Amplitude Modulation Suppression in FM Systems

By C. L. RUTHROFF

(Manuscript received January 8, 1958)

Inadequacies of some limiter concepts are discussed and a limiter analysis in terms of low-index modulation theory is presented. The analysis of a particular circuit proves the possibility of perfect amplitude modulation suppression with small loss to the frequency modulated signal. Experimental data are presented which verify the theory and demonstrate the practicability of the device analyzed.

I. INTRODUCTION

Since the early days of frequency modulation (FM) it has been known that to realize the full advantage of FM with respect to signal-to-noise ratio it is necessary to suppress any amplitude modulation (AM) which may be present on the signal. Various kinds of limiters have been used for this purpose: overloaded amplifiers, grid and plate clippers and vacuum diode clippers, to mention a few. In recent years, the trend has been to use germanium or silicon diodes for this function, particularly in broadband systems such as the TD-2 and TJ microwave systems and in systems using transistors.

There is some ambiguity in the use of the word "limiter." In this discussion a limiter is a device whose function is to reduce the index of amplitude modulation of the input signal. This function is called limiting or AM suppression and the amount of limiting is the ratio of the index of AM in the output signal to the index of AM of the input signal.

The operation of limiters has been only partially understood and as a consequence the operation is not efficient either in the sense of loss to the FM signal or in the amount of amplitude modulation suppression (limiting).

If it is recognized that limiting is a modulation process, it becomes clear that modulation theory can be used to analyze limiter circuits. This is done here for a diode limiter. Theoretical and experimental curves are presented, as are equivalent circuits.

II. GENERAL

Before launching into the modulation theory of limiters, it is first necessary to point out some of the inadequacies of some current concepts.

It might be surmised that, if the carrier output voltage is plotted as a function of the input voltage, the amount of limiting is given by the slope of this curve. In Appendix B it is shown that this is not necessarily true. In particular, it is demonstrated that the limiting characteristic can be altered without altering the carrier transfer function.

A common "clipper" limiter and its dual are shown in Figs. 1(a) and 1(b). The customary explanation of limiting by the circuit in Fig. 1(a) is accomplished with the aid of Fig. 2. The diodes are assumed ideal in the sense that the back resistance is infinite and the forward resistance is zero. The bias current I_0 is assumed to come from a constant current source and to bias both diodes in the forward or low-resistance direction. The input signal voltage e shown in Fig. 2(a) is a sine-wave amplitude-modulated carrier. The output voltage across R_l follows the input voltage until the input voltage reaches the clipping level A, Fig. 2(b). At this point, diode D_1 switches to its back resistance and further increase in input voltage does not appear in the output, since the back resistance of D_1 is infinite. Now, when the input voltage is reduced again to value A, D_1 switches to its forward resistance and the output voltage is again a replica of the input. A similar sequence is followed on the

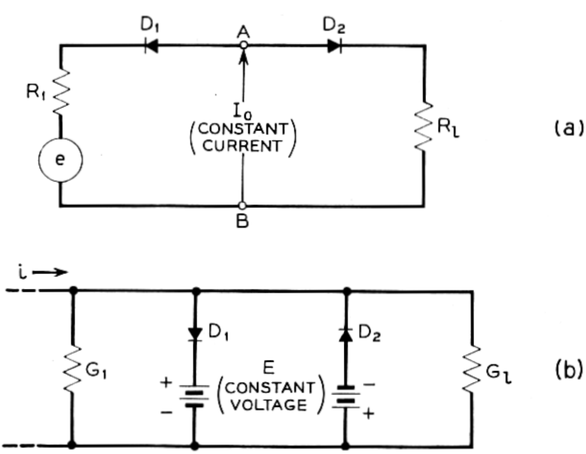

Fig. 1 — A common idealized limiter circuit (a) and its dual (b). The diodes are assumed to have zero forward resistance and infinite back resistance.

negative half of the cycle except that, in this case, it is the diode D_2 that switches at level $-A$. The net result is an output voltage of the form of Fig. 2(c). If the clipping level A were reduced and made to approach zero, only the zero crossings of the original signal would be preserved and the amplitude modulation would be completely suppressed.

The above explanation is useful for some purposes, but it has these serious limitations:

1. It is not obvious from Fig. 2(c) that any AM remains in the signal when, in fact, there may be a considerable amount. The fact that the usual envelope structure has disappeared does not prove the absence of AM.

2. The output is as shown in Fig. 2(c) only if the limiter circuit has a bandwidth which includes many harmonics of the carrier frequency and the baseband frequencies of interest. These harmonics are not negligible in amplitude and have important effects on limiting. In most practical circuits only the fundamental is passed and the simplified explanation is not accurate enough. The effects of preventing the flow of

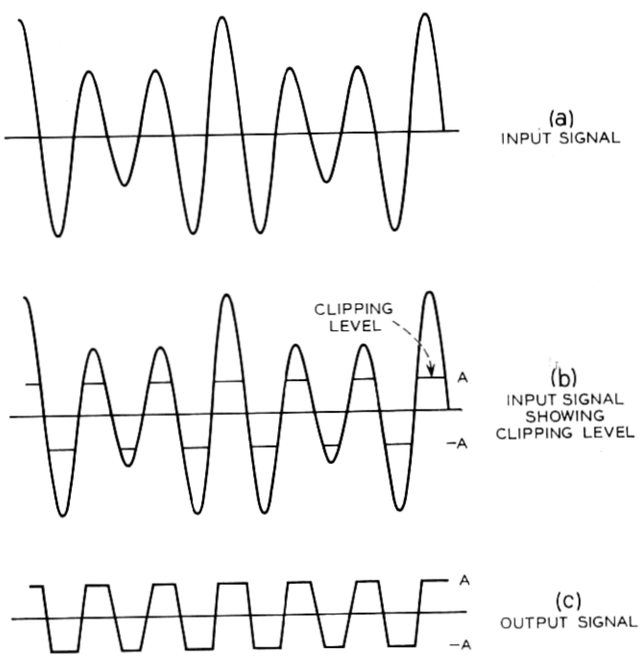

Fig. 2 — Input and output signals for the limiters of Fig. 1.

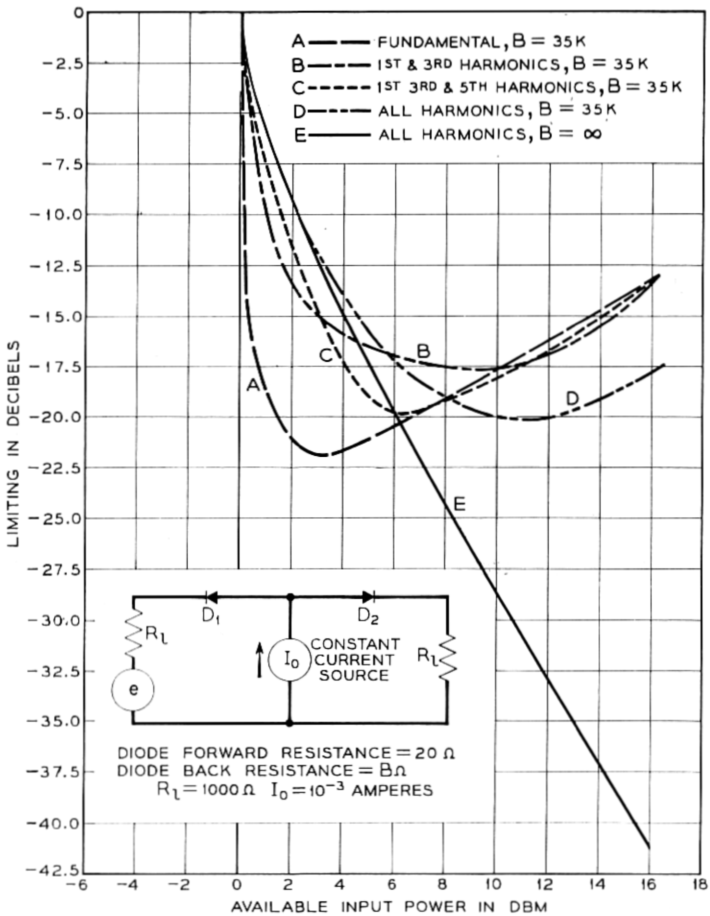

Fig. 3 — Calculated data for "clipper" limiter passing various harmonics of the carrier frequency.

harmonics of the carrier are shown in Fig. 3. This figure gives limiting as a function of available input power level. As mentioned previously, limiting is defined as the ratio of amplitude modulation index of the output signal to that in the input. Curve A shows the usual practical case, i.e., where only the band of frequencies about the carrier and the baseband frequencies are allowed. All harmonics of the carrier and sidebands are suppressed. Curves B and C include all harmonics up to the third and fifth respectively. Curve D includes all harmonics of the carrier and sidebands. Curve E is for diodes with infinite back resistance.

This is the case usually analyzed. These curves were calculated from formulae derived in Appendix A; the circuit constants used are indicated on the figure.

3. The third limitation has to do with the assumptions concerning the diodes. From Fig. 2(c) it can be seen that, if the clipping level is reduced (this is equivalent to increasing input amplitude and keeping clipping level constant), the AM suppression increases, as does the loss at the carrier frequency. In the limit, the AM suppression is perfect and the loss to the carrier is infinite. It is commonly thought that, for this reason, increasing the input level increases the AM suppression. In a practical limiter this is not true. One important reason is that the back resistance of a diode is never infinite. In Fig. 4, which is similar to Fig. 2 except that the diodes are assumed to have a finite back resistance, note that the output waveform is slightly rounded instead of being clipped sharply. The portions are rounded because, even when one diode is on its back resistance, the input voltage influences the output voltage. In fact, during the clipping interval the output waveform is simply a greatly reduced replica of the input waveform in that interval. Now, if the clipping level is reduced to zero, the output waveform is exactly the same shape as the input waveform. At every instant, one diode is on its back resistance and the other on its forward resistance or, in other words, the circuit is linear and there is no limiting whatever. The loss, however, is finite. This means that there is some one input level at which the limiting is best and that, contrary to common belief, merely increasing the input level does not necessarily improve the limiting. This is also illustrated in Fig. 3.

For these three reasons, the simple clipping picture does not give a suitable explanation of the limiting action. A more accurate description

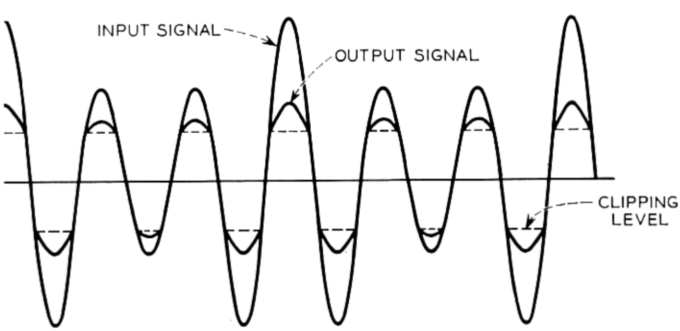

Fig. 4 — Input and output signals for a clipper limiter using diodes with finite back resistance.

is given by analyzing the circuit in terms of modulation theory. To justify the modulation theoretical approach, a simple argument can be given.

A simple example of an FM signal is a sine-wave modulated carrier with low index of modulation. If the index is sufficiently low, only the first pair of sidebands is important. The only difference between this FM wave and an AM wave modulated with the same signal is in the phase relationships between the sidebands and the carrier frequency. An AM wave can be superimposed on the FM wave so that the AM sidebands occur at exactly the same frequencies as do the FM sidebands. This situation is illustrated as follows:

$$\begin{aligned}
\text{AM:}\ & E(1 + k \cos pt) \cos \omega t = E \cos \omega t \\
&\qquad + \frac{kE}{2} [\cos (\omega + p)t + \cos (\omega - p)t], \\
\text{FM:}\ & E \cos \left(\omega t + \frac{\Delta\omega}{p} \sin pt \right) \approx E \cos \omega t \\
&\qquad + E \frac{\Delta\omega}{2p} [\cos (\omega + p)t - \cos (\omega - p)t], \\
&\qquad k \ll 1; \qquad \frac{\Delta\omega}{p} \ll 1,
\end{aligned} \tag{1}$$

where $\Delta\omega$ is the peak deviation of the carrier
ρ is the modulating frequency
ω is the carrier frequency
k is the AM index of modulation.

The composite FM signal with AM present is:

$$\begin{aligned}
& E \cos \omega t + E \frac{\Delta\omega}{2p} [\cos (\omega + p)t - \cos (\omega - p)t] \\
&\qquad + k \frac{E}{2} [\cos (\omega + p)t + \cos (\omega - p)t],
\end{aligned} \tag{2}$$

where the AM and FM sidebands can be identified from (1).

The purpose of the limiter is to eliminate only those sidebands representing AM, i.e., the second pair in (2). Clearly this cannot be accomplished by any linear circuit, since whatever happens to the AM sidebands happens also to the FM sidebands, which are at the same frequencies. The limiter must accomplish two things. First, it must identify the AM sidebands; second, it must eliminate them. And it must do these things without interfering with the FM sidebands.

The identification can be made with an amplitude detector. Such a detector does not interfere with an FM wave but does detect whatever AM is present. The output of the detector can then be used to generate a pair of sidebands in the output circuit similar to the original AM but opposite in phase. If care is taken to adjust the amplitude properly, these remodulated sidebands will completely cancel the original sidebands, and the output signal will be free of AM. It is important to note that neither the detection nor modulation affects the FM signal except to introduce some loss. The output of the limiter contains only the desired FM signal.

This argument, although heuristic, indicates three important results. It shows that limiting is necessarily a nonlinear process. More important, it indicates the possibility, at least theoretically, of perfect limiting using diodes with finite forward and reverse resistances and with a finite loss to the FM signal. In addition, it points the way to a mathematical approach using low-index modulation theory.

The limiter configuration of Fig. 5, based on the principles discussed above, has been analyzed and the details are included in Appendix A.

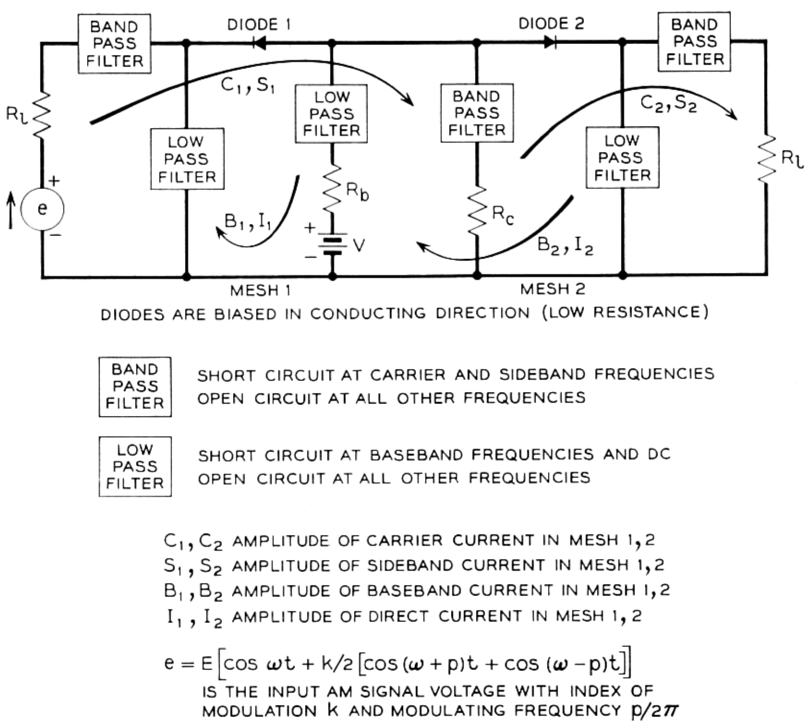

Fig. 5 — Limiter circuit with which perfect limiting can be obtained.

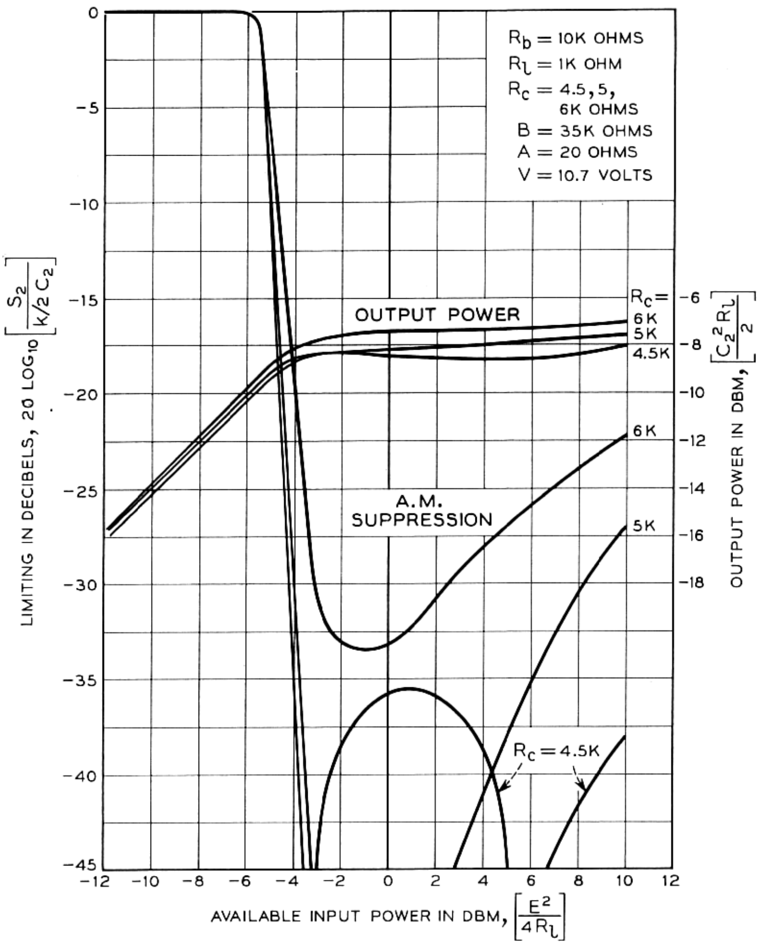

Fig. 6 — Calculated results for limiter of Fig. 5.

The limiter of Fig. 5 is similar to that of Fig. 1(a), with three important differences. In Fig. 1(a), the impedance in branch A-B is infinite at all frequencies; in Fig. 5, this impedance is finite at all frequencies of interest. Also, the circuit in Fig. 5 is arranged so that the resistance R_c, at frequencies in a band centered at the carrier frequency, can be controlled independently of the resistance R_b at baseband frequencies. (The diodes are assumed to have finite resistance.) In addition, only two bands of frequencies are allowed: a band including the carrier and sideband frequencies and another band extending from dc up to the highest baseband frequency of interest.

The resistances R_b and R_c play a vital role in the operation of the limiter. By properly proportioning these resistances, the limiting can be made theoretically perfect at one or two input levels. If good diodes are used (diodes with a high ratio of reverse to forward resistance; most germanium and silicon diodes are good by this definition) perfect limiting can be achieved when $R_b = R_c$. From formulae derived in Appendix

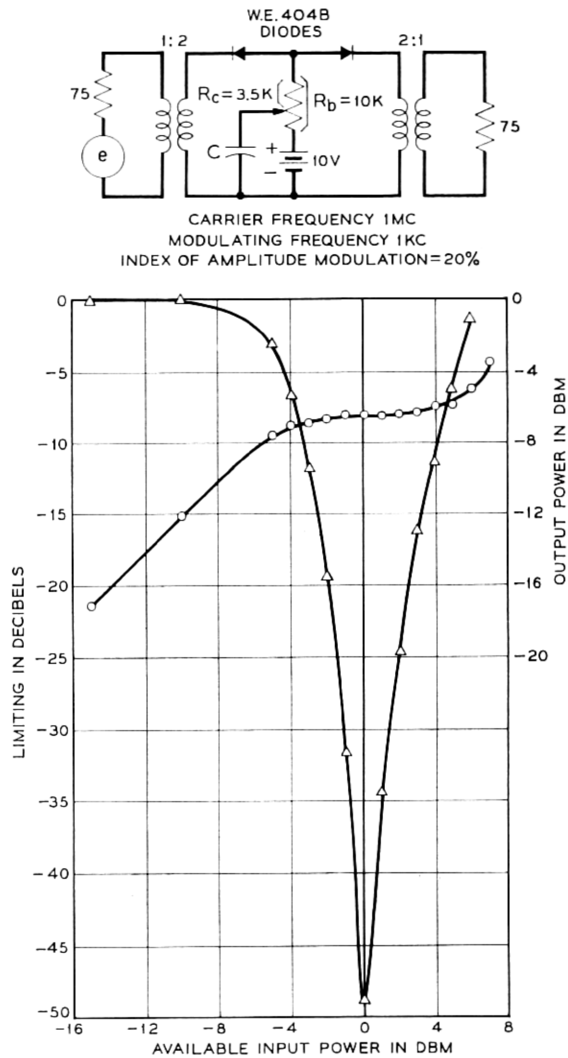

Fig. 7 — Experimental results for practical limiter based on Fig. 5, with R_c adjusted for one limiting peak.

A, several curves of limiting versus input carrier level have been calculated and are plotted in Fig. 6. For all the curves R_b = 10,000 ohms, but R_c is different for each curve. For all such data, limiting is defined as the ratio of the AM index of modulation of the output to the index of modulation of the input signal. In Fig. 6, the curve for R_c = 4500 ohms shows two points of perfect limiting, and this behavior is verified in experimental data presented in Figs. 7 and 8. It is possible to get either one or two peaks, depending upon the value of R_c .

The theory indicates that, if two points of perfect limiting exist, the phase of the AM sidebands reverses at the first point and again at the second. That this actually occurs has been verified by experiment.

Figs. 9 and 10 show similar results obtained for the condition R_b = R_c . Where possible, this is the preferred condition since it requires less by-passing of R_c and R_b , a problem which is especially important in broadband systems.

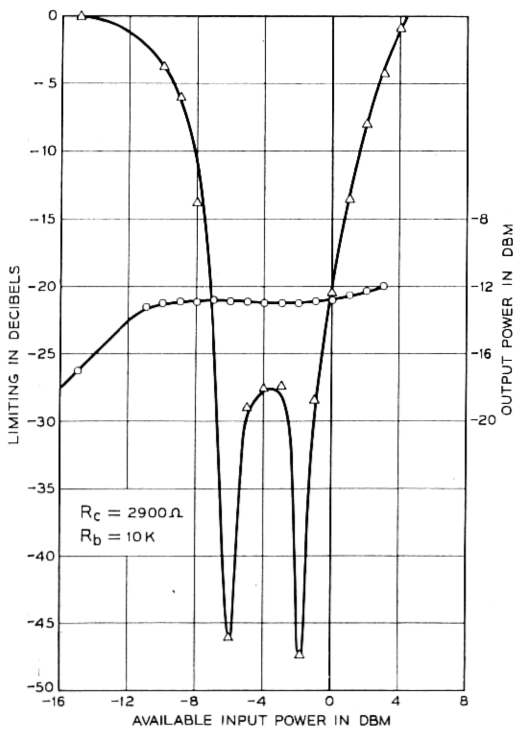

Fig. 8 — Experimental results for practical limiter based on Fig. 5, with R_c adjusted for two limiting peaks.

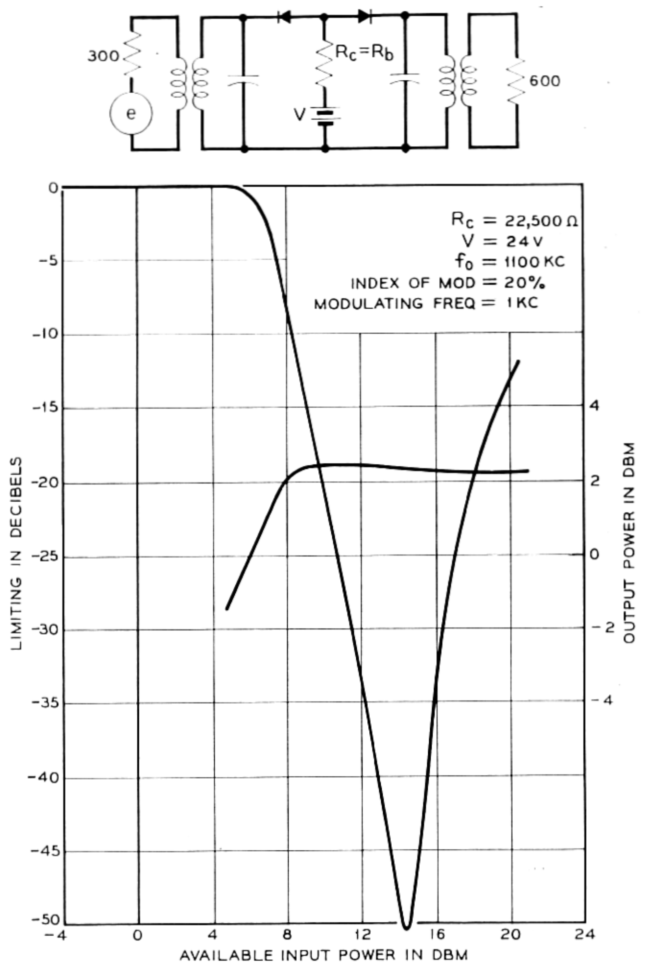

Fig. 9 — Experimental results for practical limiter based on Fig. 5, with $R_b = R_c$ and adjusted for one limiting peak.

Figs. 6 through 10 indicate that perfect limiting occurs only at isolated values of input power for this limiter. No practical difficulty arises, however, because perfect limiting can be approximated as closely as desired over any finite range of input power levels by cascading several stages similar to Fig. 5 and arranging the limiting peaks properly. In Fig. 6 the hump between the two peaks of perfect limiting can be reduced to any desired level by increasing R_c. This will have the effect of reducing the separation of the peaks. After choosing any finite amount

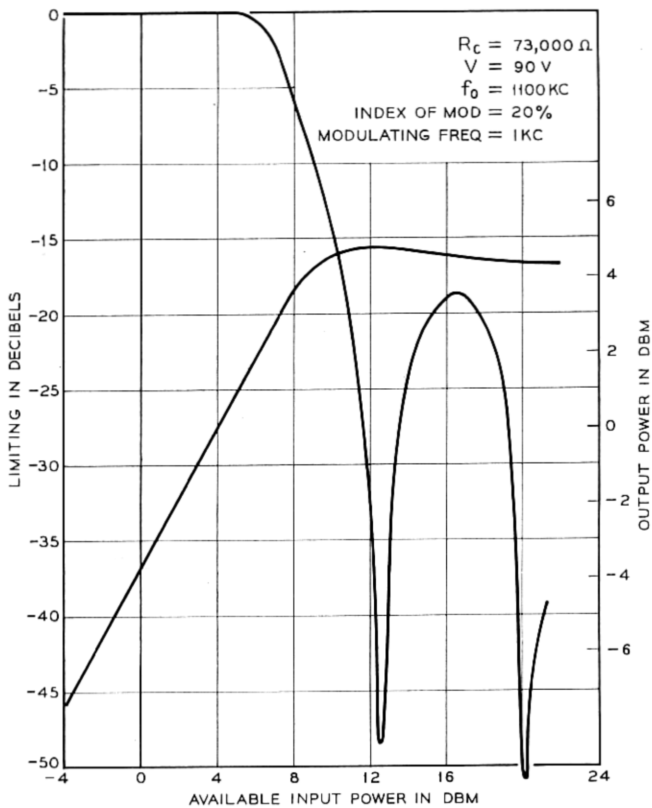

Fig. 10 — Experimental results for practical limiter based on Fig. 5, with $R_b = R_c$ and adjusted for two limiting peaks.

of limiting, R_c can be adjusted so that the limiting at the maximum point of the hump exceeds this value. (The minus sign is usually omitted and we say that 40 db of limiting is more than, or exceeds, 30 db, i.e., the decrease in AM index is greater for 40 db of limiting than for 30 db.) Then, for a continuous range of input levels, the limiting is more than the desired amount. Another stage can be adjusted so that its limiting peaks fall near but to the right of those of the first stage, in such a way that the limiting between the peaks always exceeds the predetermined number. Control of bias current and R_c will be required for this adjustment. If this process is continued it is evident that perfect limiting can be approximated as closely as desired over any finite range of input levels.

In many practical situations, one stage is sufficient because the AGC

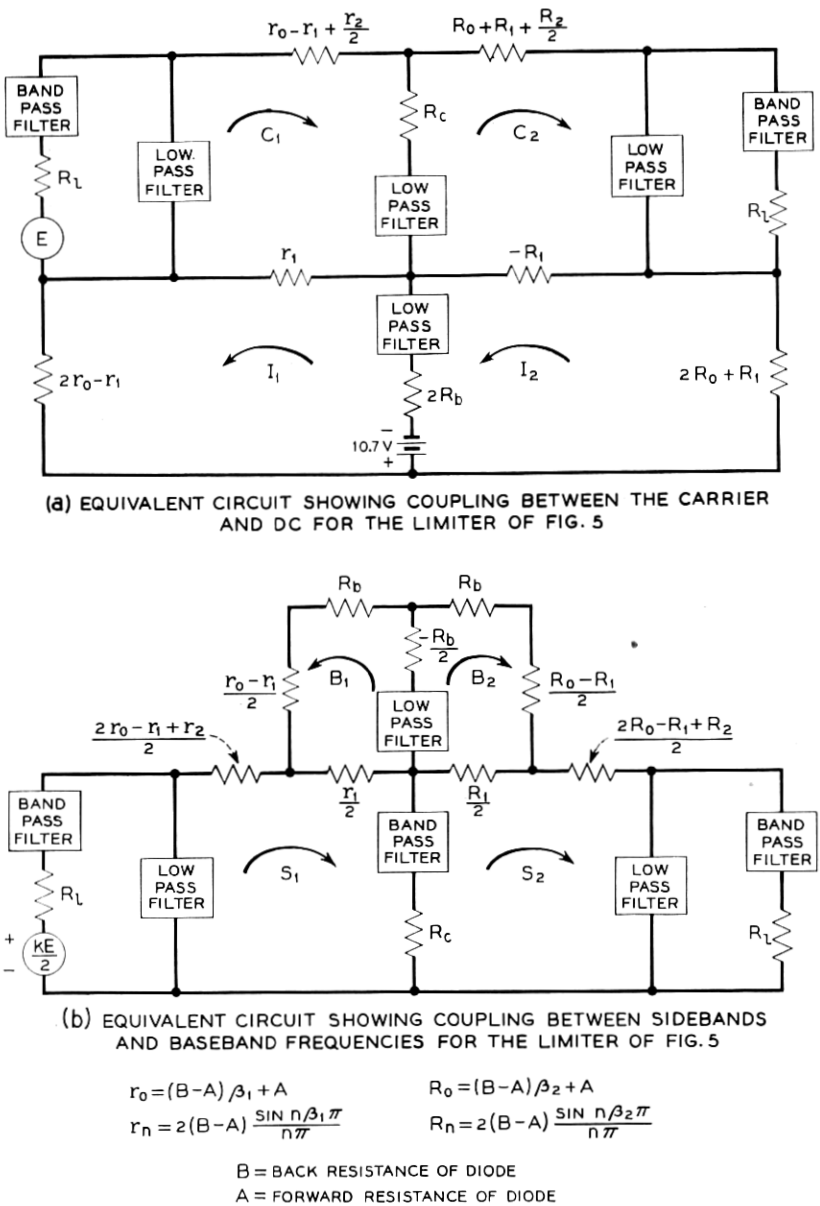

Fig. 11 — Equivalent circuits for limiter of Fig. 5.

of the IF amplifier acts to keep the carrier input level near the value for best limiting.

III. EQUIVALENT CIRCUITS

From equation (13) derived in Appendix A, equivalent circuits for the limiter of Fig. 5 can be drawn. These are shown in Fig. 11. Fig. 12 is the dual of Fig. 5. In Fig. 11 all resistances with numbered subscripts are coefficients in the Fourier expansion of the resistance function of the diodes; lower case r's indicate diode D_1. Resistances with letter subscripts are as shown in Fig. 5(a).

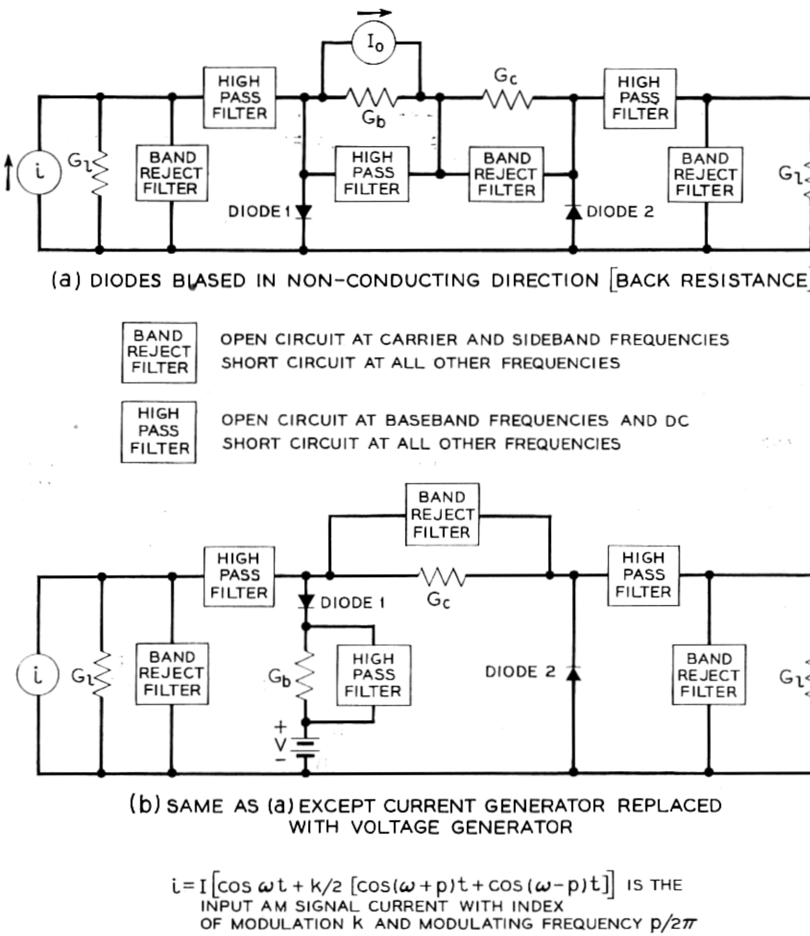

Fig. 12 — Dual of limiter circuit shown in Fig. 5.

The limiting mechanism can be understood by consideration of the circuit involving AM sidebands [Fig. 11(b)]. Some AM sideband energy is coupled from the input mesh 1 to the output mesh 2 through R_c. Some of the remaining sideband energy is coupled to mesh 3 through the $r_1/2$ where it appears at the baseband frequency. This baseband energy in mesh 3 is coupled in turn through $-R_b/2$ into mesh 4, where it is still at the baseband frequency. From mesh 4 it is coupled through $R_1/2$ into the output mesh 2, where it appears at the sideband frequencies in phase opposition to the energy coupled through R_c. By properly proportioning R_b and R_c, the energies coupled by R_c and $-R_b/2$ can be made to cancel, leaving no sideband currents in the output mesh and therefore achieving perfect limiting. The FM signal behaves in the same manner as the carrier and suffers loss only in passing through the limiter.

APPENDIX A

The material in this section is based on Refs. 1 and 2. In Ref. 1 Caruthers develops the superposition principle for low-index modulation; i.e., he shows that circuits containing nonlinear resistive elements can be resolved into equivalent linear systems using a large carrier and small signal amplitude. In Ref. 2 it is shown that a nonlinear resistance can be represented by a Fourier series, the coefficients of which have the dimensions of resistance. If the carrier is large compared to the signal, this resistance is a function of carrier frequency and harmonics only and, to a good approximation, is independent of the signal.

Based on these ideas, the limiter circuit can be solved by usual linear circuit techniques.

A.1 *Superposition*

Any nonlinear resistance has a current voltage characteristic that can be expressed by a power series of the form:

$$e = a_1 i + a_2 i^2 + a_3 i^3 + \cdots + a_n i^n + \cdots. \tag{3}$$

If a carrier and signal current are applied to this nonlinear resistance, each term after the first will produce voltages at new frequencies consisting of the intermodulation products of the carrier and signal. If finite external impedances are presented to these voltages, currents will flow at these new frequencies and produce still more new frequencies. The number of currents can be minimized by presenting infinite impedance to some of the modulation products.

If a carrier current $C \cos ct$ and two signal currents $S_1 \cos s_1 t$ and

$S_2 \cos s_2t$ are applied to the nonlinear resistance the voltage due to the nth term of (3) is

$$\begin{aligned} C_n &= A_n(C \cos ct + S_1 \cos s_1t + S_2 \cos s_2t)^n \\ &= A_nC^n\left(\cos ct + \frac{S_1}{C} \cos s_1t + \frac{S_2}{C} \cos s_2t\right)^n. \end{aligned} \tag{4}$$

We now make the usual assumption of low-index modulation theory. This assumption is that S_1/C and S_2/C are small enough so that terms in the expansion of (4) containing products and powers of S_1/C and S_2/C are negligible. The expansion of (4) is then:

$$C_n \approx A_nC^n\left(\cos^n ct + n \cos^{n-1} ct \frac{S_1}{C} \cos s_1t + n \cos^{n-1} ct \frac{S_2}{C} \cos s_2t\right). \tag{5}$$

The first term in (5) is the dc and the harmonics of the carrier. The second and third terms contain the input signal, the output signal (usually the first order sidebands) and other unwanted modulation products. The important thing to note about (5) is that the input and output voltages (second and third terms) are the same whether the input signals are applied separately and the outputs summed or whether the inputs are applied simultaneously. Therefore, the superposition principle holds and is a direct consequence of the fact that the carrier amplitude is much greater than the signal amplitude.

A.2 *Fourier Series Representation of Nonlinear Resistance*

The power series expansion for the nonlinear resistance as a function of current can be obtained by differentiating (3) with respect to i. The nonlinear resistance is:

$$r(i) = \frac{de}{di} = a_1 + 2a_2i + 3a_3i^2 + \cdots + nA_ni^{n-1} + \cdots, \tag{6}$$

where $r(i)$ is the nonlinear resistance.

The usual assumption made in low-index amplitude modulation theory and the one made here is that only the carrier current is significant in determining the nonlinear resistance. If the carrier current $i = I \cos \omega t$ is substituted into (6) and the series rearranged, the result is a Fourier series. Such a series can be written as

$$r(t) = r_0 + \sum_{n=1}^{\infty} r_n \cos n\omega t. \tag{7}$$

If the current-voltage characteristics of the nonlinear element were expanded in terms of voltage instead of current, an analogous analysis would lead to a representation of the element as a conductance:

$$g(t) = g_0 + \sum_{n=1}^{\infty} g_n \cos n\omega t. \tag{8}$$

Either (7) or (8) can be used to represent a nonlinear resistance in low-index modulation problems. The accuracy of the representation depends upon the index of modulation; the lower the index, the better the representation.

A.3 *Limiter Analysis*

With superposition established and a simple expression for the nonlinear resistance, the rest of the analysis can be done by usual linear circuit techniques. The procedure is as follows:

i. An amplitude-modulated sine wave voltage is applied to the circuit of Fig. 5. No FM need be included, since it is not affected in any way by the limiter except to suffer some attenuation.

ii. A solution is assumed; i.e., currents at certain frequencies are assumed to flow in the two meshes. The magnitudes of these currents are unknown and are to be determined. Currents at all other frequencies are not allowed to flow.

iii. The mesh equations are set up and solved for pertinent currents.

With reference to Fig. 5, the input voltage E is a carrier, amplitude-modulated with a sine wave:

$$e = E \cos \omega t + \frac{k}{2} E \left[\cos (\omega + p)t + \cos (\omega - p)t\right], \tag{9}$$

where k = index of modulation
$\omega/2\pi$ = carrier frequency
$p/2\pi$ = modulating frequency.

The currents that are assumed to flow are:

$$\begin{aligned} i_1 &= C_1 \cos \omega t + S_1 [\cos (\omega + p)t + \cos (\omega - p)t] + B_1 \cos pt + I_1 , \\ i_2 &= C_2 \cos \omega t + S_2 [\cos (\omega + p)t + \cos (\omega - p)t] + B_2 \cos pt + I_2 . \end{aligned} \tag{10}$$

The circuit of Fig. 5 is arranged so that no other currents are allowed to flow. With these currents the IR drops across the nonlinear elements can be calculated. Let the diodes have the following expansions:

$$\begin{aligned} &\text{Diode } D_1\text{: } r_1(t) = r_0 + \sum_{n=1}^{\infty} r_n \cos n\omega t, \\ &\text{Diode } D_2\text{: } r_2(t) = R_0 + \sum_{n=1}^{\infty} (-1)^n R_n \cos n\omega t. \end{aligned} \tag{11}$$

The factor $(-1)^n$ in $r_2(t)$ is due to the fact that the diodes are poled in opposite directions. The voltage drops across the diodes are found by multiplying (10) and (11):

$$\begin{aligned} v_1 = i_1 r_1(t) &= C_1\left(r_0 + \frac{r_2}{2}\right)\cos\omega t \\ &\quad + \left[r_0 S_1 + \frac{r_2}{2} S_1 + \frac{r_1}{2} B_1\right][\cos(\omega + p)t + \cos(\omega - p)t] \\ &\quad + (r_0 B_1 + r_1 S_1)\cos pt + \frac{r_1}{2} C_1 + I_1 r_0 + I_1 r_1 \cos\omega t, \\ v_2 = i_2 r_2(t) &= C_2\left(R_0 + \frac{R_2}{2}\right)\cos\omega t \\ &\quad + \left[R_0 S_2 + \frac{R_2}{2} S_2 - \frac{R_1}{2} B_2\right][\cos(\omega + p)t + \cos(\omega - p)t] \\ &\quad + (R_0 B_2 - R_1 S_2)\cos pt - \frac{R_1}{2} C_2 + I_2 R_0 - I_2 R_1 \cos\omega t. \end{aligned} \tag{12}$$

The next step is to equate the voltage drops at each frequency to the generator emf's in each mesh:

$$\begin{aligned} &\left.\begin{aligned} E &= R_m C_1 - R_c C_2 + I_1 r_1 \\ 0 &= -R_c C_1 + R_x C_2 - I_2 R_1 \end{aligned}\right\} \text{carrier} \\ &\left.\begin{aligned} -V &= \frac{r_1}{2} C_1 + I_1(r_0 + R_b) - I_2 R_b \\ V &= \frac{-R_1}{2} C_2 + I_2(R_0 + R_b) - I_1 R_b \end{aligned}\right\} \text{dc} \\ &\left.\begin{aligned} \frac{kE}{2} &= R_m S_1 - R_c S_2 + \frac{r_1}{2} B_1 \\ 0 &= -R_c S_1 + R_x S_2 - \frac{R_1}{2} B_2 \end{aligned}\right\} \text{sidebands} \\ &\left.\begin{aligned} 0 &= r_1 S_1 + (r_0 + R_b) B_1 - R_b B_2 \\ 0 &= -R_1 S_2 - R_b B_1 + (R_0 + R_b) B_2 \end{aligned}\right\} \text{baseband,} \end{aligned} \tag{13}$$

where

$$R_m = r_0 + \frac{r_2}{2} + R_c + R_l$$

$$R_x = R_0 + \frac{R_2}{2} + R_c + R_l\,.$$

The set of equations (13) can be solved for the pertinent currents:

$$C_1 = \frac{E\left\{R_x[(r_0 + R_b)(R_0 + R_b) - R_b^2] - \frac{R_1^2}{2}(r_0 + R_b)\right\} + V\left(R_1R_cr_0 + r_1R_xR_0 - \frac{r_1R_1^2}{2}\right)}{D},$$

$$C_2 = \frac{E\left\{R_c[(r_0 + R_b)(R_0 + R_b) - R_b^2] - \frac{r_1}{2}R_1R_b\right\} + V\left(R_mR_1r_0 + r_1R_cR_0 - \frac{r_1^2R_1}{2}\right)}{D}, \tag{14}$$

$$S_2 = \frac{\frac{k}{2}E\left\{R_c[(r_0 + R_b)(R_0 + R_b) - R_b^2] - \frac{r_1R_1R_b}{2}\right\}}{D},$$

where

$$D = [(r_0 + R_b)(R_0 + R_b) - R_b^2](R_mR_x - R_c^2) - \frac{R_1^2}{2}R_m(r_0 + R_b) - \frac{r_1^2}{2}R_x(R_0 + R_b) + r_1R_1R_cR_b + \left(\frac{r_1R_1}{2}\right)^2.$$

The insertion loss ratio is

$$\frac{P_{\text{out}}}{P_{\text{avail}}} = \frac{C_2^2R_l}{\frac{E^2}{4R_l}} \tag{15}$$

and the limiting, or amplitude modulation suppression, is

$$\text{limiting} = \frac{S_2}{\frac{k}{2}C_2}. \tag{16}$$

Expression (16) shows that perfect limiting occurs when $S_2 = 0$. The numerator of S_2 as given in (14) indicates the possibility of perfect limiting if R_c and R_b can be adjusted independently.

At this point it is convenient to place some further restrictions on the diode characteristics so that the calculations can be made more easily. The diodes are assumed to switch from a constant forward resistance A to a constant backward resistance B. The fraction of the carrier cycle that the diodes remain on the back resistance is designated β_1, β_2.

With these simplifications, the diode characteristics given in (3) are:

$$r_1(t) = r_0 + \sum_{n=1}^{\infty} r_n \cos n\omega t$$
$$r_2(t) = R_0 + \sum_{n=1}^{\infty} (-1)^n R_n \cos n\omega t, \tag{17}$$

where

$$r_0 = (B - A)\beta_1 + A, \qquad R_0 = (B - A)\beta_2 + A,$$
$$r_n = \frac{2(B - A)}{n\pi} \sin n\beta_1\pi, \qquad R_n = \frac{2(B - A)}{n\pi} \sin n\beta_2\pi,$$

B = back resistance of the diode,
A = forward resistance of the diode,
β = the fraction of the cycle during which the diode is on its back resistance.

All of the currents in (14) are functions of E, β_1 and β_2, where E is the amplitude of the input carrier voltage. In order to make calculations for (14), relations among E, β_1 and β_2 must be derived.

When the input to the limiter is very low, both diodes are on their forward resistances, due to the forward bias current I_0. At these low levels β_1, β_2 are both zero and, from (14),

$$\frac{C_2}{C_1} = \frac{R_c}{R_c + R_l}. \tag{18}$$

Because the circuit is resistive and because only the carrier current is involved in switching the diodes, (18) holds for all instantaneous loop currents less than I_0. From the definition of β, the diodes switch to their back resistances when

$$C_1 \cos(-\beta_1\pi) = I_0, \qquad C_1 \geqq I_0,$$
$$C_2 \cos(-\beta_2\pi) = I_0, \qquad C_2 \geqq I_0. \tag{19}$$

At the switching point, both (18) and (19) hold. Therefore, by substituting (18) into (19), a relation between β_1, β_2 follows:

$$\cos \beta_1\pi = \frac{R_c}{R_c + R_l} \cos \beta_2\pi. \tag{20}$$

In a similar manner, a relation can be derived for E, β_1. When $C_1 = I_0$, $\beta_1 = \beta_2 = 0$ and, from (14),

$$E_0 = I_0 \left. \frac{(R_m R_x - R_c^2)}{R_x} \right|_{(\beta_1 = \beta_2 = 0)}. \tag{21}$$

Now, for any $E \geqq E_0$, diode D_1 switches to its back resistance when

$$E \cos \beta_1 \pi = E_0 = I_0 \left. \frac{[R_m R_x - R_c^2]}{R_x} \right| (\beta_1 = \beta_2 = 0), \tag{22}$$

where I_0 is the bias current in each diode and is given by (13) for $\beta_1 = \beta_2 = 0$. Thus,

$$I_0 = \frac{V}{2R_b + A}, \tag{23}$$

where

V = dc bias voltage
A = forward resistance of diode.

Curves of power output and amplitude modulation suppression have been calculated and are shown in Fig. 6.

A.4 *Circuit Simplification*

A natural question to ask about the limiter is, "Can perfect limiting be obtained when $R_b = R_c$?" If so, the circuit is much simpler. The answer to this question is yes, and this can be demonstrated by considering the numerator of S_2. Perfect limiting is achieved when this numerator is zero. Setting the numerator of S_2 equal to zero and letting $R_b = R_c$, the resulting expression is solved for R_c:

$$R_c = \frac{2r_0 R_0 \left(\frac{r_1 R_1}{2r_0 2R_0} - \frac{1}{2} \right)}{r_0 + R_0}, \tag{24}$$

where

$$R_c \geqq 0. \tag{25}$$

Using this fact in (24), it follows that perfect limiting can be obtained when $R_b = R_c$ if the following inequality is satisfied:

$$\frac{r_1}{2r_0} \frac{R_1}{2R_0} \geqq \frac{1}{2}. \tag{26}$$

Let us examine this further by using the idealized diode characteristics of (17). Substituting these expressions into (26), we get

$$\left(\frac{\sin \beta_1 \pi}{\beta_1 \pi} \right) \frac{1}{\left(1 + \frac{A}{2(B - A)\beta_1} \right)} \left(\frac{\sin \beta_2 \pi}{\beta_2 \pi} \right) \frac{1}{\left(1 + \frac{A}{2(B - A)\beta_2} \right)} \geqq \frac{1}{2}. \tag{27}$$

This inequality can be met easily with good diodes, i.e., diodes in which $B/A \gg 1$.

Figs. 9 and 10 show a limiter and experimental results for the case $R_b = R_c$. It should be noted that, even with poor diodes, the limiting can be perfect if R_b and R_c are allowed to differ. This can be seen in the expression for S_2.

A.5 *The "Clipper" Limiter*

A well-known limiter has the configuration of Fig. 1. The behavior of such a "clipper" limiter can be examined by setting $R_b = R_c$ and letting R_b and V increase without bound in such a way as to maintain the initial bias current I_0 unchanged. This will also insure that $\beta_1 = \beta_2$ and $r_i = R_i$.

If these exercises are performed on (14) and the results substituted into (16), the limiting is found to be

$$\text{limiting} = \frac{E}{E + 2r_1 I_0}, \tag{28}$$

where

$$E = \text{carrier voltage amplitude,}$$

$$I_0 = \text{constant current dc bias,}$$

$$r_1 = [2(B - A)/\pi] \sin \beta_1 \pi.$$

Several remarks can be made concerning the "clipper limiter" of (28):

i. Before the diodes begin to switch, $r_1 = 0$ and there is no limiting. This is to be expected.

ii. If $I_0 = 0$ there is no limiting. This too is to be expected because, with no bias, one diode has its back resistance in the circuit while the other one has its forward resistance in the circuit, and vice versa. The circuit is linear under these conditions and cannot limit.

iii. Expression (28) indicates that the larger r_1, the better will be the limiting for a given I_0. But r_1 is proportional to $(B - A)$, so this indicates a good diode for a limiter of this type is one with a high back resistance B and a high ratio of back-to-front resistance.

iv. If the diodes have a finite back resistance, perfect limiting cannot be achieved. Furthermore, as β_1 approaches 1/2, the limiting tends to zero. The reason for this is that r_1 is proportional to $\sin \beta_1 \pi$, which tends to a constant value as β_1 approaches 1/2 while E increases without limit.

The output carrier current can be calculated in the same way as (28). It is

$$C_2 = \frac{E + 2r_1 I_0}{2r_0 + r_2 + R_l}. \tag{29}$$

Comments can also be made concerning this equation:

i. At low levels of input voltage the diodes do not switch and $r_1 = r_2 = 0$, and $r_0 = A$. The output is then $C_2 = E/[2(R_l + A)]$, as expected.

ii. If $I_1 = 0$, then $\beta_1 = \beta_2 = 1/2$, resulting in $r_0 = (B + A)/2$ and $r_2 = 0$. In this case $C_2 = E/(B + A + 2R_l)$ and the circuit is linear.

In the above discussion of the "clipper" limiter, all harmonics of the carrier were suppressed. Two other cases have been considered: in one case, the third harmonic was allowed and, in the other, the third and fifth harmonics were allowed (even harmonics cancel). The analysis follows the same pattern as previously demonstrated except that currents at all allowed frequencies are assumed. Only the results will be shown.

When the third harmonic is allowed, limiting is given by:

$$L = \frac{1}{1 + \dfrac{\cos \beta\pi}{R_l + A}\left[r_1 - \dfrac{r_3(r_2 + r_4)}{2R_l + 2r_0 + r_6}\right]}. \tag{30}$$

Similarly, when the fifth harmonic is added,

$$\begin{aligned} L = \Bigg(1 + \frac{\cos \beta\pi}{R_l + A}\, \{r_1 - r_3[(r_2 + r_4)(2R_l + 2r_0 + r_{10}) \\ - (r_2 + r_8)(r_4 + r_6)] \\ - r_5[(r_2 + r_4)(r_2 + r_8) - (r_4 + r_6)(2R_l + 2r_0 + r_6)]\}\Bigg)^{-1} \end{aligned} \tag{31}$$

Equations (30) and (31) are plotted in Fig. 3. The curve for all harmonics was calculated by a Fourier analysis of the "clipped" wave.

APPENDIX B

In this Appendix it will be shown that the amount of limiting is not necessarily given by the slope of the carrier transfer function.

Assume that the limiter of Fig. 5 is adjusted for perfect limiting for the carrier amplitude $E = E_1$. In the equivalent circuits of Fig. 11, it is noted that the output carrier current C_2 depends upon the dc coupling through R_b. Note further that the sideband current S_2 depends upon the baseband coupling through R_b.

By proper filtering, the dc and baseband currents can be made to flow through separate resistors without changing the behavior of the limiting or carrier transfer functions. With respect to R_b, the equivalent circuits of Fig. 11 are then completely independent. If the baseband resistance in Fig. 11(b) is changed, S_2 will clearly change. But the carrier transfer function is undisturbed, since no change was made in Fig. 11(a). Therefore, the limiting is not completely determined by the slope of the carrier transfer function.

This behavior has been verified experimentally using a limiter similar to that of Fig. 7.

REFERENCES

1. Caruthers, R. S., Copper Oxide Modulators in Carrier Telephone Systems, B.S.T.J., **18,** April 1939, p. 315.
2. Peterson, E. and Hussey, L. W., Equivalent Modulator Circuits, B.S.T.J., **18,** January 1939, p. 32.
3. Arguimbau, L. B. and Granlund, J., *Interference in Frequency-Modulation Reception*, Res. Lab. Elec., Mass. Inst. of Tech., Report 42, Cambridge, Mass., 1949.
4. Kahn, L. R., Analysis of a Limiter as a Variable-Gain Device, Elec. Eng., **72,** December 1953, p. 1106.
5. Tucker, D. G., Linear Rectifiers and Limiters, Wireless Engr., **29,** May 1952, p. 128.
6. Seeley, S. W. and Avins, J., The Ratio Detector, RCA Review, **8,** June 1947, p. 201.