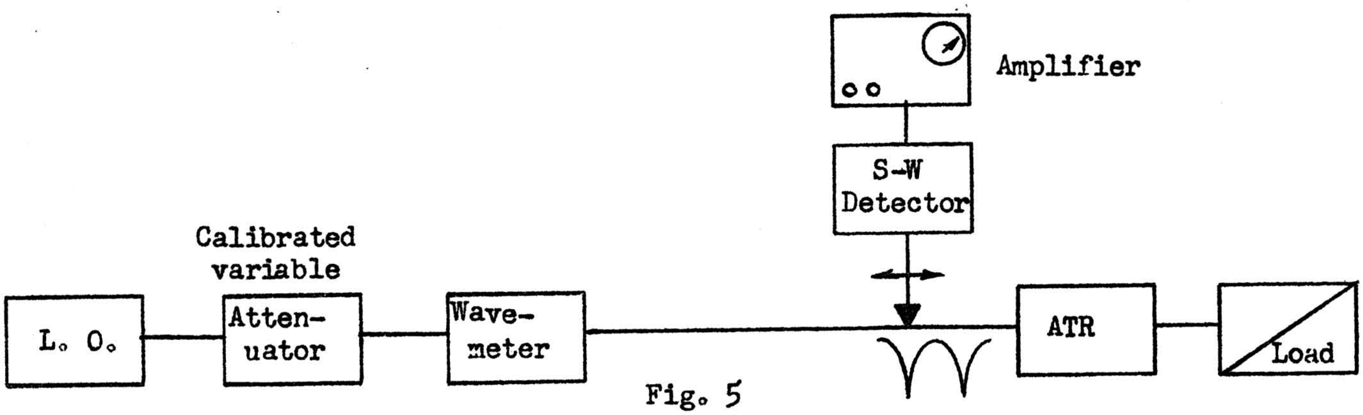


Fig. 5

Signal is provided by a local oscillator. The load may be a matched attenuator load or a matched crystal load.

The setup essentially measures the magnitude of the standing wave. Usually the standing-wave-ratio is too high to be measured accurately on the amplifier in series with the standing-wave detector. Instead, the db difference between minimum and maximum of the standing wave is converted into conductance.

It is possible, but not advisable, to measure the value of the SWR directly on an instrument such as a Galvanometer. The noise of vibration, making precision reading difficult, and the relatively long period of the instrument, prolonging the test, discourages this practice.

Other design checks given the tubes are "drop" and "glass strain" tests. These essentially are physical tests. The various electrical characteristics are measured before and after these tests to check the mechanical stability of the tube.

Type Approval Tests are those tests required by the Services on a certain number of tubes before acceptance for purchase. Among these tests usually are those for vibration, high-level VSWR, loaded-Q. Sometimes Life Tests are in this category.

The vibration test is a matter of putting the tube into a machine vibrating at a certain frequency, and subjecting the ATR to a certain number of G's to determine whether or not the tube will go out of tune, or have particles vibrating freely inside. In other words, it is another physical characteristic check, before and after which certain electrical characteristics are measured.

The high level VSWR measurement will indicate whether or not the ATR under firing conditions will set up so great a mismatch that the magnetron is pulled off frequency, or otherwise operates poorly. The high level VSWR may be performed in one of the following two methods:

1. Under actual ionization conditions at high power.

2. With simulated ionization condition at low power.

The high power setup is shown in Fig. 6a.

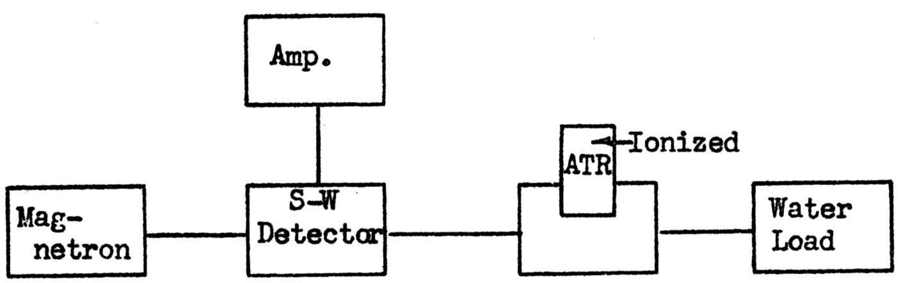


Fig. 6a
High Level VSWR Test

The load must present a very good match to the line, of the order of less than 1.05. The magnetron output is turned up to the required power as measured by the load. The ATR fires and the SWR looking into the tube is measured on the amplifier. Usually, however, the tube is operated at such high powers that there is too much noise on the amplifier to permit ease of reading. The simulated ionization condition at low power is more frequently employed.

The low power test bench is shown in Fig. 6b.

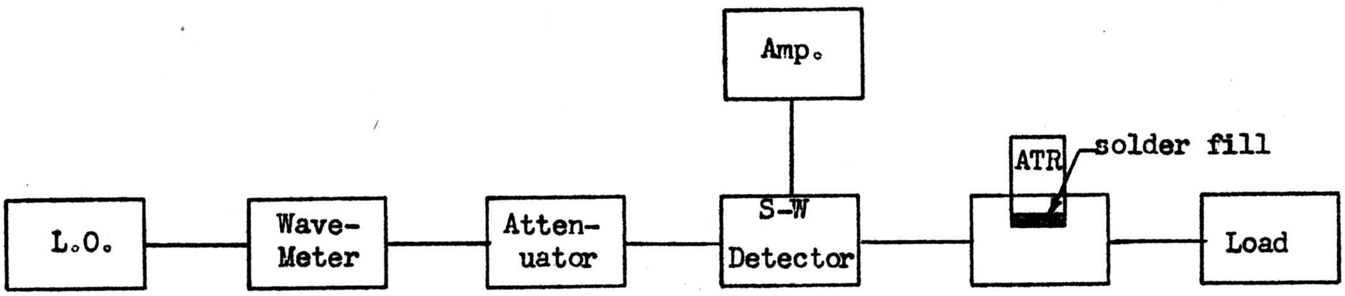


Fig. 6b
Simulated High Level VSWR Test

A special ATR tube is inserted in the tube mount. This special tube has a solder fill completely covering the internal side of the glass window so that the tube appears to be ionized. The load in this setup, too, must present a very good match to the line, of the order of less than 1.05. The standing-wave ratio looking into the tube is measured on the amplifier.

The loaded Q measurement, which usually is a type approval test, is indicative of the bandpass. This test is generally performed in a low level setup. The standing wave reflected by the ATR is measured at three frequencies; at center frequency, and two other frequencies, one on each side of center, usually at ± .5%. These readings give the change in tune, or the change in susceptance with respect to frequency, which, when substituted in formulae, give the loaded Q.

The formulas for calculating the various characteristics can be found on the test specifications.

Other tests, which are sometimes type approval tests, are life test (generally a 1000 hour period of operation) and life test end point. Tubes are checked at the end of the specified time to determine whether or not they have gone out of tune, or beyond the limits specified.

Temperature cycle life test generally is the temperature cycle test repeated 50 times. Electrical characteristics are measured again after this test.

These ATR tests are obviously all quite helpful to the equipment designer, and in this discussion we have attempted to make them self-explanatory.

The preceding descriptions of tests define the ATR from the standpoint of the testing engineer. The TR tests will be presented from the viewpoint of their significance to the equipment design engineer. We are interested principally in discussing broad band TR's since they require a greater number of tests. Generally, there are nine tests required for a broad band TR, whereas only five are necessary for a narrow band TR.

The tests are made at two power levels, low and high, and are here grouped under these classifications.

Major Low Level Measurements, TR Tubes:

1. VSWR. The voltage standing wave ratio plot indicates the bandpass of the TR. That is, the equipment designer will know how much of a band the unit can cover with negligible loss under low level or receiving condition of the radar system.

2. K-A Voltage Drop. This test and the following (interaction) are concerned with the keep-alive lead and the behavior of the gap in which the keep-alive has been inserted. The voltage measured is the drop that occurs in the gap when the keep-alive is fired. This potential has to be taken into account by the designer of the K-A circuit. The normal K-A circuit is designed with a 700 ohm input. We try to keep the voltage drop within the range of 250-300 V. Specifications allow a range in values of 150 V. That is, for some types the voltage drop is from 250 - 400 V.; for other tubes, 225-375 V.

3. Insertion Loss is that loss of r-f signal due to the tube construction alone.

4. Interaction. Interaction measures the loss of r-f signal resulting from the fired condition of the gap in which the keep-alive is placed.

Insertion loss is measured first. Then the keep-alive is fired and the difference between the second loss and the insertion loss is that due to ignition of the K-A gap, or interaction.

5. Firing Time measures the length of time required to fire the K-A gap. The normal limit is five seconds. Firing time is a function of the gas fill in the tube. It is a parameter to be considered in designing the circuit which delays turning the magnetron on until the TR switch has fired.

The above are the major low level measurements made on TR tubes. For the narrow band TR there is also included a type approval test, (loaded Q), which indicates the bandwidth of the tube.

High level measurements are of the order of 40 W. average, (40 Kw., peak), as compared to the few milliwatts in low level. The following are the major high level tests.

Major High Level Measurements, TR Tubes:

6. Leakage Power is one of the primary measurements. The radar system is designed around the ability of the TR to hold off transmitted power. That power which passes through the TR at high level (or during the transmitting cycle) is the leakage. The functioning of the crystal mixer is dependent on the hold-off power of the TR.

The leakage power of a TR is composed of two parts, called "spike" and "flat". The idealized pulse is shown in Fig. 7, where the duration of the spike is comparatively short with respect to the flat. This is pretty much the picture of the leakage pulse of a narrow band TR, so that the flat power is of major importance in narrow band TR tubes.

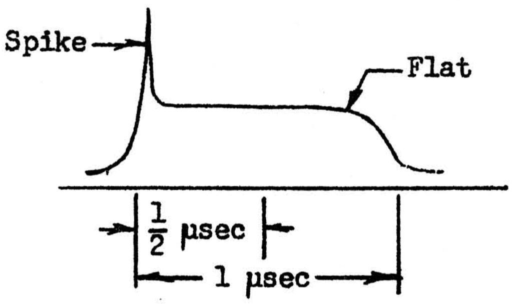


Fig. 7 Idealized Leakage Pulse

The shape of the pulse in a broadband tube (See Fig. 8) differs considerably from that of the narrow band TR. The spike energy is of longer duration for this tube and of more complicated structure. The radar system design engineer, however, is interested not so much in the shape of the pulse as in the amount of spike energy leaking through the TR tube.

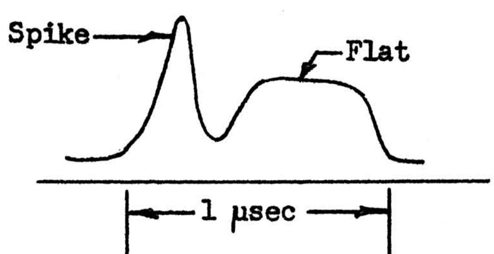


Fig. 8
Leakage Pulse, Broad Band TR Tube

A spike energy of .9 erg can be disastrous to the crystal, whereas a flat power in the neighborhood of 100 mw. is relatively easy for the crystal to take.

Because of the more complicated pattern of the pulse there is at present no completely accurate method of evaluating the spike. Usually, power measurements are made at 1 μsec. and at .5 μsec pulse widths. These two readings are combined in formulas to give values that are called "spike energy" and "flat power" more for convenient comparison than because they actually are the evaluation of those components.

For the spike energy, W_s, in ergs

$$W_s = W_1 - \frac{\Delta W}{\Delta t_p}$$

where W_1 is the power in ergs at 1 μsec,

ΔW is the difference in ergs between the two readings

Δt_p is the difference in the pulse widths

For flat power, P_f, usually in milliwatts,

$$P_f = \frac{t_p \, \Delta P}{\Delta t_p \, Du}$$

where t_p is the full pulse width, taken as 1 μsec.

P is the difference between the power at 1 μsec. and that at .5 μsec.

Du is the duty cycle of the measuring pulse, generally .001

Δt_p is the difference between the two pulse widths

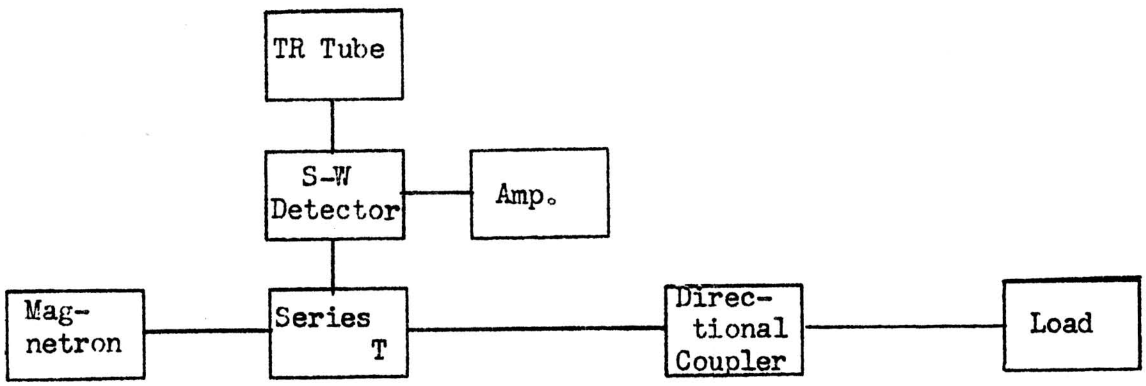
By referring to Fig. 7 it is evident that for the narrow band TR the formulas are sufficiently accurate for spike energy and flat power. It is seen from Fig. 8 that the formulas can yield but a crude approximation of the component values of the leakage pulse in a broad band TR. However, the formulas provide some means of measuring the leakage power and are useful until a better method is devised.

Specifications limit the values of spike energy and flat power as obtained by the above formulas. Usually, the maximum flat power allowable is of the order of 30 mw., while the maximum allowable for spike energy is .2 ergs.

7. <u>The Position of Short</u> test measures the distance back from the entrance flange of the TR at which the gas plasma effects a short to the line. The position of this effective short is of importance to the system design engineer, especially if he is using a rat race or balanced duplexer,

where spacing of the tubes is critical. If two TR's are put in a rat race, whose position of short vary considerably, the mismatch will be magnified, and cause trouble in performance of the system.

The position of short is measured by a relative method. The test engineer tries to simulate the actual condition of the TR operating in a duplexer system.



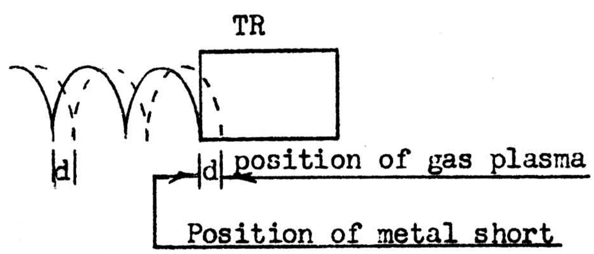


Fig. 9
Measurement of Position of Short

To simulate the conditions of the duplexer system in which the TR is to operate, a series T is used to transmit power from a high level source to the TR tube. (See Fig. 9) The magnetron supplies high level power, firing the TR tube, which reflects a very large standing wave down the line of the side branch of the series T. The roving probe of the standing-wave detector picks up the reflected signal and transmits it to the amplifier, where the dip in the meter needle indicates a minimum of the standing wave. The dip is generally so sharp that it is not possible to read accurately on the meter. Two readings are usually taken, one on either side of the minimum, and the mean of the two readings is taken as the position of minimum. This is compared with the position of minimum in the line found previously with a standard metal shorting plate substituted for the TR tube. The distance "d" between the two minima is the actual distance of the gas plasma from the entrance flange of the tube.

The position of short measurement is somewhat inaccurate, primarily because it is made at relatively high power. When high power is employed the probe depth in the line must be kept comparatively shallow to avoid burning out the bolometer.

8. Arc Loss readings are taken at a relatively high power - 4 W., average, 4 Kw., peak. The arc loss is that amount of transmitted signal used up to create the arc of a fired TR.

9. Recovery Time is the time necessary for a tube to change from an ionized state after firing to a de-ionized state--i.e., the return of the tube to the condition in which it can be fired. The mechanisms of recovery time have been set forth in Lectures 8 and 9. It is sufficient to say here that the recovery time measurement is a critical one.

Recovery time is important to the systems design engineer since it tells him the minimum range of the radar system. A 4 μsec recovery limits the short range of the system to 600 yards. Anything closer can not be seen by the system.

Special Test Equipment:

The preceding particulars cover rather briefly the various tests run on a broad-band TR tube. Some of the equipment used may be unfamiliar. A brief descriptive summary of the more unusual units is in order.

Triple Pipper. One of the unique pieces of equipment utilized is the panoramic impedance bridge, less formally known as the "triple pipper." (see Fig. 10) This testing unit has been referred to in the previous lectures, especially that on tuning TR tubes. It is an invaluable aid in obtaining

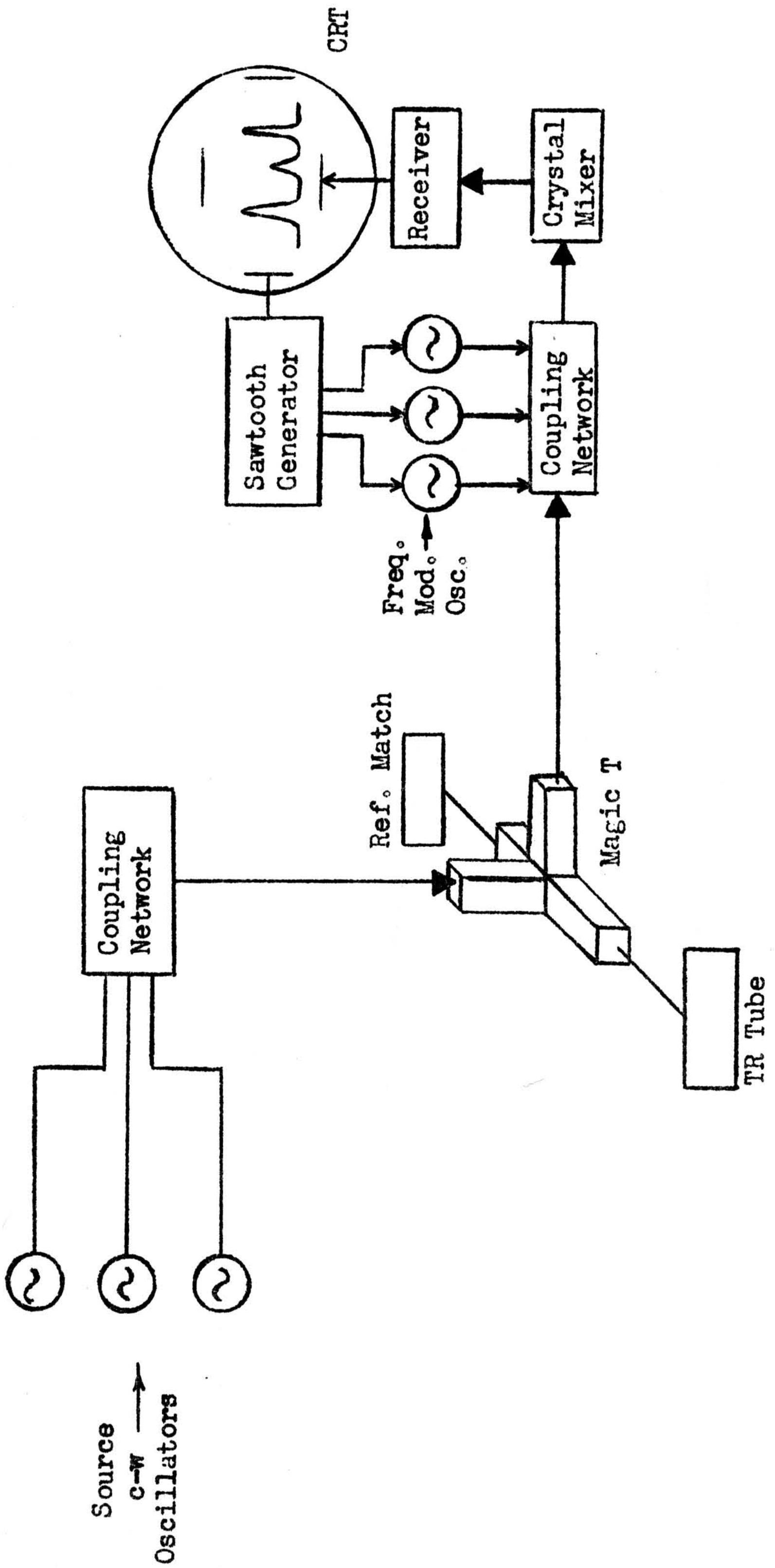


Fig. 10

Panoramic Impedance Bridge

(Triple Pipper)

bandpass characteristics of four-element TR's. As defined before, this impedance bridge essentially is a device which reads the SWR of the TR tube at three frequencies simultaneously. The triple pipper utilizes principles basic to the industry. Of these the well-known heterodyne and the crystal mixer theory have been thoroughly covered in electronic literature. With the employment of the Magic T (which has been discussed in an earlier lecture), it is possible to view on a CRT screen the tube SWR characteristic at the three frequencies being checked.

The system is composed of three component groups: r-f sources, the bridge element and a detector-indicator. Referring to the network connections of the Magic T in Fig. 10, it will be noted (recalling the theory) that the c-w signal entering the arm of the T through the coupling network, will split equally between the two side-arms (which are shown going respectively to "Ref. Match" and "TR Tube.") No power will go through the arm which connects via the coupling network to the crystal mixer. On the other hand, one-half of any energy reflected back from the TR tube will go through the outlet to the detector-indicator section. The "Ref. Match" balances the bridge looking back from the TR tube.

The triple pipper system employs six oscillators, three frequency- modulated, and three, c-w. Normally, the receiver signals are f-m, the source c-w. Each source oscillator is at a different frequency, and for each source oscillator there is a corresponding receiver oscillator. The three source signals are transmitted to the TR tube and reflected back modified by the attenuation characteristic of the tube at each frequency. They travel out of the Magic T to the detector-indicator networks. At the same time, the f-m signals are introduced into the detector-indicator networks. The tuning of each of the set of paired-off f-m and c-w oscillators is such that at some time in the modulation scale the frequency difference between each pair is equal to the i-f at which the receiver will pick up signals. These different frequencies vary with time. Thus i-f pulses are generated and video output pulses result. The pulses appear on the screen of the CRT as three pips. The amplitude of each pip is a function of the test tube's SWR for the r-f source frequency of that pip.

There are other types of multi-frequency bridges, such as the pulse modulated impedance bridge, and the mechanically modulated bridges which rely on fairly accurate switching of tubes. None of these bridges are absolutely accurate since there are errors inherent in the Magic T construction, in the directivity of the directional couplers, and also in termination. They all are, however, extremely convenient devices, since they give the SWR at more than one frequency at a time.

An extension of this multifrequency impedance bridge, mentioned in the lecture on tuning, is the sweep oscillator which covers the entire band simultaneously, instead of just at three frequencies. The Magic T, again, is the bridge element. Low frequency-modulation is employed, which depends primarily on the reflector voltage of the single oscillator in the unit. A variable motor operating at a multiple of 60 cycles sweeps the reflector voltage and the external cavity plunger, producing on the oscilloscope screen a pattern of the entire band. This sweep oscillator has been developed only for S band to date.

Power measurement bridges in general use include either a thermistor bead or a barretter mount.

The Thermistor Bridge is akin to the Wheatstone Bridge. In the fourth arm of the Thermistor Bridge is the thermistor bead which is composed of two strands of platinum wire beaded with a semi-conducting dielectric (manganese oxide, fine green copper and a few other alloys) and encased in glass. Average power is measured as a function of the resistive change in the bead with the application of r-f. R-F power entering the bridge raises the bead's temperature, which varies its resistance.

The Barretter Mount has a thin strand of platinum wire instead of the beading. This type of bridge works on the same principle, the difference between the barretter mount and the thermistor bead being that the former has a positive temperature coefficient of resistance, and the latter, a negative. The barretter type is more sensitive and accurate at low power levels. The thermistor bead type is more useful at high levels of power.

Typical Test Bench Setups:

The Low Level Bench Kit (See Fig. 11) normally includes an oscillator and attenuator. The attenuator is generally helpful in SWR measurement, and is used for padding. A wavemeter determines the frequency of the oscillator. In addition to the oscillator tube mount, there is a standing-wave detector and a load which can be of various types. The oscillator is driven with a power supply of Type TVN-7 which has a built-in modulation section. The signal from the standing-wave detector goes into an audio amplifier of Type TAA-16. Both power supply and amplifier were originally designed by Radiation Lab.

The High Level Bench Kit (See Fig. 12.) includes a magnetron, a series T, a TR mount, a directional coupler, and a load. The directional coupler transmits signal to a power monitor. The load is a water or sand load, or some type which is able to handle the increased power.

Accuracy of Tests:

The accuracy of the answers in all of the foregoing tests is dependent on

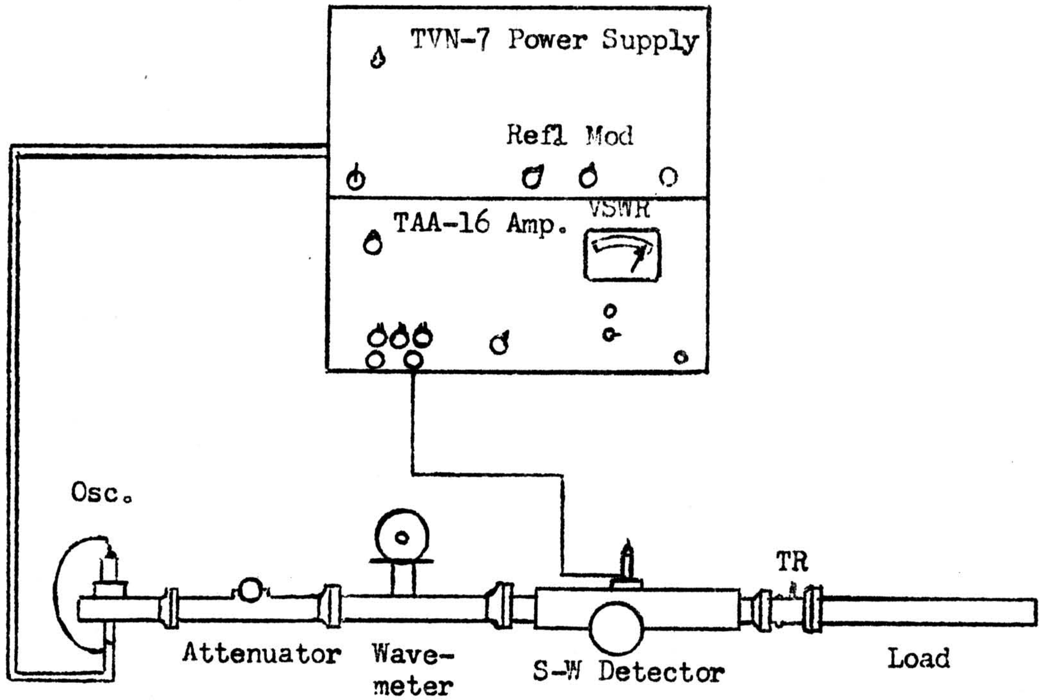


Fig. 11

Low Level Bench Kit

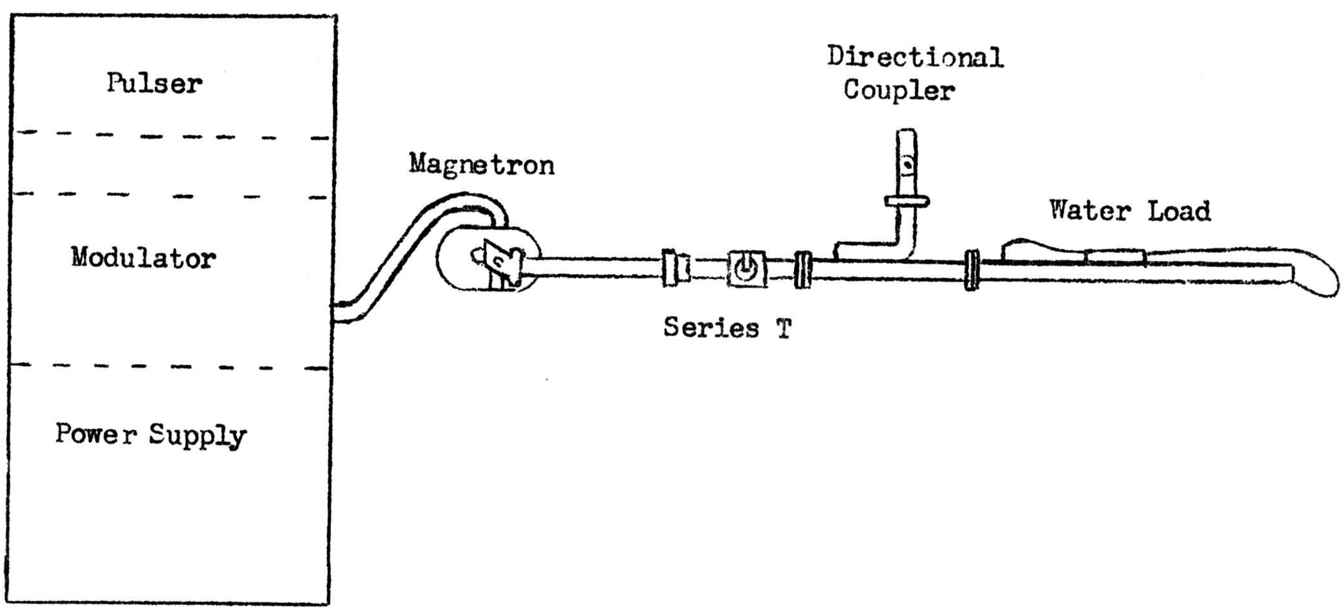


Fig. 12

High Level Bench Kit

the accuracy of the operator combined with the accuracy of the equipment involved. There are various errors in the different types of equipment, not obvious to the naked eye. The additive effect of all the errors can provide some wild answers.

Most of the measurements tend to depend on one of two basic phenomena: power or SWR. The fundamental pieces of equipment include a standing-wave detector, f-m paraphernalia, and power measuring devices such as a thermistor. Since the thermistor bead changes resistance with temperature, it is obviously affected by thermal changes in ambient temperature of the surroundings. There should also be taken into account the bridge sensitivity, that is, the inaccuracies in the other elements used to pick up the power. As a result of these errors, power measurements are not apt to be better than ± 2 μwatts. That can mean an error of as much as .05 erg in spike energy for certain applications.

In the standing-wave ratio measurements we have, however, a different proposition. The SWR is at a low level, and we are not concerned with changes in temperatures affecting the various measurements. We are interested in probe depths. If the probe is in too far, the probe itself sets up standing waves. And if the probe is too far out of the waveguide, the sleeve itself acts more or less as a resonant circuit, changing the SWR. Combined with this difficulty is the error introduced by deviation from the square law as introduced by the system---crystal and amplifier together. Square law losses are a function of crystal construction and the departure from linearity of the amplifier. The correction factor of readings on test equipment has been in the neighborhood of 1.7-1.8, and sometimes up to 2.3, 2.4, as compared to the theoretical factor of 2.0. Normally, we try to stay within the range 1.85-2.1, and most of our equipment at the present time does that.

Test Department Coverage in Power and Bands:

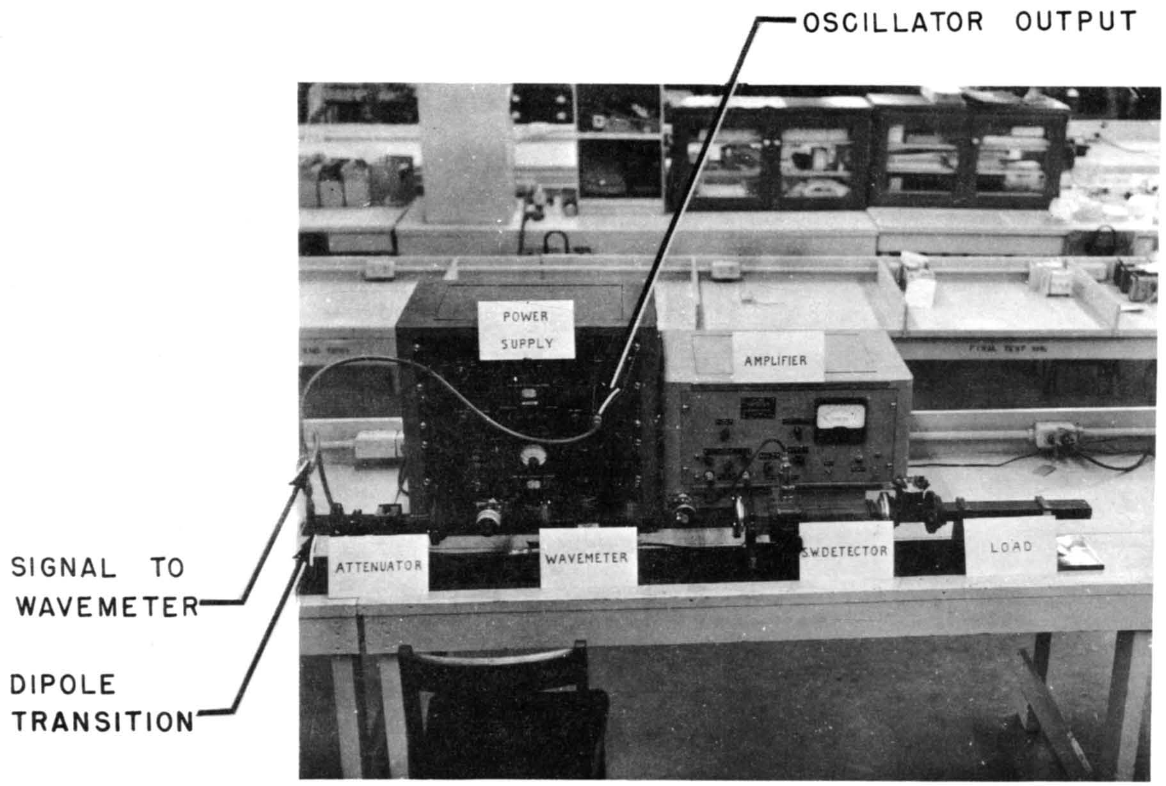
At the present time the Test Department can handle tests for wavelength ranges from 10 cm down to .8 cm, and power ranges from milliwatts up to megawatts, peak.

Equipment is available for various bands. The 10 cm band can be handled with 2K28 oscillators at low power, and the 4J- magnetron series at high power. (4J31, 4J32, 4J48, etc.)

Tests in the C-band can not yet be made at megawatt power. With the QK238 magnetron, recently added to the equipment, it is possible to obtain 300 Kw., peak.

In the X-band, where most of the work is done in testing, there is considerable range of frequency and power. For low levels, there is the velocity modulated oscillator, 723, 2K25 Pierce-Shepherd Type. In the medium power range there are the 2J42 and 2J51 type magnetrons, operating in the neighborhood of 50 W., average, 50 Kw., peak. There are also the 4J50 and 4J78 series, operating at 250 W., average, 250 Kw., peak.

For the K-band, the 1.2 cm region, the department is able to cover only up to medium power. In this vicinity, the region 1.5 - 1.8 cm is called the K_{Λ} -band. The small region around .8 cm wavelengths is known as the K_1 -band. The K_1 -band is relatively new, used by the Government primarily for weather experimentations.


OSCILLATOR OUTPUT
SIGNAL TO WAVEMETER
DIPOLE TRANSITION
POWER SUPPLY
AMPLIFIER
ATTENUATOR
WAVEMETER
S.W.DETECTOR
LOAD

OPERATION AND DESIGN

OF TR AND ATR TUBES

Lecture No. 12

Testing Techniques & Equipment, Pt. II

During the last session we considered the various tests performed on TR and ATR tubes. This session will be devoted to answering questions on test procedures or equipment.

To begin the discussion we first present one of the star performers of the Test Department—the C-band test kit. This test kit, operating at low level power in the region of 5300 Mc., measures such tubes as the TR-361, TR-331 and ATR-332.

Components of C-Band Test Kit: (See Photograph)

The oscillator is a Sperry klystron, type 2K43, or 2K39. A square wave modulated r-f signal is fed into the waveguide by means of a waveguide transition, that is, a small antenna. This antenna is a dipole which radiates into the waveguide with very low transfer of impedance.

The attenuator section following is a resistance card type, variable from 0 to 40 db, or thereabouts. It is a Sperry type flap attenuator. The resistance card travels in an arc across the guide. It presents a reasonable match to the line, of the order of 1.5, and is fairly linear up to about 30 db.

The next section is the wavemeter. It is a resonant cavity type which is tuned by changing its dimensions. It obtains energy from the main line by means of aperture coupling to the main line.

Behind the wavemeter is a crystal pickup for feeding wavemeter signal to the amplifier, for measuring frequency.

Following the wavemeter in the main line is the standing-wave detector section. This section consists of a probe traveling in a slotted section to extract small portions of energy of the standing wave. The energy is then fed to the crystal (crystal holder is visible) and thence to the amplifier for which a scale has been provided to indicate directly the standing wave ratio.

An adaptor section follows the standing-wave detector section, since the latter has a flange joint and TR's must generally be measured between two choke joints.

The TR tube is inserted between the adaptor section and a matched load. The load on this kit is durez. Durez and resistance cards are most satisfactory for low level measurements. Durez is favored here because it seems to be more adaptable to machine work. The disadvantage of resistance cards is that the resistance coating cracks and peels along the edge. For high level measurements, water and sand loads are used. For such power measurements as insertion loss and interaction, a crystal load is employed.

On top of the kit is set the keep-alive supply (not shown in photograph) to provide voltage for the TR keep-alive in the measurement of interaction at low level.

<u>Frequency Setting & VSWR Determination of Line</u>:

Q: What is the procedure for determining the VSWR of the line alone, so that the accuracy of measurements on the tube is known?

We can answer that question by going through the process of setting up the line, prior to the insertion of the tube.

In most specifications for measuring a TR, it is required that the line itself have a VSWR less than or equal to 1.05. The method for checking this characteristic of the kit is the same for practically all the types of guide in use--S-band, X-band, K-band, etc.

The filaments of the various tubes (power supply tubes, oscillator, amplifier) and high voltages are switched on. After a warming up period of about one minute, a square wave is applied to the reflector of the oscillator. This provides a square wave modulated signal in the line.

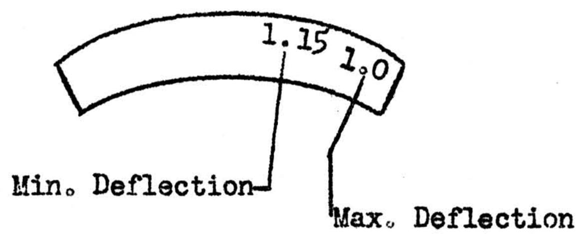
Now that we have the signal we need to find out what its frequency is, and what the SWR is in the line. We begin by determining the peak of the oscillator mode in which we wish to operate. From the three or four possible modes of operation, we normally choose the mode at the high end to obtain the maximum power. The amplifier is adjusted until the signal picked up by the crystal detector deflects the scale on the amplifier panel. The signal does not have to be large. It can be between 2 and 3 on the amplifier scale. Having obtained this indication of signal on the scale, we adjust the reflector voltage until the oscillator energy is peaked (as noted on the amplifier scale) and then adjust the square wave amplitude until a maximum deflection is reached on the scale. This operation has peaked the signal into the line.

The next step is to adjust the probe pick-up until it is matched to the signal in the main line. The probe section contains a variable piece of coaxial waveguide which changes impedance with changing position in the probe channel. The probe is moved along the channel until it reaches a point at which the needle swings radically high. The probe position that gives maximum swing is the position of maximum impedance match. In other words, the condition at which the maximum amount of power is obtained from the detector.

Once the detector has been set for maximum power, the amplifier is switched over to the channel in which the wavemeter is located. The wavemeter control is varied until the needle of the amplifier rises in scale position. The wavemeter on the C-band kit is a transmission type with a crystal behind it. The power absorbed into the cavity energizes the crystal, giving an increase in power. At the peak of the needle rise, the wavemeter scale is read and converted into frequency by means of a calibrated chart.

If this frequency is not the desired one, the oscillator must be retuned and the entire frequency measurement repeated, after switching the amplifier back to the detector. The oscillator is reset in the direction it seems necessary; the reflector voltage and modulating square wave are re-adjusted. Again the line is matched all the way through, and the frequency remeasured on the wavemeter. The entire process is repeated until the frequency specified is obtained.

When the oscillator is on the proper frequency and the detector is peaked for maximum power pickup, then the load is coupled directly to the waveguide. Generally, the load is coupled onto the line with the adaptor section omitted. The choke joint of the load is bolted onto the flange joint of the detector section. The standing-wave detector probe is moved along the channel until the needle on the amplifier meter reaches a point closest to the 1.0 on the scale. Coarse and fine controls of the amplifier are varied until the needle reads 10 on the low scale, or 1.0 on the high scale. This is the point that the Test Dept. calls the maximum. Although it is generally considered the minimum of the standing wave ratio, the Test Dept. refers to it as maximum, meaning the maximum swing of the needle or maximum pick-up. The probe is then moved along the channel in one direction or another, with the rest of the system held constant. The shifting position of the probe will be coincident with a change in reading of the needle, the latter traveling from 1.0 to the higher numbers on the scale,



back to 1.0, etc. The highest number the needle reaches, which, say, is 1.15, is the standing wave ratio of the line and load.

Locating Source of Error in Setup:

Q: Suppose that the SWR of the line is greater than the limit specified, how do you locate the source of trouble and bring the SWR down to the required ratio?

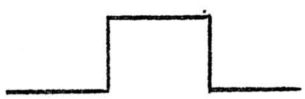
In measuring the SWR of the line, we are reading the power reflected from the load. The loads we use in our equipment usually have SWR's of the order of 1.02, and in some cases are as low as 1.005. Sometimes we do get readings of loads which run as high as 1.15, 1.2, 1.25, and even higher. There are various possible causes for such high readings.

The first step, in locating the trouble, is to determine whether the load itself is setting up this high standing wave. The simplest means of checking this is to fix the probe detector position at some arbitrary point in the channel, and then move the load back and forth in the waveguide. If the SWR changes radically with the shifting of the load, then the load itself is at fault. It is possible sometimes to shift the position of a slightly mismatched load until it presents a good match to the line at a given frequency. This practice is not an approved one. We prefer to make broadband loads, reasonably flat across the band. By broad band is meant at least a 10% bandpass, so that at any frequency in this 10% bandpass, a shift in location of the load results in a SWR deviation of not more than .01.

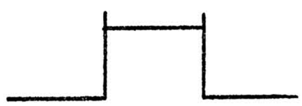
However, let us assume that the load is not responsible for the line SWR reading of 1.15. That is, changing the location of the load does not affect the SWR reading one way or another. The next point of investigation, then, is the oscillator.

There are a number of ways in which the oscillator can be responsible for a high SWR of the line. One of the fundamental difficulties is that the oscillator may not be operating at peak. The response of the oscillator may be viewed on an oscilloscope. A normal good pulse is a clean square wave. If fringing of the square wave is evident, then the oscillator is not directly on peak operation. The fringe condition is cleared by adjusting either the amplitude of the modulating square wave, or the level of the reflector voltage. When two distinct bars are apparent in the oscillator response, it is a partial condition of modiness.

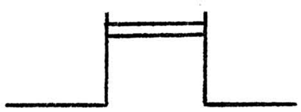
Oscillator Response



good pulse



operation not at peak

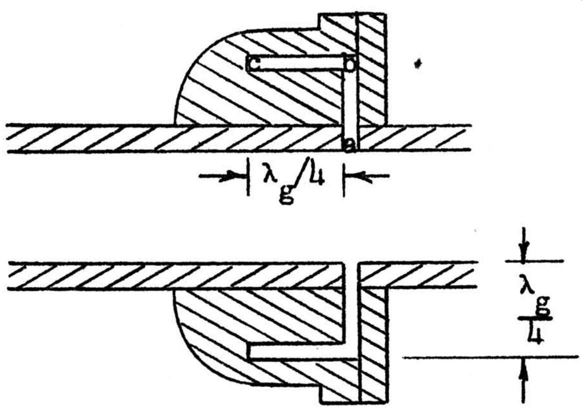


indication of modiness

The oscillator response and that of other components preceding the standing-wave detector in the line, will have no direct effect on the SWR of the line, which is a measure of energy reflected from the load, since no energy from components preceding the probe will reflect beyond the probe. However, a poor oscillator—operating off peak, or off mode, or in two modes—will result in a peculiar standing wave in the line, and hence a troublesome SWR. This is a natural consequence of a distorted signal, picked up by the probe, being considerably amplified by the crystal.

If the oscillator is checked and eliminated as the cause of trouble, there remains possible the error due to faulty contact. In certain cases, error due to faulty contact has been investigated rather rigorously, particularly at X-band. It has been found that the best possible joint is a butt joint—

flange-to-flange connection, with both flanges absolutely flat. That means lapped surfaces. However, it is not always advisable to devote machine capacity to lapping surfaces. The next best coupling is a choke-flange connection. The choke consists of a kind of series branched waveguide,



Choke-Flange Coupling

one-half wavelength long, broken into two equal lengths. (See figure inset.) The short at c reflects minimum impedance at the gap between the two sections of main line being coupled. Electrical contact is broken at b, a low current, high voltage point. The minimum impedance, or high current point is at a. Thus, although the mechanical gap at point a may be something of the order of .030", electrical contact is established between the two sections of the main line.

After consideration of faulty contact, we have more or less taken care of most sources of error that give high readings of the SWR. This leaves to be considered the operator of the equipment. In the last analysis the accuracy of measurements obtainable on a piece of equipment is dependent on the ability of the operator to use the equipment. The finest equipment in the laboratory operated by some person who cannot use it will yield the incorrect answer 50% of the time.

<u>Probe Depth and SWR Error:</u>

Q: Won't the insertion of the probe be a contributing cause of error?

Yes. The insertion of the probe beyond a certain depth in the channel can adversely affect the measured SWR. That, however, is more or less a function of the operator. He can tell whether or not he reads the wrong SWR by adjusting the probe level. If decreasing the depth of the probe in the channel decreases the VSWR reading, the larger reading is likely to be a result of the probe setting up its own SWR. In this connection, it is fairly obvious that, if the surfaces involved in the travel of the probe are not flat, so that the carriage of the probe across the length of the slotted section is not even, there will be a variation in probe depth which will give excessive deviations in SWR. Once a condition is reached where two consecutive readings at the same point agree with each other, it is fairly safe to assume that the probe is not setting up enough SWR to interfere with the measurements.

Variations from Square Law Operation:

Q: How much error can be contributed by a crystal whose response deviates from that expected by square law?

The accuracy of the measurement being dependent on the response of the crystal and amplifier together, we calibrate the output of the combination to get a correction factor. There is a formula for this correction factor which can be referred to for actual usage. To get an idea of the amount of error contributed by the crystal, let us assume that the VSWR measured is 1.1. Say that the response of the detector unit is 1.8, as compared to the expected 2.0 of straight square law. Then the measured VSWR will differ from the actual value by something like ±.03. If the detector unit response is 2.2 as compared to the expected 2.0, the deviation from the actual quantity will be of the order of ±.05. Although the magnitude of deviation of the measured from the actual SWR is greater for departures from square law at the high end of 2.0, the percentage difference is about the same, since a plot on log paper of such differences is linear.

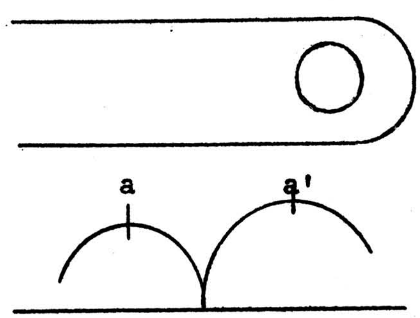
Q: How can you tell whether the error in the readings is due to a faulty crystal or to some other difficulty in the setup?

We are not considering now a change from 1.05 to 1.07 which might easily be a result of normal square law response. Assuming, instead, that the change in the VSWR is from the normal operating characteristic of 1.05 to 1.2 or 1.3, there is no quick diagnosis for the trouble. Where there is a marked change in VSWR, the only sure method of locating the source of maximum error is that outlined previously— the step by step checking and correcting, where necessary, of components, starting with the load. If the load, the oscillator, etc., are operating satisfactorily, and the detector section appears to be in order, then, the next step is to check the square law characteristic of the crystal, if nothing else appears to be responsible for the confused answer.

<u>End Effect, Slotted Section:</u>

Q: In measuring the VSWR over a bandpass, as the wavelength gets longer, how can you tell when the reading is being modified by the end effect of the slotted section?

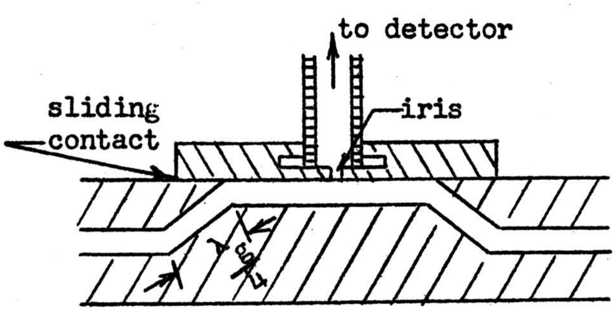
The end effect, as it is usually understood, is evidenced by that change on the amplifier scale (increasing or decreasing) which bears no relation to the actual SWR, and which occurs when the probe approaches a terminus of the slot. If the slot is a full wavelength in electrical length, it is possible



to determine whether end effect enters into the reading. Two maxima (a and a')— or two minima—can be compared. A noticeable difference in values between corresponding points of the standing wave values indicates the the reflection from the slot end is contributing to the SWR. If the electrical length of the slot is too short to permit comparison between two maximum or two minimum points, there is no way of determining whether end effect is causing inaccuracy of measurements.

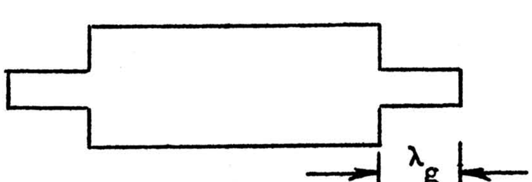
It is not a simple matter to get away from end effect entirely. A certain amount of it has been eliminated by insulating the probe holder from the waveguide slot. It is advisable in making measurements to keep as far from the end of the slot as possible. It is observed, though, that some measurements, such as the position of minimum of ATR tubes, are best made at locations nearest the tube. Common sense dictates the distance from the end of the slot.

There have been standing wave detector units made which eliminate the end effect fringe. Instead of a probe an iris couples energy from the slotted channel to the detector. One such type is designed expressly for high power.



In it the iris is a part of a branch guide set in a "cover plate" or top portion of the waveguide, which slides back and forth across the main body of the waveguide. The slot has two pairs of bends to bring it up to the level of the "cover plate". The length of each portion of the slot that is at an acute angle to the cover plate by virtue of these bends is made $\lambda_g/4$ to minimize reflection at the bends.

Another type of slot (which may be used for low power measurements, also) is stepped at each end. The narrowed portions of the slot formed by the steps is a quarter wavelength long. Thus the contacting component of the detector is always a minimum of a quarter wavelength from each terminal.



Q: Does the end effect alter the position only, or the value, too, of maxima and minima?

The end effect changes the value as well as the position of minimum and maximum. When the probe arrives at the end of the slot, the SWR value read will change. On one end it will be high, on the other end, low. The slots are made sufficiently long (at least two full guided wavelengths) to minimize this difficulty.

Q: Why couldn't the slots be made longer and a mechanical stop inserted to prevent the probe from entering the region where fringing would occur?

That kind of arrangement is used on the S-band detector. It is not too practical for general usage because the mechanical problem of avoiding torque enters the picture for lines longer than two wavelengths. It does not take too much pressure to "squeeze" the elongated slot and make it into a line stretcher.

It can, however, be done and has been done at a cost of excessive weight. For a special purpose a section four wavelengths long has been made in the industry. It weighs about 200 lbs.

OPERATION AND DESIGN OF TR AND ATR TUBES

Lecture No. 13

General Aspects of TR Design

The various discussions of tube performance giving insight into the properties of the tube as a bandpass filter, the gas fill considerations and the type of tests made, have by now given a fairly definitive analysis of the TR and ATR tube. In this presentation we shall look at the tube from a slightly different angle to show the general trend of, or the procedure followed in, the preliminary design and cost considerations required for bidding on a contract.

Initiation of Tube Request

Let us suppose that some company, or some branch of the Services, or one of the laboratories associated with a Service Group, wishes to procure a TR tube. We'll pick on the Naval Research Laboratory. Say that they have become interested in a radar system to operate at 15,000 Mc. For their radar set they want to use the modern type of duplexing. That more or less automatically means that they want a broad band TR box. We'll follow this tube through the stages from the bid proposal to the cost estimate.

The Naval Research Laboratory writes up the requirements for constructing this particular radar system. There will be a number of tube components, system components and equipment components necessary. They submit their requests to the Bureau of Shipments, from where they are issued by the Contracting Office in the form of bid proposals. Bid proposals for tube components are sent to tube manufacturers, for indicating systems, to manufacturers of them, etc. The proposal for the TR tube, drawn up in legal form, and liberally sprinkled with the legal terminology of whereas's, whereby's, states that what is wanted is a 15,000 Mc. broad band TR tube (12% bandpass) with insertion loss held to a practical minimum, the VSWR to be as small as possible, the tube to withstand 100 Kw. power, to operate at a duty cycle of .001, 1 μsec pulse, and to be used with a 1N26 crystal. The proposal asks the manufacturer interested in bidding for the contract how he is going to make the tube, how long it is going to take for delivery and how much money he is going to charge.

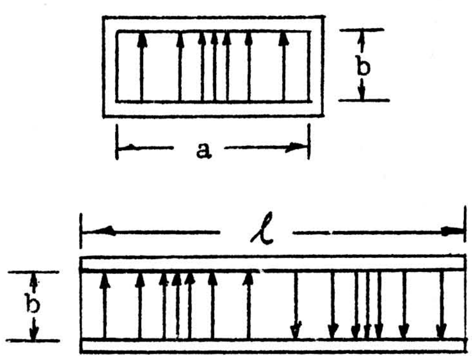
The contract to be awarded may be in one of several forms: cost-plus-a-fixed-fee, fixed-price, fixed-price-negotiable. Most of the contracts that we get from the Navy are cost-plus-a-fixed-fee -- that is, to the cost of producing and delivering the items a fee of 7-1/2% is allowed for profit.

The bid comes in through the Sales Dept. Several copies are made for internal circulation. The copy going to the Engineering Department contains pertinent specifications. It is then up to the Engineering Department to tell how it is proposed to construct the tube, and estimate how long it would take to complete development. A cost analysis is made. The information is collected and returned in the prescribed form as an answer to the bid proposal. The Navy, pleased with our planned design, and impressed with our speed and economy, awards us the handsome contract, and we are in business. Simple, enough? Let's first look into the work involved in the proposed design.

Structural Considerations of Envelope

In proposing a design, or setting up a development program, the first thing that we have to decide is what the tube is going to look like. Since it is a broad band tube we would expect it to be made of waveguide, with normal mounting flanges to mate properly to those in use in the system. If no flanges are specified, we can be fairly arbitrary as to the provision for coupling the tube into the system.

It will be recalled that the TR tubes are effectively bandpass filters set in waveguide. We have to choose the waveguide. Various groups of people have gone to some trouble to tabulate possible waveguide sizes in terms of frequency coverage. Frequency coverage is determined primarily by the cutoff frequency and the frequency at which other modes than the fundamental may interfere. Each condition can be represented in terms of simple formulae. The type of propagation with which we will be concerned is cutoff. We generally choose rectangular waveguide, and operate with the direction of the E field running up and down, with one variation in directivity down the length of the guide and none in the b-dimension. This is known variously as the TE_{10}, TE_1, or TE_{01} mode.



From the published information on rectangular waveguide sizes vs frequency coverage, we choose one with dimensions .622" x .311" x .040". The .040" wall thickness is not critical but we'll keep it anyway. A rectangular guide of the internal dimensions given is capable of propagating frequencies ranging from 12,400 Mc. to 18,000 Mc.

We're quite fortunate, so far, since we are practically at mid-band. Sometimes in selecting a waveguide size we find that our frequency of operation is up near one end or the other of the range of propagation. Then we have the headache, complicated by other considerations, of deciding which of two sizes would be less bad. So far, however, we're doing all right, and we'll go on to the next step.

Here we are very much concerned with the type of material in which we can obtain the waveguide. We find ourselves pretty high up in frequency, so that it might be pleasant to make the tube body of 90% silver --pleasant from the standpoint of conductivity and appearance. But we'd better find out how big the tube is going to be. There might be so much silver involved that the cost would be disagreeable. Let's say that silver is an attractive thought, but we'd better look around for other metals.

One of the most likely metals, meeting requirements of rigidity and absence of contaminants, is 90/10 bronze. If we find it made in the size we want, it will be an excellent choice. If it is not available and we have to go to something like commercial brass, of which there are many varieties, we have a contamination worry. In the process of brazing commercial brass we run the danger of contaminating our furnaces with zinc and other unwelcome metals which will cause much sorrow and havoc when we process parts of magnetrons and other finicky tubes through these furnaces.

Not possessing a multitude of furnaces, we are somewhat skeptical about working with brass. If we have to use brass, there is a trick which will help out. The tube with the necessary jigs and fixtures is placed inside a box small enough to go through the furnace, and processing is completed in this fashion. It has been used rather effectively by others in the field, but we have tended to shy away from this expedient since there is danger of leaks in the box.

We could make the body of steel if manufacturers happened to be drawing steel this size. It is an off size, though, so that there would be little likelihood of obtaining it. It just so happens (to make things easier for us) that we can get 90/10 bronze in this particular size. We are ready, then, to go ahead with the design.

Next we decide how long the tube body is to be. This involves knowing the number of elements necessary for the bandpass we need. Some knowledge of general filter considerations in waveguides is required. The Rad. Lab. series is very useful in finding this information. "Microwave Duplexers," Vol. 14 of the series, edited by Smullin and Montgomery, has many formulae for the power loss in the tube. In general, the loss is expressed as a logarithmic function

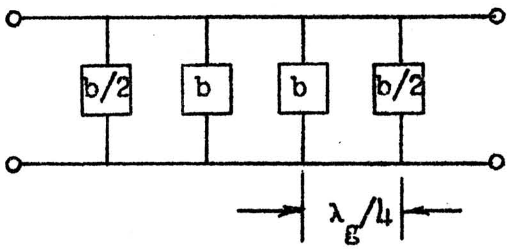
$$\text{Loss} = \log_{10} F(b, \ell, \lambda_g)$$

where b is the susceptance of the tube elements

and ℓ is the distance between the elements.

and λ_g the guided wavelength.

This formula predicts the bandpass plot, which will have a number of bumps in it, depending on the number of elements used. The methods for solving bandpass filters in terms of the susceptance of the elements employed and the number of elements, has already been indicated in previous lectures. After we go through the calculations, we find that what we want is a tube with two gap structures. The susceptance of each gap is twice that of each window. The elements are spaced a quarter guided wavelength apart. We could if we wished put in a third susceptance. That is, have a triple-gap tube. But the four element tube will probably give us the bandspread we need at 15,000 Mc.



Continuing from this point, we determine the length of the tube. We find a formula which will tell us how long the guided wavelength is. This formula is in terms of the cutoff wavelength, λ_c, and the wavelength in free space, λ_o

$$\lambda_g = \frac{\lambda_o}{\sqrt{1 - (\lambda_o/\lambda_c)^2}}$$

The wavelength in free space is readily obtainable from the familiar relationship among the terms frequency (f), wavelength (λ_o) and velocity of light (c). ($\lambda_o = c/f$) Since $c = 2.998 \times 10^{10}$ cm/sec, and $f = 15 \times 10^9$ c.p.s., λ_o is approximately 2 cm.

For the value of the cutoff wavelength, we have its relation to the dimensions of the guide

$$\lambda_c = \frac{2\ ab}{\sqrt{(ma)^2 + (nb)^2}}$$

where m is the number of variations of the field along the up and down direction (a to a) t the guide, and n is the number of variations along the wide dimension (ℓ). Utilizing the knowledge that we have the TE_{01} mode, we can substitute 0 for m, and 1 for n. This reduces the expression for λ_c to 2a, or, (2 x .511 x 2.54) cm. Now, substituting the values of λ_o and λ_c in the formula for λ_g, we find that one guide wavelength at 15,000 Mc. is approximately 2.6 cm. Our tube is to be made of three quarter guided wavelengths, so that the length will be

$$\ell = (3/4)(2.6) \doteq 2 \text{ cm.}$$

The tube is slightly less than 2 cm. long, and the spacing between elements is slightly over .65 cm.

In designing the tube for actual construction we would make our calculations less approximate. Even so more exact evaluations of the formulae will not necessarily give the proper tube length for best band-spread, for we have to take into account fringing. Fringing is the general heading we use to cover a few effects on bandpass that we don't know too much about. Into this classification we toss such things as the interaction existing between the various internal resonant elements, and the discontinuities in the guide sides. The fringing phenomenon tends to throw the calculations off somewhat. Usually the effect can be neglected; but if the bandpass is seriously narrowed by fringing, the spacing between the elements may be improved with the use of correction factors available in papers published in the company.

<u>Resonant Gap Configuration</u>

We know where to place the resonant elements in the tube body. It is time now to decide what the elements will look alike. All we have decided about them so far is that the susceptance of the windows will be one-half that of the gaps. It will be more convenient to discuss these elements in terms of the doubly loaded Q, rather than their susceptance. By definition we have

$$Q_{2L} = \frac{\omega_o}{\omega_2 - \omega_1}$$

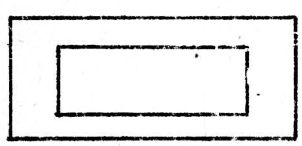
where ω_o is the frequency of operation

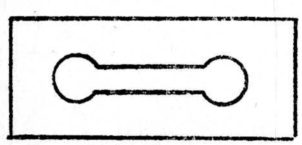
and ω_2, ω_1 occur where the total susceptance $B = \pm G$

Bearing in mind this definition of the doubly loaded Q, we find, through the use of the Rad. Lab. books, or other convenient publications, that if our gap Q's vary between 4 and 5, and the window Q's are of the order of 2, we will obtain a good bandpass structure when the elements are spaced a quarter guided wavelength apart.

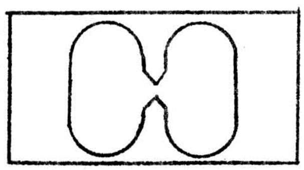
Our problem at this stage is to find some form or construction of the resonant elements which will yield a Q_{2L} between 4 and 5. Most likely we will try for a pair of elements with Q_{2L} of 4, another pair with Q_{2L} of 5, and probably one in between. A resonant gap or iris cut in diaphragm or baffle is the standard method of providing resonance at these high frequencies. The catch is that we don't know too much about calculating the theoretical performance of any one iris type.

An iris being merely an opening in a baffle, so shaped as to transmit energy of a given frequency, one of the earliest apertures was a simple rectangular slot. The capacitance resulting from the length of this opening is rather high, giving too large a doubly loaded Q. To keep the gap small without sacrificing the doubly loaded Q, the rectangular shape of the opening was changed into that of a dumbbell, where the length of the capacitative opening is small, and the circular holes provide inductance.

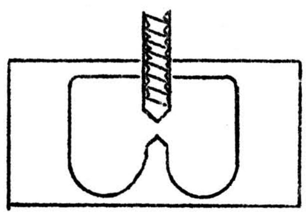




The capacitance of this pattern is still on the high side, however, for the low value of Q_{2L} needed, and the leakage energy far in excess of the limit specified. The gap had to be narrowed still further, yet at the same time the capacitance decreased. This was effected by narrowing to points the portions providing capacitance. Since the gap spacing



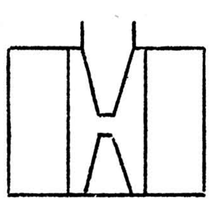
is critical and has to be controlled to tens of thousandths, and even to hundreds of thousandths, it was found advisable to make at least one member of the capacitive circuit components movable. This brought about



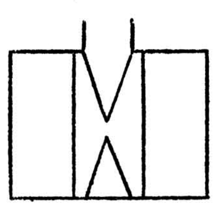
the adjustable threaded cone or post.

Once you consider cones and posts, you have more variations **available** in the geometry of the resonant element. The capacitive unit need no longer be mechanically a part of the baffle. Also straight line surfaces can be set in the guide to provide inductance. Thus we have the separate cone and straight baffle combinations.

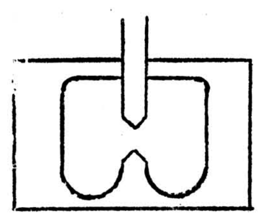
Of the various possible combinations, let us narrow the choice down to one of three most commonly employed



1
X-Band



2
C-Band



3
S-Band

Type 1 is used at X-Band --e.g., the 1B63A. It has hollow truncated cones.

Type 2 we see used in most C-Band tubes. Note the TR361 and TR331. The cones are pointed and solid, except for the one in which the keep-alive is inserted.

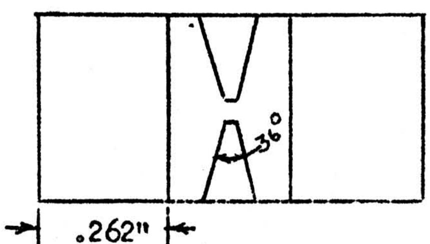
Type 3 is employed at S-Band.

A deciding factor in choosing one of the three types is tolerance requirements. The higher the frequency, the more important tolerances become. Let's take up Type 3. We look back at our experience with this type at 3000 Mc. We recall that we had quite a bit of trouble there, and we can be fairly certain that these troubles will be amplified at least by a factor of 5 at 15,000 Mc. We turn our attention to the other two types.

Say that we can make the straight plates that form the baffle, and locate them properly in the tube body. We then have to select one of the two cone types. Suppose we become intrigued with the truncated cone. To make hollow cones with cut-off ends we have to have a complete set of about 18 special drawing dies. The 1B63A cones are .200" high and have a 36° taper. The cones we will need for our tube will have to be something like .150" high, so that we cannot utilize the dies already on hand for X-Band. We would have to take into account for the bid request whether or not we could set up someone to make the drawing dies, and, if we get the contract, whether or not the expense of the drawings would be allowed. Assuming that the expense will not be allowed, we are left with the second type.

Our next problem is to find the dimensions of the parts. We can go back to the susceptances and find out how dimension variations affect susceptance. Knowing what susceptance we want, we can arrive at approximate values for the size of the structure. The results, however, are apt to be no less approximate than those obtained by scaling from a tube that has already been developed. We might take the tube at X-Band, for example, and reproduce the cones on a smaller scale, cutting them down by the ratio of the wavelengths.

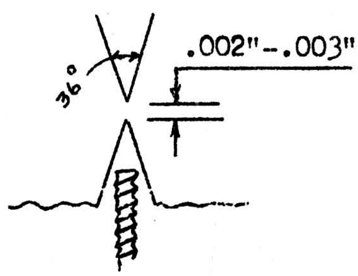
Direct scaling is a practical short cut in design that is effective when the ratio of the guide dimensions is the same both for the tube to be made and the tube used as a model. If we decide to copy the cone-baffle configuration of the 1B63A, we note that its internal guide dimensions are in the ratio of 9/4. Since the waveguide inside the dimensions of the 15,000 Mc. tube are in the ratio of 2/1, we would remark that direct scaling does not seem advisable. If we wanted to design a K_1 TR tube, observing that the inside dimensions of the tube body are .140 x .280, we could scale directly from the 15,000 Mc tube, as soon as the latter was completed and working correctly.



1B63A (3.33 cm)

Suppose for the 15,000 Mc. tube we try to scale from the 1B63A, anyway. It is possible that the end results will not be too far off from that required. We can begin by getting the iris opening. The baffle width of the 1B63A tube is .262". To estimate the baffle width of the 15,000 Mc. tube, we multiply this dimension of the tube set up as the model, by the ratio of the wavelengths. The wavelength of the 1B63A is 3.33 cm; that of the tube under design, 2 cm. Thus the width we want is about (2/3.33)(.262"). With this dimension and the internal width of the guide, we have fixed the iris opening.

We'll copy the 36° taper of the cones making them solid pointed instead of hollow truncated. That gives us sufficient information to determine the cone geometry. We want four solid cones, and we'llmake them of OFHC copper to eliminate conductivity worries. In one of the cones we'll drill a hole for keep-alive lead insertion. One cone in each resonant circuit is to be adjustable, so we will provide a tuner assembly for

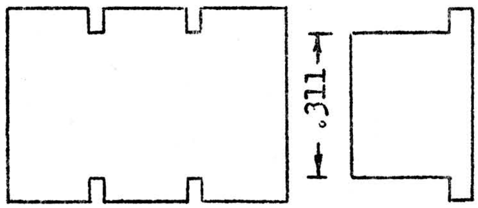


it, consisting of a diaphragm at the base of the cone and a tuning screw or plug inserted through an axial hole drilled from the base. The design can be checked by someone in the plant skilled in the mechanical aspects, who will be able to tell whether it is feasible, or will suggest improvements.

Assembly of Resonant Elements:

Now that we have the tentative design of the resonant elements, we can consider how we are going to place them in the tube body. After we've cut the waveguide the calculated length, we usually slot the body at the points determined by the electrical spacings calculated. Through these slots we slide the baffles. Similarly, through properly located openings in the wide dimension of the waveguide we can insert the cones.

Clearance is going to be one of our worries. It is desirable that the fit of the baffles in the guide be as close as can be tolerated.



We could make the baffle height greater than .311" and score the guide, but that is likely to get us into trouble. As a matter of fact, we may already be in trouble because a tolerance of ±.005" may have to be combined with the .311". We can overcome that obstacle by assuming that we can obtain the waveguide with a tolerance of ±.002", and that we can buy or make the baffles to fit satisfactorily into the guide. We then can slide the baffles in so that they are reasonably straight, and anchor them to the body by some means. When we have assembled two cone-baffle structures in the body, we can place the tube on a test bench and obtain a VSWR plot. From this plot we determine the resonant frequency for a given gap separation.

Since the spike energy is, among other things, a function of the square of the gap spacing, we would like to maintain the separation between cone points to a small value. From our experience with other tubes, or from published tabular information, we guess that what we want is a gap spacing between cones of .002" to .003". We might find as a result of the unorthodox scaling that the resonant frequency is obtained with too large a gap. Then we would have to correct the iris opening.

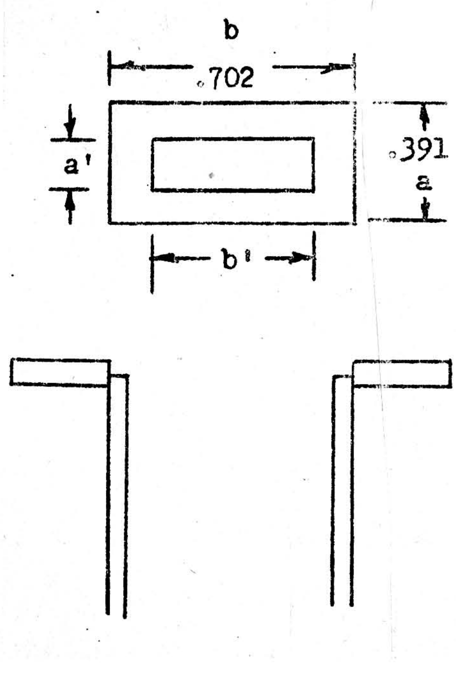
If initially we soft soldered the baffle plates into the waveguide, we can unsolder them and readjust the baffle spacing. If we had cut the slots too deep for the iris opening we need, and then pulled the plates out, the gap left would be a good spot for leakage. What we would probably have to do is cut several depths of slots, and keep adjusting the iris width until the cone-gap separation was what we wanted. Then we would make a VSWR vs frequency plot and calculate the Q's. The chances are that our bandpass coverage is about 8%. We'll take that as a fairly good start and consider the next step in assembly.

Window Design

Putting windows on the tube is a delicate undertaking requiring fine techniques to avoid cracks and bubbles in the glass and poor seals. We usually combine a kovar frame with glass and choose the latter to give the best seal to kovar and the least r-f loss. Glass for windows falls into three principal categories, 704, 7052 and 707. After checking the dielectric properties we select 707 glass because it has the best r-f properties for our purpose, and at the same time we hope we have not made for ourselves too much of a sealing problem.

We may seal the glass to oxidized kovar, or, with new techniques, to unoxidized kovar. The former has been used more by us. The sealing process may be effected either by r-f or by furnace techniques. Where the window is relatively small, the r-f technique is necessary.

Having decided on the component parts we turn our attention to the proportions. The size of the frame or kovar blank is fixed by the dimensions of the waveguide, since the window is to terminate the tube body. Thus it would have to be at least .622 + 2(.040) by .311 + 2(.040) for best fit. Locating the window on the tube body will not concern us too much at this point. We could easily enough put a rectangular flange on the body in such a way that supporting surfaces are provided for the window.



The window, in operation, is itself to be a resonant element in the tube. Note that we have implied a rectangular opening. We have to find the dimensions appropriate for transmitting the r-f of operation. We can make use of an empirical formula to get approximately the a' and b' dimensions of the opening. This formula relates the a' and b' of the window with the a and b of the blank in terms of λ_o.

$$\frac{a}{b}\sqrt{1 - (\lambda_o/2a)^2} = \frac{a'}{b'}\sqrt{1 - (\lambda_o/2a')^2}$$

The a' value can be "guessed" fairly well by the fact that a Q_{2L} of the order of 2 is to be the final result. Here, again, knowledge gained from earlier developments is drawn upon. We can examine the window of the 1B63A where the Q is 1.7, and the window of the X-7109 where the Q is 2.1 and estimate what fraction of the waveguide each window is. On the same basis we can estimate or choose the a' as being of the order of .070". The b' then will come out to be of the order of .300".

An opening of these dimensions will give proper resonance if the diaphragm is very thin. But the thickness of the glass will change the characteristic impedance of the opening and, hence, its resonant wavelength. There are formulae available in handbooks for correcting the opening and for determining the approximate glass thickness to give a good match. We will find that .020" is about the right thickness.

We would check the performance of this preliminary design and probably discover it is a poor match. There are two measures we take to correct the impedance. We may grind down the thickness of the glass and we may decrease the iris size. The latter is conveniently accomplished by removing the oxide from the kovar, painting strips of the glass along the boundaries with silver oxide paste and firing the window at a reasonable temperature. This forms metallic strips that diminish the rectangular opening to the size suitable for resonance with the glass thickness chosen. Finer tuning is accomplished by grinding down the thickness.

Mechanical performance is then checked by various tests. After our window has passed the vibration, crack-cycling and other tests of mechanical strength, we are faced with the problem of anchoring the window to the tube. We would prefer hard soldering, because then more can be done in the way of high temperature processing to remove contaminants. Techniques of hard soldering on a production basis have not as yet been perfected, however; and this problem should be kept in mind when estimating developments costs.

Leakage Power Controls

In addition to the resonant element designs for the bandpass, we have to give some thought to the protection afforded the receiver crystal at high power levels. Assuming that we maintain the .002" - .003" gap, we must put in a gas fill to counteract the high spike energy obtained with low-Q gaps. If the crystal burns out when 1/3 erg is applied to it, then we have to keep the spike to a maximum of .1 erg. We need an easily ionizable gas, such as argon. But at the same time, we aim for a rapid recovery time, so that we have to sacrifice some of the protection by adding water vapor to the gas fill. Investigation has for a long time been, and currently still is, in process to find an element or mixture which will give reasonable recovery time and adequate crystal protection.

Meanwhile, prior experience has taught us that an argon-water vapor fill of about 10-12 mm pressure Hg is a good starting point. We can experiment with the percentages and the total pressure of the combination until the tube performance is satisfactory.

Also we have to put in the tube the keep-alive lead that we've been hinting about all along. Even with the very low Q's, the resonant build-up across the gap is not sufficient to establish a firing condition soon enough to block r-f energy from the crystal. With the keep-alive lead we always have one of the gaps "idling." A comparatively small increase in r-f voltage on the tube will fire the gap more rapidly.

Estimate of Development Costs

We have generally blueprinted the plans for developing the tube. The only item remaining to consider is the cost of the development.

For this tube we can guess that the advisory services of a Sr. Engineer will be needed for from 3 to 6 months. The Sr. Engineer steers the Jr. Engineer, who may often be a new employee, onto the right track, a path often found the hard way. The Jr. Engineer will have to take care of life tests, or use testing time in the Test Department. A technician's time, as well, will be needed. And some estimate should be made of the cost of engineering paper work.

To complete this discussion we can give some sample figures. If we charge

1/3 year for Sr. Engineer, say,	$2,000.00
1 year for Jr. Engineer, say	3,500.00
Technician time	1,000.00
then, engineering labor costs	$6,500.00

To engineering labor costs add 150% overhead, about $5,000 for material and a 7-1/2% fee for profit, and we have fairly good insight into the cost and extent in time of developing the TR tube.

It should be kept in mind that if the contract is fixed-price, amounts spent on development over that estimated will be paid for out of our own pockets. If the contract is cost-plus-a-fixed-fee, it sometimes is possible to get more money to complete development, but the fixed fee remains fixed, cutting the percentage profit. So, even on the so-called "safe" contracts, a risk is involved. That this risk shall be taken with full awareness and as a matter of definite consistent policy, is one of the most important jobs of engineering management.

OPERATION AND DESIGN OF TR AND ATR TUBES

Lecture No. 14

Some Background Notes on Radar Systems

To conclude this series of lectures, which has focussed on various features of the TR and ATR tubes, we consider the systems for which the switch tubes were developed. A brief survey of the early history and developments in radar, in addition to an indication of current trends in the field will be given.

Origins of Radar

Observations of the reflections of radio waves from objects directed attention to the principles of radar both here and in England in the early part of this century. The reflection phenomenon was mainly of academic interest until scientists used a radio-pulse echo technique to measure ionospheric heights in 1925. Within the next ten years, individuals in various countries independently noted that echoes could be received from objects smaller in extent than the ionosphere.

The notes and observations of the principles evolved into the notion of detecting aircraft and ships at about the same time in America, England, France and Germany. By 1935 development of radar as a practical working system was underway in England. Being pressed by the urgency of immediate defense needs, the British had radar systems in production at the start of the war. A system of early warning radars ringed the coast of England in time for the shooting part of the war. America proceeded slowly in early experimentings, but accelerated the pace of development after it became directly involved in the fracas.

Advent of Pulsed Microwave Radar

The early warning radar systems operated at frequencies of from 100 Mc. to 600 Mc. This band was used because it was the highest frequency range permitted by the limited techniques of tube construction, communications circuitry and associated propagation methods. The equipment engineered to perform detection was tremendous in size and cost. From thiz point of view alone adventure into higher frequencies for extending the function of radar systems would naturally be tempting to engineers knowing that size of components varied inversely with the frequency. A greater incentive for exploring possibility of the frequencies above u-h-f was the desire for accurately determining direction, logically added to the need for finding the range of an object. This implies the necessity of focussing the energy propagated into sharp beams. The width of the beam depends upon the ratio of the antenna dimensions to the wavelength of the energy radiated. Given an antenna of fixed size, the beam width varies directly with the wavelength.

The greatest push toward expanding radar developments was given by the arrival of the improved magnetron. The magnetron had been a laboratory plaything in the early 20's of this century. When British scientists in 1940 produced the cavity magnetron as a source of high power microwave pulses, it provided the impetus into microwave research and brought into view innumerable refinements and extensions of radar functionings, in addition to the portability feature.

The microwave region for radar purposes ranges from 1000 Mc. up. The lower frequency or upper wavelength limit being set by the size of an antenna practical for the size of the airplane. The upper frequency or lower wavelength limit is set by the propagation characteristics of microwaves and the atmospheric conditions surrounding our earth. Microwaves are not reflected from the ionosphere, therefore limiting detection of targets to ranges within line of sight.

During the enthusiastic and successful conquests of the higher frequencies for diverse applications, systems were developed and built to operate at K-band around 24,000 Mc., creating a new high in compactness and definition of information presented. One of the saddest discoveries, though, after these sets were put in operation, was the phenomenon of absorption of this wave energy by water vapor, hampering target search on this band in moist atmospheres. Rain or fog washed out the picture on the indicator. To get beyond the water absorption point research was pushed to the shorter wavelengths - - the K_1-band (35,000 Mc.) More recently, in an effort to keep K-band definition and completely avoid moisture-absorption, the trend is to the lower frequency side of K-band, or K_A, about 16,000 Mc.

Radar Range Limitations

In regard to range coverage of radars an interesting limitations is a blind region caused by reflection of waves from the surface of the earth or sea. This is neatly illustrated by the trick played on a group making an evaluation test of early warning defenses. Thirty planes were used in the test. They were flown in toward the defense system in three groups. The first group flew in at an altitude of 30,000 feet and were detected about 120 miles distance from site. They were theoretically conquered. The second group flew to the objective at an altitude of about 10,000 feet. They were picked up on the radar system at something like 40 or 50 miles away and theoretically conquered. The third group, however, flew in at about 150 feet above sea level. The first warning the radar operator got of this group was seeing them overhead.

The absence of signal in the last case resulted from interference due to reflection from the earth. If we consider an antenna on a tower, beaming signal at an object, we note that the energy reaches the object by two paths, one directly, the other, by reflection, or effectively, from an image antenna. (Fig. I) The returning echo energy is thus composed of two signals, with a phase difference that might cause disturbing interference, depending on distance and altitude of the object. The extent of interference is readily seen by noting the effect on the radiation pattern of the antenna. The contour of the directive antenna's coverage, without a ground plane, which is that of Fig. II, becomes the lobe structure of Fig. III. The elevation angle from the ground of the lobe is inversely proportional to the distance between the antenna and its

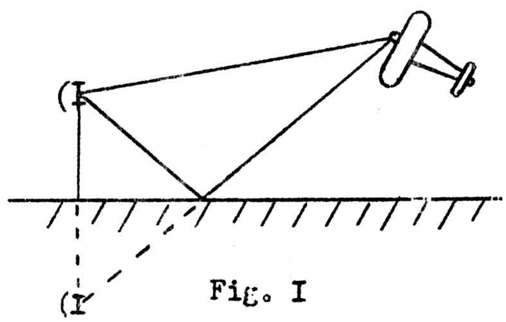


Fig. I

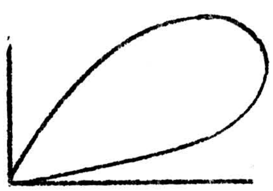


Fig. II

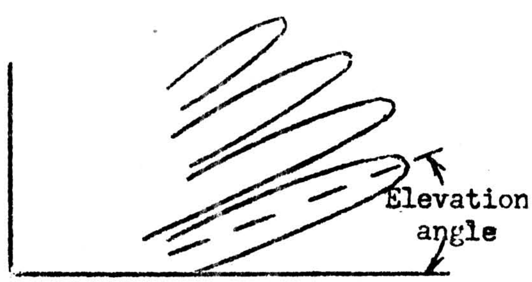


Fig. III

image. If this distance is measured in terms of the wavelength of propagation, it can be seen that shorter wavelengths of propagation will bring the lowest lobe closer to the reflecting surface, allowing coverage at lower altitudes. That principle is one of the reasons for shifting warning radars to higher frequencies. A diagram of field strength may be made to take into account the curvature of the earth, as in Fig. IV, which indicates possible radar coverage in terms of azimuth and range.

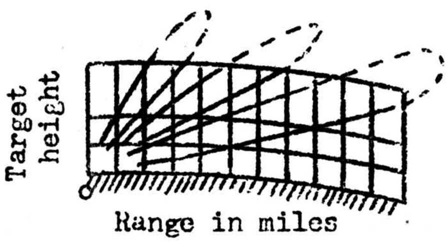


Fig. IV

Since microwaves are neither reflected from the ionosphere to the extent observed at frequencies below 30 Mc., nor spread around the curvature of the earth by the diffraction peculiar to lower frequencies, their coverage is more closely limited to line-of-sight range. There is, however, a special condition of the atmosphere which will permit range beyond the horizon, known as the duct effect, on which investigation is currently in process.

This duct action is due to a temperature gradient which can produce a rapidly changing index of refraction. Over a relatively smooth sea, where the duct action is more prevalent, the layer of air in contact with the surface of water is very nearly saturated with water vapor. The layer of air above is not. That is, the continual evaporation of water vapor from the sea is diffused into the region of air above the sea. This would mean a vertical gradient of water vapor concentration, heaviest at the surface of the water and diminishing with respect to height above the surface. Over land, warm masses of air radiating heat from the earth at night, may cause to be formed a thin layer of cold (dense) air just above the surface. Again, there is a vertical gradient, the refractive index decreasing rapidly with distance from the ground.

A sufficiently sharp change in the index of refraction will bend rays making an angle of incidence greater than a critical angle for total reflection, so that they follow the curvature of the earth. The distance from the surface at which this bending occurs is the top of the duct, which, together with the surface, form a kind of waveguide in which waves are trapped and transmitted around the earth for unusual distances beyond line-of-light. (See Fig. V.) As far as has been observed, the duct action usually is confined to the microwave region. For, similar to metal waveguides, ducts have a cut-off frequency which apparently is a function of the height of the duct and the steepness of the refractive index within the duct region. The height of the ducts ranges from a few tens of feet to a few hundreds of feet.

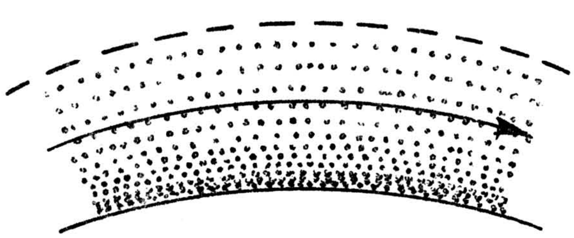


Fig. V

Categories of Radar Systems

Radar systems may be classified by function and by size. The primary functions of radar, in its current stage, are search, fire control and navigation. Other usages have come into existence with the need for counter measures against the enemy; and still further applications are being found to satisfy the exiggncies of modern expanding navigation. The microwave bands have been allocated in accordance with the most satisfactory power fulfillment of the various functional requirements, but there is no rigid demarcation between bands, principally due to size consideration.

The desirability of moving systems with ease to the area of operation soon brought into prominence the engineering of movable units. Generally, units that can be moved are (1) either portable with handles, (2) or transportable, truck borne, (3) airborne or (4) shipborne. In a system destined to be portable, weight is a prime factor. To get maximum performance for the limited weight allowable, it is a natural course to go to a wavelength shorter than used for a permanent land system of the same function. Thus, the overlapping of allocations is readily understandable, where a band ordinarily used for fire control may be used for search as well.

The following table indicates the general trend of employment of the various bands for radar purposes:

Band	Frequency in Mc.	Available Power	Approximate Range in Miles	Usual Function
L	1,000	10 Megawatts	350	Early warning, search
S	3,000	About 1 Megawatt	200	" " "
C	5,800	About 350 Kw.	80	Search
X	10,000	250-300 Watts	100	Fire control, some search
K_A	16,000	Current experimental for navigation		
K	24,000	Up to 100 W.	25	High definition navigation (limited by center absorption), and weather tracking
K_1	35,000	About 50 Kw.	15	Weather tracking

Range is determined by frequency and such other factors as power output, antenna gain, and receiver sensitivity. The range figures given are on the basis of the heights above ground obtained with aircraft.

Radar Countermeasures

The principles of radar were known to the enemy as well as to the Allies. As a matter of fact, the enemy early warning system in the vicinity of 500 Mc. was exceedingly effective. The northern coast of Europe was quite thoroughly covered with the systems. It became necessary to devise means of getting through the screen safely. Radar countermeasures were developed to "jam" enemy equipment.

An extremely effective method for disturbing and confusing reception on radar indicators consisted of transmitting random noise at the operating frequency of the equipment to be jammed. One such jamming system developed as part of the program at Radio Research Laboratory was built around a specially designed oscillator tube, the resnatron. The resnatron is a tetrode, operated as a Class C oscillator, in the frequency range 340-625 Mc., and capable of delivering up to 60 Kw. continuous power at high efficiencies. A noisy gas tube provides source noise, which is amplified to modulate the transmitter with 5 Kw. of random noise.

One of the important factors in designing the tube was that it be demountable for field use. Interestingly enough, mobility of the entire system was achieved by housing the components of the units in ten trucks and a trailer. The assembled resnatron tube weighed in the neighborhood of 250 lbs. Two resnatron oscillators could be operated simultaneously. Two others were standing-by for immediate use. At the first installation in Southern England, the antenna design had not been completed and a temporary antenna was rigged up. This was a huge horn made of chicken wire, and about a quarter-mile long. The permanent antenna was designed for truck mounting. Two accompanied each installation, one for the 460-625 Mc. operation, the other for 340-520 Mc. operation. The performance of the system was so effective that the enemy became furiously busy with changing the frequency of their warning radars. As fast as the frequency of warning systems was changed, the frequency of the jamming system was changed. For about a year there was an energetic game of hide-and-seek.

A rudimentary technique of jamming radar systems utilized "window" . These were pieces of aluminum or tin foil cut to be half-wave dipoles. Bales of them were thrown out of planes to mask by a cloud of noise the information about the plane returned to a radar system. When "window" was falling downward, the "grass" or noise level (Fig. VI) of the radar scope went wild, making it almost impossible to detect a target.

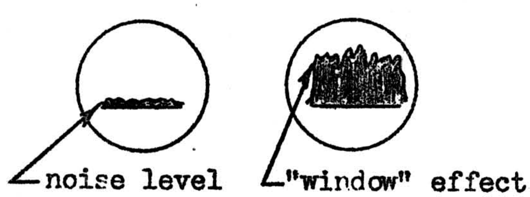


Fig. VI

During heavy air raids, several tons of "window" were thrown downward. One of the amusing results was that Ardenne Forest in Belgium, which is largely evergreen, the day after a raid was turned into a forest of predecorated Christmas trees.

F-M Radar

Although the major portion of work done on radar involved systems using pulse techniques, it should be noted that echo signals are obtainable when transmitted signals are other than pulsed. The pulse technique was heavily favored since it provides the simplest method of measuring elapsed time between transmitted and received signal, and, hence, distance between receiver and target. For distinguishing between two or more targets, and obtaining maximum information at high rates, the pulse technique is superior to other techniques. But, since echoes cannot be received until transmitted pulses end, there is a definite minimum on the range of the system. Where requirements call for working down to zero range, and the rapid rate of collecting information is of secondary importance, c-w radar systems can be used successfully. The radio altimeter is a good illustration of this case, where only one target, the earth, is of interest, and the rate of information needed about it per second is relatively low.

There are a number of c-w systems in use for various purposes. The radio altimeter utilizes the f-m technique of range determination. The transmitter operates continuously and is frequency-modulated. Briefly, range of the target is computed by measuring the frequency of returned signal and comparing it with the known variations with time of the transmitter.

Another application of f-m radar is the tail-warning device. This is a system placed in the tail of an airplane to keep the pilot informed as to whether or not somebody is behind him.

Refinements in Radar

The urgency of developing radar systems as aids to warfare accelerated the investigation of possible improvements. The rate of developments was so rapid that radar became one of the fastest growing engineering industries ever known. It would not be possible within the scope of this lecture to note the various types of systems constructed and put into use. Of prime importance, perhaps, are those which were worked on to increase the range coverage of the system, and to simplify the interpretation of information received. The airplane early served as an expedient for adding to the range. Many systems were found for presenting more exact information, among these the PPI, the various methods for azimuth determination (such as lobe or beam switching) and the clutter elimination contrivances.

One of the height finding methods made practical use of the lobe pattern resulting from interference of direct and reflected waves, (See Fig. IV). When a target lies well within the coverage pattern, the detection of such a target depends upon its altitude as well as its range. If the target is a plane approaching the radar at a constant altitude, the echo from the plane will vanish, reappear, vanish, etc. as it crosses the lobe pattern. The interference pattern can be calibrated with a target plane at various heights so that range and altitude of a target is given by the manner in which it fits into the lobe pattern. Each such installation has to be calibrated, but the results are well worth the effort. This method of control was the forerunner of the more accurate tracking of aircraft by ground control of approach (GCA), currently being discussed for controlling traffic at airports.

Moving-Target Indication

In ground installations where the primary function is obtaining information on airplanes, a major nuisance is the echo information returned from fixed landmarks. The radar display becomes cluttered with signals from nearby mountains, stationary structures, etc., which are of no interest to the operator. To present on the display solely that which is of value, the moving-target indicator (MTI) is so constructed as to select only moving targets. For this purpose the response of the system to moving targets differs from the response to fixed targets. The Doppler effect, or shift in frequency, or change in phase, resulting from reflection from a moving object, is taken advantage of in differentiating between the fixed and mobile. By means of a delay system and the subtraction of one sweep from its succeeding sweep, echoes of fixed frequency response are cancelled. The radar display, then, is relatively clean for a clear picture of such targets that are moving with respect to the radar installation.

A further application is the detection of convoys traveling on the road in darkness. The MTI is installed in aircraft to improve coverage. Compensation is made for the motion of the antenna and the speed of flight to cancel out the effect of stationary objects on the ground, moving with respect to the aircraft. In this manner early detection is possible of trucks moving down the highways at night.

Radar Relay

One of the means of extending coverage of radar systems was the development of relay links. Considering the line-of-sight limitations resulting from the earth's curvature, it would seem that a target would have to be at an unusually high altitude normally to be detected at 350 miles from an L-band system. The unusual range is more readily obtainable, though, by transmitting information from one site to another. Taking advantage of the greater line-of-sight

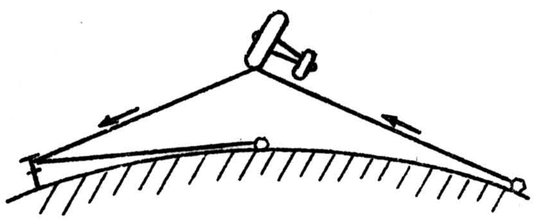


Fig. VII

conditions, or extended horizon, realized from aircraft, a relay link is set up between airborne systems and the surface installations. (Fig. VII) Information collected by the airborne system is relayed to the surface-installed system by a specially developed transmission system. In this manner better air and sea coverage is obtained.

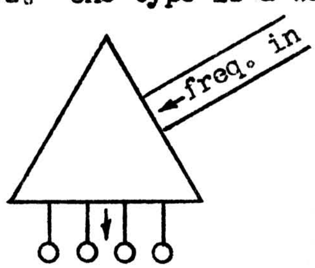
Recent Trends

The various refinements to the radar systems have steadily multiplied the applications possible. There is a great deal of literature published on the numerous applications found for radar since the early warning systems were first put into operation. A few of these we have noted briefly above. All the questions, however, have not yet been answered; all the problems not yet solved. As a matter of interest, we will outline briefly some of the more recent programs to show the trends in development.

Search Receivers

Under development in the countermeasures divisions are automatic search receivers. The radar system, controlled by a motor device, will investigate a given band. Whenever it picks up a signal of another radar system, it stops scanning and gives warning, by such means as the ringing of a bell. The operator by using auxiliary equipment can discover a large amount of information about the radar system such as the detected system's pulse length and p.r.f., and the speed of the antenna's rotation.

Also under development are instantaneous all-band search receivers. One type is a waveguide prism which works very much on the principle of an optical prism. That is, several bands of frequency can be fed into the waveguide and separated by the prism effect. Information received at one frequency band is channeled to the proper detector-receiver system. (Fig. VIII) Thus, a large number of frequencies can be monitored without the need for manual or automatic tuning. However, the number of bands that can be covered is limited by the waveguide.



Separate Detectors
Fig. VIII

Correlation

Improvements brought about by technical skills, though, are not the sole hope of future developments. In recent years there has been a re-examination of the basic problem of the transmission of information, especially in the presence of noise. Investigations are being made into fundamental problems such as: what is the best method of transmitting information; how long does it take to transfer a given amount of information; and what can be done to separate the information from noise. These problems can be more clearly defined when they are related to familiar factors. For example, the transmission or gathering of information is dependent on time. The amount of signal transmitted and the amount detected in noise are related to the bandwidth of the system.

The approach of the more recent information theory differs from the earlier in that it combines many phenomena previously not treated. Included in communication theory are such subjects, formerly considered unrelated, as thermodynamics and statistics. Certain principles of statistics are being adapted in interesting fashion and will be touched upon here as they bear on the subject.

An important notion in statistics is that of correlation. Correlation is a method of comparing two or more items to find how they agree. Signals used in radar are functions of time. One particular type of correlation is known as autocorrelation. Inspection of the term leads us to expect a method of relating some class of items with itself; and that is exactly what is done to clarify the presentation of a piece of information. Over a period of time many return signals can be received from the same object. We can take one signal and compare it with itself at some later period of time.

The comparison method consists in sampling the signal and the signal delayed, multiplying them, and taking an average. Theoretically, it would take a great deal of time to perform this averaging. Actually, the time of taking the average can be limited and still permit a fairly good approximation of the signal.

There are two time factors introduced in the comparison method: the amount of time delay between signals to be compared and the amount of time taken to perform the averaging process. These will be considered as they enter into results obtained with two types of signal -- periodic and random.

If a signal is more or less periodic, it is found that the autocorrelation value is sometimes large and sometimes small. As the delay is increased it will be found that the autocorrelation at first is large in value, then decreases to a small value, increases to a large value, etc. In other words, the autocorrelation value of a periodic signal is itself nearly periodic. (See Fig. IX.)

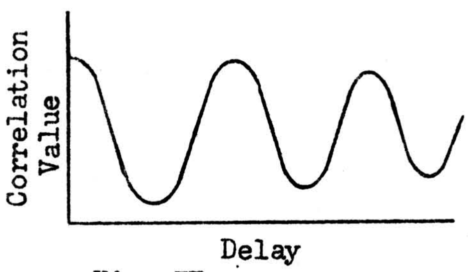


Fig. IX

Now, suppose that the autocorrelation of a random signal, such as noise is made. For a very small delay there is an appreciable correlation since the bandwidth is limited. With increase in delay, the value of autocorrelation decreases. (Fig. X)

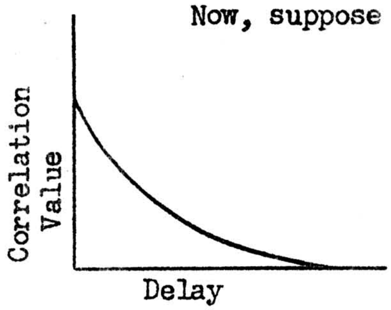


Fig. X

Let us consider the signal as it is being received by the system, where it shows up as a composite of both noise and signal.

The autocorrelation of the combined signal will be a combination of the signal with signal; correlation of noise with noise; and finally, signal with noise. The latter two follow the curve in Fig. X. So that as the delay is increased, the autocorrelation values obtained for comparison of noise with noise and for signal with noise continually decreases. Therefore, if there is enough time to perform the autocorrelation process so that a relatively large delay can be introduced between signals being compared, it becomes possible to detect the signal in the noise, that is, improve the S/N ratio.

There is another fundamental consideration in the problem of improving the S/N ratio. The question arises whether maximum use is made of the information being received. A definite store of facts is on hand about the signal transmitted, and quite a bit of information which has not been used is known about the echo signal. The idea is to devise a system for utilizing this information. If, instead of comparing the echo signal with itself, the returning signal were compared with known information, such as the transmitted pulse, the correlation might prove of considerable aid in detecting signal amid noise clutter. This type of correlation is known as "cross-correlation."

For example, with a pulsed radar system, a transmitted pulse will at a certain time later, if there is a target, return as a reflected pulse. The transmitted pulse already contains some valuable information, so that if it is compared with the received pulse, the S/N ratio resulting is better even than that obtained with autocorrelation of the received signal.

We can illustrate this principle by a simplified system such as diagrammed in Fig. XI. At the same time that a pulse is sent out from the transmitter to the target, a transmitter pulse is fed through a delay unit and reaches the receiver simultaneously with the returned echo signal. The improvement in eliminating noise is considerable. If the delay system is fixed in time, then the cross-correlation will work only for a fixed range. But the delay box can be made variable to cover information gathered at various points within the range of the radar system.

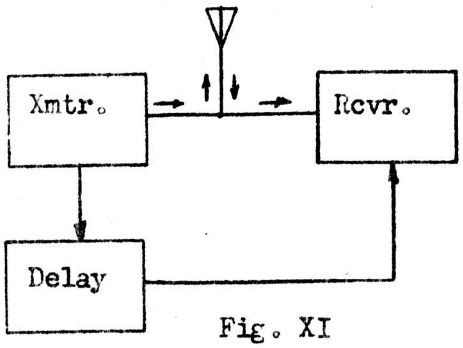


Fig. XI

Cross-correlation is more promising in results and would probably be simpler than autocorrelation.

There are then, many types of radar systems that have been used for different applications. Others are being or will be developed. Each type demands its own "tailor-made" components, including TR's and, sometimes, ATR's. Sometimes these requirements overlap. More often they do not. Consequently, not only are there now many different TR's and ATR's, but we may expect that other types will be required in the future. These new types, it is safe to predict, will fully reflect our expanding experience in their ability to handle higher frequencies, greater power, wider or flatter bands, or other peculiar specifications. The design and development of these tubes will continue to demand the utmost in ingenuity and ability.

As a final recommendation, therefore, we shall simply suggest the continued maintenance of a large inventory of aspirin tablets.