

# The Transient Response of Damped Linear Networks with Particular Regard to Wideband Amplifiers

W. C. ELMORE\*

University of California, Los Alamos Scientific Laboratory, Santa Fe, New Mexico

(Received April 11, 1947)

When the transient response of a linear network to an applied unit step function consists of a monotonic rise to a final constant value, it is found possible to define *delay time* and *rise time* in such a way that these quantities can be computed very simply from the Laplace system function of the network. The usefulness of the new definitions is illustrated by applications to low pass, multi-stage wideband amplifiers for which a number of general theorems are proved. In addition, an investigation of a certain class of two-terminal interstage networks is made in an endeavor to find the network giving the highest possible gain—rise time quotient consistent with a monotonic transient response to a step function.

How to exactly define  $G_2(s)$ ?

## 1. INTRODUCTION

THE transient behavior of any linear system (or network) is contained implicitly in the system function  $F(s)$  which expresses directly the steady-state (sinusoidal) response of the system. The variable in the system function,  $s = \sigma + j\omega$ , is the complex angular frequency;  $\omega$  is the ordinary (real) angular frequency, and  $\sigma$  is a real variable introduced for the purpose of facilitating the transient analysis of the system.<sup>1</sup> In the present paper we shall be concerned primarily with the class of linear systems in which the transient response to a unit step function (the so-called *indicial admittance*) consists of a *monotonic* rise to a final constant value. For simplicity in presentation only the transient response of a low pass, wideband amplifier will be considered. Many of the results obtained, however, apply equally well to other electrical systems, as well as to mechanical, acoustical, thermal, and to mixed systems, provided only that they are linear and have a monotonic transient response to a unit step function.

The most important system function of an amplifier is the complex gain,  $G(s)$ , connecting input and output voltages of the form  $Ee^{st}$ . In the case of a low pass amplifier,  $G(s)$  can always be separated into two factors,  $G_1(s)$ , which governs the response at low frequencies, and  $G_2(s)$ , which governs the response at high frequencies. In an

unfedback amplifier,  $G_1(s)$  owes its origin to various  $RC$  networks which couple the plate of one tube to the grid of the next, and which furnish bias voltages to various points in the amplifier. The system function  $G_2(s)$  owes its origin primarily to parasitic interstage capacitances which shunt the signal-carrying leads. Since we shall be interested in the problem of obtaining the greatest possible *gain—rise time quotient* for an amplifier,<sup>2</sup>  $G_2(s)$  may reflect the presence of compensating inductances, of feedback, or of any other circuit arrangements used to shorten the rise time or to improve the transient properties of the amplifier. The portion  $G_2(s)$  of the system function may be considered as that of an equivalent amplifier idealized to have perfect low frequency response.

For convenience in analysis, we shall use the *normalized* system function  $g_2(s) = G_2(s)/G_2(0)$ , where  $G_2(0)$  is the gain of the idealized amplifier at zero frequency. Normalization evidently makes the final value of the response to a unit step function (given by the final-value theorem of the Laplace transformation) also unity.

It is not difficult to show that the normalized system function  $g_2(s)$  of a stable amplifier containing a finite number of lumped circuit elements takes the form

$$g_2(s) = \frac{1 + a_1s + a_2s^2 + \dots + a_ns^n}{1 + b_1s + b_2s^2 + \dots + b_ms^m} \quad (1)$$

<sup>2</sup> The gain—rise time quotient is analogous to the more familiar gain-band width product, but appears to be a more useful measure of amplifier performance in the case of amplifiers designed to amplify fast transients. The definition of rise time is considered in Section 2.

\* Now at Swarthmore College, Swarthmore, Pennsylvania.

<sup>1</sup> The notation and terminology adopted here is that found in M. F. Gardner and J. L. Barnes, *Transients in Linear Systems*, (John Wiley and Sons, Inc., New York, 1942), Vol. 1.

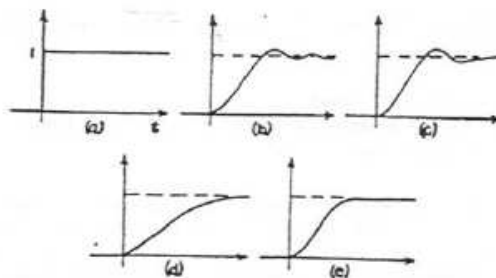


FIG. 1. Some typical transient response curves.

where the coefficients  $a$ , and  $b$ , are real,  $m > n$  and the poles of  $g_2(s)$  all lie in the left half of the complex  $s$ -plane. The normalized transient response of the amplifier to the unit step function  $u(t)$  can be computed by means of the inverse Laplace transformation

$$e(t) = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} \frac{1}{s} g_2(s) e^{st} ds, \quad c > 0. \quad (2)$$

Transient response curves computed from Eq. (2) for various amplifiers have a variety of shapes, some common forms of which are illustrated in Fig. 1. The input signal,  $u(t)$ , is shown in (a). The transient response shown in (b) consists of a delayed rise, followed by a train of damped oscillations. The response shown in (c) is similar to that in (b) except that only a finite number of oscillations occur, preceding a gradual approach of the curve to the final value unity. In (d) and (e) are illustrated monotonic transient response curves having different amounts of damping. The response in (e) is supposed to be that of an amplifier having certain adjustable circuit parameters which have been chosen to achieve the shortest possible monotonic rise for a given amplifier gain.

Any circuit elements introduced in an amplifier for the purpose of controlling the shape of the transient response curve may be termed *compensating elements*. In the present instance they afford high frequency compensation to the response of the amplifier. When the fastest possible monotonic rise has been obtained with the particular type of compensation used, the amplifier is said to be *critically compensated*. If the transient response is monotonic but the rise is slower than can be obtained by suitably adjusting the com-

pensating elements, the amplifier will be referred to as *under-compensated*. Finally, if the transient response is not monotonic, usually as the result of increasing the compensating elements beyond the point giving critical compensation, the amplifier will be referred to as *over-compensated*.<sup>3</sup>

It is evident that the various types of transient curves illustrated in Fig. 1 possess certain common features, in particular, a delay which occurs before the response is well under way, and a finite time of rise. For many purposes each curve can be sufficiently well characterized by its *delay time* and *rise time*, which can be defined in several different, but approximately equivalent ways. One of the purposes of the present paper is to propose useful definitions for these quantities, with a view to facilitating their computation from the system function  $g_2(s)$ . The new definitions, unfortunately, are of such a nature that they apply only to systems which are not over-compensated. Their utility for all systems having a monotonic transient response, however, appears to be great enough to outweigh this defect. It is possible that an equally useful method for treating the over-compensated case can be discovered.

## 2. THE DEFINITION OF DELAY TIME AND OF RISE TIME

A number of definitions of delay time and a rise time appear to be in practical use. Two of these will be illustrated by reference to Fig. 2 which shows the transient response  $e(t)$  to the unit step function, and its derivative,  $e'(t)$ , of an under-compensated amplifier.<sup>4</sup>

The delay time,  $T_D$ , is usually defined as the time required for the response to reach half its final value, as illustrated in Fig. 2a. The rise time,  $T_R$ , is sometimes defined as the reciprocal of the slope of the tangent drawn to the response curve at its half-value point, again as illustrated in Fig. 2a. A somewhat more practical definition

<sup>3</sup> For many applications it is important to avoid over-compensation in an amplifier. This is particularly true for pulse amplifiers used in nuclear physics (to amplify pulses obtained from an electrical detector of radiation) and for wideband amplifiers used in studying fast electrical transients (such as amplifiers for cathode-ray oscillographs). Video amplifiers used in television applications are not as critical in this respect since small oscillations resulting from over-compensation do not impair the quality of television pictures.

<sup>4</sup> The curve  $e'(t)$  may be considered to be the response of the amplifier to a unit impulse applied at time  $t=0$ .

results if  $T_R$  is taken to be the time required for the response to increase from 10 to 90 percent of its final value.<sup>6</sup> Although these definitions are useful in the laboratory, they are extremely awkward for making computations, or for entering upon a theoretical investigation of the relative merits of various methods of compensating an amplifier to reduce its rise time. The difficulty, of course, lies in the necessity for computing the transient response curve for each case under consideration, a formidable undertaking. It is practically impossible to obtain values of  $T_D$  and  $T_R$ , as defined, by any simple method of analysis.

Let us now consider alternative definitions for delay time and rise time. Evidently the delay time should be measured from  $t=0$  to some time at which the transient rise is about one-half over. It is reasonable, therefore, to measure  $T_D$  to the centroid of area of the curve  $e'(t)$ , that is,

$$T_D = \int_0^{\infty} t e'(t) dt. \quad (3)$$

The formula for the centroid takes this simple form since

$$\int_0^{\infty} e'(t) dt = 1.$$

This definition of delay time is illustrated in Fig. 2b, and it is seen to give a result which differs but little from that obtained from the customary definition. The two values of delay time depart most markedly in the case of a very asymmetrical response curve. It is easy to convince oneself that the new definition becomes meaningless if the curve  $e'(t)$  possesses a negative portion, i.e.,  $e(t)$  is not monotonic. It will be shown presently that it is a simple matter to obtain a value of the integral in Eq. (3) directly from the system function  $g_2(s)$ .

The rise time  $T_R$  should express in a prescribed manner the time required for the transient rise to occur. Now the shorter the rise time, the narrower (and higher) the curve of  $e'(t)$ . It is reasonable, therefore, to define  $T_R$  as proportional to the radius of gyration of the area under the curve,

that is,

$$T_R^2 = \text{Const.} \int_0^{\infty} (t - T_D)^2 e'(t) dt. \quad (4)$$

In other words, the rise time is proportional to the standard deviation of the response  $e'(t)$ . The constant of proportionality is chosen to be  $2\pi$  for the following reason: it is possible to show that the curve  $e'(t)$  for any  $n$ -stage amplifier<sup>7</sup> approaches more and more closely the form of a Gauss error curve with increasing  $n$ .<sup>7</sup> To make the new definition of rise time agree with the definition based on the slope of the transient response curve (Fig. 2a), the value of  $T_R$  should therefore be

$$T_R = \frac{1}{e'(t)|_{\max}} = (2\pi)^{1/2} [\text{radius of gyration of } e'(t)],$$

which expresses the relation between the height and the radius of gyration of a Gauss error curve of unit area, here denoted temporarily by  $e'(t)$ . Equation (4) can now be written

$$T_R = \left\{ 2\pi \left[ \int_0^{\infty} t^2 e'(t) dt - T_D^2 \right] \right\}^{1/2}, \quad (5)$$

where the integral has been expressed in terms of moments about the time origin. It is found in most instances that rise times computed from Eq. (5) differ by less than ten percent from the

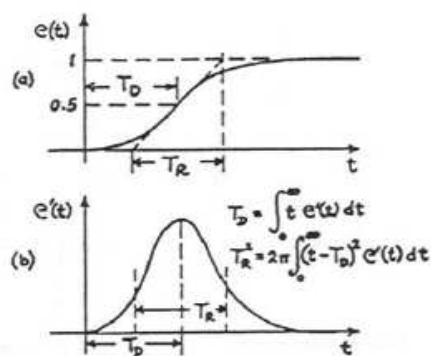


FIG. 2. Curves illustrating the definitions of delay time and of rise time.

<sup>6</sup> The individual stages in the amplifier must each have a monotonic transient response to the unit step function.

<sup>7</sup> This result appears to have been first noticed by Henry Wallman, and will be discussed in Chap. 7, Vol. 18 of the *Radiation Laboratory Series* (McGraw-Hill Book Company, Inc., New York, in press).

<sup>8</sup> H. E. Kallman and R. E. Spencer, Proc. I.R.E. 33, 169-195 (1945).

rise times defined earlier, which can continue to be used for most laboratory work.

The great usefulness of the new definitions of delay time and rise time will now be demonstrated. The system function  $g_2(s)$  and the transient response  $e'(t)$  are related by the direct Laplace transformation

$$g_2(s) = \int_0^{\infty} e'(t) e^{-st} dt, \quad (6)$$

where  $\sigma$ , the real part of  $s$ , must be greater than  $\sigma_1$ , the real part of the pole  $s_1$  of  $g_2(s)$  lying farthest to the right in the  $s$ -plane. In the case of a stable amplifier,  $\sigma_1$  is *negative*, in fact, for a monotonic transient response, the poles of  $g_2(s)$  all lie on the negative real axis of the  $s$ -plane. Let us now expand the Laplace integral (6) in a power series in  $s$ , which will be a valid expansion of  $g_2(s)$  for values of  $s$  lying within the circle of convergence  $|s| = |s_1|$ . We do this by first expanding  $e^{-st}$  in a power series in  $st$  and then integrating term by term, obtaining the development

$$g_2(s) = 1 - s \int_0^{\infty} t e'(t) dt + \frac{s^2}{2!} \int_0^{\infty} t^2 e'(t) dt - \dots \quad (7)$$

It follows that if a given system function is expanded in ascending powers of  $s$ , it is a simple matter to obtain by inspection the first and second moments of  $e'(t)$  about the time origin,<sup>8</sup>

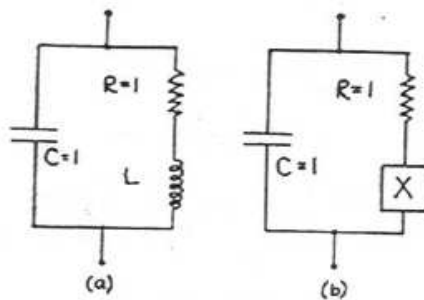


FIG. 3. Shunt-compensated interstage networks.

<sup>8</sup> It has been called to the attention of the writer that this method of computing moments is closely related to methods used in mathematical statistics, the method of the moment generating function, and the method of the characteristic function. See, for instance, S. S. Wilks, *Mathematical Statistics* (Princeton University Press, Princeton, New Jersey, 1943).

The mathematical steps leading to Eq. (7) can be made more rigorous by noticing that the series for the integrand

and therefore to obtain values of  $T_D$  and  $T_R$  defined by Eqs. (3) and (5), respectively. Part of the virtue of the proposed definitions lies in the ready way in which delay times and rise times can be computed. Other advantages of the definitions will be made use of in Section 3.

It is useful to obtain expressions for  $T_D$  and  $T_R$  for a system function of the form given by Eq. (1). By expanding Eq. (1) in ascending powers of  $s$ , it is found that

$$T_D = b_1 - a_1, \quad (8)$$

and that

$$\frac{T_R^2}{2\pi} = b_1^2 - a_1^2 + 2(a_2 - b_2). \quad (9)$$

Before considering other matters, let us compute values of  $T_D$  and  $T_R$  for a single-stage amplifier having a two-terminal plate load impedance of the type shown in Fig. 3a.<sup>9</sup> Such an amplifier stage is said to be *shunt compensated*. The system function of the single stage is identical to the driving point impedance of the plate load, since it can be assumed that the amplifying tube is equivalent to a constant current generator. Hence we have that

now  
to get  
this?

$$g_2(s) = \frac{1 + Ls}{1 + s + Ls^2} \quad (10)$$

In order that no transient oscillation of the type shown in Fig. 1b shall exist, the poles of  $g_2(s)$  must lie on the negative real axis of the  $s$ -plane. This requires that in Eq. (10)  $L \leq 1/4$ . The values of  $T_D$  and  $T_R$  (computed using Eqs. (8) and (9)) are

$$\left. \begin{aligned} T_D &= 1 - L, \\ T_R &= (2\pi(1 - 2L - L^2))^{1/2}. \end{aligned} \right\} \quad (11)$$

When  $L=0$ , corresponding to a simple resistance-coupled amplifier stage,  $T_D=1$ , and  $T_R=(2\pi)^{1/2}=2.51$ . When  $L=1/4$ , corresponding to critical shunt compensation,  $T_D=3/4$ , and  $T_R=(2\pi)^{1/2}(7/16)^{1/2}=1.66$ . To express the improve-

converges *uniformly*, thus permitting term-wise integration between zero and a finite upper limit  $T$ . Since the integrated series, considered as a function of  $T$ , is also uniformly convergent when  $s$  is restricted to the region near the origin, the series must converge, as  $T$  approaches infinity, to the Laplace integral (6) from which it is derived.

<sup>9</sup> By setting  $C=1$ ,  $R=1$ , and expressing  $L$  in units of  $RC$ , values of  $T_D$  and  $T_R$  are obtained in units of  $RC$ . This device enables the system function to be written immediately in a simple, normalized form.

ment realized by compensating the stage it is convenient to define the *rise-time-figure-of-merit*  $S$ , which is the ratio of the *rise time* of an  $R$ -coupled stage to the rise time of the same stage (i.e., a stage with the same value of  $R$  and  $C$ ) compensated to reduce its rise time. Evidently  $S = (16/7)^{1/2} = 1.51$  for a critically shunt-compensated stage. In Section 4 an attempt will be made to discover an interstage network which gives the smallest value of  $T_R$  with a given interstage parasitic capacitance and load resistance, that is, the largest value of  $S$ . The problem is somewhat analogous to that of discovering the network which leads to the maximum band width (without regard to good transient response).<sup>10</sup>

### 3. SOME THEOREMS REGARDING MULTI-STAGE AMPLIFIERS

We have just seen how the delay time and rise time of a single amplifier stage can be computed. Let us now consider how the delay time and rise time of an unfeedback multi-stage amplifier depends on the properties of individual stages in it.

If the amplifier contains  $n$  stages in tandem, the system function of the entire amplifier is the product of the system functions of the individual stages.<sup>11</sup> Let the system function of the  $i$ th stage be  $g_{1i}(s)$ , and let the corresponding values of delay time and rise time be  $T_{D_i}$  and  $T_{R_i}$ , respectively. The function  $g_{1i}(s)$  can be expanded in the series

$$g_{1i}(s) = 1 - sT_{D_i} + \frac{s^2}{2} \left( \frac{T_{R_i}^2}{2\pi} + T_{D_i}^2 \right) - \dots, \quad (12)$$

which is obtained directly from Eqs. (3), (5) and (7). The system function of the entire amplifier therefore becomes

$$g_1(s) = \prod_1^n g_{1i}(s) \\ = 1 - s \sum T_{D_i} + \frac{s^2}{2} \left[ \sum \frac{T_{R_i}^2}{2\pi} + \sum T_{D_i}^2 + 2 \sum_{i > j} T_{D_i} T_{D_j} \right] - \dots$$

<sup>10</sup> See, for instance, H. W. Bode, *Network Analysis and Feedback Amplifier Design*, (D. Van Nostrand Company, Inc., New York, 1945), pp. 408, et seq.

<sup>11</sup> This statement is true provided that no coupling between stages exists except through the electron stream in the constituent amplifying tubes. This situation can be realized in practice if the tubes in the amplifier are pentodes.

By again using Eqs. (3), (5) and (7), the delay time and the rise time of the entire amplifier are found to be

$$T_D = \sum_1^n T_{D_i}, \quad (13)$$

and

$$T_R = \left( \sum_1^n T_{R_i}^2 \right)^{1/2}. \quad (14)$$

The result expressed by Eq. (13) is intuitively obvious, since it is to be expected that the total delay is the sum of the delays of the individual stages. The manner of combining rise times indicated by Eq. (14) is not as evident, although the fact that this simple mode of combination is the correct one has been proposed by several of the author's colleagues prior to the present proof of the theorem.

Another theorem of practical importance concerns the manner in which the gain of an  $n$ -stage amplifier should be distributed among the individual stages in order to achieve the shortest possible over-all rise time for a given over-all gain. Now the rise time of any stage in the amplifier varies directly with the gain of the stage, since both quantities are proportional to the value of the plate load resistor. It is desired, therefore, to minimize the expression (14) subject to the condition that

$$\prod_1^n T_{R_i} = \text{Constant}. \quad (15)$$

It is easy to prove from Eqs. (14) and (15) that the over-all rise time is a minimum when the rise times of all stages are made the same.<sup>12</sup> If  $T_{R_1}$  is the rise time of each stage, the rise time of an  $n$ -stage amplifier becomes

$$T_R = T_{R_1} n^{1/2}. \quad (16)$$

Let us now consider certain matters regarding the design of an amplifier consisting of  $n$  identical stages. We shall treat the simple case where the interstage couplings are of the general type

<sup>12</sup> For instance, by using Lagrange's method of undetermined multipliers, the differential of Eq. (14) must be zero,  $\sum T_{R_i} dT_{R_i} = 0$ , subject to the condition  $(\prod T_{R_i}) \sum dT_{R_i} / T_{R_i} = 0$  which is the differential of Eq. (15). After multiplying the latter equation by the undetermined multiplier  $\alpha$  and adding it to the former, each coefficient of  $dT_{R_i}$  must be identically zero, giving  $T_{R_1} = T_{R_2} = \dots$ . A proof that this condition leads to a *minimum* rise time is scarcely needed.

illustrated in Fig. 3, i.e., a parasitic capacitance  $C$ , and a resistance  $R$  in series with some sort of compensating reactance whose impedance becomes zero at zero frequency. At frequencies where  $1/\omega C \ll R$ , the gain of each stage is

$$G_1 = g_m R, \quad (17)$$

where  $g_m$  is the transconductance of the amplifying tube.<sup>13</sup> The rise time of each stage can be written in the form

$$T_1 = \frac{1}{S} (2\pi)^{\frac{1}{2}} RC, \quad (18)$$

where  $S$  is the rise-time figure-of-merit of the stage. By definition  $S=1$  for a simple  $R$ -coupled stage and we have already shown, for example, that  $S=(16/7)^{\frac{1}{2}}$  for a critically shunt-compensated stage. Eliminating the resistance  $R$  between Eqs. (17) and (18), we have that

$$G_1 = S \left( \frac{g_m}{(2\pi)^{\frac{1}{2}} C} \right) T_1. \quad (19)$$

The quantity  $g_m/(2\pi)^{\frac{1}{2}} C$  expresses the figure-of-merit of the amplifier tubes, and may be conveniently stated as so much gain per microsecond rise time.

If  $T$  is the rise time of the  $n$ -stage amplifier, then according to Eq. (16) the rise time of each stage must be  $T_1 = T/n^{\frac{1}{2}}$ , requiring a gain for each stage

$$G_1 = S \left( \frac{g_m}{(2\pi)^{\frac{1}{2}} C} \right) \frac{T}{n^{\frac{1}{2}}}. \quad (20)$$

Equation (20) can be written as a pair of equations,

$$\left. \begin{aligned} G_1 &= \frac{S}{n^{\frac{1}{2}}} G_0, \\ G_0 &= \left( \frac{g_m}{(2\pi)^{\frac{1}{2}} C} \right) T. \end{aligned} \right\} \quad (20a)$$

where

The quantity  $G_0$  is the gain of a single  $R$ -coupled stage of rise time  $T$ , whereas  $G_1$  is the gain that each stage of the  $n$ -stage amplifier must have in

<sup>13</sup> It is assumed that  $R \ll r_p$ , the plate resistance of the tube.

order that the entire amplifier shall have the rise time  $T$ .<sup>14</sup>

The total gain of the amplifier is

$$G_t = G_1^n. \quad (21)$$

Let us now investigate what gain per stage will result in the shortest rise time for a given total amplifier gain. From Eqs. (20) and (21), we have that

$$T = \frac{1}{S} \left( \frac{(2\pi)^{\frac{1}{2}} C}{g_m} \right) n^{\frac{1}{2}} G_t^{1/n}. \quad (22)$$

On minimizing  $T$  with respect to  $n$  while keeping  $G_t$  constant, it is found that  $n = 2 \ln G_t$ , or that

$$G_1 = e^{\frac{1}{2}} = (2.72 \dots)^{\frac{1}{2}} = 1.65 \dots. \quad (23)$$

This result is independent of the degree of compensation used, provided, of course, that critical compensation is not exceeded. The minimum rise time which can be obtained for a total gain  $G_t$  is found from Eqs. (20), (21), and (23) to be

$$T_{\min} = \frac{1}{S} \left( \frac{(2\pi)^{\frac{1}{2}} C}{g_m} \right) (2e \ln G_t)^{\frac{1}{2}}, \quad (24)$$

requiring a total of  $n = 2 \ln G_t$  stages.

#### 4. SOME CRITICALLY COMPENSATED INTERSTAGE NETWORKS

There are two matters of considerable interest concerning interstage networks of a critically compensated wideband amplifier. The first is primarily of theoretical importance and concerns the maximum value that can be obtained for the quantity  $S$  (the ratio of the rise time obtained with a simple RC network to that obtained with a compensated network). The second matter is of practical importance and concerns the design of

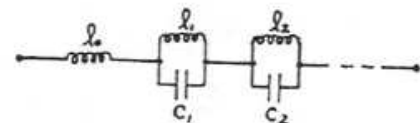


FIG. 4. A general driving point reactance having zero reactance at zero frequency. The inductance  $l_0$  will vanish if the reactance must be zero at infinite frequency.

<sup>14</sup> The pair of equations (20a) can be made the basis of a convenient nomograph to aid in the design of an amplifier of assigned rise time and total gain.

networks whose performance approaches as nearly as possible the theoretical limit.

Interstage coupling networks of two types must be distinguished, two-terminal and four-terminal.<sup>15</sup> This distinction is necessary since it is possible to separate the parasitic interstage capacitance into two portions, the output capacitance of one stage, and the input capacitance of the following stage. If a critically compensated four-terminal network is based on the two capacitances, as separate entities, it would seem likely that a shorter rise time can be achieved than for the two-terminal case. Only two-terminal networks of a simple type will be discussed in the present paper, mainly because a treatment of other cases is beset with algebraic difficulties.

Let us then consider the generalized, shunt-compensated interstage network illustrated in Fig. 3b, where the pure reactance  $X$  has a value zero at zero frequency, but is otherwise unrestricted in form. According to Foster's reactance theorem,<sup>16</sup> a possible formula for any reactance of this type can be written

$$X(s) = k \frac{s(s^2 - s_2^2)(s^2 - s_4^2) \cdots (s^2 - s_m^2)}{(s^2 - s_1^2)(s^2 - s_3^2) \cdots (s^2 - s_{m-1}^2)} \quad (25)$$

where the  $s_i^2$  ( $i = 1 \cdots m$ ) are negative real numbers,  $k$  is a positive constant, and  $m$  is an even integer. The general reactance can be realized physically by a variety of equivalent networks

$$Z(s) = \frac{1 + d_1 s + c_2 s^2 + d_3 s^3 + \cdots + c_{n-2} s^{n-2} + d_{n-1} s^{n-1}}{1 + s + (d_1 + c_2) s^2 + c_3 s^3 + \cdots + c_{n-3} s^{n-1} + d_{n-1} s^n} \quad (26)$$

where  $n = m + 2$ . It should be noticed that when  $l_0 = 0$ , the coefficient  $d_{m+1} = d_{n-1}$  vanishes. Equation (26), of course, has the form of Eq. (1).

To realize a monotonic transient response to the unit step function, it appears necessary to require that the poles of  $Z(s)$  all lie on the negative real axis of the  $s$ -plane. (Otherwise the transient response will contain oscillatory terms.) We shall assume that the most desirable arrangement of poles is to have but one multiple pole, and then show that this assumption leads to useful results.

<sup>15</sup> Strictly speaking, two-terminal and three-terminal networks.

<sup>16</sup> See reference 10, pp. 177-181 for a discussion of Foster's reactance theorem and of the various networks which can be used to realize an arbitrary reactance.

TABLE I. Some critically compensated networks.

$p$	Circuit constants	$S$	Response $e'(t)$ to unit impulse
1	$X=0$	1.000	$e^{-t}$
2	$l_0=1/4$	1.512	$e^{-2t}(1+2t)$
3	$l_1=8/27; c_1=1/8$	1.769	$e^{-3t}(1+2t+6t^2)$
4	$l_0=1/4; l_1=1/16; c_1=1$	1.899	$e^{-4t}(1+4t+(64/3)t^2)$
5	$l_1=(4/125)(5+(5)^{1/2}); c_1=(1/16)(3+(5)^{1/2}); l_2=(4/125)(5-(5)^{1/2}); c_2=(1/16)(3-(5)^{1/2})$	1.970	$e^{-5t}(1+4t+20t^2-(100/3)t^3+(250/3)t^4)$
$\infty$	$l_0=2/\pi^{2k}; c_k=1/2; (k=1, 2, 3, \dots)$	2.121	$\begin{cases} 1-t/2, & 0 \leq t \leq 2 \\ 0 & 2 \leq t \end{cases}$

made up of inductances and capacitances.<sup>16</sup> It is convenient here to adopt the form of network shown in Fig. 4 to represent the general reactance  $X(s)$ . If the inductance  $l_0$  vanishes, i.e., the general reactance becomes zero at infinite frequency, it is necessary to omit the factor  $(s^2 - s_m^2)$  from the right-hand member of Eq. (25).

By writing Eq. (25) in the form

$$X(s) = \frac{d_1 s + d_2 s^2 + \cdots + d_{m+1} s^{m+1}}{1 + c_2 s^2 + c_4 s^4 + \cdots + c_m s^m}$$

where the new, real, positive constants,  $c_i$  and  $d_i$ , are uniquely related to the constants appearing in Eq. (25), and to the circuit constants defined in Fig. 4, we find that the driving point impedance becomes

Indeed it can be proved at once for the network under consideration that a single multiple pole must lead to the shortest rise time. Thus, the rise time

$$T_R = [2\pi(2 - 2d_1 - d_1^2)]^{1/2}$$

computed from Eq. (26) using Eq. (9), is a minimum when the real, positive coefficient  $d_1$  is maximum. The negative real roots of the denominator of  $Z(s)$ ,  $-s_1, -s_2, \dots, -s_n$ , must satisfy the relations

$$\sum (1/s_i) = 1 \quad \text{why?}$$

and

$$d_1 = \sum_{i < j} (1/s_i s_j) - \sum_{i < j < k} (1/s_i s_j s_k)$$

found by multiplying together the  $n$  factors

( $1+s/s_i$ ) of the denominator and comparing the resulting expression with the denominator in Eq. (26). Using Lagrange's method of undetermined multipliers it is found that  $d_1$  is a maximum, i.e.,  $T_R$  is a minimum when  $s_1 = s_2 = s_3 = \dots$ .

The denominator of Eq. (26) can be an even or an odd degree polynomial of degree  $n$ , or  $n-1$ , respectively, depending on whether or not  $l_0$  occurs in the network of Fig. 4. The treatment for both cases follows similar lines, and will be illustrated for the case where  $l_0 \neq 0$ . In this case we require that  $Z(s)$  have one multiple pole of order  $n$  ( $n$  is always even), and the denominator of Eq. (26), accordingly, must be the binomial expansion of  $[1+(s/s_0)]^n$ , giving a set of  $n$  equations from which the  $n$  quantities  $s_0, d_1, c_2, d_3, c_4, \dots, d_{n-1}$  can be determined. The values of the components in the network of Fig. 4 can then be computed, as well as a value for the rise-time figure-of-merit  $S$ , and an expression for the transient response to a unit impulse applied at  $t=0$ .

The computation suggested has been carried out for cases where  $Z(s)$  has poles of order  $p=1, 2, 3, 4$ , and 5, as well as for the limiting case where  $p \rightarrow \infty$ . The following general expressions are found from Eqs. (8) and (9) for the delay time and for the rise-time figure-of-merit

$$\left. \begin{aligned} T_D &= \frac{2}{3} + \frac{1}{3p^2} \\ \text{and} \\ S &= \frac{3}{(2+8/p^2-1/p^4)^{1/2}} \end{aligned} \right\} \quad (27)$$

where  $p$  is the order of the multiple pole of  $Z(s)$ . A summary of the results derived from the computations is presented in Table I. The analysis employed for the limiting case is given in Appendix I.

The  $RC$  network ( $p=1$ ) has been included in Table I to serve as a basis for comparing the other critically compensated networks. The network for  $p=2$  is the well-known, shunt-compensated network, used as an illustration in Section 2. By increasing the value of the inductance from 0.25 to 0.296, and shunting it with a capacitance of 0.125 ( $p=3$ ), a decrease in rise time of about

17 percent is obtained. Adding a second inductance ( $p=4$ ) results in a further decrease in rise time of about 7 percent. By adding more and more components, the limiting value for the rise-time figure-of-merit,  $S_{\max} = 3/\sqrt{2} = 2.12$  is approached. The remaining improvement possible in the transient behavior after a few inductances and capacitances are incorporated in the network is not very marked. These cases, therefore, are not of great practical importance. The limiting case ( $p \rightarrow \infty$ ) is of interest primarily because it possesses the greatest figure-of-merit possible with a network of the type under consideration. It is conjectured that this network has the greatest figure-of-merit possible for a low pass two-terminal interstage network. No completely adequate proof, however, has been found for this theorem.

The transient response to a unit step function for all the cases listed in Table I has a monotonic form, which, of course, is necessary in order that the method used for computing delay time and rise time be applicable. The general proof that the transient response is monotonic for arbitrary values of  $p$  appears to present considerable algebraic difficulties.

#### APPENDIX I

##### Case Where $p \rightarrow \infty$

The analysis for the case where the reactance  $X(s)$  in Fig. 3b has an infinite number of poles can be made by setting the denominator in Eq. (26) equal to  $[1+(s/p)]^p$  and then writing the resulting expression in the algebraically equivalent form

$$Z(s) = 1/s - 1/2s^2 \left[ 1 - \left( 1 - \frac{s}{p} / 1 + \frac{s}{p} \right)^p \right].$$

In the limit where  $p \rightarrow \infty$ , this expression becomes

$$Z(s) = 1/s - 1/2s^2(1 - e^{-2s}) \quad (28)$$

which has the inverse Laplace transform given in Table I.<sup>17</sup>

From Eq. (28) and the network of Fig. 3b, the reactance  $X(s)$  is found to have the form

$$X(s) = \coth s - 1/s. \quad (29)$$

<sup>17</sup> It is of interest to note that if a switch is inserted in series with the capacitor  $C$  in Fig. 3b, and the capacitor is initially given a unit charge, then Eq. (28) is the Laplace transform of the voltage developed across the network when the switch is closed at  $t=0$ . Since the voltage across the capacitor decreases linearly while it is being discharged into the remaining branch of the network, the current flowing through the resistor must have the form of a rectangular pulse (of amplitude  $1$ ). The network can evidently be used (ideally, at any rate) to convert either a current impulse, or the sudden discharge of a capacitor, into a rectangular voltage pulse across a resistive load.



The zeros and poles of  $X(s)$  are located, of course, on the real frequency axis, and are given by the roots of  $\tan \omega = \omega$ , and  $\sin \omega / \omega = 0$ , respectively.

To determine values of  $l_k$  and  $c_k$ , the expression for the reactance, Eq. (29), can be expanded in the infinite series:<sup>18</sup>

$$X(s) = \coth s - 1/s = \sum_{k=1}^{\infty} \frac{2s}{s^2 + k^2\pi^2} \quad (30)$$

Each term in the infinite series can be interpreted as the reactance of a parallel combination of inductance and

<sup>18</sup> See, for instance, E. T. Whittaker and G. N. Watson, *Modern Analysis*, (Cambridge University Press, Teddington, England, 1927), fourth edition, p. 136, example 7.

capacitance, where

$$\left. \begin{aligned} l_k &= 2/\pi^2 k^2, \\ c_k &= 1/2. \end{aligned} \right\} (k=1, 2, 3, \dots). \quad (31)$$

It is evident from the nature of the terms occurring in the infinite series that the inductance  $l_k$  must vanish.

The formulas for delay time and rise-time figure-of-merit, Eqs. (27), hold in the limit when  $p \rightarrow \infty$ , so no separate computation need be made for these quantities.

This paper is based on work performed under Contract No. W-7405-Eng-36 with the Manhattan Project at the Los Alamos Scientific Laboratory of the University of California.

## A General Divergence Formula\*

H. J. RIBLET\*\* AND C. B. BARKER\*\*\*

*Radiation Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts*

(Received April 28, 1947)

A divergence expression for the ratio of energy per steradian reflected from a smooth curved surface to that incident on the surface is derived. It generalizes previous results in that the source and point of observation may both be at finite distances from the reflecting surface. No restrictions are placed on the angles of incidence and reflection except that they be equal. The only limitation placed on the analytical accuracy of the geometrical result is that the surface be sufficiently smooth so that the principal radii of curvature are defined at the point of reflection. It is required, of course, that the wave-length of the energy shall be small compared to the principal radii of curvature of the surface, in order that the geometrical result may be interpreted as a divergence formula. All of the previous results on this problem known to the authors are derived as special cases. Application of the result in connection with the spreading of radio rays by the curvature of the earth leads to somewhat simpler formulas than now available.

### INTRODUCTION

A PROBLEM of basic importance in the radio and radar art is that of determining the amount of energy reflected in any direction from a given object placed in a plane or spherical electromagnetic wave. The exact solution of this type of reflection problem is known for only a few special cases. However, if the wave-length is small compared to the radii of curvature of the surface of the reflecting object, it is possible to handle this question by means of geometrical

optics. The cases of normal incidence on a general curved surface for a plane and spherical wave have been discussed, respectively, by Goudsmit and Carlson<sup>1</sup> and Silver.<sup>2</sup> Spencer<sup>3</sup> has extended the analysis to arbitrary angles of incidence with the restriction that the wave be plane, while Barker and Riblet<sup>4</sup> have discussed the general case of a spherical wave under the assumption that the point of observation is infinitely far from the reflecting object. This

\* This paper is based in part on work done for the Office of Scientific Research and Development under Contract OEMsr-262. It was originally submitted under the title "Reflection from curved surfaces" having Eq. 18 as its principal result. A question of accuracy raised by the reviewer led us to the general result of Eq. 13.

\*\* Now at the Submarine Signal Company, Boston, Massachusetts.

\*\*\* Now at Naval Research Laboratory, Washington, District of Columbia.

<sup>1</sup> S. A. Goudsmit and J. F. Carlson, "Microwave radar reflections," Radiation Laboratory Report, 43-13, February 20, 1943.

<sup>2</sup> S. Silver, "Contribution of the dish to the impedance of an antenna," Radiation Laboratory Report, 442, September 17, 1943.

<sup>3</sup> R. C. Spencer, "Reflections from smooth curved surfaces," Radiation Laboratory Report, 661, January 26, 1945.

<sup>4</sup> C. B. Barker and H. J. Riblet, "Reflections from curved surfaces," Radiation Laboratory Report, 976, February 1, 1946.