Relations Between Attenuation and Phase in Feedback Amplifier Design

By H. W. BODE

INTRODUCTION

THE engineer who embarks upon the design of a feedback amplifier must be a creature of mixed emotions. On the one hand, he can rejoice in the improvements in the characteristics of the structure which feedback promises to secure him.[1] On the other hand, he knows that unless he can finally adjust the phase and attenuation characteristics around the feedback loop so the amplifier will not spontaneously burst into uncontrollable singing, none of these advantages can actually be realized. The emotional situation is much like that of an impecunious young man who has impetuously invited the lady of his heart to see a play, unmindful, for the moment, of the limitations of the $2.65 in his pockets. The rapturous comments of the girl on the way to the theater would be very pleasant if they were not shadowed by his private speculation about the cost of the tickets.

In many designs, particularly those requiring only moderate amounts of feedback, the bogy of instability turns out not to be serious after all. In others, however, the situation is like that of the young man who has just arrived at the box office and finds that his worst fears are realized. But the young man at least knows where he stands. The engineer's experience is more tantalizing. In typical designs the loop characteristic is always satisfactory—except for one little point. When the engineer changes the circuit to correct that point, however, difficulties appear somewhere else, and so on ad infinitum. The solution is always just around the corner.

Although the engineer absorbed in chasing this rainbow may not realize it, such an experience is almost as strong an indication of the existence of some fundamental physical limitation as the census which the young man takes of his pockets. It reminds one of the experience of the inventor of a perpetual motion machine. The perpetual motion machine, likewise, always works—except for one little factor. Evidently, this sort of frustration and lost motion is inevitable in

---

[1] A general acquaintance with feedback circuits and the uses of feedback is assumed in this paper. As a broad reference, see H. S. Black, "Stabilized Feedback Amplifiers," *B. S. T. J.*, January, 1934.

feedback amplifier design as long as the problem is attacked blindly. To avoid it, we must have some way of determining in advance when we are either attempting something which is beyond our resources, like the young man on the way to the theater, or something which is literally impossible, like the perpetual motion enthusiast.

This paper is written to call attention to several simple relations between the gain around an amplifier loop, and the phase change around the loop, which impose limits to what can and cannot be done in a feedback design. The relations are mathematical laws, which in their sphere have the same inviolable character as the physical law which forbids the building of a perpetual motion machine. They show that the attempt to build amplifiers with certain types of loop characteristics *must* fail. They permit other types of characteristic, but only at the cost of certain consequences which can be calculated. In particular, they show that the loop gain cannot be reduced too abruptly outside the frequency range which is to be transmitted if we wish to secure an unconditionally stable amplifier. It is necessary to allow at least a certain minimum interval before the loop gain can be reduced to zero.

The question of the rate at which the loop gain is reduced is an important one, because it measures the actual magnitude of the problem confronting both the designer and the manufacturer of the feedback structure. Until the loop gain is zero, the amplifier will sing unless the loop phase shift is of a prescribed type. The cutoff interval as well as the useful transmission band is therefore a region in which the characteristics of the apparatus must be controlled. The interval represents, in engineering terms, the price of the ticket.

The price turns out to be surprisingly high. It can be minimized by accepting an amplifier which is only conditionally stable.[2] For the customary absolutely stable amplifier, with ordinary margins against singing, however, the price in terms of cutoff interval is roughly one octave for each ten db of feedback in the useful band. In practice, an additional allowance of an octave or so, which can perhaps be regarded as the tip to the hat check girl, must be made to insure that the amplifier, having once cut off, will stay put. Thus in an amplifier with 30 db feedback, the frequency interval over which effective control of the loop transmission characteristics is necessary is at least four octaves, or sixteen times, broader than the useful band. If we raise the feedback to 60 db, the effective range must be more than a hundred times the useful range. If the useful band is itself large these factors

---

[2] Definitions of conditionally and unconditionally stable amplifiers are given on page 432.

may lead to enormous effective ranges. For example, in a 4 megacycle amplifier they indicate an effective range of about 60 megacycles for 30 db feedback, or of more than 400 megacycles if the feedback is 60 db.

The general engineering implications of this result are obvious. It evidently places a burden upon the designer far in excess of that which one might anticipate from a consideration of the useful band alone. In fact, if the required total range exceeds the band over which effective control of the amplifier loop characteristics is physically possible, because of parasitic effects, he is helpless. Like the young man, he simply can't pay for his ticket. The manufacturer, who must construct and test the apparatus to realize a prescribed characteristic over such wide bands, has perhaps a still more difficult problem. Unfortunately, the situation appears to be an inevitable one. The mathematical laws are inexorable.

Aside from sounding this warning, the relations between loop gain and loop phase can also be used to establish a definite method of design. The method depends upon the development of overall loop characteristics which give the optimum result, in a certain sense, consistent with the general laws. This reduces actual design procedure to the simulation of these characteristics by processes which are essentially equivalent to routine equalizer design. The laws may also be used to show how the characteristics should be modified when the cutoff interval approaches the limiting band width established by the parasitic elements of the circuit, and to determine how the maximum realizable feedback in any given situation can be calculated. These methods are developed at some length in the writer's U. S. Patent No. 2,123,178 and are explained in somewhat briefer terms here.

RELATIONS BETWEEN ATTENUATION AND PHASE IN PHYSICAL NETWORKS [3]

The amplifier design theory advanced here depends upon a study of the transmission around the feedback loop in terms of a number of general laws relating the attenuation and phase characteristics of physical networks. In attacking this problem an immediate difficulty presents itself. It is apparent that no entirely definite and universal

[3] Network literature includes a long list of relations between attenuation and phase discovered by a variety of authors. They are derived typically from a Fourier analysis of the transient response of assumed structures and are frequently ambiguous, because of failure to recognize the minimum phase shift condition. No attempt is made to review this work here, although special mention should be made of Y. W. Lee's paper in the *Journal for Mathematics and Physics* for June, 1932. The proof of the relations given in the present paper depends upon a contour integration in the complex frequency plane and can be understood from the disclosure in the patent referred to previously.

relation between the attenuation and the phase shift of a physical structure can exist. For example, we can always change the phase shift of a circuit without affecting its loss by adding either an ideal transmission line or an all-pass section. Any attenuation characteristic can thus correspond to a vast variety of phase characteristics.

For the purposes of amplifier design this ambiguity is fortunately unimportant. While no unique relation between attenuation and phase can be stated for a general circuit, a unique relation does exist between any given loss characteristic and the *minimum* phase shift which must be associated with it. In other words, we can always add a line or all-pass network to the circuit but we can never subtract such a structure, unless, of course, it happens to be part of the circuit originally. If the circuit includes no surplus lines or all-pass sections, it will have at every frequency the least phase shift (algebraically) which can be obtained from any physical structure having the given attenuation characteristic. The least condition, since it is the most favorable one, is, of course, of particular interest in feedback amplifier design.

For the sake of precision it may be desirable to restate the situations in which this minimum condition fails to occur. The first situation is found when the circuit includes an all-pass network either as an individual structure or as a portion of a network which can be replaced by an all-pass section in combination with some other physical structure.[4] The second situation is found when the circuit includes a transmission line. The third situation occurs when the frequency is so high that the tubes, network elements and wiring cannot be considered to obey a lumped constant analysis. This situation may be found, for example, at frequencies for which the transit time of the tubes is important or for which the distance around the feedback loop is an appreciable part of a wave-length. The third situation is, in many respects, substantially the same as the second, but it is mentioned separately here as a matter of emphasis. Since the effective band of a feedback amplifier is much greater than its useful band, as the introduction pointed out, the considerations it reflects may be worth taking into account even when they would be trivial in the useful band alone.

It will be assumed here that none of these exceptional situations is found. For the minimum phase condition, then, it is possible to derive

[4] Analytically this condition can be stated as follows: Let it be supposed that the transmission takes place between mesh 1 and mesh 2. The circuit will include an all-pass network, explicit or concealed, if any of the roots of the minor Δ_{12} of the principal circuit determinant lie below the real axis in the complex frequency plane. This can happen in bridge configurations, but not in series-shunt configurations, so that all ladder networks are automatically of minimum phase type.

a large number of relations between the attenuation and phase characteristics of a physical network. One of the simplest is

$$\int_{-\infty}^{\infty} B\,du = \frac{\pi}{2}(A_\infty - A_0),\tag{1}$$

where u represents $\log f/f_0$, f_0 being an arbitrary reference frequency, B is the phase shift in radians, and A_0 and A_∞ are the attenuations in nepers at zero and infinite frequency, respectively. The theorem states, in effect, that the total area under the phase characteristic plotted on a logarithmic frequency scale depends only upon the difference between the attenuations at zero and infinite frequency, and not upon the course of the attenuation between these limits. Nor does it depend upon the physical configuration of the network unless a non-minimum phase structure is chosen, in which case the area is necessarily increased. The equality of phase areas for attenuation characteristics of different types is illustrated by the sketches of Fig. 1.
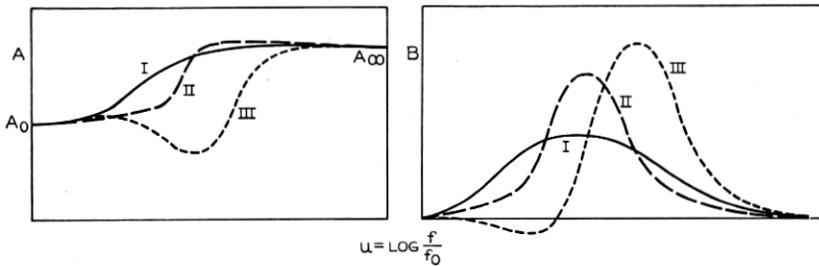

Fig. 1—Diagram to illustrate relation between phase area and change in attenuation.

The significance of the phase area relation for feedback amplifier design can be understood by supposing that the practical transmission range of the amplifier extends from zero to some given finite frequency. The quantity $A_0 - A_\infty$ can then be identified with the change in gain around the feedback loop required to secure a cut-off. Associated with it must be a certain definite phase area. If we suppose that the maximum phase shift at any frequency is limited to some rather low value the total area must be spread out over a proportionately broad interval on the frequency scale. This must correspond roughly to the cut-off region, although the possibility that some of the area may be found above or below the cut-off range prevents us from determining the necessary interval with precision.

A more detailed statement of the relationship between phase shift and change in attenuation can be obtained by turning to a second

theorem. It reads as follows:

$$B(f_c) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{dA}{du} \log \coth \frac{|u|}{2} du, \tag{2}$$

where $B(f_c)$ represents the phase shift at any arbitrarily chosen frequency f_c and $u = \log f/f_c$. This equation, like (1), holds only for the minimum phase shift case.

Although equation (2) is somewhat more complicated than its predecessor, it lends itself to an equally simple physical interpretation. It is clear, to begin with, that the equation implies broadly that the
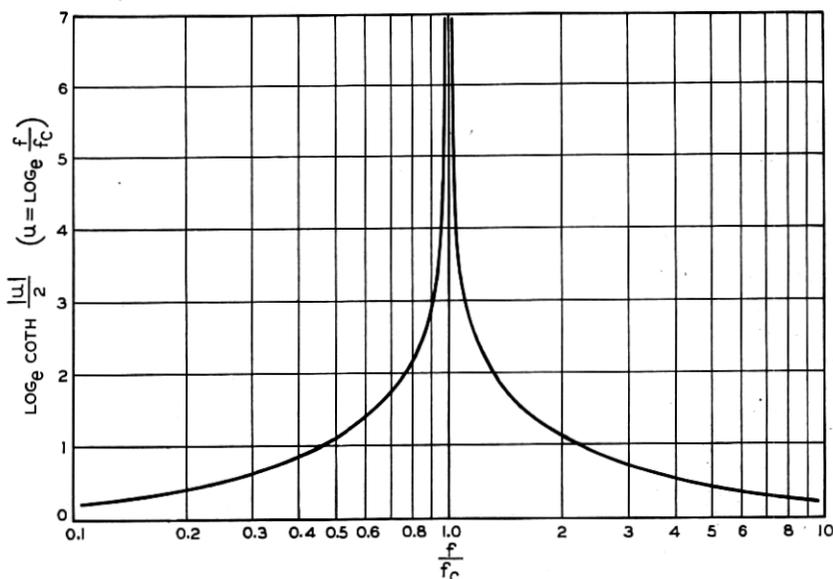
Fig. 2—Weighting function in loss-phase formula.

phase shift at any frequency is proportional to the derivative of the attenuation on a logarithmic frequency scale. For example, if dA/du is doubled B will also be doubled. The phase shift at any particular frequency, however, does not depend upon the derivative of attenuation at that frequency alone, but upon the derivative at all frequencies, since it involves a summing up, or integration, of contributions from the complete frequency spectrum. Finally, we notice that the contributions to the total phase shift from the various portions of the frequency spectrum do not add up equally, but rather in accordance with the function $\log \coth |u|/2$. This quantity, therefore, acts as a weighting function. It is plotted in Fig. 2. As we might expect physically

it is much larger near the point $u = 0$ than it is in other regions. We can, therefore, conclude that while the derivative of attenuation at all frequencies enters into the phase shift at any particular frequency $f = f_c$ the derivative in the neighborhood of f_c is relatively much more important than the derivative in remote parts of the spectrum.

As an illustration of (2), let it be supposed that $A = ku$, which corresponds to an attenuation having a constant slope of 6 k db per octave. The associated phase shift is easily evaluated. It turns out, as we might expect, to be constant, and is equal numerically to $k\pi/2$ radians. This is illustrated by Fig. 3. As a second example, we may consider
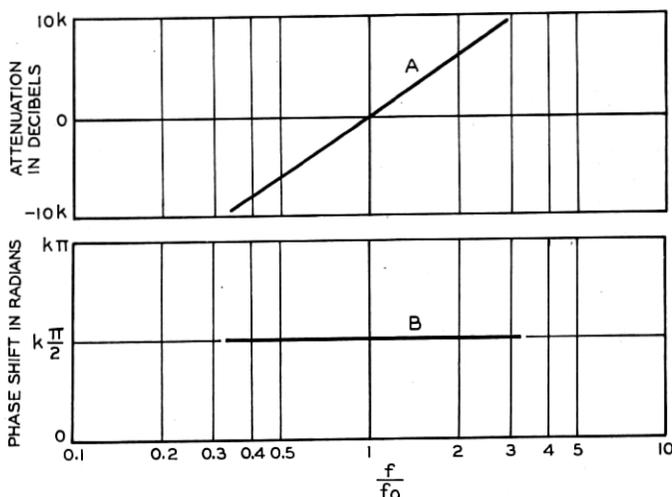
Fig. 3—Phase characteristic corresponding to a constant slope attenuation.

a discontinuous attenuation characteristic such as that shown in Fig. 4. The associated phase characteristic, also shown in Fig. 4, is proportional to the weighting function of Fig. 2.

The final example is shown by Fig. 5. It consists of an attenuation characteristic which is constant below a specified frequency f_b and has a constant slope of 6 k db per octave above f_b. The associated phase characteristic is symmetrical about the transition point between the two ranges. At sufficiently high frequencies, the phase shift approaches the limiting $k\pi/2$ radians which would be realized if the constant slope were maintained over the complete spectrum. At low frequencies the phase shift is substantially proportional to frequency and is given by the equation

$$B = \frac{2k}{\pi} \frac{f}{f_b} \cdot \qquad (3)$$

Solutions developed in this way can be added together, since it is apparent from the general relation upon which they are based that the phase characteristic corresponding to the sum of two attenuation
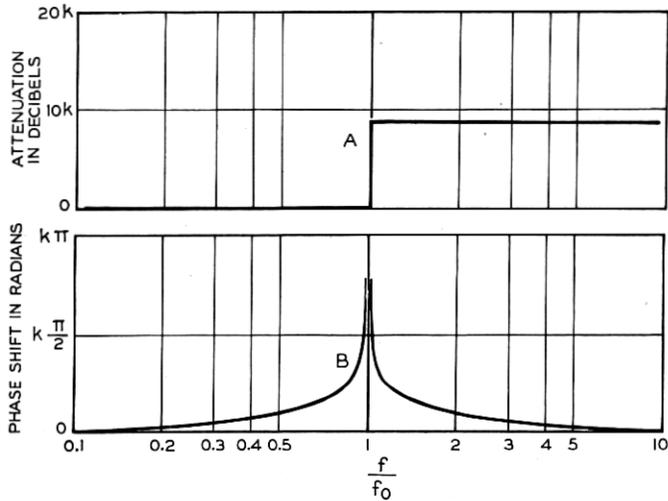

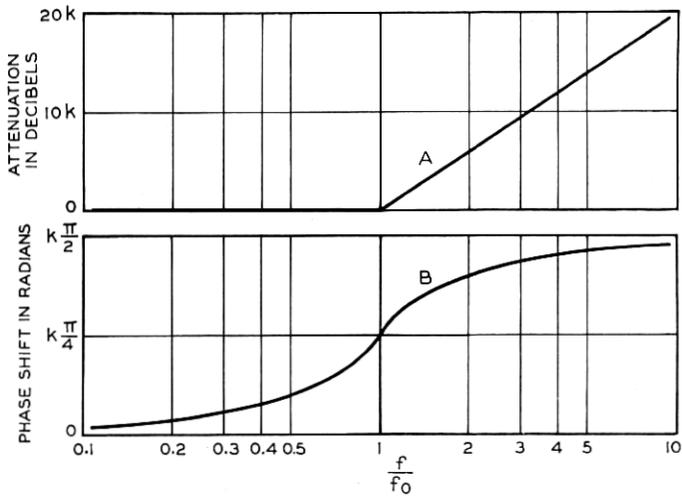Fig. 4—Phase characteristic corresponding to a discontinuity in attenuation.

Fig. 5—Phase characteristic corresponding to an attenuation which is constant below a prescribed frequency and has a constant slope above it.

characteristics will be equal to the sum of the phase characteristics corresponding to the two attenuation characteristics separately. We can therefore combine elementary solutions to secure more complicated

characteristics. An example is furnished by Fig. 6, which is built up from three solutions of the type shown by Fig. 5. By proceeding sufficiently far in this way, an approximate computation of the phase characteristic associated with almost any attenuation characteristic can be made, without the labor of actually performing the integration in (2).
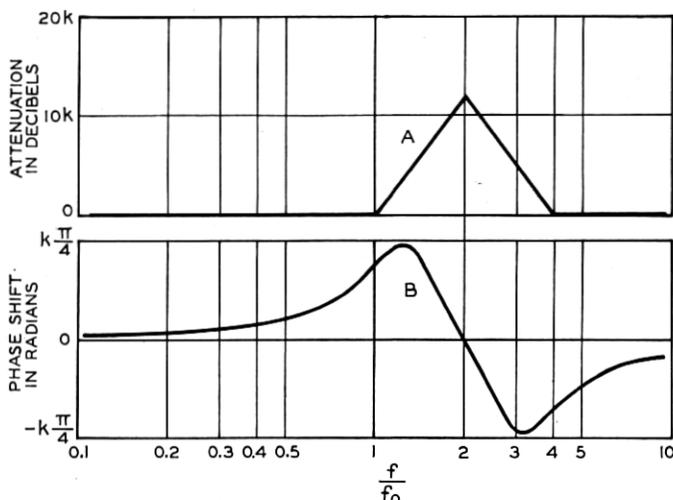

Fig. 6—Diagram to illustrate addition of elementary attenuation and phase characteristics to produce more elaborate solutions of the loss-phase formula.

Equations (1) and (2) are the most satisfactory expressions to use in studying the relation between loss and phase in a broad physical sense. The mechanics of constructing detailed loop cut-off characteristics, however, are simplified by the inclusion of one other, somewhat more complicated, formula. It appears as

$$\int_0^{f_0} \frac{A\,df}{\sqrt{f_0^2 - f^2}\,(f^2 - f_c^2)} + \int_{f_0}^{\infty} \frac{B\,df}{\sqrt{f^2 - f_0^2}\,(f^2 - f_c^2)}$$

$$= \frac{\pi}{2} \frac{B(f_c)}{f_c\sqrt{f_0^2 - f_c^2}}, \qquad f_c < f_0$$

$$= -\frac{\pi}{2} \frac{A(f_c)}{f_c\sqrt{f_c^2 - f_0^2}}, \qquad f_c > f_0, \quad (4)$$

where f_0 is some arbitrarily chosen frequency and the other symbols have their previous significance.

The meaning of (4) can be understood if it is recalled that (2) implies that the minimum phase shift at any frequency can be computed if the

attenuation is prescribed at all frequencies. In the same way (4) shows how the complete attenuation and phase characteristics can be determined if we begin by prescribing the attenuation below f_0 and the phase shift above f_0. Since f_0 can be chosen arbitrarily large or small this is evidently a more general formula than either (1) or (2), while it can itself be generalized, by the introduction of additional irrational factors, to provide for more elaborate patterns of bands in which A and B are specified alternately.

As an example of this formula, let it be assumed that $A = K$ for $f < f_0$ and that $B = k\pi/2$ for $f > f_0$. These are shown by the solid lines in Fig. 7. Substitution in (4) gives the A and B characteristics in the rest of the spectrum as

$$B = k \sin^{-1}\frac{f}{f_0}, \qquad\qquad\qquad f < f_0$$

$$A = K + k \log\left[\sqrt{\frac{f^2}{f_0^2} - 1} + \frac{f}{f_0}\right], \qquad f > f_0. \qquad (5)$$

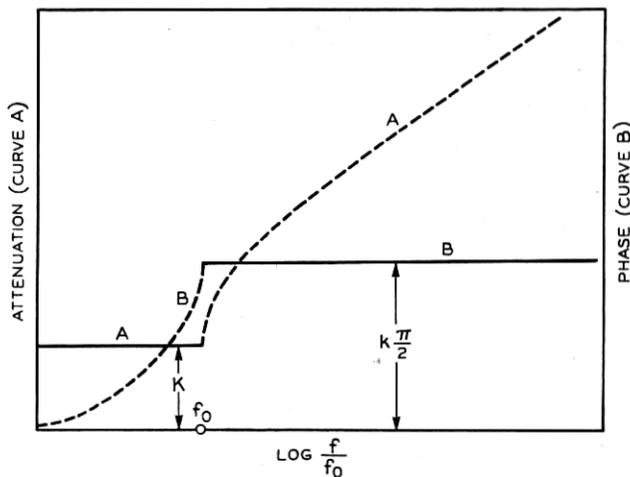These are indicated by broken lines in Fig. 7. In this particularly

Fig. 7—Construction of complete characteristics from an attenuation characteristic specified below a certain frequency and a phase characteristic above it. The solid lines represent the specified attenuation and phase characteristics, and the broken lines their computed extensions to the rest of the spectrum.

simple case all four fragments can be combined into the single analytic formula

$$A + iB = K + k \log\left[\sqrt{1 - \frac{f^2}{f_0^2}} + i\frac{f}{f_0}\right]. \qquad (6)$$

This expression will be used as the fundamental formula for the loop cut-off characteristic in the next section.

OVERALL FEEDBACK LOOP CHARACTERISTICS

The survey just concluded shows what combinations of attenuation and phase characteristics are physically possible. We have next to determine which of the available combinations is to be regarded as representing the transmission around the overall feedback loop. The choice will naturally depend somewhat upon exactly what we assume that the amplifier ought to do, but with any given set of assumptions it is possible, at least in theory, to determine what combination is most appropriate.
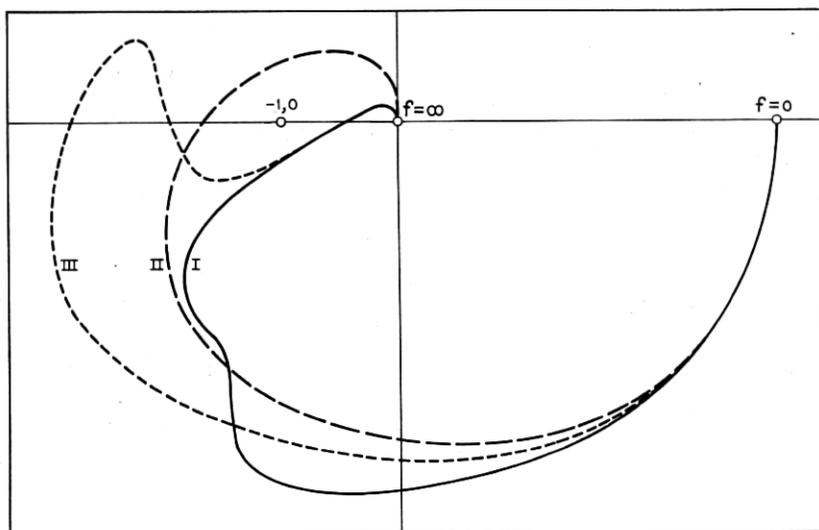

FIG. 8—Nyquist stability diagrams for various amplifiers. Curve I represents "absolute" stability, Curve II instability, and Curve III "conditional" stability. In accordance with the convention used in this paper the diagram is rotated through 180° from its normal position so that the critical point occurs at − 1, 0 rather than + 1, 0.

The situation is conveniently investigated by means of the Nyquist stability diagram [5] illustrated by Fig. 8. The diagram gives the path

[5] *Bell System Technical Journal*, July, 1932. See also Peterson, Kreer, and Ware, *Bell System Technical Journal*, October, 1934. The Nyquist diagrams in the present paper are rotated through 180° from the positions in which they are usually drawn, turning the diagrams in reality into plots of − $\mu\beta$. In a normal amplifier there is one net phase reversal due to the tubes in addition to any phase shifts chargeable directly to the passive networks in the circuit. The rotation of the diagram allows this phase reversal to be ignored, so that the phase shifts actually shown are the same as those which are directly of design interest.

traced by the vector representing the transmission around the feedback loop as the frequency is assigned all possible real values. In accordance with Nyquist's results a path such as II, which encircles the point − 1, 0, indicates an unstable circuit and must be avoided. A stable amplifier is obtained if the path resembles either I or III, neither of which encircles − 1, 0. The stability represented by Curve III, how-ever, is only "Nyquist" or "conditional." The path will enclose the critical point if it is merely reduced in scale, which may correspond physically to a reduction in tube gain. Thus the circuit may sing when the tubes begin to lose their gain because of age, and it may also sing, instead of behaving as it should, when the tube gain increases from zero as power is first applied to the circuit. Because of these possibilities conditional stability is usually regarded as undesirable and the present discussion will consequently be restricted to "abso-lutely" or "unconditionally" stable amplifiers having Nyquist diagrams of the type resembling Curve I.

The condition that the amplifier be absolutely stable is evidently that the loop phase shift should not exceed 180° until the gain around the loop has been reduced to zero or less. A theoretical characteristic which just met this requirement, however, would be unsatisfactory, since it is inevitable that the limiting phase would be exceeded in fact by minor deviations introduced either in the detailed design of the amplifier or in its construction. It will therefore be assumed that the limiting phase is taken as 180° less some definite margin. This is illustrated by Fig. 9, the phase margin being indicated as $y\pi$ radians. At frequencies remote from the band it is physically impossible, in most circuits, to restrict the phase within these limits. As a supple-ment, therefore, it will be assumed that larger phase shifts are permis-sible if the loop gain is x db below zero. This is illustrated by the broken circular arc in Fig. 9. A theoretical loop characteristic meeting both requirements will be developed for an amplifier transmitting between zero and some prescribed limiting frequency with a constant feedback, and cutting off thereafter as rapidly as possible. This basic characteristic can be adapted to amplifiers with varying feedback in the useful range or with useful ranges lying in other parts of the spec-trum by comparatively simple modifications which are described at a later point. It is, of course, contemplated that the gain and phase margins x and y will be chosen arbitrarily in advance. If we choose large values we can permit correspondingly large tolerances in the detailed design and construction of the apparatus without risk of in-stability. It turns out, however, that with a prescribed width of cutoff interval the amount of feedback which can be realized in the

useful range is decreased as the assumed margins are increased, so that it is generally desirable to choose as small margins as is safe.

The essential feature in this situation is the requirement that the diminution of the loop gain in the cutoff region should not be accompanied by a phase shift exceeding some prescribed amount. In view of the close connection between phase shift and the slope of the attenuation characteristic evidenced by (2) this evidently demands that the amplifier should cut off, on the whole, at a well defined rate which is not too fast. As a first approximation, in fact, we can choose the cutoff
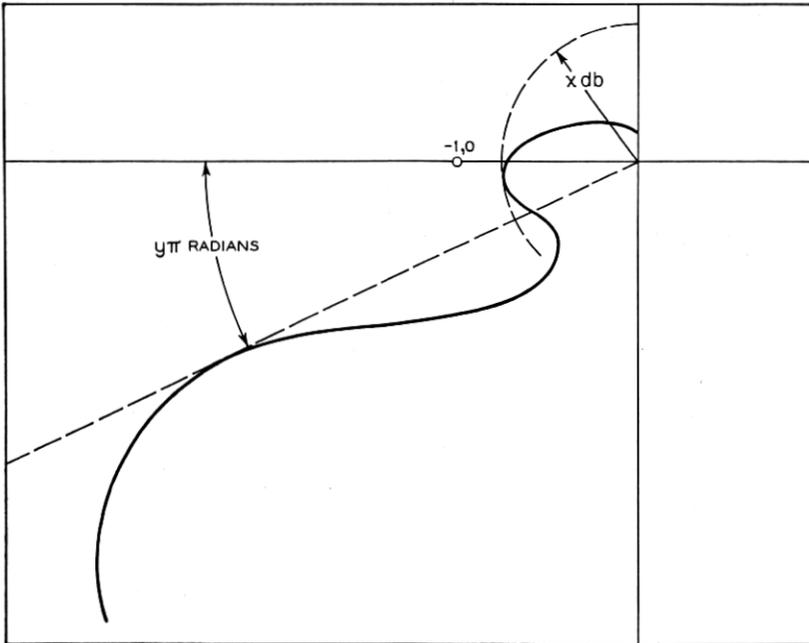
Fig. 9—Diagram to illustrate definitions of phase and gain margins for the feedback loop.

characteristic as an exactly constant slope from the edge of the useful band outward. Such a characteristic has already been illustrated by Fig. 5 and is shown, replotted,[6] by the broken lines in Fig. 10. If we choose the parameter corresponding to k in Fig. 5 as 2 the cutoff rate is 12 db per octave and the phase shift is substantially 180° at high frequencies. This choice thus leads to zero phase margin. By choosing a somewhat smaller k on the other hand, we can provide a definite

[6] To prevent confusion it should be noticed that the general attenuation-phase diagrams are plotted in terms of relative loss while loop cutoff characteristics, here and at later points, are plotted in terms of relative gain.

margin against singing, at the cost of a less rapid cutoff. For example, if we choose $k = 1.5$ the limiting phase shift in the $\mu\beta$ loop becomes 135°, which provides a margin of 45° against instability, while the rate of cutoff is reduced to 9 db per octave. The value $k = 1.67$, which corresponds to a cutoff rate of 10 db per octave and a phase margin of 30°, has been chosen for illustrative purposes in preparing Fig. 10. The loss margin depends upon considerations which will appear at a later point.

Although characteristics of the type shown by Fig. 5 are reasonably satisfactory as amplifier cutoffs they evidently provide a greater phase
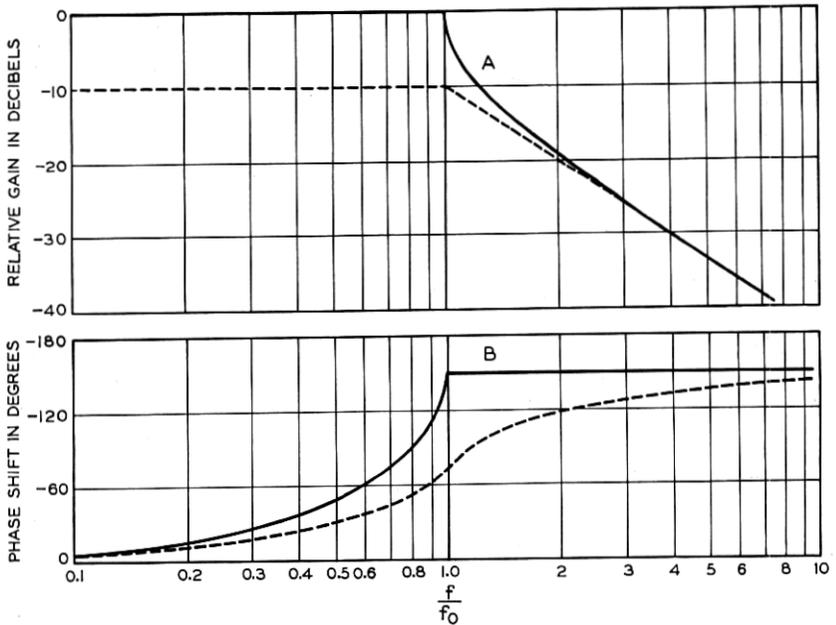

Fig. 10—Ideal loop cutoff characteristics. Drawn for a 30° phase margin.

margin against instability in the region just beyond the useful band than they do at high frequencies. In virtue of the phase area law this must be inefficient if, as is supposed here, the optimum characteristic is one which would provide a constant margin throughout the cutoff interval. The relation between the phase and the slope of the attenuation suggests that a constant phase margin can be obtained by increasing the slope of the cutoff characteristic near the edge of the band, leaving its slope at more remote frequencies unchanged, as shown by the solid lines in Fig. 10. The exact expression for the required curve can be found from (6), where the problem of determining such a characteristic appeared as an example of the use of the general formula (4).

At high frequencies the new phase and attenuation characteristics merge with those obtained from the preceding straight line cutoff, as Fig. 10 indicates. In this region the relation between phase margin and cutoff slope is fixed by the k in the equation (6) in the manner already described for the more elementary cutoff. At low frequencies, however, the increased slope near the edge of the band permits 6 k db more feedback.

It is worth while to pause here to consider what may be said, on the basis of these characteristics, concerning the breadth of cutoff interval required for a given feedback, or the "price of the ticket," as it was expressed in the introduction. If we adopt the straight line cutoff and assume the k used in Fig. 10 the interval between the edge of the useful band and the intersection of the characteristic with the zero gain axis is evidently exactly 1 octave for each 10 db of low frequency feedback. The increased efficiency of the solid line characteristic saves one octave of this total if the feedback is reasonably large to begin with. This apparently leads to a net interval one or two octaves narrower than the estimates made in the introduction. The additional interval is required to bridge the gap between a purely mathematical formula such as (6), which implies that the loop characteristics follow a prescribed law up to indefinitely high frequencies, and a physical amplifier, whose ultimate loop characteristics vary in some uncontrollable way. This will be discussed later. It is evident, of course, that the cutoff interval will depend slightly upon the margins assumed. For example, if the phase margin is allowed to vanish the cutoff rate can be increased from 10 to 12 db per octave. This, however, is not sufficient to affect the order of magnitude of the result. Since the diminished margin is accompanied by a corresponding increase in the precision with which the apparatus must be manufactured such an economy is, in fact, a Pyrrhic victory unless it is dictated by some such compelling consideration as that described in the next section.

Maximum Obtainable Feedback

A particularly interesting consequence of the relation between feedback and cutoff interval is the fact that it shows why we cannot obtain unconditionally stable amplifiers with as much feedback as we please. So far as the purely theoretical construction of curves such as those in Fig. 10 is concerned, there is clearly no limit to the feedback which can be postulated. As the feedback is increased, however, the cutoff interval extends to higher and higher frequencies. The process reaches a physical limit when the frequency becomes so high that parasitic effects in the circuit are controlling and do not permit the prescribed cutoff

characteristic to be simulated with sufficient precision.   For example, we are obviously in physical difficulties if the cutoff characteristic specifies a net gain around the loop at a frequency so high that the tubes themselves working into their own parasitic capacitances do not give a gain.

This limitation is studied most easily if the effects of the parasitic elements are lumped together by representing them in terms of the asymptotic characteristic of the loop as a whole at extremely high frequencies.   An example is shown by Fig. 11.   The structure is a
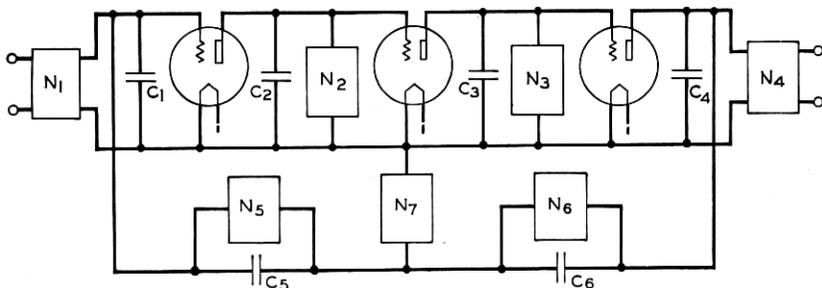
Fig. 11—Elements which determine the asymptotic loop transmission characteristic in a typical amplifier.

shunt feedback amplifier.   The β circuit is represented by the T composed of networks N_5, N_6 and N_7.   The input and output circuits are represented by N_1 and N_4 and the interstage impedances by N_2 and N_3. The C's are parasitic capacitances with the exception of C_5 and C_6, which may be regarded as design elements added deliberately to N_5 and N_6 to obtain an efficient high frequency transmission path from output to input.   At sufficiently high frequencies the loop transmission will depend only upon these various capacitances, without regard to the N's.   Thus, if the transconductances of the tubes are represented by G_1, G_2, and G_3 the asymptotic gains of the first two tubes are $G_1/\omega C_1$ and $G_2/\omega C_3$.   The rest of the loop includes the third tube and the potentiometer formed by the capacitances C_1, C_4, C_5 and C_6.   Its asymptotic transmission can be written as $G_3/\omega C$, where

$$C = C_1 + C_4 + \frac{C_1 C_4}{C_5 C_6} (C_5 + C_6).$$

Each of these terms diminishes at a rate of 6 db per octave.   The complete asymptote is $G_1 G_2 G_3/\omega^3 C C_2 C_3$.   It appears as a straight line with a slope of 18 db per octave when plotted on logarithmic frequency paper.

A similar analysis can evidently be made for any amplifier. In the particular circuit shown by Fig. 11 the slope of the asymptote, in units of 6 db per octave, is the same as the number of tubes in the circuit. The slope can evidently not be less than the number of tubes but it may be greater in some circuits. For example if C_5 and C_6 were omitted in Fig. 11 and N_5 and N_6 were regarded as degenerating into resistances the asymptote would have a slope of 24 db per octave and would lie below the present asymptote at any reasonably high frequency. In any event the asymptote will depend only upon the parasitic elements of the circuit and perhaps a few of the most significant design elements. It can thus be determined from a skeletonized version of the final structure. If waste of time in false starts is to be avoided such a determination should be made as early as possible, and certainly in advance of any detailed design.

The effect of the asymptote on the overall feedback characteristic is illustrated by Fig. 12. The curve $ABEF$ is a reproduction of the ideal
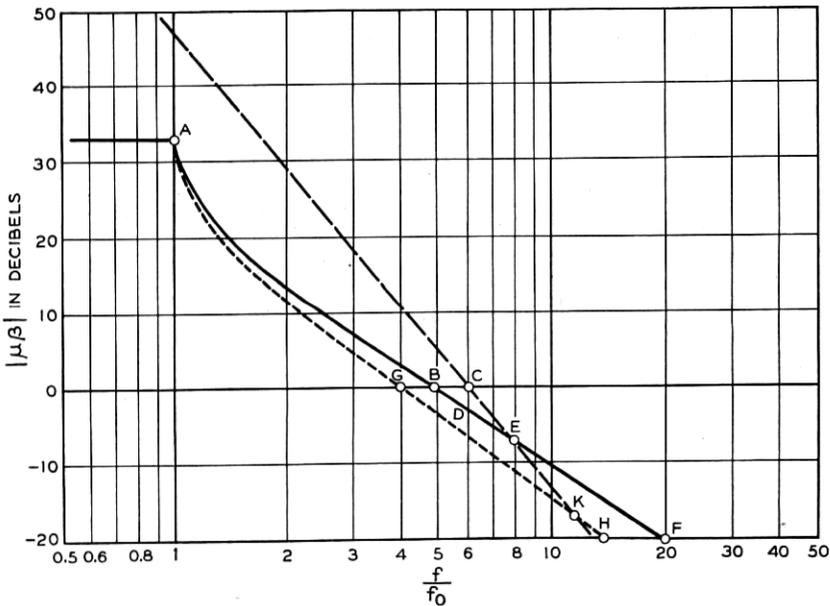

Fig. 12—Combination of asymptotic characteristic and ideal cutoff characteristic.

cutoff characteristic originally given by the solid lines in Fig. 10. It will be recalled that the curve was drawn for the choice $k = 5/3$, which corresponds to a phase margin of 30° and an almost constant slope, for the portion DEF of the characteristic, of about 10 db per octave. The

straight line *CEK* represents an asymptote of the type just described, with a slope of 18 db per octave.   Since the asymptote may be assumed to represent the practical upper limit of gain in the high-frequency region, the effect of the parasitic elements can be obtained by replacing the theoretical cutoff by the broken line characteristic *ABDEK*.   In an actual circuit the corner at *E* would, of course, be rounded off, but this is of negligible quantitative importance.   Since *EF* and *EK* diverge by 8 db per octave the effect can be studied by adding curves of the type shown by Fig. 5 to the original cutoff characteristic.
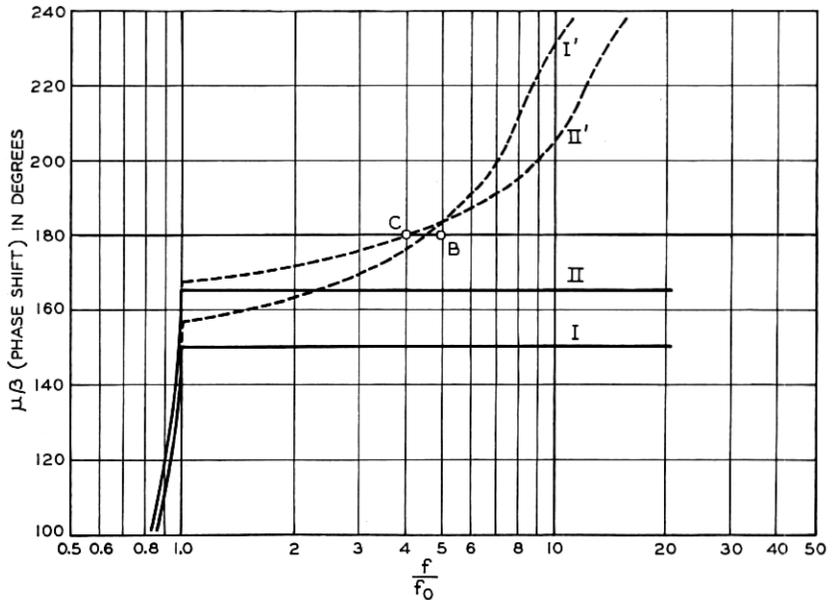
Fig. 13—Phase characteristics corresponding to gain characteristics of Fig. 12.

The phase shift in the ideal case is shown by Curve I of Fig. 13. The addition of the phase corresponding to the extra slope of 8 db per octave at high frequencies produces the total phase characteristic shown by Curve I'.   At the point *B* where $|\mu\beta| = 1$, the additional phase shift amounts to 35 degrees.   Since this is greater than the original phase margin of 30 degrees the amplifier is unstable when parasitic elements are considered.   In the present instance stability can be regained by increasing the coefficient *k* to 1-5/6, which leads to the broken line characteristic *AGKH* in Fig. 12.   This reduces the nominal phase margin to 15 degrees, but the frequency interval between *G* and *K* is so much greater than that between *B* and *E* that the added phase is reduced still more and is just less than 15° at the new

cross over point G. This is illustrated by II and II' in Fig. 13. On the other hand, if the zero gain intercept of the asymptote CEK had occurred at a slightly lower frequency, no change in k alone would have been sufficient. It would have been necessary to reduce the amount of feedback in the transmitted range in order to secure stability.

The final characteristic in Fig. 13 reaches the limiting phase shift of 180° only at the crossover point. It is evident that a somewhat more efficient solution for the extreme case is obtained if the limiting 180° is approximated throughout the cutoff interval. This result is attained by the cutoff characteristic shown in Fig. 14. The characteristic con-
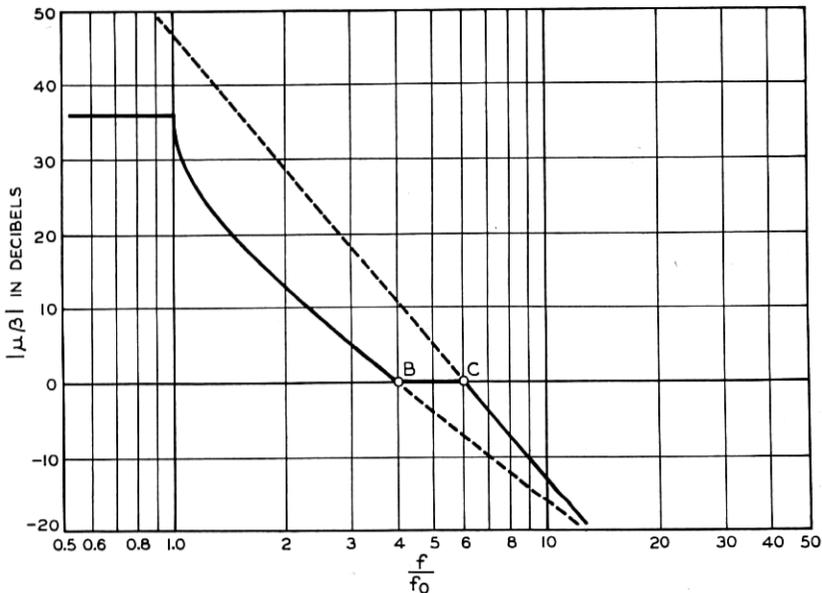
Fig. 14—Ideal cutoff modified to take account of asymptotic characteristic. Drawn for zero gain and phase margins.

sists of the original theoretical characteristic, drawn for $k = 2$, from the edge of the useful band to its intercept with the zero gain axis, the zero gain axis from this frequency to the intercept with the high-frequency asymptote, and the asymptote thereafter. It can be regarded as a combination of the ideal cutoff characteristic and two characteristics of the type shown by Fig. 5. One of the added characteristics starts at B and has a positive slope of 12 db per octave, since the ideal cutoff was drawn for the limiting value of k. The other starts at C and has the negative slope, -18 db per octave, of the asymptote itself. As (3) shows, the added slopes correspond at lower frequencies to ap-

proximately linear phase characteristics of opposite sign. If the frequencies B and C at which the slopes begin are in the same ratio, 12 : 18, as the slopes themselves the contributions of the added slopes will substantially cancel each other and the net phase shift throughout the cutoff interval will be almost the same as that of the ideal curve alone. The exact phase characteristic is shown by Fig. 15. It dips
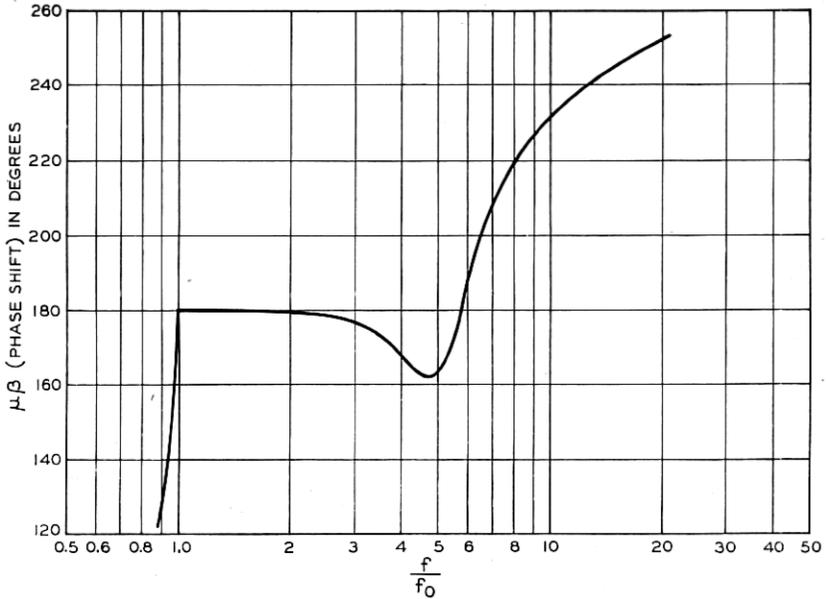

Fig. 15—Phase characteristic corresponding to gain characteristic of Fig. 14.

slightly below 180° at the point at which the characteristic reaches the zero gain axis, so that the circuit is in fact stable.

The same analysis can evidently be applied to asymptotes of any other slope. This makes it easy to compute the maximum feedback obtainable under any asymptotic conditions. If f_0 and f_a are respectively the edge of the useful band and the intercept (C in Figs. 12 and 14) of the asymptote with the zero gain axis, and n is the asymptotic slope, in units of 6 db per octave, the result appears as

$$A_m = 40 \log_{10} \frac{4f_a}{f_0}, \tag{7}$$

where A_m is the maximum feedback in db.[7]

---

[7] The formulæ for maximum feedback given here and in the later equation (8) are slightly conservative. It follows from the phase area law that more feedback should be obtained if the phase shift were exactly 180° below the crossover and rose

For the sake of generality it is convenient to extend this formula to include also situations in which there exists some further linear phase characteristic in addition to those already taken into account. In exceptional circuits, the final asymptotic characteristic may not be completely established by the time the curve reaches the zero gain axis and the additional phase characteristic may be used to represent the effect of subsequent changes in the asymptotic slope. Such a situation might occur in the circuit of Fig. 11, for example, if C_5 or C_6 were made extremely small. The additional term may also be used to represent departures from a lumped constant analysis in high-frequency amplifiers, as discussed earlier, If we specify the added phase characteristic, from whatever source, by means of the frequency f_d at which it would equal $2n/\pi$ radians, if extrapolated, the general formula corresponding to (7) becomes

$$A_m = 40 \log_{10} \frac{4}{nf_0} \frac{f_a f_d}{f_a + f_d}. \tag{8}$$

It is interesting to notice that equations (7) and (8) take no explicit account of the final external gain of the amplifier. Naturally, if the external gain is too high the available μ circuit gain may not be sufficient to provide it and also the feedback which these formulæ promise. This, however, is an elementary question which requires no further discussion. In other circumstances, the external gain may enter the situation indirectly, by affecting the asymptotic characteristics of the β path, but in a well chosen β circuit this is usually a minor consideration. The external gain does, however, affect the parts of the circuit upon which reliance must be placed in controlling the overall loop characteristic. For example, if the external gain is high the μ circuit will ordinarily be sharply tuned and will drop off rapidly in gain beyond the useful band. The β circuit must therefore provide a decreasing loss to bring the overall cutoff rate within the required limit. Since the β circuit must have initially a high loss to correspond to the high final gain of the complete amplifier, this is possible. Conversely, if the gain of the amplifier is low the μ circuit will be relatively flexible and the β circuit relatively inflexible.

rapidly to its ultimate value thereafter. These possibilities can be exploited approximately by various slight changes in the slope of the cutoff characteristic in the neighborhood of the crossover region, or a theoretical solution can be obtained by introducing a prescribed phase shift of this type in the general formula (4). The theoretical solution gives a Nyquist path which, after dropping below the critical point with a phase shift slightly less than 180°, rises again with a phase shift slightly greater than 180° and continues for some time with a large amplitude and increasing phase before it finally approaches the origin. These possibilities are not considered seriously here because they lead to only a few db increase in feedback, at least for moderate n's, and the degree of design control which they envisage is scarcely feasible in a frequency region where, by definition, parasitic effects are almost controlling.

In setting up (7) and (8) it has been assumed that the amplifier will, if necessary, be built with zero margins against singing. Any surplus which the equations indicate over the actual feedback required can, of course, be used to provide a cutoff characteristic having definite phase and gain margins. For example, if we begin with a lower feedback in the useful band the derivative of the attenuation between this region and the crossover can be proportionately reduced, with a corresponding decrease in phase shift. We can also carry the flat portion of the characteristic below the zero gain axis, thus providing a gain margin when the phase characteristic crosses 180°. In reproportioning the characteristic to suit these conditions, use may be made of the approximate formula

$$A_m - A = (A_m + 17.4)y + \frac{n-2}{n}x + \frac{2}{n}xy, \qquad (9)$$

where A_m is the maximum obtainable feedback (in db), A is the actual feedback, and x and y are the gain and phase margins in the notation of Fig. 9. Once the available margin has been divided between the x and
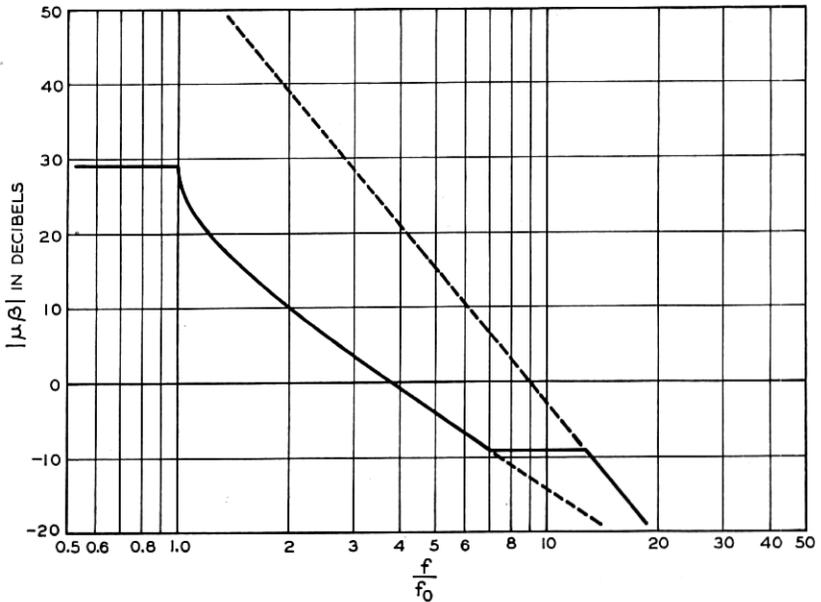

Fig. 16—Modified cutoff permitting 30° phase margin and 9 db gain margin.

y components by means of this formula the cutoff characteristic is, of course, readily drawn in. An example is furnished by Figs. 16 and 17,

where it is assumed that $A_m = 43$ db, $A = 29$ db, $x = 9$ db, $n = 3$ and $y = 1/6$. The Nyquist diagram for the structure is shown by Fig. 18. It evidently coincides almost exactly with the diagram postulated originally in Fig. 9.
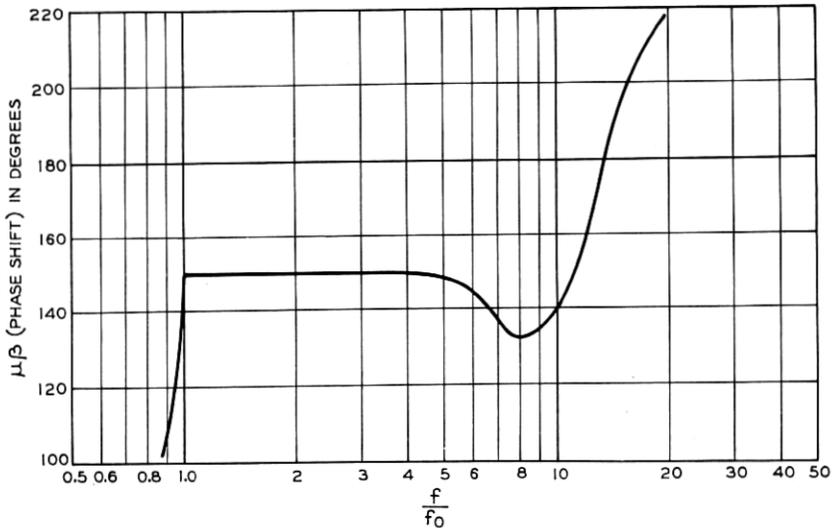

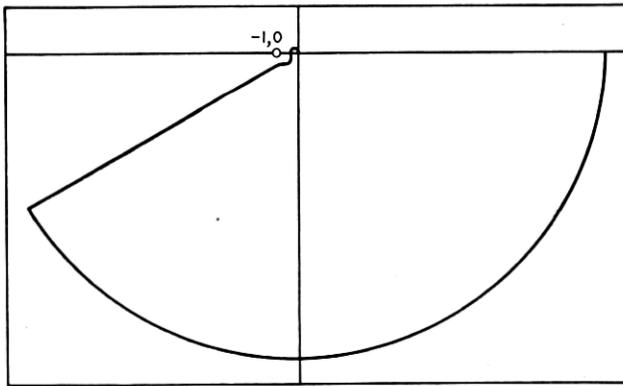Fig. 17—Phase characteristic corresponding to gain characteristic of Fig. 16.

Fig. 18—Nyquist diagram corresponding to gain and phase characteristics of Figs. 16 and 17. As in Fig. 8 the diagram is rotated to place the critical point at − 1, 0 rather than + 1, 0.

With the characteristic of Fig. 16 at hand, we can return once more to the calculation of the total design range corresponding to any given feedback. From the useful band to the intersection of the cutoff

characteristic with the zero gain axis the calculation is the same as that made previously in connection with Fig. 10. From the zero gain intercept to the junction with the asymptote, where we can say that design control is finally relaxed, there is, however, an additional interval of nearly two octaves. Although Fig. 16 is fairly typical, the exact breadth of the additional interval will depend somewhat on circumstances. It is increased by an increase in the asymptotic slope and reduced by decreasing the gain margin.

RELATIVE IMPORTANCE OF TUBES AND CIRCUIT IN LIMITING FEEDBACK [8]

The discussion just finished leads to the general conclusion that the feedback which can be obtained in any given amplifier depends ultimately upon the high-frequency asymptote of the feedback loop. It is a matter of some importance, then, to determine what fixes the asymptote and how it can be improved. Evidently, the asymptote is finally restricted by the gains of the tubes alone. We can scarcely improve upon the result secured by connecting the output plate directly to the input grid. Within this limit, however, the actual asymptotic characteristic will depend upon the configuration and type of feedback employed, since a given distribution of parasitic elements may evidently affect one arrangement more than another. The salient circuit problem is therefore that of choosing a general configuration for the feedback circuit which will allow the maximum efficiency of transmission at high frequencies.

The relative importance of tube limitations and circuit limitations is most easily studied if we replace (7) by

$$A_m = 40 \log_{10} \frac{4f_t}{n f_0} - \frac{2A_t}{n}, \tag{10}$$

where f_t is the frequency at which the tubes themselves working into their own parasitic capacitances have zero gain [9] and A_t is the asymptotic loss of the complete feedback loop in db at $f = f_t$. The first term

---

[8] The material of this section was largely inspired by comments due to Messrs. G. H. Stevenson and J. M. West.

[9] I.e., $f_t = \frac{G_m}{2\pi C}$, where G_m and C are respectively the transconductance and capacitance of a representative tube. The ratio $\frac{G_m}{C}$ is the so-called "figure of merit" of the tube. The analysis assumes that the interstage network is a simple shunt impedance, so that the parasitic capacitance does correctly represent its asymptotic behavior. More complicated four-terminal interstage networks, such as transformer coupling circuits and the like, are generally inadmissible in a feedback amplifier because of the high asymptotic losses and consequent high-phase shifts which they introduce.

of (10) shows how the feedback depends upon the intrinsic band width of the available tubes. In low-power tubes especially designed for the purpose f_t may be 50 mc or more, but if f_0 is small the first term will be substantial even if tubes with much lower values of f_t are selected. The second term gives the loss in feedback which can be ascribed to the rest of the circuit. It is evidently not possible to provide input and output circuits and a β-path without making some contribution to the asymptotic loss, so that A_t cannot be zero. In an amplifier designed with particular attention to this question, however, it is frequently possible to assign A_t a comparatively low value, of the order of 20 to 30 db or less. Without such special attention, on the other hand, A_t is likely to be very much larger, with a consequent diminution in available feedback.

In addition to f_t and A_t, (10) includes the quantity n, which represents the final asymptotic slope in multiples of 6 db per octave. Since the tubes make no contribution to the asymptotic loss at $f = f_t$ we can vary n without affecting A_t by changing the number of tubes in the circuit. This makes it possible to compute the optimum number of tubes which should be used in any given situation in order to provide the maximum possible feedback. If A_t is small the first term of (10) will be the dominant one and it is evidently desirable to have a small number of stages. The limit may be taken as $n = 2$ since with only one stage the feedback is restricted by the available forward gain, which is not taken into account in this analysis. On the other hand since the second term varies more rapidly than the first with n, the optimum number of stages will increase as A_t is increased. It is given generally by

$$n = \frac{A_t}{8.68} \tag{11}$$

or in other words the optimum n is equal to the asymptotic loss at the tube crossover in nepers.

This relation is of particular interest for high-power circuits, such as radio transmitters, where circuit limitations are usually severe but the cost of additional tubes, at least in low-power stages, is relatively unimportant. As an extreme example, we may consider the problem of providing envelope feedback around a transmitter. With the relatively sharp tuning ordinarily used in the high-frequency circuits of a transmitter the asymptotic characteristics of the feedback path will be comparatively unfavorable. For illustrative purposes we may assume that $f_a = 40$ kc. and $n = 6$. In accordance with (7) this would provide a maximum available feedback over a 10 kc. voice band of 17 db. It

will also be assumed that the additional tubes for the low-power portions of the circuit have an f_t of 10 mc.[10]   The corresponding A_t is 33 nepers [11] so that equation (11) would say that the feedback would be increased by the addition of as many as 27 tubes to the circuit.   Naturally in such an extreme case this result can be looked upon only as a qualitative indication of the direction in which to proceed.   If we add only 4 tubes, however, the available feedback becomes 46 db while if we add 10 tubes it reaches 60 db.   It is to be observed that only a small part of the available gain of the added tubes is used in directly increasing the feedback.   The remainder is consumed in compensating for the unfortunate phase shifts introduced by the rest of the circuit.

Amplifiers of Other Types

The amplifier considered thus far is of a rather special type.   It has a useful band extending from zero up to some prescribed frequency f_0, constant feedback in the useful band, and it is absolutely stable.   Departures from absolute stability are rather unusual in practical amplifiers and will not be considered here.   It is apparent from the phase area relation that a conditionally stable amplifier may be expected to have a greater feedback for a cut-off interval of given breadth than a structure which is unconditionally stable, but a detailed discussion of the problem is beyond the scope of this paper.

Departures from the other assumptions are easily treated.   For example, if a varying feedback in the useful band is desired, as it may be in occasional amplifiers, an appropriate cut-off characteristic can be constructed by returning to the general formula (4), performing the integrations graphically, if necessary.   If the phase requirement in the cut-off region is left unchanged only the first integral need be modified.   The most important question, for ordinary purposes, is that of determining how high the varying feedback can be, in comparison with a corresponding constant feedback characteristic, for any given asymptote.   This can be answered by observing the form to which the first integral in (4) reduces when f_c is made very large.   It is easily seen that the asymptotic conditions will remain the same provided the

---

[10] In tubes operating at a high-power level f_t may, of course, be quite low.   It is evident, however, that only the tubes added to the circuit are significant in interpreting (11).   The additional tubes may be inserted directly in the feedback path if they are made substantially linear in the voice range by subsidiary feedback of their own. This will not affect the essential result of the present analysis.

[11] It is, of course, not to be expected that the actual asymptotic slope will be constant from 40 kc. to 10 mc.   Since only the region extending a few octaves above 40 kc. is of interest in the final design, however, the apparent A_t can be obtained by extrapolating the slope in this region.

feedback in the useful band satisfies a relation of the form

$$\int_0^{\pi/2} A\, d\phi = \text{constant}, \tag{12}$$

where $\phi = \sin^{-1} f/f_0$. Thus the area under the varying characteristic, when plotted against ϕ, should be the same as that under a corresponding constant characteristic having the same phase and gain margins and the same final asymptote. This is exemplified by Fig. 19, the
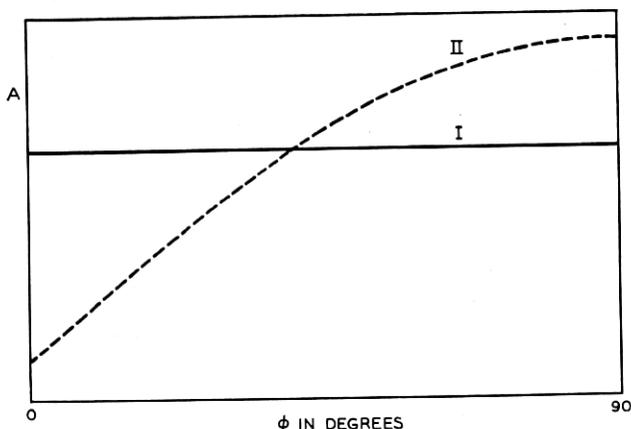

Fig. 19—Diagram to illustrate the computation of available feedback when the required feedback in the useful band is not the same at all frequencies.

varying characteristic being chosen for illustrative purposes as a straight line on an arithmetic frequency scale.

The most important question has to do with the assumption that the useful transmission band extends down to zero frequency. In most amplifiers, of course, this is not true. It is consequently necessary to provide a cut-off characteristic on the lower as well as the upper side of the band. The requisite characteristics are easily obtained from the ones which have been described by means of frequency transformations of a type familiar in filter theory. Thus if the cut-off characteristics studied thus far are regarded as being of the "low-pass" type the characteristics obtained from them by replacing f/f_0 by its reciprocal may be regarded as being of the "high-pass" type. If the band width of the amplifier is relatively broad it is usually simplest to treat the upper and lower cut-offs as independent characteristics of low-pass and high-pass types. In this event, the asymptote for the lower cut-off is furnished by such elements as blocking condensers and choke coils in the plate supply leads. The low-frequency asymptote is usually not so

serious a problem as the high-frequency asymptote since it can be placed as far from the band as we need by using large enough elements in the power supply circuits. The superposition of a low-frequency cutoff on the idealized loop gain and phase characteristics of a "low-pass" circuit is illustrated by the broken lines in Fig. 20.
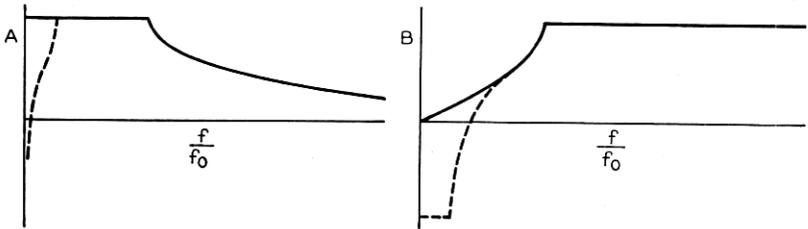

Fig. 20—Modification of loop characteristics to provide a lower cutoff in a broad-band amplifier.

If the band width is relatively narrow it is more efficient to use the transformation in filter theory which relates a low-pass to a symmetrical band-pass structure. The transformation is obtained by replacing f/f_0 in the low-pass case by $(f^2 - f_1 f_2/f(f_2 - f_1))$, where f_1 and f_2 are the edges of the prescribed band. It substitutes resonant and anti-resonant circuits tuned to the center of the band for the coils and condensers in the low-pass circuit. In particular each parasitic inductance is tuned by the addition of a series condenser and each parasitic capacity is tuned by a shunt coil. The parameters of the transformation must, of course, be so chosen that the parasitic elements have the correct values for use in the new branches.

This leads to a simple but important result. If the inductance of a series resonant circuit is fixed, the interval represented by $f_b - f_a$ in Fig. 21, between the frequencies at which the absolute value of the reactance reaches some prescribed limit X_0, is always constant and equal to the frequency at which the untuned inductance would exhibit the reactance X_0, whatever the tuning frequency may be. The same relation holds for the capacity in an anti-resonant circuit. Thus the frequency range over which the branches containing parasitic elements exhibit comparable impedance variations is the same in the band-pass structure and in the prototype low-pass structure. But since the transformation does not affect the relative impedance levels of the various branches in the circuit, this result can be extended to the complete $\mu\beta$ characteristic. We can therefore conclude that *the feedback which is obtainable in an amplifier of given general configuration and with given parasitic elements and given margins depends only upon the breadth*
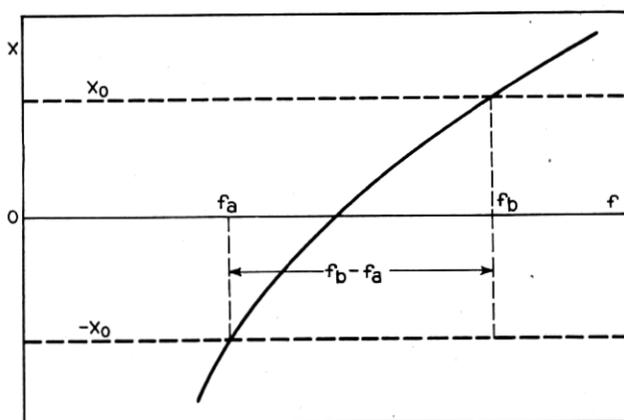
Fig. 21—Frequency interval between prescribed reactances of opposite sign in a resonant circuit with fixed inductance.

of the band in cycles and is independent of the location of the band in the frequency spectrum.

These relations are exemplified by the plots of a low-pass cutoff characteristic and the equivalent band-pass characteristic shown by Fig. 22. The equality of corresponding frequency intervals is indicated by the horizontal lines *A*, *B* and *C*.
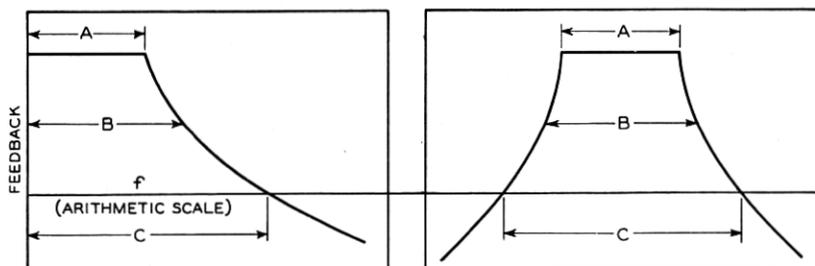

Fig. 22—Diagram to illustrate the conservation of band width in the low-pass to band-pass transformation with fixed parasitic elements. *A*, *B* and *C* represent typical corresponding intervals of equal breadth.

EXAMPLE

An example showing the application of the method in an actual design is furnished by Fig. 23. The structure is a feedback amplifier intended to serve as a repeater in a 72-ohm coaxial line.[12] The useful frequency range extends from 60 to 2,000 kc. Coupling to the line is

[12] The author's personal contact with this amplifier was limited to the evolution of a paper schematic for the high frequency design. The other aspects of the problem are the work of Messrs. K. C. Black, J. M. West and C. H. Elmendorf.
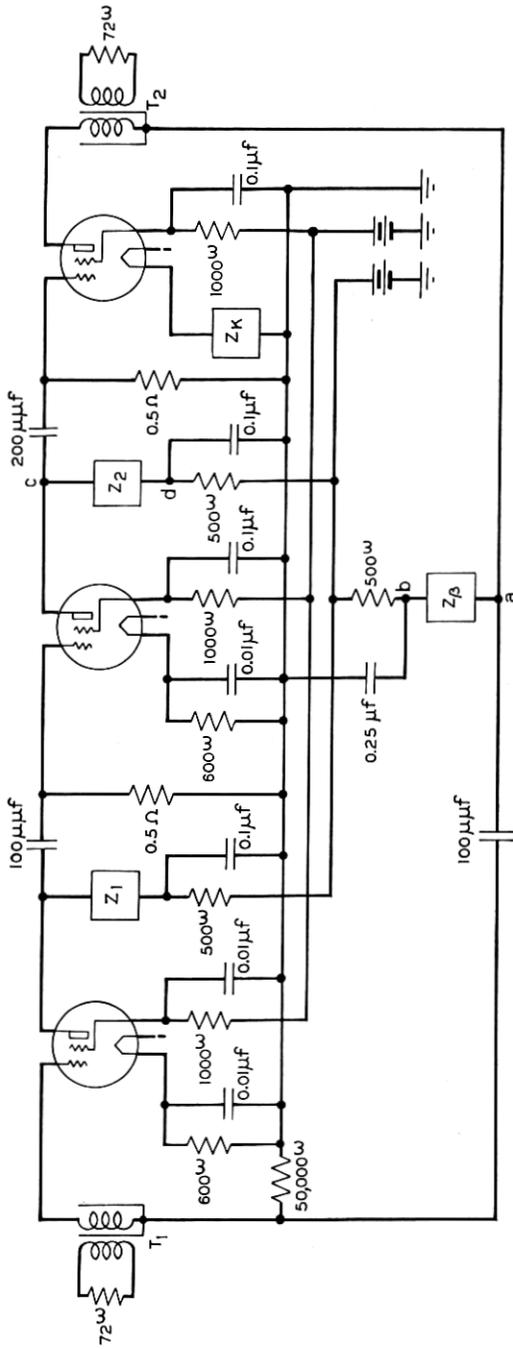
Fig. 23—General schematic of a 2 mc. feedback amplifier.

obtained through the shielded input and output transformers T_1 and T_2. The three stages in the μ circuit are represented in Fig. 23 as single tubes. Physically, however, each stage employes two tubes in parallel, the transconductances of the individual tubes being about 2000 micromhos. The principal feedback is obtained through the impedance Z_β. There is in addition a subsidiary local feedback on the power stage through the impedance Z_K. This is advantageous in producing a further reduction in the effects of modulation in this stage but it does not materially affect the feedback available around the principal loop.

The elements shown explicitly include resistance-capacitance filters in the power supply leads to the plates and screens, cathode resistances and by-pass condensers to provide grid bias potentials, and blocking condenser-grid-leak combinations for the several tubes. In addition to serving these functions, the various resistance-capacitance combinations are also used to provide the cutoff characteristic below the useful band. The low-frequency asymptote is established by the grid leak resistances and the associated coupling condensers and the approach of the feedback characteristic to the asymptote is controlled mainly by the cathode impedances and the resistance-capacitance filters in the power supply leads to the plates. The principal parts of the circuit entering into the $\mu\beta$ characteristic at high frequencies are the interstage impedances Z_1 and Z_2, the feedback impedance Z_β,[13] the cathode impedance Z_K, and the two transformers. The four network designs are shown in detail in Figs. 24, 25, 26, and 27.

The joint transconductance, 4000 micromhos, of two tubes in parallel operating into an average interstage capacity of 14 mmf, as indicated by Figs. 24 and 25, gives an f_t of about 50 mc. The parasitic capacities (chiefly transformer high side and ground capacities) in the other parts of the feedback loop provide a net loss, A_t, of about 18 db at this frequency. Since the asymptotic slope is 18 db per octave the intercept of the complete asymptote with the zero gain axis occurs about one octave lower, at slightly less than 25 mc. This is a relatively high intercept and may be attributed in part to the high gain of the vacuum tubes. The care used in minimizing parasitic capacities in the construction of the amplifier and the general circuit arrangement, including in particular the use of single shunt impedances for the coupling and feedback networks, are also helpful.

---

[13] The relative complexity of this network is explained by the fact that it actually serves as a regulator to compensate for the effects of changes in the line temperature. (See H. W. Bode, "Variable Equalizers," *Bell System Technical Journal*, April, 1938.) The present discussion assumes that the controlling element is at its normal setting. For this setting the network is approximately equal to a resistance in series with a small inductance. The fact that the amplifier must remain stable over a regulation range may serve to explain why the design includes such large stability margins.
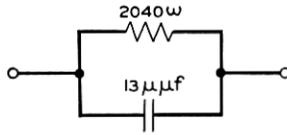
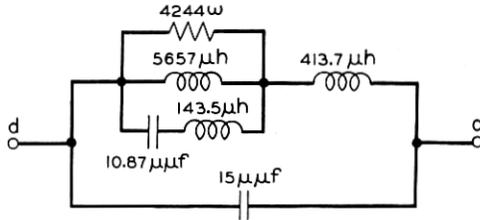Fig. 24—First interstage for the amplifier of Fig. 23.

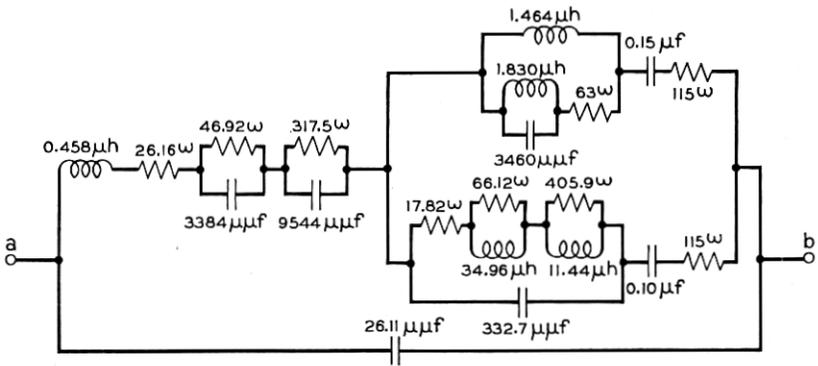Fig. 25—Second interstage for the amplifier of Fig. 23.

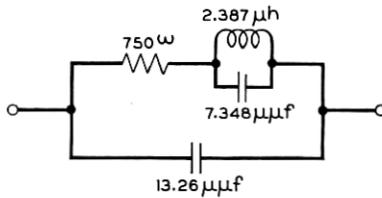Fig. 26—β circuit impedance for the amplifier of Fig. 23.

Fig. 27—Cathode impedance for the amplifier of Fig. 23.

In accordance with (7) the maximum available feedback A_m is 48 db. For design purposes, however, x and y in (9) were chosen as 15 db and 1/5 respectively. This reduces the actual feedback A to about 28 db. The theoretical cutoff characteristic corresponding to
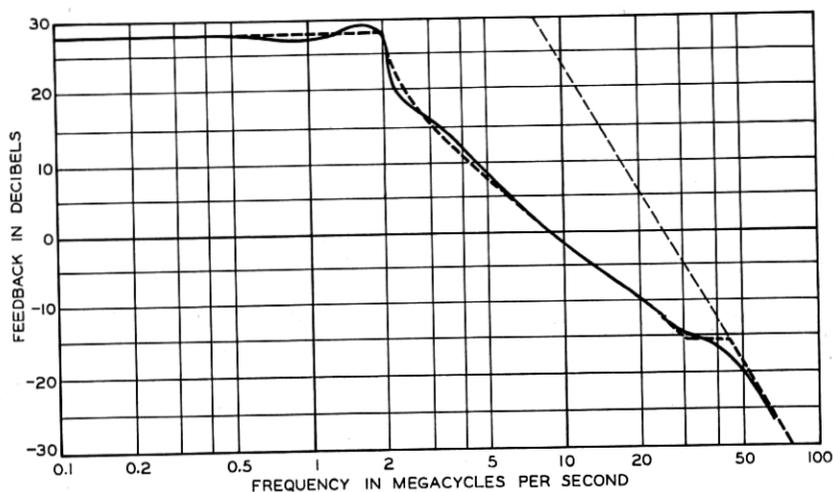

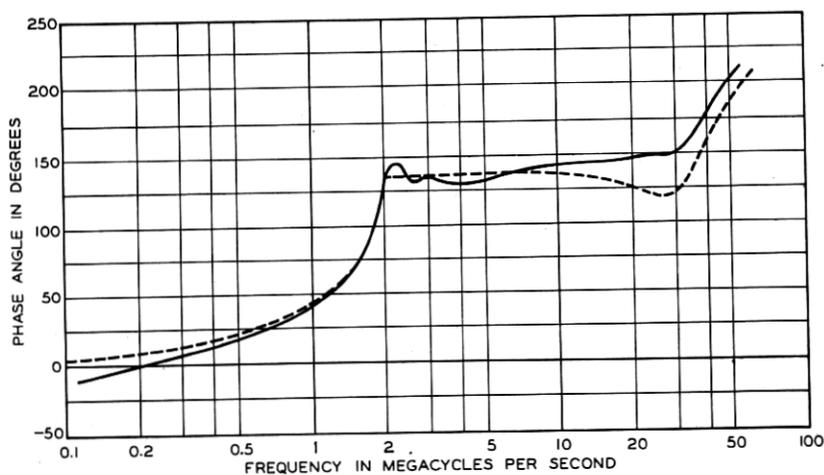Fig. 28—Loop gain characteristic for the amplifier of Fig. 23.

Fig. 29—Loop phase characteristic for the amplifier of Fig. 23.

these parameters is shown by the broken lines in Figs. 28 and 29, and the actual design characteristic by the solid lines. Since this is a structure in which the required forward gain is considerably less than the maximum available gain, the general course of the cutoff character-

istic is controlled, in accordance with the procedure outlined previously, by the elements in the μ circuit.   The sharp slope just beyond the edge of the useful band is obtained from a transformer anti-resonance.   The relatively flat portion of the characteristic near its intersection with the asymptote is due partly to an anti-resonance of the β circuit with its distributed capacitance and partly to an increase in the gain of the third tube because of the filter-like action of the elements of Z_K in cutting out the local feedback on the tube in this region.

The large margins in the design made it possible to secure a substantial increase in feedback without instability.   For example, with a loss margin as great as 15 db the feedback can be increased by adjusting the screen and plate voltages to increase the tube gains.   A higher feedback can also be obtained by adjusting the resistance in the first interstage.   As this interstage was designed, an increase in the resistance results in an increased amplifier gain and a correspondingly increased feedback which follows a new theoretical characteristic with a somewhat reduced phase margin.   The adjustment, in effect, produces a change in the value of the constant k in equation (6).   With this adjustment the feedback can be increased to about 40 db before the amplifier sings.