THEORY OF OSCILLATOR DESIGN

by

Erich Hafner

U. S. ARMY ELECTRONICS RESEARCH & DEVELOPMENT LABORATORIES

I. INTRODUCTION:

The approach to the design of harmonic oscillators which will be reported on in this paper has been developed in the course of a program aimed principally at establishing the fundamental limitations on the frequency stability attainable by means of mechanical resonators. A major portion of the paper therefore, deals with the criteria which have to be met in order to assure that the frequency stabilizing characteristics of a crystal unit are fully utilized and that the influence of changes in the various circuit parameters on the output frequency are kept to a minimum.

However, it is the formulation of the conditions for oscillation in an apparently new manner which is of pivotal importance for the developments in this paper. This new form, though only slightly different from previous usage, opens the way to achieving a very thorough qualitative understanding of oscillator performance which will be found useful in a wide variety of applications.

Many different methods and techniques for the design of harmonic oscillators have been reported on in the literature [1-4] and under Signal Corps Contracts[5-8], Some may be considered more important than others, but each one has added significantly to the present state-of-the-art. A detailed discussion of their relative merits however, would undoubtedly lead too far afield.

The success of any analysis of oscillator performance depends in no small measure on an adequate appreciation of the phenomena which determine the amplitude of the steady state oscillations. Within the concept of the equivalent linearization which, incidentally, will be used throughout this paper, amplitude limiting occurs because the equivalent linear parameters of an actually nonlinear element depend upon the amplitude of the signal. Dasher and Witt[9] have investigated some aspects of this phenomenon in connection with oscillator design considerations, while Reich [10] has emphasized its major significance for the understanding of oscillator behaviour an a broader basis.

II. DISCUSSIONS:

1. THE BASIC OSCILLATOR EQUATIONS

An idealized oscillator circuit is shown in the upper part of Fig. 1. The general impedances Z_1, Z_2, and Z_3 form the feedback network and the electron tube supplies the energy necessary to sustain oscillations. If the concept of the equivalent linearization is used, the vacuum tube can be treated as though it were a linear element whose a.c. output current

i_1 is proportional to the a.c. input voltage v_g and 180° out of phase with it:

$$i_1 = -g_m v_g \qquad (1)$$

Since the tube, however, is actually a non-linear element, the value of g_m depends upon the amplitude of the signal v_g. A thorough discussion of this subject, though most interesting in itself, is outside the scope of this paper and we will assume here only that g_m as a function of A, the amplitude of the fundamental component of the oscillator signal, gives a steadily decreasing curve such as shown in the lower part of Fig. 1. This assumption meets with the most commonly encountered behavior and does not impose a serious restriction on the developments of this paper. The value of g_{mo}, which represents the transconductance of the tube for infinitely small signals, as well as the exact shape of the g_m vs amplitude curve depends upon the type of tube and upon the d.c. potentials on all the tube electrodes However, this curve is independent from the impedances Z_1, Z_2, and Z_3 in the feedback network.

The voltage v_g in Fig. 1 is related to the input current i_1 by the relation

$$v_g = i_1 \frac{Z_1 \; Z_2}{Z_s} \qquad (2)$$

with

$$Z_s = Z_1 + Z_2 + Z_3. \qquad (2.a)$$

The relation (2), because of (1), reduces to the conditions for oscillations in the form

$$g_m \frac{Z_1 Z_2}{Z_s} = -1. \qquad (3)$$

If we set

$$Z_i = Z_i \; e^{j\theta_i} \quad (i = 1, 2, 3, s) \qquad (4)$$

and separate the real from the imaginary part in (3) we obtain the two conditions

$$\theta_1 + \theta_2 - \theta_s = \pi \qquad (5)$$

$$\frac{|Z_s|}{|Z_1| \; |Z_2|} = g_m < g_{mo} \qquad (6)$$

which have to be satisfied simultaneously for steady state oscillations to occur. The condition (6) requires that the ratio of the impedance magnitudes be smaller than g_{mo}. The value of this ratio determines the g_m which must be exhibited by the active device during steady state oscillations and hence it determines the amplitude of the oscillations by way of the amplitude dependence of g_m shown in the lower part of Fig. 1. If large amplitude oscillations are required, therefore, it is necessary to make the difference $g_{mo} - g_m$ as large as possible. To obtain small amplitudes the opposite is true, of course.

It might be well to point out in this connection, that an incidental change in any one of the impedance magnitudes will change the value of the impedance ratio in (6) and hence cause g_m to change. Since, however, a variation in g_m can take place only if the amplitude of oscillations changes, the incidental variation in one of the impedance magnitudes will cause a change in amplitude which can be quite appreciable if the slope of the g_m vs amplitude curve is small. In order to steepen the slope of this curve, particularly for low amplitude operation, one or more of the d.c. bias potentials are frequently made a function of amplitude through the use of AGC circuits.

While equation (6) determines then the amplitude of oscillation, the equation (5) determines the frequency at which the oscillations take place.

Since all three impedances Z_1, Z_2, and Z_3 in Fig. 1 and therefore Z_s as well are, in general, functions of frequency, one or more of the phase angles in (5) will also be functions of frequency and the equation can be satisfied only at one, or more often, at several discrete points. If there are several, stable oscillations will occur at that root of (5) at which (6) requires the lowest value of g_m.

Any analytic method to solve (5) is impractical for all but the most idealized circuits, however, it is possible and indeed most instructive to solve this equation without any additional approximations graphically in the impedance plane.

While this technique is by no means restricted to quartz crystal oscillators, we will use in the following the Pierce and Miller circuits to illustrate the approach.

Before we go into it, however, we believe it quite useful to review in some detail the impedance diagram of a crystal unit [12] because of its fundamental importance for crystal oscillator operation and performance.

2. THE IMPEDANCE DIAGRAM OF A QUARTZ CRYSTAL UNIT

The familiar equivalent electrical circuit of a crystal unit with load capacitor is shown on the upper righthand side of Fig. 2 and next to it, it is drawn again to identify the symbols which will be used in describing its properties. Only the narrow frequency range of a particular crystal response is of interest and it will be assumed that X_1, the reactance of the motional arm, is the only quantity which changes appreciably with frequency in this range, i.e., X_o, X_L, and R_1 are assumed to be constants.

If the impedance $Z_x = R + jX$ of a crystal unit is measured point by point as a function of frequency, it will be found to describe a circle in the R-X plane. Such an impedance diagram of a crystal unit is shown as the heavily drawn circle in Fig. 2. It can be described by the equation

$$\left(R - \frac{X_o^2}{2R_1} \right)^2 + (X - X_o - X_L)^2 = \left(\frac{X_o^2}{2R_1} \right)^2 \qquad (7)$$

The vector $Z_x = |Z_x| \, e^{j\theta_x}$ representing the impedance between the two terminals of the crystal network follows this circle in a clockwise direction as the frequency of operation is changed upward through the range of the crystal response. Each point of this circle corresponds to a different frequency and a measure of the change in Z_x with frequency can be obtained by providing the circle with a frequency calibration.

This is the purpose of the set of circles drawn in Fig. 2 with thin lines. Each one of these circles, which follow the equation

$$R^2 + \left(X - X_o - X_L + \frac{X_o^2}{2\,(X_o + X_1)} \right)^2 = \left(\frac{X_o^2}{2\,(X_o + X_1)} \right)^2, \qquad (8)$$

belongs to a different value for X_1 and hence to a different frequency. Those drawn into the diagram, Fig. 2, are separated from one another by an equal increment in X_1, and hence by an equal number of cycles, over a certain range of X_1 around $X_1 = 0$. Since the intercepts of these circles with the impedance circle of the crystal unit identify the frequencies at which Z_x assumes the corresponding values, a diagram such as shown in Fig. 2 provides a rather illustrative picture of the behavior of the crystal impedance throughout the response range.

As an example we will consider the effects of a variation in the load impedance X_L. From the equations (7) and (8) it can be seen that any change in X_L effects a simple translation of the entire diagram along the imaginary axis. If X_L and X_o are both capacitive as has been assumed here, an increase in the magnitude of X_L will cause the entire set of circles to move downward. Consequently, since the frequency calibration on the impedance circle is not affected thereby, the frequency at which Z_x has a given phase angle, say $\theta_x = 0$, will increase. Z_x will thereby move into a range of the impedance circle where the frequency calibration becomes more open. In an oscillator which should be frequency modulated, therefore, a given amount of modulation in X_L will cause a correspondingly smaller modulation of the output frequency. On the other hand, if X_L is made inductive, for example, by using a condenser and coil in series in place of C_L, the distance between the center of the impedance circle and the real axis can be made quite small. Since now the resonance frequency of the network, i.e., the frequency at which $\theta_x = 0$, is in a range of the impedance circle where the frequency calibration points are closer together, the same amount of modulation in X_L will cause a much larger modulation of the output frequency.

To help avoid a trivial trap in the use of the impedance diagram later on, it should be emphasized that the impedance vector Z_X starts of course at the origin of the R-X coordinate system, while the impedance circle touches the imaginary axis at $X = X_o + X_L$.

3. THE PIERCE OSCILLATOR

In an oscillator such as shown in Fig. 1, the crystal network can take the place of any one of the three impedances Z_1, Z_2, and Z_3. If it is Z_3 and the impedances Z_1 and Z_2 are capacitive, the resulting configuration is that of the well known Pierce oscillator, shown schematically in Fig. 3.

For the moment, it will be assumed that the impedances Z_1 and Z_2 have been chosen initially and that it shall be determined if this configuration will oscillate with a particular crystal unit and at what point on the impedance circle the crystal will be operated. According to the condition (5), oscillations can take place only if θ_s, the phase angle of the sum vector $Z_s = Z_1 + Z_2 + Z_3$, is given by

$$\theta_s = \theta_1 + \theta_2 + \pi$$

We can plot, therefore, the impedances Z_1 and Z_2 in the impedance plane and find the angle $\theta_1 + \theta_2$ as well as $\theta_s = \theta_1 + \theta_2 + 180^o$, such as shown in the lefthand diagram of Fig. 3. Hence, for the phase relation to be satisfied, the sum vector Z_s must fall along the broken line in the first quadrant.

To find the magnitude of Z_s, it is only necessary to first obtain $Z_1 + Z_2$ and to use this impedance as the origin for the impedance diagram of the crystal network. This is shown in the righthand diagram of Fig. 3. The broken line circle denotes all possible values for Z_3 while Z_s can only assume values along the line θ_s = const. The intercept of these two lines therefore determines Z_s and Z_3. Apparently Z_3, the impedance of the crystal network, must always have an inductive reactance which is larger than the sum of the capacitive reactances of Z_1 and Z_2 unless the resistive components of the latter are zero, i.e. unless θ_1 and θ_2 are both -90^o.

If the diagram has been drawn to scale, the length of Z_s can be measured and the ratio g_m

$$\frac{|Z_s|}{|Z_1| \; |Z_2|} = g_m$$

determined. If g_m is smaller than g_{mo}, oscillations will take place and, from the g_m vs amplitude curve, the amplitude of these oscillations can be found. Their frequency follows from the value of Z_3, together with the equation (8). The latter, incidentally, will generally be found much easier to use, either in the form shown or solved for X_1, than most other relations for this purpose.

From the diagram in Fig. 3, it can be readily appreciated that, for example, a decrease in $X_o + X_L$ will cause the impedance circle to intercept the line θ_s = const at a lower point. The magnitude of Z_s, therefore, decreases and with it the ratio g_m, calling for higher amplitude oscillation. This effect of course becomes more prcnounced as the curvature of the crystal impedance circle increases either due to a larger value of R_1 or a smaller X_o. Drawing a set of constant frequency circles (8) into the diagram of Fig. 3 provides furthermore a very direct impression of the interdependence of frequency of oscillation, crystal resistance R and load reactance X_L. It is well to take note thereby of the fact that the constant frequency circles (8) are independent of R_1.

4. THE MILLER OSCILLATOR

If the crystal network takes the place of Z_2 in Fig. 1, we obtain the basic diagram of the Miller oscillator as shown in Fig. 4. In order to explain the qualitative features of this oscillator, we will assume now that Z_1 and Z_3 have been chosen initially and that we wish to determine the impedance, Z_2 in this case, which the crystal network is required to exhibit during steady-state oscillations - provided oscillations are possible.

The phase relation (5) requires now

$$\theta_s - \theta_2 = \pi + \theta_1. \tag{9}$$

Since, however, θ_2 and θ_s are of course not independent from one another, the problems of solving (5) is now slightly more complicated, though not essentially different from the previous case. In Section 3 we had to determine at first the loci of all values Z_s for which (5) is satisfied. This was done in the lefthand diagram in Fig. 3 and the curve happened to be the straight line θ_s = const. In the righthand diagram of this figure we then found the actual operating point as the intercept of this curve with the one describing those values of Z_3 which the crystal network is capable of assuming.

An identical procedure can be followed to solve (5) if Z_2 is the unknown impedance. Disregarding for the moment the physical significance of the θ's, solving the equation (9) becomes a purely geometric problem. The solution describes a circle in the impedance plane. Part of this circle is drawn into the lefthand diagram in Fig. 4. It goes through the origin of the impedance plane and through the end point of the vector $Z_1 + Z_3$. A third point on this circle is found most conveniently by assuming $\theta_2 = 90^o$. According to (5), θ_s is then given by $\theta_s = \theta_1 - \frac{\pi}{2}$ which is readily obtained from the graph. The circle is then drawn through these three points, however, only the solid portion of it corresponds to physically realizable values of Z_2 and hence Z_s. In the righthand picture of Fig. 4 the impedance diagram of the crystal unit is shown superimposed on the graph just obtained, with its origin at $Z_1 + Z_3$. Again the intercepts of these two curves identify those values of Z_s and Z_2 which satisfy the phase relation and at the same time are permitted for the crystal network. Of the two intercepts, only the upper one is of interest because it corresponds to a substantially lower value of g_m. If the latter is smaller than g_{mo}, the circuit will oscillate at the frequency at which the crystal exhibits the required impedance Z_2.

The diagram in Fig. 4 shows quite clearly that even though the impedance Z_2 is inductive, the crystal unit is still operated in its low resistive region. By no means, therefore, should the fact that the Miller oscillator is known as a parallel resonance oscillator be taken to imply that the crystal unit is operated near its anti-resonance point, i.e., in the high resistance range of its impedance circle. If the crystal is operated correctly, that is, according to specifications, the Pierce oscillator as well as the Miller oscillator will have the same output frequency which is only a different way of stating that the crystal is operated on the same point of its impedance circle in both cases.

Considerations such as were made for the Pierce oscillator concerning the effects of a change in series load reactance X_L can of course easily be translated to apply for the Miller oscillator. With slightly more effort, the graphical construction in Fig. 4 together with the equations (7) and (8) can be used to illustrate the effects of a change in X_o, such as introduced by a load element in parallel with the static capacitance of the crystal unit. X_L thereby may or may not be left to go to zero.

5. THE STABILITY RELATIONS

Principally for the purpose of demonstrating some of the features of crystal operation in an oscillator we have assumed in Sections 3 and 4 that the operating point of the crystal unit is what has to be determined while the other impedances in the circuit are known. However, one problem in designing an oscillator is frequently to assure, by proper choice of the other circuit components, that the crystal unit is operated at its specified operating point in order to obtain frequency correlation. This means that the impedances Z_1 and Z_2 in the case of the Pierce oscillator or Z_1 and Z_3 for the Miller oscillator have to be determined, together with the proper values of X_L and X_o, to obtain the desired operation.

Since the conditions for oscillation can obviously be met with a very wide range of circuit impedances, the decisive question which has to be answered is what are the optimum values.

For frequency control applications, the most pertinent criterion for the performance is undoubtedly the degree to which the desired frequency of oscillation is maintained. It can easily be appreciated from either Fig. 3 or Fig. 4 that any change in, for example, Z_1, will require either Z_2 or Z_3 or both to change if the conditions for oscillations should again be satisfied for the new value of Z_1, i.e., for $Z_1 + \Delta Z_1$.

The relationship between the various ΔZ's can be found quite generally by differentiating the condition for oscillation (3). We find

$$\frac{\Delta Z_s}{Z_s} - \frac{\Delta Z_1}{Z_1} - \frac{\Delta Z_2}{Z_2} = - \frac{\Delta g_m}{g_m} \qquad (10)$$

Because the Z's are generally functions of a number of parameters such as resistors, condensers, and inductors as well as of frequency, i.e.,

$$Z_1 = Z_1\,(\alpha_{11}, \alpha_{12}, \alpha_{13} - - - -; \omega), \qquad (11)$$

the differentials are

$$\Delta Z_1 = \sum_k \frac{\partial Z_1}{\partial \alpha_{1k}} \Delta \alpha_{1k} + \frac{\partial Z_1}{\partial \omega} \Delta \omega \tag{12}$$

with like expressions for Z_2, Z_3 and ΔZ_2, ΔZ_3 respectively.

Each one of the ΔZ's has, as a complex quantity a magnitude and a phase angle

$$\Delta Z = |\Delta Z| e^{j\theta_{\Delta Z}} \tag{13}$$

If (4) and (13) are substituted into (10), the equation can be separated into real and imaginary parts. The real part of (10) determines the changes in g_m and hence in the amplitude of oscillation. The imaginary part, however, contains the equation for $\Delta\omega$, the change in frequency. With $\Delta Z_s = \Delta Z_1 + \Delta Z_2 + \Delta Z_3$, which follows from (2a), the imaginary part of (10) becomes

$$\frac{|\Delta Z_3|}{|Z_s|} \sin(\theta_{\Delta Z_3} - \theta_s) + \frac{|\Delta Z_1|}{|Z_s|} \sin(\theta_{\Delta Z_1} - \theta_s) - \frac{|\Delta Z_1|}{|Z_1|} \sin(\theta_{\Delta Z_1} - \theta_1) + \frac{|\Delta Z_2|}{|Z_s|} \sin(\theta_{\Delta Z_2} - \theta_s) - \frac{|\Delta Z_2|}{|Z_2|} \sin(\theta_{\Delta Z_2} - \theta_2) = 0 \tag{14}$$

With (12) and (13) it is possible to compute from (14) the frequency shift $\Delta\omega/\omega$ caused by a given set of parameter changes ($\Delta\alpha$). While the equation looks formidable in its general form, it frequently reduces to quite manageable expressions when a particular circuit is being considered.

Using the diagrams of Fig 3 or Fig 4 it will be realized that each one of the terms in (14) represents a phase angle change. For example, the second term is the ratio of that component of Z_1 which is normal to Z_s to the magnitude of Z_s and hence represents the change in θ_s due to the action of ΔZ_1. Similarily, the third term is the change in θ_1 due to ΔZ_1 a.s.o. The information contained in (14) could have been derived therefore in principle by appropriately differentiating (5), however, the resulting equation would be of very limited use and would again have to be transformed into the form (14) for practical application. Nevertheless the mere possibility is of great value conceptually in interpreting the relation (14). For example it demonstrates quite clearly that the sum of all phase angle changes caused by ΔZ_1 and ΔZ_2 must be compensated by an equal and opposite change in θ_s due to the action of a ΔZ_3 of appropriate magnitude. The change in θ_3, the phase angle of the impedance Z_3, does not enter the stability relation directly.

An important aspect of (14) is the general type of design information which can be extracted from it. We can use again the Pierce oscillator to illustrate this point.

In the case of the Pierce oscillator Z_3 represents the crystal network and $\partial Z_3/\partial\omega$ will ordinarily be so large compared to $\partial Z_2/\partial\omega$ and $\partial Z_1/\partial\omega$ that the latter two can be neglected compared to the first. ΔZ_1 and ΔZ_2 therefore represent changes in Z_1 and Z_2 which are caused by variations in any one or more of the parameters α_{1k} and α_{2k} of (11). Since in this type oscillator Z_1 and Z_2 are nearly always the impedances of networks consisting of several elements in parallel, it is more convenient to replace

$|\Delta Z_1|$ by $|Z_1|^2|\Delta Y_1|$ and $|\Delta Z_2|$ by $|Z_2|^2|\Delta Y_2|$. If in addition it is assumed that the parameters of the crystal network, α_{3k} remain constant, the equation (14) becomes

$$\frac{\Delta\omega}{|Z_s|}\left|\frac{\partial Z_3}{\partial\omega}\right|\sin(\theta_{\Delta Z_3}-\theta_s)+\frac{|Z_1|^2}{|Z_s|}|\Delta Y_1|\sin(\theta_{\Delta Z_1}-\theta_s)-|Z_1||\Delta Y_1|\sin(\theta_{\Delta Z_1}-\theta_1)+$$
$$+\frac{|Z_2|^2}{|Z_s|}|\Delta Y_2|\sin(\theta_{\Delta Z_2}-\theta_s)-|Z_2||\Delta Y_2|\sin(\theta_{\Delta Z_2}-\theta_2)=0 \quad (15)$$

The first term in (15) can be transformed further by using an expression for the quality factor of the crystal network which is valid over the entire range of the crystal response. Such a formula for the crystal Q has been found to be given by

$$Q_o = \frac{\omega}{2\ \text{Real}\ Z_3}\left|\frac{\partial Z_3}{\partial\omega}\right| \quad (16)$$

With (16) and (6) the equation (15) can be written in the following form:

$$\frac{2\Delta\omega}{\omega}Q_{eff}+\frac{1}{g_m}\frac{|Z_1|}{|Z_2|}|\Delta Y_1|\sin(\theta_{\Delta Z_1}-\theta_s)-|Z_1||\Delta Y_1|\sin(\theta_{\Delta Z_1}-\theta_1)+$$
$$+\frac{1}{g_m}\frac{|Z_2|}{|Z_1|}|\Delta Y_2|\sin(\theta_{\Delta Z_2}-\theta_s)-|Z_2||\Delta Y_2|\sin(\theta_{\Delta Z_2}-\theta_2)=0 \quad (17)$$

whereby

$$Q_{eff} = Q_o\frac{\text{Real}\ Z_3}{|Z_s|}\sin(\theta_{\Delta Z_3}-\theta_s) \quad (18)$$

Evidently, a given set of $|\Delta Y_1|$ and $|\Delta Y_2|$ will cause the smallest change in frequency if Q_{eff} and g_m are as large as possible and $|Z_1|$ and $|Z_2|$ are small.

Q_{eff} is the effective quality factor of the crystal network in the oscillator circuit. It always is smaller than Q_o, the quality factor of the crystal network alone, with the factors contributing to the degradation apparent from (18). Evidently, the maximum value for Q_{eff} is Q_o and to attain it, it would be necessary that $|Z_s| = \text{Real}\ Z_3$ and $(\theta_{\Delta Z_3}-\theta_s) = 90^\circ$. From Fig. 3 it is seen that, approximately, $|Z_s| = \text{Real}\ Z_1 + \text{Real}\ Z_2 + \text{Real}\ Z_3$. Hence, the condition $|Z_s| = \text{Real}\ Z_3$ would require the impedances Z_1 and Z_2 to be purely imaginary. This fact has of course long been recognized and can frequently be approximated fairly well in vacuum tube circuits. Fig 3 also shows that the phase angle difference $(\theta_{\Delta Z_3}-\theta_s)$ will be 90° only if Z_s intersects the impedance circle of the crystal at right angle - a situation which is not attainable in a Pierce oscillator, even if the first condition is met, unless C_o cancellation is used. Since the sin function, however, is quite flat around 90°, this source of Q degradation, by itself, is frequently small, particularly at lower frequencies. A more extensive discussion of the effective quality factor will be found in Section 7.

Since Z_1, Z_2 and Z_3 are related to one another and the effective transconductance of the tube, g_m, by (6), the lower limit for $|Z_1|$ and $|Z_2|$ is set by the available g_{mo} of the active device. To obtain the highest frequency stability, therefore, requires that g_m is as close to g_{mo} as possible. This means by implication that the amplitude of the oscillations will be small unless the g_m vs amplitude curve is extremely flat. The latter of course is undesirable because of poor amplitude stability as pointed out in Section 1. As the crystal frequency is sensitive to amplitude variations there is an optimum point of operation on the g_m vs amplitude curve which will give the highest frequency stability and this applies also if the shape of this curve is modified by AGC action. Particularly in connection with AGC it is important to note that the frequency deviation $\Delta\omega/\omega$ according to (17) depends actually upon the impedance magnitudes and only because of (6) on the effective g_m of the active device. AGC should therefore be applied sparingly so as not to reduce the operating point of the active device and hence the available g_m more than necessary.

It has been assumed so far that the parameters of the crystal network α_{3k}, remain constant while the terminating impedances Z_1 and Z_2 are subject to change. According to (17) it is principally the effective quality factor of the crystal unit which determines the sensitivity of the oscillator frequency to the variations in Z_1 and Z_2. An essentially different type of condition, however, exists if $\Delta Z_1 \equiv \Delta Z_2 = 0$ is assumed and the parameters of the crystal network are varied. In this case the equation (14) reduces to the equation

$$\theta_{\Delta Z_3} - \theta_s = 0 \tag{19}$$

which states that Z_3 can only change along the line θ_s = const. This latter fact can easily be verified by using the diagram in Fig. 3. Since Z_1 and Z_2 are assumed constant, θ_s is not affected by a variation in any one of the paramenters α_{3k} and hence remains constant. A specific example of the situation existing now has already been considered in Section 3 where the effects of a variation in X_L were discussed. It can be shown that the change in frequency of oscillations in response to a variation of the parameters α_{3k} is controlled primarily by the capacitance ratio C_o/C_1 of the crystal unit and is independent to a large measure of the quality factor Q_o.

All the considerations made so far for the Pierce oscillator apply, with some modifications, to the Miller oscillator as well, however, in practical applications it will be found that the latter is more difficult to design properly and in addition its impedance level appears to be restricted to a rather unfavorable range.

6. Generalization of the Oscillator Equations.

When dealing with vacuum tube oscillators, particularly those employing pentodes, the approach taken in Section 1 might be found acceptable for most engineering applications without much hesitation. The nature of presently available transistors however leaves no choice but to examine the problem more carefully in order to determine how much of the foregoing analysis needs to be modified and amended for it to apply to transistor oscillators as well.

A perfectly general representation of a harmonic oscillator is shown in Fig. 5. It consists of two four terminal networks in a cyclic arrangement. The Network I shall contain the active element, or elements if more than one is used, while the Network II contains all the elements of the biasing network and the a-c feedback loop. The load is naturally a part of Network II.

Within the concept of equivalent linearization it is possible, using matrix theory, to completely characterize the small signal performance of the Network I by a set of four constants, the most appropriate for the configuration being the "a" parameters. The input-output relations for the Network I are thereby given by

$$\begin{aligned} V_1 &= a_{11}V_2 - a_{12}\, i_2 \\ i_1 &= a_{21}\, V_2 - a_{22}\, i_2 \end{aligned} \qquad (20)$$

It is important to emphasize that the parameters (a_{jk}) are entirely independent from the properties of the Network II. Their values do of course depend upon the biasing conditions and upon the signal level in a manner similar to that explained in Section 1 for g_m.

In input-output relations for the Network II are

$$\begin{aligned} V_3 &= a'_{11}\, V_4 - a'_{12}\, i_4 \\ i_3 &= a'_{21}\, V_4 - a'_{22}\, i_4 \end{aligned} \qquad (21)$$

Because of the cyclic arrangement of the networks shown in Fig. 5. $i_3 = -i_2$, $V_3 = V_2$, $i_1 = -i_4$ and $V_4 = V_1$ must hold during steady state oscillations. It can be verified quite readily with (20) and (21) that for these relations to be possible, it is required that

$$1 - a_{21}a'_{12} - a_{12}a'_{21} - a_{22}a'_{22} - a_{11}a'_{11} + \Delta^{a}\Delta^{a'} = 0 \qquad (22)$$

with

$$\Delta^{a} = a_{11}a_{22} - a_{12}a_{21} \qquad ; \qquad \Delta^{a'} = a'_{11}a'_{22} - a'_{12}a'_{21}$$

Equation (22) therefore expresses the conditions for oscillation in general form, subject only to the limitations imposed by the concept of equivalent linearization. It applies to vacuum tube and transistor oscillators alike.

If a feedback loop in the form of a π network is assumed, as suggested in Fig. 5, the parameters (a'_{ik}) are given by

$$(a'_{ik}) = \begin{pmatrix} 1 + \frac{Z_3}{Z_2} & Z_3 \\ \frac{Z_s}{Z_1 Z_2} & 1 + \frac{Z_3}{Z_1} \end{pmatrix} \qquad \Delta^{a'} = 1 \qquad Z_s = Z_1 + Z_2 + Z_3 \tag{23}$$

Since every linear passive network can be represented in this form, the above assumption does not restrict the generality of the following considerations.

The "a" parameters of a transistor are not widely used and it will be more convenient to express the equations in terms of the common emitter "h" parameters. The latter are less abstract in their meaning and are usually obtained by direct measurement. Again, the fact that the common emitter parameters are chosen here is only a matter of convention and does not impose any limitations on the validity of the equations. The two four terminal networks in Fig. 5 can always be defined such as to conform to this convention.

The well known relations between the "a" and "h" parameters are given by

$$(a_{ik}) = \begin{pmatrix} -\frac{\Delta^h}{h_{21}} & -\frac{h_{11}}{h_{21}} \\ -\frac{h_{22}}{h_{21}} & -\frac{1}{h_{21}} \end{pmatrix} \qquad \Delta a = -\frac{h_{12}}{h_{21}} \qquad \Delta h = h_{11} h_{22} - h_{12} h_{21} \tag{24}$$

If (23) and (24) are inserted into (22) it will be found possible, after some re-arrangement of terms but without approximations, to write the general condition for oscillations in the form

$$g_m \frac{\overline{Z}_1 \overline{Z}_2}{\overline{Z}_s} = -1 \tag{25}$$

and thus to reduce it to the basic equation discussed in the foregoing sections of this paper. The symbols in (25) are defined as follows:

$$g_m = \frac{h_{21}}{h_{11}} - h_{12} + \frac{h_{12} h_{21}}{h_{11}^2} Z_3 \tag{26}$$

$$\overline{Z}_s = \overline{Z}_1 + \overline{Z}_2 + Z_3 \tag{27}$$

$$\frac{1}{\overline{Z}_1} = \frac{1}{Z_1} + \frac{\Delta^h}{h_{11}} \tag{28}$$

$$\frac{1}{\overline{Z}_2} = \frac{1}{Z_2} + \frac{1}{h_{11}} \tag{29}$$

Clearly, if the elements of the passive feedback network in Fig.5 are suitably redefined to include the effective input and output impedances of the active network all the considerations based before on an idealized oscillator circuit remain valid, save for the modifications required by the fact that the equivalent transconductance g_m cannot ordinarily be assumed to be a real quantity.

The correction term involving Z_3 in (26) is likely to prove negligible in all but extreme cases. If it is not sufficiently small, an iteration process has to be used in solving (25).

It is observed, that according to (26), (28) and (29) the transistor in an oscillator is equivalent to a current generator of internal impedance h_{11}/Δ^h and an input impedance h_{11}. Its strength i_1 is related to the input voltage by $i_1 = g_m v_b$. Since the transistor thereby is represented by a black box, this statement applies equally well to any other type of amplifier, no matter how simple or complicated its structure. To apply, therefore, the methods of oscillator analysis developed for an ideal current generator to any actual circuit configuration it is only necessary to consider the internal impedance as well as the input impedance of the equivalent generator as part of the feedback network.

The parameters of the active device are still separated from those of the feedback network in the expressions (28) and (29). The basic advantage of dividing the oscillator into an active and a passive network as shown in Figure 5 is still retained therefore, namely, that the parameter of each one can be measured independently from the other.

The amplitude and phase relations in this general case are obtained from (25) by separating the real and imaginary parts just as (5) and (6) were found from (3). They are

$$\bar{\theta}_s - \bar{\theta}_1 - \bar{\theta}_2 - \theta_g = \pi \tag{30}$$

$$\frac{|\bar{Z}_s|}{|Z_1||Z_2|} = |g_m| < |g_{mo}| \tag{31}$$

whereby

$$g_m = |g_m| e^{j\theta_g} \tag{32}$$

Though the existence of a finite phase angle θ_g does require some minor modification, the graphical method explained previously is readily applied to the solution of (30). The effects of a finite θ_g are most clearly demonstrated qualitatively by using the corresponding diagrams. For example since θ_g is usually negative, i.e., of the same sign as θ_1 and θ_2 in a Pierce oscillator, it is noted that a θ_g of proper value can cause θ_s to become zero even if $\theta_1 + \theta_2$ is less than 180°. Under certain conditions therefore, it is possible that a finite θ_g will result in improved oscillator performance. To identify these conditions, however, becomes increasingly more difficult as the number of variables increases and their relative tendency to change in value with time and environmental conditions is considered.

7. The Generalized Stability Relations:

The fact that g_m in (25) is no longer a real quantity in the general case also requires some modifications of the stability relations derived previously in Section 5. To simplify writing we will, throughout the following, leave off the bars over the symbols in (25) and stipulate that from now on the impedances Z_1 and Z_2 are to include the internal impedance and the input impedance of the equivalent current generator respectively.

Differentiating (25) leads to the following relation between the changes in the quantities Z_1, Z_2, Z_3 and g_m

$$\frac{g_m}{\Delta g_m}\left(\frac{\Delta Z_1}{Z_s} - \frac{\Delta Z_1}{Z_1} + \frac{\Delta Z_2}{Z_s} - \frac{\Delta Z_2}{Z_2} + \frac{\Delta Z_3}{Z_s}\right) = 1 \tag{33}$$

As in (10), the imaginary part of (33) determines the changes in frequency, the real part can be used to find the concurrent change in amplitude of oscillation, provided the dependence of $|g_m|$ on amplitude is known.

The imaginary part of (33) is given by

$$\begin{aligned}&\frac{|\Delta Z_3|}{|Z_s|}\sin(\theta_{\Delta Z_3} - \theta_s - \theta_{\Delta g} + \theta_g) + \\ &\frac{|\Delta Z_1|}{|Z_s|}\left(\sin(\theta_{\Delta Z_1} - \theta_s - \theta_{\Delta g} + \theta_g) - \frac{|Z_s|}{|Z_1|}\sin(\theta_{\Delta Z_1} - \theta_1 - \theta_{\Delta g} + \theta_g)\right) + \\ &\frac{|\Delta Z_2|}{|Z_s|}\left(\sin(\theta_{\Delta Z_2} - \theta_s - \theta_{\Delta g} + \theta_g) - \frac{|Z_s|}{|Z_2|}\sin(\theta_{\Delta Z_2} - \theta_2 - \theta_{\Delta g} + \theta_g)\right) = 0\end{aligned} \tag{34}$$

Evidently, this equation reverts to (14) if the vector Δg_m has the same direction as the vector g_m, i.e., if $\theta_{\Delta g} = \theta_g$. In the general case, however, it is necessary to know both, θ_g and $\theta_{\Delta g}$ to use the stability relation (34). Δg_m, its magnitude and phase angle, can be found by differentiating (26). The change of g_m with amplitude, however, is best obtained from direct measurements. The difference vectors ΔZ_1, ΔZ_2 and ΔZ_3 are again of the form (12) whereby the α's now include the pertinent parameters of the active device.

The stability relation (34) is in a form which is immediately useful to evaluate the frequency change caused by a specific change in any one single component in a specific circuit. To attempt a discussion of the full equation in general terms, however, is pointless. The relation is sufficiently complicated and the number of variables so large that an almost limitless variety of situations can be devised with apparently no criterion available to eliminate those which are not likely to occur in practical circuits.

Yet, it is precisely the arbitrary nature of the variations that can take place, which enables us to extract generally valid design criteria from (34). This will be illustrated using again the Pierce oscillator as an example. The term Pierce oscillator is thereby used in a broad sense and applied to any oscillator in which the crystal unit is part of Z_3, even if several of such oscillators are known under different names.

If the crystal unit is part of Z_3 and all the network parameters in Z_3 are assumed constant, ΔZ_3 becomes according to (12):

$$|\Delta Z_3| = \left|\frac{\partial Z_3}{\partial \omega}\right| \Delta\omega$$

Because of the large differences in Q, the change in Z_1 and Z_2 with frequency can be neglected compared to $\partial Z_3/\partial\omega$ and therefore, the second and the third terms in (34) do not contain $\Delta\omega$. The first term, however, can under these conditions be written with (16) as

$$\frac{|\Delta Z_3|}{|Z_s|} \sin(\theta_{\Delta Z_3} - \theta_s - \theta_{\Delta g} + \theta_g) = \frac{2\Delta\omega}{\omega} Q_{eff} \tag{35}$$

whereby

$$Q_{eff} = \frac{Q_o \mathrm{Real} Z_3}{|Z_s|} \sin(\theta_{\Delta Z_3} - \theta_s - \theta_{\Delta g} + \theta_g). \tag{36}$$

Whatever the second and third terms in (34) might be, therefore, that is, whatever the change in the terminating impedances Z_1 and Z_2, the resulting change in frequency will be lowest if the effective quality factor, Qeff, is as large as possible.

The truth of this statement has of course long been recognized even if the expression (36) has apparently not been available before.

The first general objective in designing an oscillator for maximum frequency stability therefore is to reduce to a minimum all those factors which tend to degrade the Q of the crystal unit once it is incorporated into the circuit.

The second criterion derives most directly from a consideration of the frequency changes caused by a change in any one of the components in Z_3. It can be shown, as mentioned already in Section 5, that the controlling parameter here is the C_o/C_1 ratio of the crystal network. A large ratio is desired if the resulting frequency change should be small. For optimum stability therefore, the C_o/C_1 ratio should be kept as large as practical. To some extent this criterion applies to the proper choice of the crystal unit to be used in the circuit, since generally, for a given Q_o, the crystal with the higher resonance resistance will have the higher C_o/C_1 ratio. However, in some cases it might be desirable to include an impedance matching network into Z_3 to transform the crystal impedance to a more favorable level.

Because of the fundamental importance of the effective quality factor for the proper design of an oscillator it might be well to bring (36) into a more convenient form. We will use thereby the following definitions and relations.

$$Z_1 = R_1 + jX_1 = \frac{G_1}{G_1^2 + B_1^2} - j\frac{B_1}{G_1^2 + B_1^2} \qquad Z_3 = R_3 + jX_3$$

$$Z_2 = R_2 + jX_2 = \frac{G_2}{G_2^2 + B_2^2} - j\frac{B_2}{G_2^2 + B_2^2} \qquad \mathrm{Real} Z_3 = R_3 \tag{37}$$

$$|Z_s| \cos\theta_s = R_1 + R_2 + R_3 \tag{38}$$

With (37) and (31) we can write

$$\frac{R_1 + R_2}{R_3} \approx \frac{1}{|g_m|}\frac{|Z_s|}{R_3}\left(G_1\frac{|Z_1|}{|Z_2|} + G_2\frac{|Z_2|}{|Z_1|}\right) \approx \frac{1}{|g_m|}\left(G_1\frac{|Z_1|}{|Z_2|} + G_2\frac{|Z_2|}{|Z_1|}\right) \qquad (39)$$

and with it obtain, except for terms which are small of second order, the following expression for Q_{eff}

$$Q_{eff} = Q_o \frac{\cos\theta_s \sin(\theta_{\Delta Z_3} - \theta_s - \theta_{\Delta g} + \theta_g)}{1 + \frac{1}{|g_m|}\left(G_1\frac{|Z_1|}{|Z_2|} + G_2\frac{|Z_2|}{|Z_1|}\right)} \qquad (40)$$

Further transformations do not appear advisable in the general case. If G_1, G_2, R_3 and g_m are given quantities, the optimum ratio $|Z_1|/|Z_2|$ can be determined from (40), however, the process is quite cumbersome since it is usually not possible to ignore the dependence of the numerator on θ_1 and θ_2 according to (30). The form (40) appears to be quite suitable for the use of an iteration process, carried out analytically, or graphically in the impedance plane.

A substantially more convenient expression for the effective quality factor can be found, however, if, as is very often the case, it can be proven justified to assume

$$\theta_{\Delta Z_3} = \pm 90^\circ \quad ; \quad \theta_{\Delta g} = \theta_g = 0 \qquad (41)$$

The numerator in (40) then reduces to

$$\cos\theta_s \sin(\theta_{\Delta Z_3} - \theta_s - \theta_{\Delta g} + \theta_g) = \cos^2\theta_s \qquad (42)$$

which can be approximated for reasonably small angles by

$$\cos^2\theta_s \approx 1 - 2\theta_s^2 = 1 - 2(\theta_1 + \frac{\pi}{2} + \theta_2 + \frac{\pi}{2})^2 \approx 1 - 2(G_1|Z_1| + G_2|Z_2|)^2 \qquad (43)$$

Because of (31)

$$(G_1|Z_1| + G_2|Z_2|)^2 \doteq \frac{R_3}{g_m}\left(G_1^2\frac{|Z_1|}{|Z_2|} + G_2^2\frac{|Z_2|}{|Z_1|} + 2G_1G_2\right) \qquad (44)$$

so that the expression (40) can be rewritten approximately in the highly useful form:

$$D = \frac{Q_o - Q_{eff}}{Q_o} = \frac{|Z_1|}{|Z_2|}\frac{G_1}{g_m}(1 + 2R_3G_1) + \frac{|Z_2|}{|Z_1|}\frac{G_2}{g_m}(1 + 2R_3G_2) + \frac{4G_1G_2R_3}{g_m} \qquad (45)$$

Except for terms which are small of higher order, the relation (45) is valid if the assumptions (41) are met and the phase angle θ_s is small enough for (43) to hold.

Most frequently the values of G_1, G_2, g_m and R_3 are already determined once an active device and a crystal unit have been selected, leaving the ratio $|Z_1|/|Z_2|$ in (45) to be adjusted for minimum degradation. Evidently this requires

$$\frac{|Z_1|^2}{|Z_2|^2} = \frac{G_2}{G_1}\frac{1 + 2R_3G_2}{1 + 2R_3G_1} \tag{46}$$

The minimum value for (45) is then given by

$$D_{min} = \left(\frac{Q_o - Q_{eff}}{Q_o}\right)_{min} = \frac{2}{g_m}\sqrt{G_1G_2(1 + 2R_3G_1)(1 + 2R_3G_2)} + \frac{4G_1G_2R_3}{g_m} . \tag{47}$$

Among the interesting features of (47) is its dependence on R_3, the resistive component of the crystal network. As R_3 is decreased from some large value, the Q degradation becomes eventually independent of R_3 and the use of a crystal unit with a still lower R_3 will not result in a improved effective quality factor. Moreover, a crystal unit with a very low resistance has a correspondingly low C_o/C_1 ratio and its use would therefore violate one of the general design criteria established before, namely that C_o/C_1 should be as large as practical.

Evidently there is an optimum value for R_3 and even without going into a detailed analysis of the complete stability relation it can be assumed that in most cases R_3 should be just large enough so as not to contribute appreciably to the Q degradation. Usually this requirement can be met with a fairly large range of R_3 values.

A specific example will illuminate this point further. Using a 2N700 transistor one might find $G_1 = 10^{-4}$ mho, $G_2 = 10^{-2}$ mho, $g_m = 4 \times 10^{-2}$ mho The minimum Q degradation, $D_{min}(R_3)$, as a function of R_3 can then be found from (47) e.g. $D_{min}(0) = 5\%$, $D_{min}(50) = 7\%$, $D_{min}(100) = 8.5\%$, $D_{min}(150) = 11.5\%$. Hence, there is evidently little to be gained from using an R_3 below 50 Ω with this particular set of values, and even an R_3 as high as 150 Ω might still be considered acceptable in some cases.

The conditions for oscillation (5) and (6), or, in the general case (30) and (31) specify the magnitude of the product $|Z_1|\,|Z_2|$ while the requirement for minimum Q degradation leads to a value for the ratio $|Z_1|/|Z_2|$. Together, therefore, these relations are sufficient to determine the impedances Z_1 and Z_2 required by an oscillator with optimum frequency stability. Since in a large number of applications it will not be necessary to solve the general expression (40) for the optimum ratio $|Z_1|/|Z_2|$, the relation (46) becomes the most important result of this section. It permits the rapid evaluation of the optimum ratio $|Z_1|/|Z_2|$ which can then be used in (31) to find $|Z_1|$ and $|Z_2|$.

If the resulting Q degradation is low enough, the phase angles θ_1 and θ_2 will frequently be close enough to $\pi/2$ to permit the approximations $|Z_1| \doteq |X_1|$ and $|Z_2| \doteq |X_2|$ so that the conditions (31) and (46) can be further simplified to

$$\frac{|X_1|^2}{|X_2|^2} = \frac{G_2(1 + 2R_3G_2)}{G_1(1 + 2R_3G_1)} \tag{48}$$

$$g_m = \frac{R_3}{|X_1|\,|X_2|} \tag{49}$$

For example, with the same values for G_1, G_2, and g_m assumed before for a 2N700 transistor one obtains $|X_1| = 208\Omega$, $|X_2| = 12\Omega$ if the crystal resistance R_3 is 100Ω. With X_1 determined from (48) and (49) it only remains to choose X_L in Z_3 such that $X_1 + X_2 + X_L$ equals the reactance of the load capacitance specified for the crystal unit to be used and the design of the oscillator is essentially completed.

One factor, however, has not yet been considered explicitly, namely the power dissipation in the various components.

8. Drive Level and Output Power:

During steady state oscillations the voltages and currents in an oscillator circuit have all to be related to one another in a very definite manner and hence their relative magnitudes are firmly established. In fact it is precisely this relationship which leads to the conditions for oscillations, no matter how they are formulated. The absolute magnitude of the voltages and currents remain undetermined unless the amplitude dependence of the nonlinear elements is known. The left side of equation (22) for example is a function of the transistor parameters (a_{ik}), all of which depend upon the amplitude of the signal. Normally there will only be one set of values (a_{ik}), assumed at a definite value for the signal amplitude, for which the right side of (22) is zero, for any other amplitude it will be larger or smaller than zero and the conditions for steady oscillations are not satisfied. An accurate calculation of this critical amplitude is possible in principle if the parameters (a'_{ik}) of the passive network are known and the parameters (a_{ik}) of the active device are available as functions of signal amplitude. Such calculations would obviously be quite complicated and, because of the inherent sources of error, are very likely not justified even in extreme cases.

A more practical approach can be taken if the phase angles of the impedances Z_1 and Z_2 as defined by (28) and (29) are in the order of or less than 10°. It is then possible to approximately consider Z_1 and Z_2 as independent of amplitude, using nominal values for h_{11} and Δh or values measured in an impedance bridge at approximately the desired amplitude if increased accuracy is required. This leaves g_m, or $|g_m|$ respectively, as the only amplitude dependent quantity in (25) and hence in (31).

If the term with Z_3 is deleted in (26), the remaining expression for g_m represents the effective transconductance of the active device with its output short circuited and a suitable modification of any one of the methods customarily used on vacuum tubes can be employed to evaluate its magnitude and phase angle.

Using reasonable care, it is generally possible to adjust the parameters of the feedback network such that the impedance ratio $|Z_s|/|Z_1||Z_2|$ falls within 10% of a predetermined value. If therefore a curve such as indicated in Fig. 1 has been found experimentally and the g_m corresponding to the desired amplitude of oscillations determined, the amplitude actually obtained can be in error by an amount which clearly depends on the slope of this curve. Particularily at low amplitudes, a 10% error in the ordinate of this curve will generally be found intolerable and the desired amplitude has to be obtained by manual or automatic fine adjustments of the bias conditions of the active device.

With the output effectively shortcircuited during the g_m measurement, the amplitude of the input voltage becomes the critical variable for the g_m vs. amplitude curves, with the bias conditions as parameters. Since the amplitudes of the currents and voltages at the various points in the oscillator are of course quite different it was left open so far which amplitude is best chosen as a reference and it is now established that the amplitude of the input voltage to the active device is the logical choice. The amplitudes of all other voltages and currents during steady state oscillations can be derived from it, using the conditions for oscillation.

For the following consideration in this section we will again use the basic circuit diagram shown in Fig. 1, keeping in mind, however, that the impedances Z_1 and Z_2 now include respectively the effective output and input impedance of the active device. If the rms value of the input voltage $\bar{v}_g$ is known, the power, P_k, dissipated in the impedance Z_k $(k = 1, 2, 3)$ can be computed to

$$\begin{aligned} P_1 &= \frac{|Z_2 + Z_3|^2}{|Z_2|^2} G_1 \bar{v}_g^2 \\ P_2 &= G_2 \bar{v}_g^2 \\ P_3 &= \frac{R_3}{|Z_2|^2} \bar{v}_g^2 \end{aligned} \tag{50}$$

The effective value $\bar{v}_1$ of the voltage across Z_1 is given by

$$\bar{v}_1^{\,2} = \frac{|Z_2 + Z_3|^2}{|Z_2|^2} \bar{v}_g^2 \tag{51}$$

Normally the load will be part of Z_1 and the power delivered into the load can be computed from P_1 for any particular case. For the present purpose it will suffice to consider P_1 as synonymous with P_L, the power into the load. If the crystal network is Z_3 and the only resistive component in this network is due to the crystal unit, P_3 will be the power dissipated in the crystal.

Since the ratio of the power delivered to the load to the power dissipated in the crystal unit is frequently of considerable concern, we will derive now an expression for P_1/P_3 which very clearly illustrates the the influence of the various circuit parameters on this ratio.

From (50)

$$\frac{P_1}{P_3} = \frac{G_1}{R_3} |Z_2 + Z_3|^2 \tag{52}$$

In the general case, the magnitude of the vector $Z_2 + Z_3$ is best obtained from diagrams such as used for the graphical analysis of the phase relations (30), which of course will be very similar to the one shown in Fig. 3. If however, the phase angle of Z_2 is close to 90^o and the phase angle of the sum vector Z_s close to zero, we can set approximately

$$Z_2 + Z_3 \approx R_3 + j(X_2 + X_3) \approx R_3 - jX_1 \tag{53}$$

since, under these conditions $X_1 + X_2 + X_3 \approx 0$. With (53) and (49), which also holds under these conditions, the power ratio (52) can be written as

$$\frac{P_1}{P_3} = G_1 (R_3 + \frac{|X_1|}{g_m |X_2|}). \tag{54}$$

In a well designed oscillator the Q degradation should be kept at a minimum, that is, the ratio $|X_1| / |X_2|$ is given by (48). Hence,

$$\frac{P_1}{P_3} = G_1 R_3 + \frac{1}{g_m} \sqrt{G_1 G_2 \frac{1 + 2R_3 G_2}{1 + 2R_3 G_1}} \tag{55}$$

is the expression for the power ratio in a Pierce type oscillator whose components are adjusted for a relative maximum in effective Q. G_1 and G_2 are the conductances of the terminating networks Z_1 and Z_2 respectively and R_3 is the resistive component of the crystal network.

Using again the approximations which lead from (52) to (55), one obtains in a similar manner the following expressions:

$$\frac{P_2}{P_3} = \frac{G_2}{g_m} \frac{|X_2|}{|X_1|} = \frac{1}{g_m} \sqrt{G_1 G_2 \frac{1 + 2R_3 G_1}{1 + 2R_3 G_2}} \tag{56}$$

$$P_3 = \bar{v}_g^2 \, g_m \frac{|X_1|}{|X_2|} = \bar{v}_g^2 \, g_m \sqrt{\frac{G_2 (1 + 2R_3 G_2)}{G_1 (1 + 2R_3 G_1)}} \tag{57}$$

$$\bar{v}_1^2 = P_3 (R_3 + \frac{|X_1|}{g_m |X_2|}) = \bar{v}_g^2 \, g_m R_3 \sqrt{\frac{G_2 (1 + 2R_3 G_2)}{G_1 (1 + 2R_3 G_1)}} + \bar{v}_g^2 \frac{G_2 (1 + 2R_3 G_2)}{G_1 (1 + 2R_3 G_1)} \tag{58}$$

In the preceeding section we had chosen the following parameters as representative of a 2N700 transistor circuit $G_1 = 10^{-4}$ mho, $G_2 = 10^{-2}$ mho $g_m = 4 \times 10^{-2}$ mho, $R_3 = 100\,\Omega$. Inserting these values into the relations (55) through (58) we find $P_1/P_3 = 5.3 \times 10^{-2}$ $P_2/P_3 = 1.4 \times 10^{-2}$, $P_3 = .69\ \bar{v}_g^2$ $\bar{v}_1 = 19\ \bar{v}_g$. Hence, if the drive level of the crystal unit is $P_3 = 10^{-6}$ watts it follows that $P_1 = 5.3 \times 10^{-8}$ watts, $P_2 = 1.4 \times 10^{-8}$ watts, $\bar{v}_g = 1.2 \times 10^{-3}$ V and $\bar{v}_1 = 2.3 \times 10^{-2}$ V. The corresponding Q degradation has previously been determined to D = 8.5%.

It is noted that even if the major part of the conductance G_1 is the load, the output power from a Pierce type oscillator is only a small fraction of the power dissipated in the crystal unit and any attempt to increase this fraction will result in increased Q degradation, even under optimum conditions. As is well known, large output power and high stability are contradictory requirements. Nevertheless, the relations (55) through (58) together with (47) will be found useful to reach an acceptable compromise for a given application.

It is noted further that for low crystal drive the voltages $\bar{v}_1$ and $\bar{v}_g$ are at an extremely low level. If the biasing voltages are adjusted very carefully, steady state operation of the oscillator can be achieved at these levels even without AGC. The amplitude stability, however is generally not adequate under these conditions and some form of artificial level control is nearly always required.

III. CONCLUSIONS:

In the foregoing discussions we have illustrated an approach to the design of oscillators which circumvents to a large measure the extreme complexity of the analytical expressions usually encountered. By representing for the most part the network impedances in polar form it was also possible to avoid approximations in the early stages of the development and thus to demonstrate the interrelations between the various elements in general terms.

It is an axiom of any analytic treatment that the content of an equation describing a physical phenomenon cannot be changed by any amount of transformation. The interpretation and solution of the equation however, very often depends on one's ability and good fortune to bring it into a form which in itself suggests already the solution. This of course is true in the present analysis as well and a case in point is the phase relation(5). The structure of this relation as shown is such as to actually demand a graphical method for it a solution. A measure of the difficulties involved in extracting the salient features of this equation by any other means are the complexities encountered if one attempts to transform (5) into the condition for oscillation as it results from representing the impedances Z_1, Z_2, and Z_3 in rectangular coordinates initially.

After choosing the basic oscillator diagram as shown in Fig. 1 and writing the conditions for oscillation in the form (3), the rest of the analysis of the idealized oscillator follows almost inevitably once the impedances are expressed in polar form. In particular, the graphical interpretation of the phase relation and further on the stability relations are direct consequences of the initial approach.

The equation (25), however, which states in effect that any arbitrary oscillator can be treated rigorously by the method developed for an idealized oscillator, is again an example of the tremendous simplification which can be achieved if sufficient effort is expended into transforming the initially obtained equations until they are most amenable to interpretation and solution.

Throughout the main body of this paper we have restricted the detailed discussions principally to those oscillators which have the crystal unit in the top section of the π network. Although this covers a fairly large class of circuits, it is admittedly a configuration which is most easily analyzed by the present technique. The more important circuits in this category are the Pierce oscillator and the CI meter circuit, as well as the transformer coupled oscillator and their various modifications.

Common to all of these circuits are three and only three essential nodes in the feedback network.

More complicated feedback circuits can, as linear passive networks, always be transferred into equivalent Π networks, but the equivalent impedances are generally not independent from one another and are often physically not realizable. While the conditions for oscillation can therefore be expressed formally through the same equations which applied for the three node circuits, the interpretation is frequently so difficult that there is no practical advantage to be derived from it. Rather, such circuits are best treated by deriving the "a" parameters of the feed back network in terms of the impedances Z_k as they actually appear between the various nodes. If the active element is then assumed at first to be an ideal current generator, the condition for oscillation involves only the parameter a_{21}, i.e. the transfer admittance of the network and the resulting equation will have the smallest number of terms possible with this type oscillator. By proper grouping of the terms an expression which resembles (25) can often be obtained and the developments of the three node oscillators can be used as a guide in the analysis.

For transistor oscillators, the approximation of the active device by an ideal current generator is rarely adequate and additional complexities have to be expected, even if only h_{11} and h_{21} are considered, i.e., if $h_{12} = h_{22} = 0$ is assumed. Nevertheless, for example in the case of a "bridged T" oscillator, very simple and surprisingly accurate relations have been obtained by expanding the basic approach.

The cardinal rule here, as in treating any oscillator, is to represent the impedances between the network nodes in general terms and to introduce their resistive and reactive components only if it can no longer be avoided and only after a clear qualitiative picture of their respective roles has been obtained. There is absolutely no need to use the R's, L's and C's of the actual network elements during the analysis. Within the frame work of the equivalent linearization, these elements serve no other purpose in the oscillator than to physically realize the resistive and reactive components of the respective impedances and their values can be determined accordingly at the very end of the analysis.

This of course, is also the principle which has been followed throughout the developments of this paper. The graphical solution of the phase relation provides a very clear interpretation of the role of the three impedances involved. While it may be desirable or even necessary to use these graphs for the quantitative analysis in extreme cases, it is principally their qualitative features which have been important in the later sections of the paper by providing guide lines to acceptable approximations.

The stability relations are still treated in general terms and only after the expression for the effective quality factor was obtained in Section 7 did we introduce the resistive and reactive components of the circuit impedance - or the corresponding conductance and susceptance if this was more convenient. Even if it was necessary in the numerical example to choose likely values for the circuit losses, there was no need at any point in the development to stipulate how the reactive components are to be realized physically.

With a given set of circuit losses, the analysis of an oscillator in the general Pierce Configuration results in specific values for $|X_1|$ and $|X_2|$ for which oscillations occur with a minimum on Q degradation. The signs of these reactances are determined by the phase relation. They must apparently be either both negative or both positive if the active element is a 180° phase shift device.

A number of factors have to be considered in determining how the impedances Z_1 and Z_2 thus specified are to be realized physically. First, they have to provide a d.c. path for biasing and one of them, generally Z_1 has to include the loading effect due to the amplifier stage following the oscillator. Second, their frequency dependence must be such as to prevent oscillations at any but the desired frequency.

The preferred way of biasing a transistor is through a large resistance in the emitter path and very small resistances in the base and collector paths[13,14]. Hence if Z_1 and Z_2 consist each of a parallel L-C combination, the inductors provide very low resistance d.c. paths for biasing, besides, incidentally, improving the noise performance of the oscillator. If overtone crystals are used, one of these L-C combinations must be resonant above the frequency of the next lower overtone so that its reactance at the frequency of this overtone is already positive while the reactance of the other L-C combination must still be negative. The other L-C combination must have its resonance frequency below the fundamental mode of the crystal to prevent oscillations at any one of the overtones below the desired one.

At the frequency of the desired overtone both L-C combinations will then be capacitive. The actual values of the L's and C's necessary to meet all these requirements will generally be found to be in a reasonable range and to meet the additional requirement that no oscillations be possible at frequencies below the lower of the two resonances, where both reactances are positive.

The condensor bypassing the emitter resistance must be chosen as large as possible but small enough for the corresponding R-C time constant to be smaller than the start up time of the oscillator to help prevent squegging[15]. The latter is more likely to occur with biasing through inductors than through resistors, however, it can readily be avoided in most cases. Nevertheless, even if it does respond frequently to simple corrective measures, the phenomenon of squegging shows a number of features which seem to defy a simple explanation. A detailed study of its causes and the conditions for its occurrence may well reveal eventually some very fundamental aspects of the behaviour of nonlinear active devices.

IV. ACKNOWLEDGEMENTS:

It is a pleasure to gratefully acknowledge the enthusiastic assistance of Mr. G. Davidson of USAELRDL, who has carried out a great many experiments and measurements on oscillators which have been indispensable for the development of the present theory. Mr. W. H. Thorn has constructed most of the circuits with painstaking care.

V. REFERENCES:

1. R. A. Heising: "Quartz Crystals for Electrical Circuits" Van Nostrand 1946.

2. W. A. Edson: "Vacuum Tube Oscillators" John Wiley 1953

3. W. Herzog: "Oszillatoren mit Schwingkristallen" Springer 1958

4. H. J. Reich: "Functional Circuits and Oscillators" Van Nostrand 1961

5. B. J. Dasher & S. N. Witt Jr: "Transistor Oscillators of Extended Frequency Range" Final Report, Signal Corps Contract DA36-039-sc-42712. Engineering Experiment Station, Georgia Institute of Technology, June 1955

6. H. E. Gruen & A. O. Plait: "A Study of Crystal Oscillator Circuits", Final Report, Signal Corps Contract DA36-039-sc-64609, Armour Research Foundation, August 1957

7. W. L. Smith, " An Ultra-Precise Standard of Frequency" Final Report, Signal Corps Contract DA36-039-sc-73078, Bell Telephone Laboratories, December 1960.

8. H. R. Meadows & D. Firth: "Quartz Crystal Oscillator Circuits" Final Report, Signal Corps Contract DA36-039-sc-88892, The Magnavox Company; December 1962.

9. B. J. Dasher & S. N. Witt op cit

10. H. J. Reich op.cit.

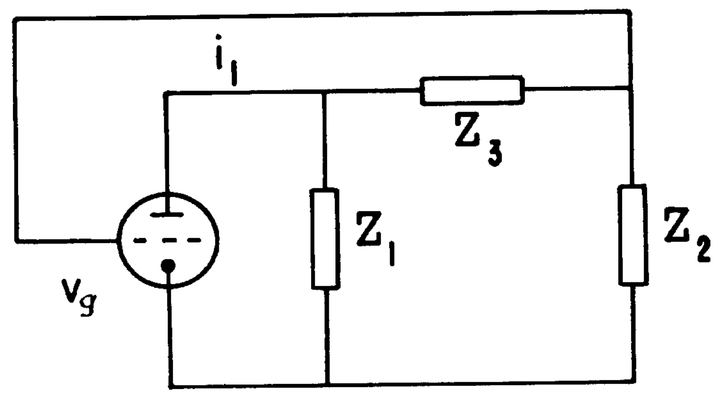
11. H. J. Reich op. cit.

12. W. G. Cady, "Piezoelectricity" McGraw Hill 1946

13. R. F. Shea Edit. "Transistor Circuit Engineering" John Wiley 1957

14. W. W. Gärtner: "Transistors: Principles, Design and Applications" Van Nostrand 1960.

15. H. J. Reich "Amplitude Instability in Oscillators", Proc IRE (Correspondence) Vol 51 P 381 Feb 1963.



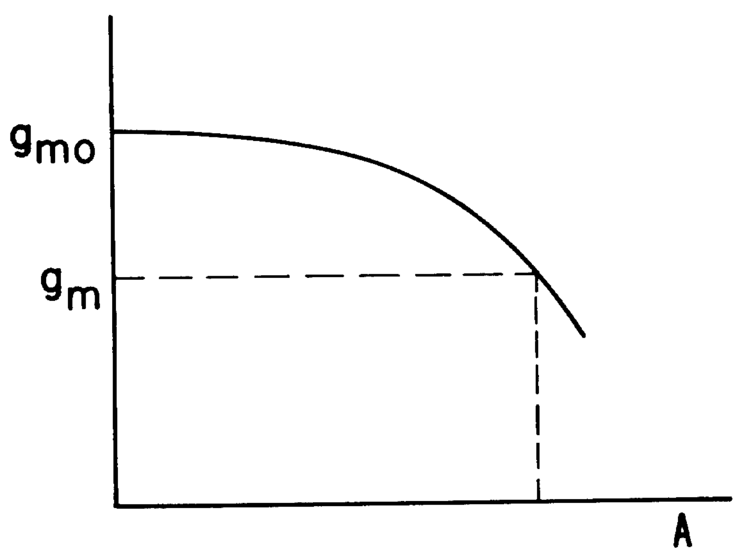
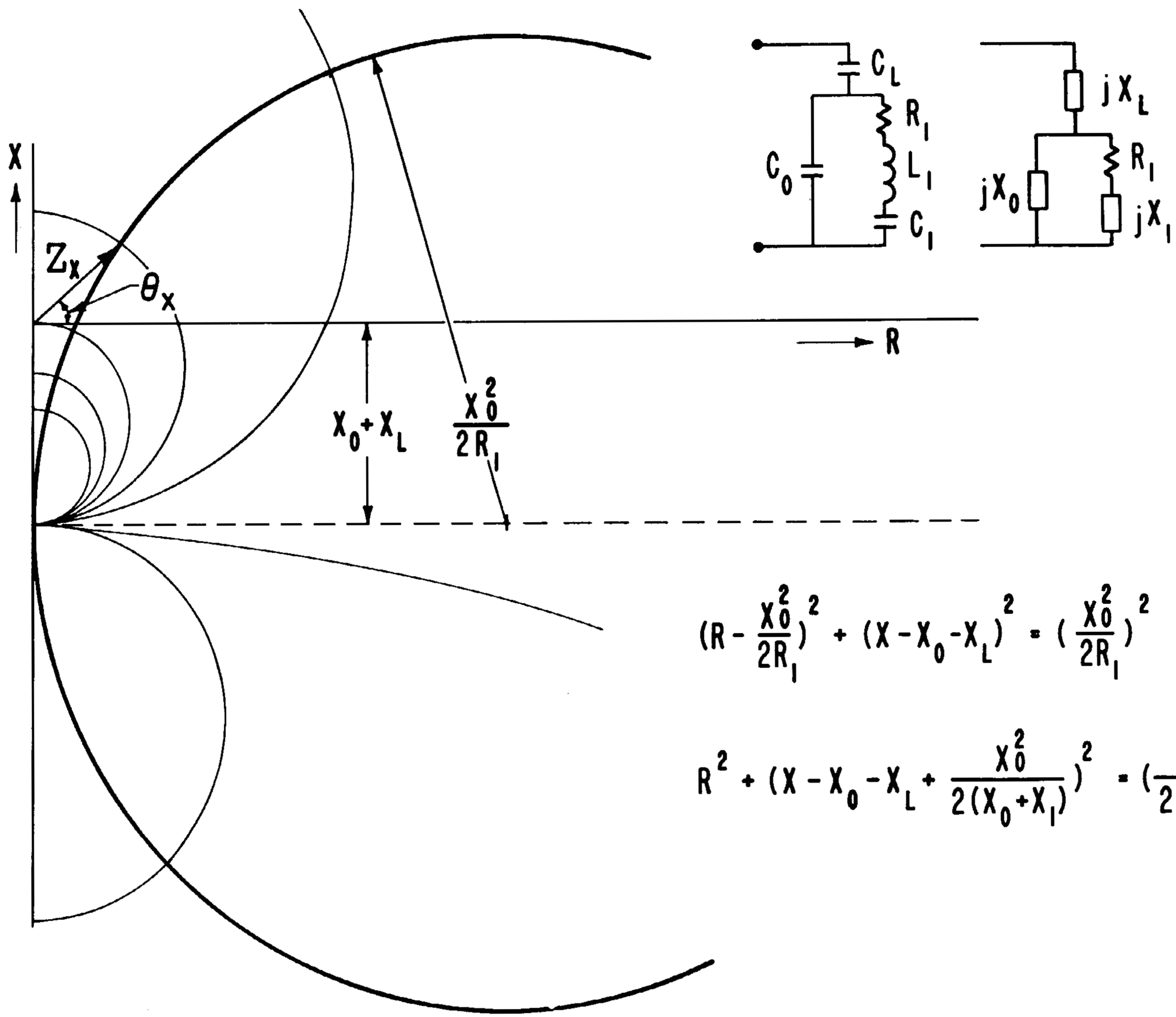


FIGURE 1



$$\left(R - \frac{X_0^2}{2R_1}\right)^2 + (X - X_0 - X_L)^2 = \left(\frac{X_0^2}{2R_1}\right)^2$$

$$R^2 + \left(X - X_0 - X_L + \frac{X_0^2}{2(X_0 + X_1)}\right)^2 = \left(\frac{X_0^2}{2(X_0 + X_1)}\right)^2$$

FIGURE 2

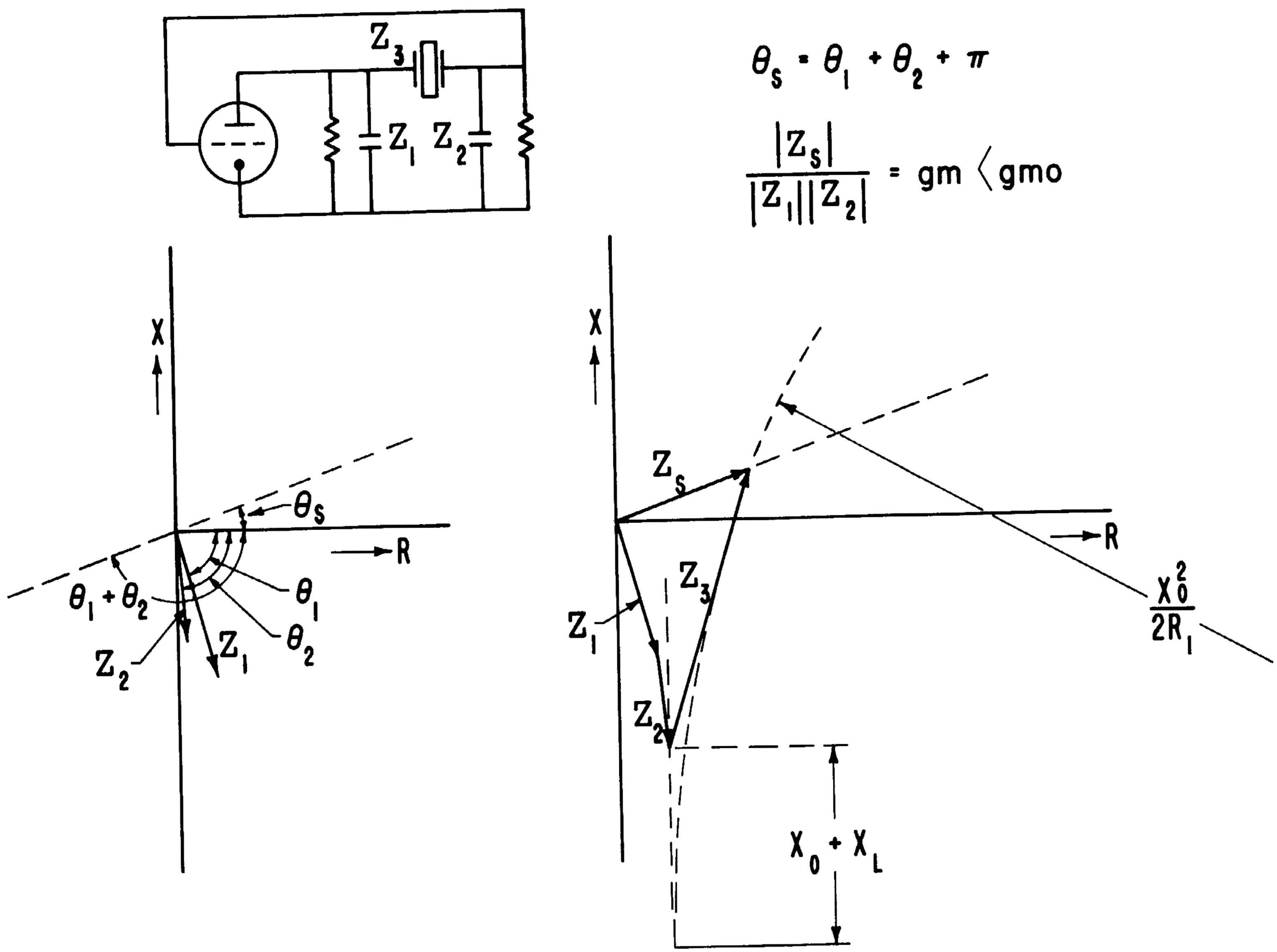

Z3
Z1
Z2
θs = θ1 + θ2 + π
|Zs| / |Z1||Z2| = gm < gmo
X
R
θs
θ1 + θ2
θ1
θ2
Z1
Z2
X
R
Zs
Z3
Z1
Z2
X0^2 / 2R1
X0 + XL

FIGURE 3

$$\theta_s - \theta_2 = \pi + \theta_1$$

$$\frac{|Z_s|}{|Z_1||Z_2|} = gm \langle gmo$$

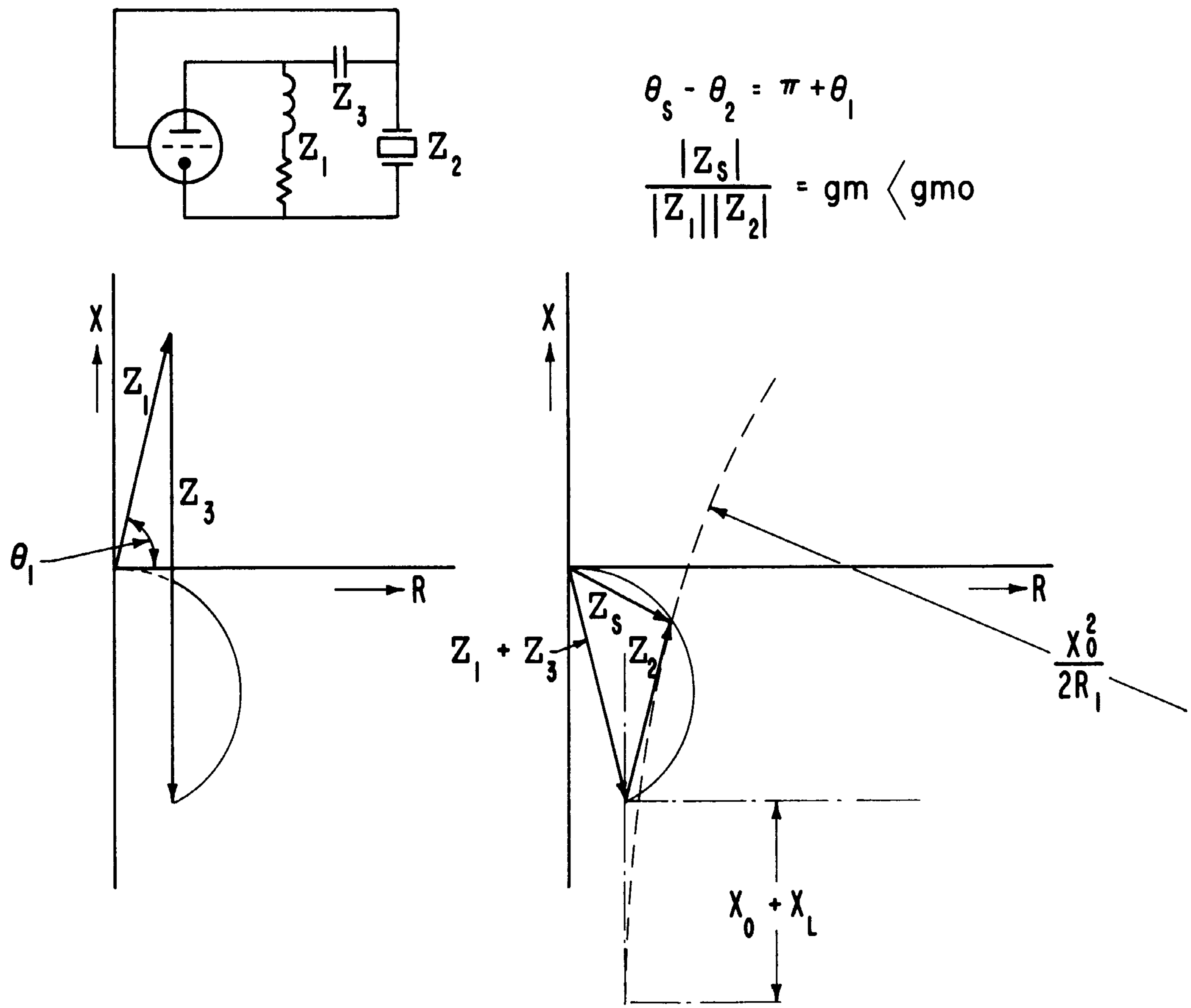
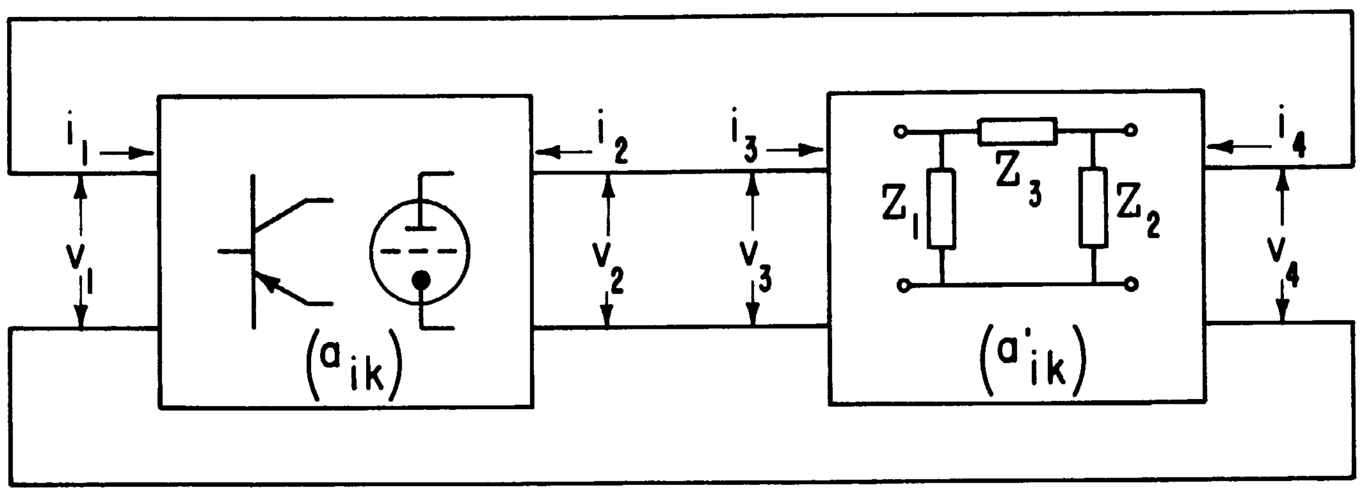


FIGURE 4



$$i_1 = -i_4$$

$$i_3 = -i_2$$

$$v_1 = v_4$$

$$v_3 = v_2$$

FIGURE 5