Theoretical Fundamentals of Pulse Transmission — II

By E. D. SUNDE

(Manuscript received September 23, 1953)

PART II.

Part I of this paper dealt with various idealized transmission characteristics and with methods of evaluating pulse distortion resulting from various system imperfections. In Part II the resultant transmission impairments or limitations on pulse transmission rates are discussed for systems with low-pass, symmetrical band-pass and asymmetrical band-pass characteristics, and a comparison made of the transmission performance of double and vestigial sideband systems. The limitation on channel capacity imposed by random imperfections in the transmission-frequency characteristic, as compared to random noise, is also discussed.

12. IMPULSE CHARACTERISTICS AND PULSE TRAIN ENVELOPES

In pulse modulation systems pulses are transmitted in various combinations to form pulse trains, and at the receiving end the envelope of the pulse train is sampled at regular intervals to determine the amplitudes of the transmitted pulses. As a result of pulse overlaps there may be appreciable distortion of the pulse train envelope, which may cause errors in reception or noise, depending on the type of system. To evaluate transmission impairments, or limitations imposed on transmission capacity to avoid excessive transmission impairments from pulse distortion, it is necessary to establish basic relations between the impulse characteristic of the system and the envelope of the received pulse train.

In Fig. 42 are shown three transmitted pulses of different peak amplitudes, A_{-1}, A_0 and A_1, transmitted at intervals τ with the first and third pulse overlapping into the middle pulse. The instantaneous amplitude of the received train at a time t_0 referred to the peak amplitude of the middle pulse is

$$\begin{aligned} W(t_0) &= A_{-1}P(t_0 - \tau) + A_0P(t_0) + A_1P(t_0 + \tau), \\ &= \sum_{n=-1}^{1} A_nP(t_0 + n\tau). \end{aligned} \tag{12.01}$$

When the sequence of pulses transmitted at uniform intervals τ extends between $n = -\infty$ and ∞, the instantaneous amplitude of the pulse train at time t_0 is

$$W(t_0) = \sum_{n=-\infty}^{\infty} A_nP(t_0 + n\tau). \tag{12.02}$$

The above equation gives the instantaneous value $W(t_0)$ for any selected combination of transmitted pulses. The transmitted pulses may have any value within certain limits, as when they represent signal samples in a pulse amplitude modulation system, or may assume two or

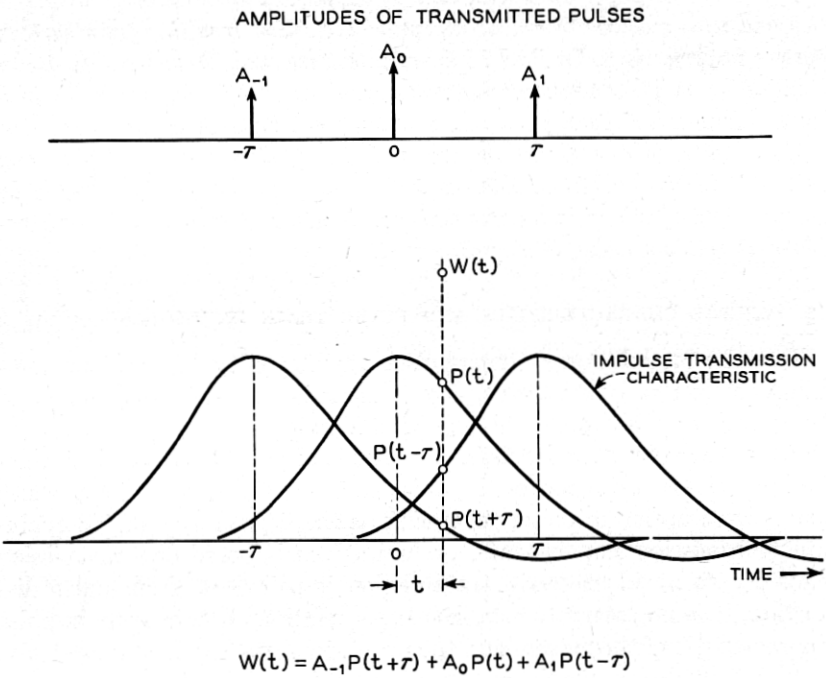

Fig. 42 — Formulation of expression for pulse train envelope in terms of impulse characteristic and amplitudes of transmitted pulses.

more discrete values as in pulse code modulation systems. In pulse position modulation, $A_n = 0$ except at the instants pulses of a given amplitude are transmitted, and n may not necessarily be an integer. In pulse duration modulation, $A_n = 1$ over the intervals $n\tau$ of varying duration in which pulses are transmitted, and zero otherwise. Equation (12.02) is thus a general formulation of the wave shape of a received pulse train, applicable to various pulse modulation methods.

Inserting (2.09) in (12.02) with $R_- + R_+ = R$ and $Q_- - Q_+ = Q$ and taking $\psi_r = 0$ without loss of generality

$$\begin{aligned} W(t_0) &\\ = \sum_{n=-\infty}^{\infty} A_n[\cos \omega_r(t_0 + n\tau)R(t_0 + n\tau) + \sin \omega_r(t_0 + n\tau)Q(t_0 + n\tau)], &\\ = \cos \omega_r t_0 \sum_{-\infty}^{\infty} A_n[\cos \omega_r n\tau R(t_0 + n\tau) + \sin \omega_r n\tau Q(t_0 + n\tau)] &\\ + \sin \omega_r t_0 \sum_{-\infty}^{\infty} A_n[\cos \omega_r n\tau Q(t_0 + n\tau) - \sin \omega_r n\tau R(t_0 + n\tau)]. & \end{aligned} \tag{12.03}$$

The envelope of the wave at the sampling instant $t_0 = 0$ is

$$\overline{W}(0) = (\bar{R}^2 + \bar{Q}^2)^{1/2}, \tag{12.04}$$

$$\begin{aligned} \bar{R} &= \sum_{-\infty}^{\infty} A_n[\cos \omega_r n\tau R(n\tau) + \sin \omega_r n\tau Q(n\tau)], \\ \bar{Q} &= \sum_{-\infty}^{\infty} A_n[\cos \omega_r n\tau Q(n\tau) - \sin \omega_r n\tau R(n\tau)]. \end{aligned} \tag{12.05}$$

For the particular case of a low-pass system

$$Q = 0, \text{ and } \omega_r = 0,$$

so that

$$W(0) = \sum_{n=-\infty}^{\infty} A_n P(n\tau). \tag{12.06}$$

A band-pass characteristic can be obtained with the aid of band-pass filters at the transmitting or receiving ends, or at both ends of a system, and the equations developed previously for the impulse characteristic tacitly assumed such an arrangement. Equivalent performance can, however, also be secured by methods which are usually employed in practice, and to which the equations also apply. Impulses can thus be applied to a low-pass pulse shaping network or filter, and the output used to modulate a carrier. There will then be a symmetrical distribution of sidebands with respect to the carrier, equivalent to a band-pass characteristic, with the spectrum of the sideband frequencies determined by the characteristic of the low-pass filter. The equivalent of an asymmetrical band-pass characteristic can be obtained by suppressing part of the upper or lower sideband with the aid of filters.

Although the mathematical formulation with both methods is essentially the same when ω_r is identified with the carrier frequency ω_c, with impulse excitation the phase of ω_r is fixed in relation to the envelope but is independent of it with carrier modulation. By proper choice of the pulse interval τ in (12.03), such that $\cos \omega_r (t_0 + n\tau) = \cos \omega_r t_0$ or $\omega_r \tau = 2\pi m$, $m = 0, 1, 2 \cdots$ it is possible with impulse excitation to obtain the same relation between the reference or carrier frequency as when the output of a low-pass filter is used to modulate a carrier. In the above case the pulses are transmitted at intervals $\tau = m/f_r = m/f_c$, corresponding to multiples of the duration of a carrier cycle. Since the duration of a carrier cycle is ordinarily small in relation to the pulse interval, there is essentially no important difference in the rate at which pulses can be transmitted with the above two methods. However, with band-pass filters the exact relationship of pulse intervals to the carrier frequency may be difficult to maintain with simple instrumentation, while this is no problem with carrier modulation. For this reason, and since the performance is otherwise equivalent, only the basic relationships with carrier modulation will be discussed further.

Assuming that $\cos \omega_r(t_0 + n\tau) = \cos \omega_r t_0$, as discussed above, equation (12.03) becomes

$$W(t_0) = \cos \omega_r t_0 \sum_{-\infty}^{\infty} A_n R(t_0 + n\tau) + \sin \omega_r t_0 \sum_{-\infty}^{\infty} A_n Q(t_0 + n\tau). \qquad (12.07)$$

The envelope at the sampling point is accordingly

$$\overline{W}(0) = \left(\left[\sum_{-\infty}^{\infty} A_n R(n\tau) \right]^2 + \left[\sum_{-\infty}^{\infty} A_n Q(n\tau) \right]^2 \right)^{1/2}. \qquad (12.08)$$

In ideal transmission systems there would be no pulse overlaps or intersymbol interference, and the amplitude of the pulse train at the sampling instant would be

$$\overline{W}(0) = A_0[R^2(0) + Q^2(0)]^{1/2}. \qquad (12.09)$$

This condition could be realized with sufficient pulse spacing. However, the objective in the design of efficient pulse systems is to determine the minimum pulse spacing consistent with tolerable intersymbol interference and thus the maximum transmission capacity or optimum performance in other respects for a given bandwidth. In the following sections this problem is discussed further.

13. TRANSMISSION LIMITATIONS IN SYMMETRICAL SYSTEMS

In a symmetrical system the amplitude characteristic has even symmetry and the phase characteristic odd symmetry with respect to a properly chosen frequency. A low-pass transmission system is thus symmetrical with respect to zero frequency, when the negative frequency range is included. A double sideband system is symmetrical if the amplitude characteristic has even and the phase characteristic odd symmetry with respect to the mid-band frequency.

Equation (12.06) applying to a low-pass system or baseband transmission may be written

$$W(0) = A_0P(0) + \sum_{n=1}^{\infty} [A_nP(n\tau) + A_{-n}P(-n\tau)]. \qquad (13.01)$$

Let it be assumed that pulses of varying but discrete amplitudes are transmitted, with a maximum peak amplitude equal to $A_{\max}$ and a minimum peak amplitude $A_{\min}$. If q pulse amplitudes are employed, the difference between peak amplitudes is then $(A_{\max} - A_{\min})/(q - 1)$. Let P^+ designate positive values of $P(n\tau)$ and P^- the absolute value of negative amplitudes.

The maximum value of $W(0)$ when a pulse of amplitude A_0 is transmitted at the sampling point $n = 0$ is then

$$\begin{aligned} W_{\max} = A_0P(0) &+ \sum_{n=1}^{\infty} A_{\max}[P^+(n\tau) + P^+(-n\tau)] \\ &- \sum_{n=1}^{\infty} A_{\min}[P^-(n\tau) + P^-(-n\tau)] \end{aligned} \qquad (13.02)$$

The minimum amplitude of $W(0)$ when a pulse of the next higher amplitude $A_0 + (A_{\max} - A_{\min})/(q - 1)$ is transmitted becomes

$$\begin{aligned} W_{\min} = \left(A_0 + \frac{A_{\max} - A_{\min}}{q - 1}\right) P(0) \\ - \sum_{n=1}^{\infty} A_{\max}\,[P^-(n\tau) + P^-(-n\tau)] \\ + \sum_{n=1}^{\infty} A_{\min}\,[P^+(n\tau) + P^+(-n\tau)]. \end{aligned} \qquad (13.03)$$

To permit distinction between the two pulse peaks it is necessary that

$W_{\min}$ be greater than $W_{\max}$. The difference $M = W_{\min} - W_{\max}$, which represents the margin for distinction between pulse amplitudes, becomes

$$M = \frac{A_{\max} - A_{\min}}{q - 1} P(0) - (A_{\max} - A_{\min}) \sum_{n=1}^{\infty} [P^{+}(n\tau) + P^{-}(n\tau) + P^{+}(-n\tau) + P^{-}(-n\tau)], \tag{13.04}$$

or:

$$M = (A_{\max} - A_{\min}) \left[\frac{P(0)}{q - 1} - \sum_{n=1}^{\infty} | P(n\tau) | + | P(-n\tau) | \right], \tag{13.05}$$

where $| P(\pm n\tau) |$ designates the absolute values of the impulse characteristic.

Equation (13.05) shows that for a given value of q the margin depends on the maximum pulse excursion $A_{\max} - A_{\min}$ and is thus the same with $A_{\max} = 1$ and $A_{\min} = 0$ as with $A_{\max} = 0.5$ and $A_{\min} = -0.5$. As an example, equation (13.05) shows that with two pulse amplitudes, $q = 2$, it is possible to distinguish between pulses and spaces, or between positive and negative pulses, if the sum of the absolute values of the impulse characteristic at all the sampling points, excluding 0, is less than the amplitude P(0) of the impulse at sampling point 0.

The maximum margin against errors is obtained without pulse overlaps, i.e. when the summation term in (13.05) is zero, and is

$$M_{\max} = (A_{\max} - A_{\min}) \frac{P(0)}{q - 1}. \tag{13.06}$$

The ratio of the margin M as given by (13.05) to the maximum margin becomes:

$$M/M_{\max} = 1 - \frac{q - 1}{P(0)} \sum_{n=1}^{\infty} | P(n\tau) | + |P(-n\tau) |. \tag{13.07}$$

This equation may be employed to determine the maximum possible pulsing rate for a given impulse characteristic and number of pulse amplitudes, obtained when $M/M_{\max} = 0$, or to determine the margin for a given pulse transmission rate. An example of the latter application is illustrated in Fig. 43, which shows the margin $M/M_{\max}$ in per cent, obtained when (13.07) with $q = 2$ is applied to the curves shown in Fig. 23 for various degrees of delay distortion. The pulse interval is taken as $\tau = 1/2f_1 = 1/f_{\max}$, where f_1 is the frequency at the 6 db down point on the amplitude characteristic and $f_{max} = 2f_1$. Under this condition there is no intersymbol interference in the absence of phase distortion.

The above equations apply to peak intersymbol interference, obtained by taking the maximum positive and negative values of the summation term in (13.01). As discussed in previous sections, certain types of transmission system imperfections give rise to pulse distortion extending over long time intervals, such as fine structure deviations over the transmission band, a low-frequency cut-off and pronounced band-edge phase deviations. Evaluation of peak intersymbol interference is then rather difficult, and a more convenient approximate method is to evaluate rms intersymbol interference, which can be related to rms deviation in the transmission frequency characteristic by methods discussed previously. Peak intersymbol interference may then be estimated by applying a peak factor between 3 and 4, depending on the type of transmission distortion.

If $P_0(n\tau)$ designates an ideal impulse characteristic, which is zero for $n = \pm 1, \pm 2$ etc., the deviation from the ideal envelope of a pulse

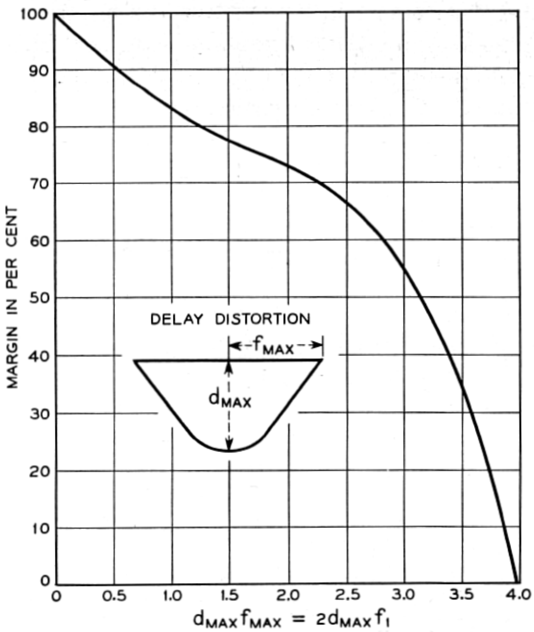

Fig. 43 — Margin against excessive peak interference in systems employing two pulse amplitudes with intervals between pulses $\tau = \tau_1 = 1/2f_1 = 1/f_{max}$. for impulse transmission characteristic as shown in Fig. 23.

train may be written

$$\Delta W(0) = W(0) - W_0(0) = \sum_{n=-\infty}^{\infty} A_n[P(n\tau) - P_0(n\tau)]. \tag{13.08}$$

The rms deviation becomes, with $\Delta P(n\tau) = P(n\tau) - P_0(n\tau)$

$$\underline{\Delta W}(0) = \underline{A} \left(\sum_{n=-\infty}^{\infty} [\Delta P(n\tau)]^2 \right)^{1/2}, \tag{13.09}$$

$$\cong \underline{A} \left(\frac{1}{\tau} \int_{-\infty}^{\infty} [\Delta P(t)]^2 \, dt \right)^{1/2}, \tag{13.10}$$

$$= \underline{A} P(0) \underline{U}. \tag{13.11}$$

$\underline{A}$ is the rms amplitude of the transmitted pulses and $\underline{U}$ the rms intersymbol interference referred to unit amplitude of the received pulses. Expressions for $\underline{U}$ applying to fine structure imperfections in the transmission frequency characteristic were given in Section 8, for a low-frequency cut-off in Section 9 and for band-edge phase deviations in Section 10.

For balanced pulse systems employing positive and negative pulses, rms intersymbol interference in the positive and negative directions will be equal. For such systems the maximum value of the summation in (13.02) becomes $k\underline{W}(0)$ and in (13.03) $-k\underline{W}(0)$, where k is the peak factor. Equation (13.04) is then replaced by

$$\begin{aligned} M &= \frac{A_{\max} - A_{\min}}{q - 1} P(0) - 2k\underline{A}P(0)\underline{U}, \\ &= 2A_{\max}P(0) \left[\frac{1}{q - 1} - k\underline{U}(\underline{A}/A_{\max}) \right], \end{aligned} \tag{13.12}$$

when $A_{\min} = -A_{\max}$.

In a balanced pulse system employing q pulse amplitudes, i.e., $q/2$ positive and $q/2$ negative amplitudes, with equal steps $2A_{\max}/(q - 1)$ between pulse amplitudes, the following relation applies if all amplitudes have equal probability.

$$\underline{A}/A_{\max} = \left[\frac{q + 1}{3(q - 1)} \right]^{1/2}. \tag{13.13}$$

Hence,

$$M = \frac{2A_{\max}P(0)}{q - 1} \left[1 - k \left(\frac{q^2 - 1}{3} \right)^{1/2} \underline{U} \right]. \tag{13.14}$$

As mentioned before, the factor k may be as high as 4, in which case the maximum tolerable rms intersymbol interference $\underline{U}$ referred to unit peak amplitude of the received pulses becomes for $M = 0$:

q =	$\underline{2}$	$\underline{4}$	$\underline{8}$
$\underline{U}$ =	0.25	0.112	0.054

In (13.14) and in the above table, $\underline{U}$ is the maximum tolerable rms intersymbol interference from all sources, such as fine structure imperfections over the transmission band, band-edge phase distortion and a low-frequency cut-off. Interference from these various sources may be combined on a root-sum-square basis.

In the above evaluation of rms intersymbol interference a balanced pulse system was assumed. An unbalanced system can be obtained by superposing on a balanced system an infinite sequence of pulses of equal amplitude and polarity at uniform intervals as indicated in Fig. 44. This superposed system will give rise to a fixed intersymbol interference or displacement of the received pulse train, which does not alter the margin for distinction between pulse amplitudes and which can be corrected by a fixed bias at the receiving end if necessary. For this reason, in the case

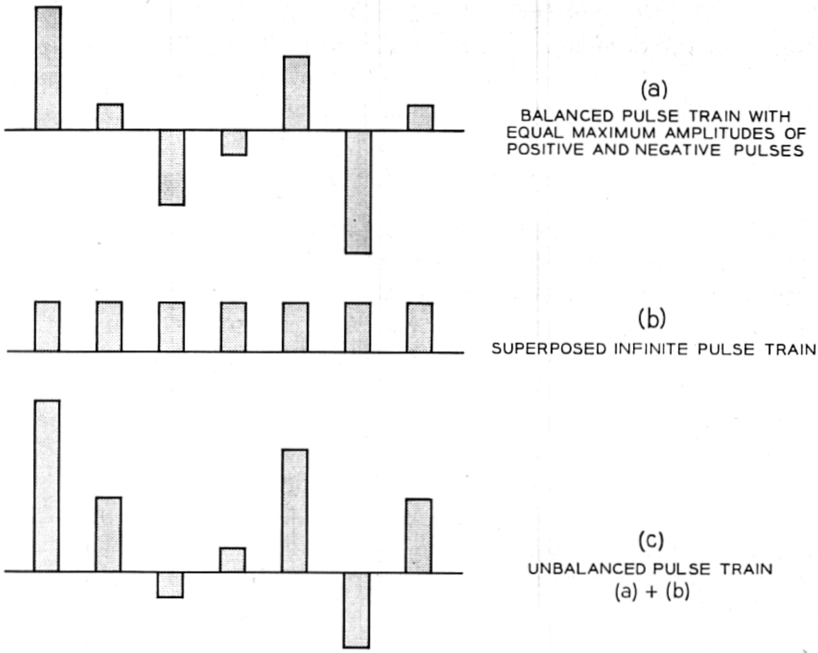

Fig. 44 Derivation of an unbalanced from a balanced pulse train by superposition of an infinite train of pulses of equal amplitude.

of an unbalanced system, only the balanced component need to be considered in evaluating rms intersymbol interference, which will thus be the same whether or not the system is balanced. As shown previously, peak intersymbol interference, or the margin for distinction between pulse amplitudes, depends only on the peak to peak pulse excursion and is thus the same for unbalanced as for balanced systems. It may be noted here that for a balanced system the transmitted power is a minimum for a given margin in pulse reception, as is the interference in other systems that may be caused by the transmitted pulses.

For a symmetrical band-pass system, rather than a low-pass system as discussed above, $Q(n\tau) = 0$ in (12.08). The envelope of the pulse train then becomes

$$\overline{W}(0) = \sum_{n=-\infty}^{\infty} A_n R(n\tau), \tag{13.15}$$

where $R(n\tau) = R_-(n\tau) + R_+(n\tau) = 2R_+(n\tau)$, with R_- and R_+ given by (2.10).

Since (13.15) is of the same form as (13.01), the relationships established above for low-pass systems also apply to symmetrical band-pass systems, with $R(n\tau)$ replacing $P(n\tau)$. $R(n\tau)$ will have the same shape as $P(n\tau)$, but will be greater by a factor 2, which will appear as a multiplier in the various expressions and hence not alter the requirements on tolerable pulse distortion or intersymbol interference.

14. TRANSMISSION LIMITATIONS IN ASYMMETRICAL SIDEBAND SYSTEMS

The formulation of transmission limitations imposed by pulse distortion in asymmetrical sideband systems is complicated by the presence of the quadrature component in the impulse transmission characteristic. Of particular interest are the transmission limitations with vestigial sideband as compared with double sideband transmission, assuming the same bandpass characteristic in both cases, a question which has been dealt with in literature for systems with a linear phase characteristic[4, 11]. Relationships (2.18) and (2.19) facilitate a comparison also for systems with phase distortion, as shown in the following.

If the envelope of the impulse characteristic with double sideband transmission is $\overline{P}(t)$, the in-phase and quadrature components with vestigial sideband transmission are given by (2.19), with $\omega_y = \omega_s$ or

$$\begin{aligned} R &= R_- + R_+ = \cos(\omega_s t - \psi_s)\,\overline{P}(t), \\ Q &= Q_- - Q_+ = \sin(\omega_s t - \psi_s)\,\overline{P}(t). \end{aligned} \tag{14.01}$$

If t is so chosen that $\omega_s t - \psi_s = 0$, and the time with respect to this value of t is designated t_0, then

$$\begin{aligned} R(t_0) &= \cos \omega_s t_0 \overline{P}(t_0), \\ Q(t_0) &= \sin \omega_s t_0 \overline{P}(t_0). \end{aligned} \tag{14.02}$$

An application of this method to the impulse characteristic shown in Fig. 23 for $b = 15$ radians is illustrated in Fig. 45.

In order to compare vestigial with double sideband transmission, it suffices to evaluate the in-phase and quadrature components at the sampling instants. With $\tau = \pi/2\omega_s = 1/4f_s$, the in-phase and quadrature components at times $m\tau$, for $m = 0 \pm 1, \pm 2$, etc., will be as illustrated in Fig. 46.

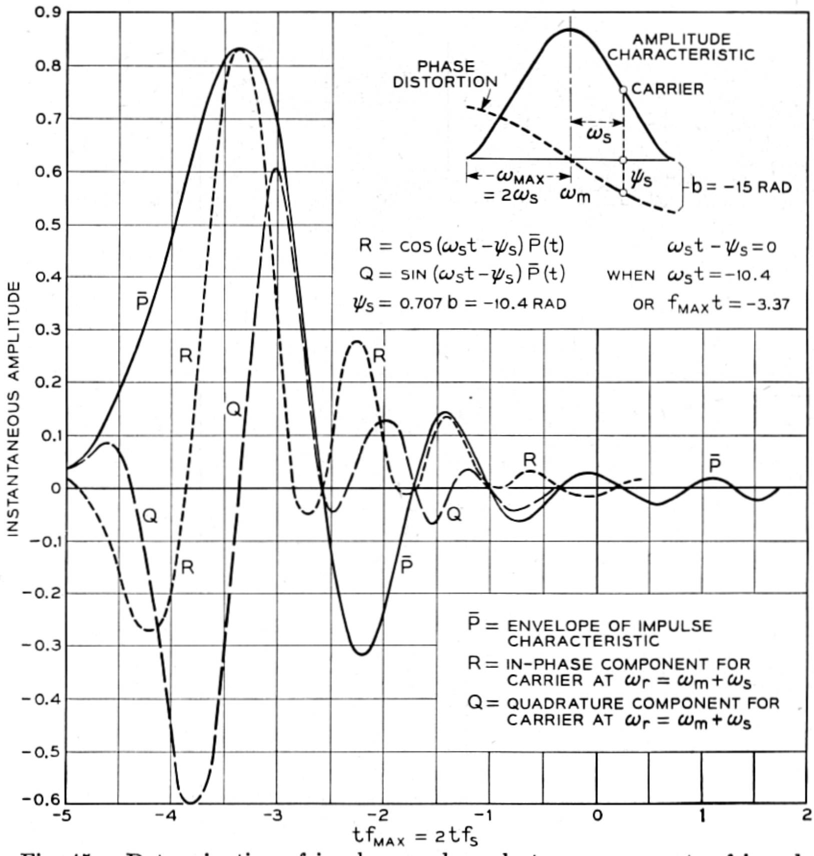

Fig. 45 — Determination of in-phase and quadrature components of impulse characteristic for vestigial side-band transmission.

With double sideband transmission, pulses would be transmitted at the points $m = 0, \pm 2, \pm 4$, etc. At these points the quadrature components vanish, as indicated in the above figure, and the in-phase components are the same in amplitude as with double sideband transmission. Thus, if pulses were transmitted at the same rate as with double sideband transmission, the sum of the absolute values of the in-phase components at the sampling points would be identical with the sum of the absolute values of the envelope with double sideband transmission. It follows from the criteria established in Section 13 that for this particular pulse transmission rate the effect of pulse distortion would be the same with both transmission methods. With an ideal transmission frequency

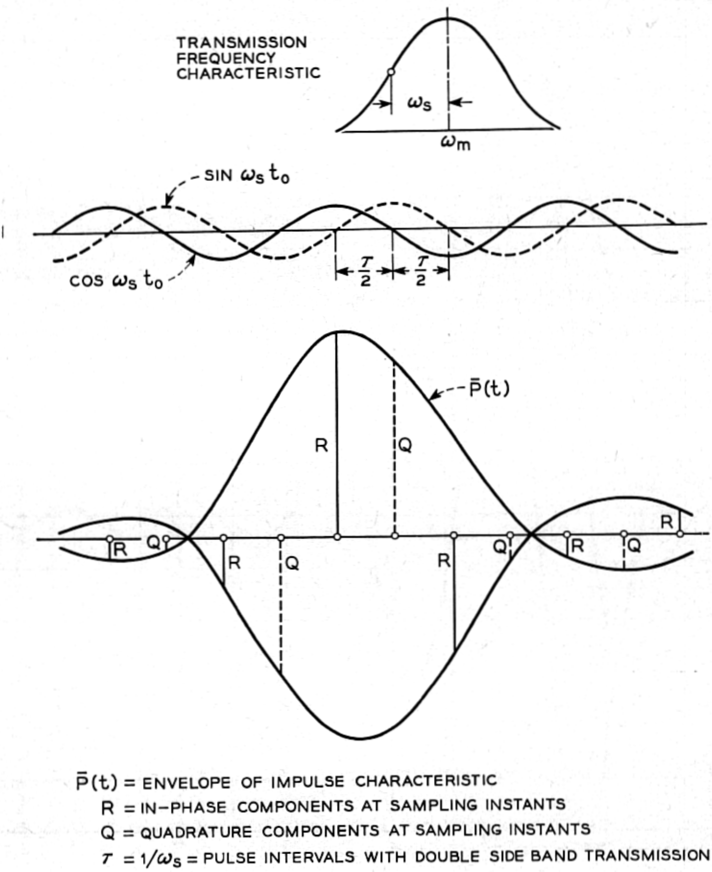

Fig. 46 — In-phase and quadrature components of impulse characteristic with vestigial side-band transmission.

characteristic having a linear phase shift, there would be no intersymbol interference with either method for the above rate of pulse transmission.

Assume next that the pulse transmission rate is doubled and that the quadrature component is eliminated. This is possible if the carrier frequency is transmitted and is derived at the receiving end with the aid of filters and applied in proper phase to a product demodulator, a method known as homodyne detection. At the points $m = 1, 3, 5$, etc., there would then be no quadrature components and no in-phase components. The sum of the absolute values of the in-phase components at the other sampling points, $m = 2, 4$, etc., would be the same as with double sideband transmission. It follows that the transmission capacity (pulsing rate) can be doubled by vestigial sideband transmission if the quadrature component is eliminated by homodyne detection, for the same margin against excessive intersymbol interference as with double sideband transmission.

An increase in transmission capacity can be realized with vestigial sideband transmission without elimination of the quadrature component by homodyne detection, although a two-fold increase is then possible only if the phase characteristic is linear, as discussed below. Vestigial sideband transmission can be employed without transmission of the carrier, or with a fixed level of carrier in the absence of pulses and a higher level in the presence of pulses. The latter method is equivalent to the transmission of two or more pulse amplitudes, with the minimum amplitude greater than zero. With this method the effect of the quadrature component on the envelope of a pulse train can be reduced, and even eliminated provided the phase characteristic is linear. In the following, vestigial sideband transmission with two pulse amplitudes at twice the double sideband pulsing rate is discussed, for the case in which the minimum pulse amplitude is finite rather than zero.

With pulses transmitted at twice the double sideband rate, i.e., with the interval between pulses equal to $\tau = \pi/2\omega_s$, equation (12.08) for the envelope becomes in view of (14.02)

$$\overline{W}(0) = \left(\left[\sum_{-\infty}^{\infty} A_n \cos \omega_s n\tau \overline{P}(n\tau)\right]^2 + \left[\sum_{-\infty}^{\infty} A_n \sin \omega_s n\tau \overline{P}(n\tau)\right]^2\right)^{1/2}. \tag{14.03}$$

At the even sampling points, i.e., $n = 0, 2, 4 \cdots$, $\cos \omega_s n\tau = \pm 1$ and the in-phase components may be written

$$R(\pm 2m\tau) = \pm \overline{P}(\pm 2m\tau), \qquad m = 0, 1, 2 \cdots$$

At the odd sampling points, i.e., $n = 1, 3, 5 \cdots$, $\sin \omega_s n\tau = \pm 1$ and the quadrature components may be written

$$Q[\pm(2m - 1)\tau] = \pm\overline{P}[\pm(2m - 1)\tau], \qquad m = 1, 2, 3$$

Let

$$\begin{aligned}
\sum R^+ &= \sum_{m=1}^{\infty} [R^+(2m\tau) + R^+(-2m\tau)], \\
\sum R^- &= \sum_{m=1}^{\infty} [R^-(2m\tau) + R^-(-2m\tau)], \\
\sum Q^+ &= \sum_{m=1}^{\infty} Q^+[(2m - 1)\tau] + Q^+[-(2m - 1)\tau], \\
\sum Q^- &= \sum_{m=1}^{\infty} Q^-[(2m - 1)\tau] + Q^-[-(2m - 1)\tau],
\end{aligned} \tag{14.04}$$

where R^+, Q^+ designate positive values and R^-, Q^- the absolute values of negative amplitudes of the in-phase and quadrature components.

Let it be assumed that two pulse amplitudes are employed, $A_{\min}$ and $A_{\max}$. When the minimum amplitude is transmitted, the maximum value of the envelope is obtained by considering the maximum positive overlaps of the in-phase components in conjunction with the maximum value of the quadrature component. The value thus obtained is

$$\begin{aligned}
\overline{W}_{\max} = [(A_{\min} R(0) + A_{\max} \textstyle\sum R^+)^2 \\
+ (A_{\max} \textstyle\sum Q^- - A_{\min} \textstyle\sum Q^+)^2]^{1/2}
\end{aligned} \tag{14.05}$$

It is assumed that $\sum Q^- > \sum Q^+$, otherwise Q^- and Q^+ would be interchanged in the last term.

When the maximum amplitude is transmitted, the minimum value of the envelope is obtained by considering the maximum negative overlaps of the in-phase components, in conjunction with the minimum value of the quadrature component, which gives

$$\begin{aligned}
\overline{W}_{\min} = [(A_{\max} R(0) - A_{\max} \textstyle\sum R^-)^2 + A^2_{\min} (\textstyle\sum Q^- \\
- \textstyle\sum Q^+)^2]^{1/2}.
\end{aligned} \tag{14.06}$$

The margin for distinction between $A_{\min}$ and $A_{\max}$ is $M = W_{\min} - W_{\max}$ and becomes

$$\begin{aligned}
M = A_{\max} [(R(0) - \textstyle\sum R^-)^2 + \mu^2(\textstyle\sum Q^+ - \textstyle\sum Q^-)^2]^{1/2} \\
- A_{\max} [(\mu R(0) + \textstyle\sum R^+)^2 + (\textstyle\sum Q^- - \mu \textstyle\sum Q^+)^2]^{1/2},
\end{aligned} \tag{14.07}$$

where

$$\mu = A_{\min}/A_{\max}.$$

The margin for a unit difference $A_{max} - A_{min}$, i.e. $M_1 = M/(A_{max} - A_{min})$ becomes:

$$M_1 = \frac{1}{1-\mu} \left([(R(0) - \sum R^-)^2 + \mu^2(\sum Q^- - \sum Q^+)^2]^{1/2} - [(\mu R(0) + \sum R^+)^2 + (\sum Q^- - \mu \sum Q^+)^2]^{1/2} \right). \tag{14.08}$$

The special case of an ideal transmission characteristic as shown in Fig. 47 will be considered first. In this case

$$\begin{array}{lll} R(0) = 1 & R(2\tau) = 0 & R(-2\tau) = 0 \\ & R(4\tau) = 0 & R(-4\tau) = 0 \\ & Q(\tau) = 0.5 & Q(-\tau) = -0.5 \\ & Q(3\tau) = 0 & Q(-3\tau) = 0 \end{array}$$

so that:

$$\begin{array}{ll} \sum R^+ = 0 & \sum R^- = 0 \\ \sum Q^+ = 0.5 & \sum Q^- = 0.5 \end{array}$$

Equation (14.08) in this case simplifies to

$$M_1 = \frac{1}{1-\mu} \left(1 - \left[\mu^2 + \frac{1}{4}(1-\mu)^2 \right]^{1/2} \right). \tag{14.09}$$

For various values of $\mu = A_{min}/A_{max}$ the margin for unit difference in

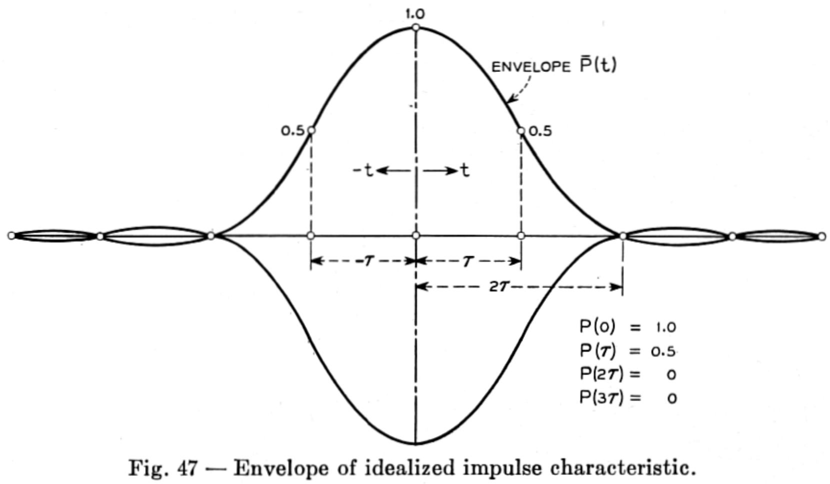

Fig. 47 — Envelope of idealized impulse characteristic.

pulse amplitudes becomes:

μ	0	0.2	0.3	0.5	0.8	0.9	1.0
M_1	0.50	0.69	0.77	0.87	0.96	0.998	1.0

Thus, for an ideal impulse characteristic as assumed above, the quadrature component gives rise to 50 per cent maximum intersymbol with $\mu = 0$, and to negligible intersymbol interference when $\mu \cong 0.8$ or greater. By way of comparison, the margin would be zero with double sideband transmission at the rate assumed above, i.e., twice the normal double sideband rate. This follows from (13.07) when it is considered that $P(\pm\tau) = \frac{1}{2} P(0)$, $P(\pm 2\tau) = 0$, so that the sum of the absolute values of the impulse characteristic at the sampling points is equal to $P(0)$ and thus $M/M_{max} = 0$.

Elimination of the effect of the quadrature component by the above method is contingent on a symmetrical impulse characteristic, i.e., $\overline{P}(n\tau) = \overline{P}(-n\tau)$, a condition which can be realized only with a linear phase shift. Furthermore, the in-phase components must vanish at the sampling points, which entails an ideal amplitude characteristic. In the presence of phase distortion the effect of the quadrature component cannot be eliminated but may be reduced by proper choice of the ratio μ, as discussed below for a transmission characteristic with moderate phase distortion.

As an example consider an impulse characteristic as shown in Fig. 23 for $b = 5$ radians. The in-phase and quadrature components at the various sampling points are in this case

$$R(0) = 0.97 \qquad R(-2\tau) = -0.09 \qquad R(2\tau) = 0.13$$

$$R(-4\tau) \cong 0 \qquad R(4\tau) \cong 0$$

$$Q(-\tau) = -0.54 \qquad Q(\tau) = 0.44$$

$$Q(-3\tau) \cong 0 \qquad Q(3\tau) = -0.03$$

$$Q(-5\tau) \cong 0 \qquad Q(5\tau) \cong 0$$

Hence

$$\sum R^+ = 0.13 \qquad \sum R^- = 0.09 \qquad \sum Q^+ = 0.44 \qquad \sum Q^- = 0.57$$

Equation (14.08) in this case becomes

$$M_1 = \frac{1}{1-\mu}\left((0.88^2 + 0.13^2\mu^2)^{1/2} - [(0.97\mu + 0.13)^2 + (0.57 - 0.44\mu)^2]^{1/2}\right).$$

For various values of $\mu = A_{min}/A_{max}$ the margin for unit difference in pulse amplitudes becomes

μ	0	0.2	0.3	0.4	0.5	0.6	0.7	0.75
M_1	0.30	0.375	0.40	0.375	0.34	0.25	0.13	0

The optimum condition is thus in the above particular case obtained with $\mu \cong 0.3$, with a comparatively small variation in transmission performance for any value of μ between 0 and 0.5.

In the above discussion of vestigial sideband transmission, modulation of a carrier was assumed, with elimination of one sideband except for the wanted vestige. The equivalent performance can be secured by application of impulses to a band-pass transmission characteristic with the proper interval between pulses in relation to the midband frequency, as discussed below:

When equation (12.03) is written with respect to the midband frequency, $\omega_r = \omega_m$, and a symmetrical amplitude characteristic is assumed so that $Q = 0$, the following relation obtained.

$$\begin{aligned} W(t_0) &= \cos \omega_m t_0 \sum_{n=-\infty}^{\infty} A_n \cos \omega_m n\tau R(t_0 + n\tau) \\ &\quad - \sin \omega_m t_0 \sum_{n=-\infty}^{\infty} A_n \sin \omega_m n\tau R(t_0 + n\tau), \end{aligned} \tag{14.10}$$

in which R may be replaced by $\overline{P}$, the envelope of the impulse characteristic.

Let it be assumed that τ is so chosen that $\cos \omega_m n\tau = \cos n\pi/2$ in which case $\sin \omega_n n\tau = \sin n\pi/2$. The above equation then becomes

$$\begin{aligned} W(t_0) &= \cos \omega_m t_0 \sum_{n=-\infty}^{\infty} A_n \overline{P}(t_0 + n\tau) \cos n\pi/2 \\ &\quad - \sin \omega_m t_0 \sum_{n=-\infty}^{\infty} A_n \overline{P}(t_0 + n\tau) \sin n\pi/2. \end{aligned} \tag{14.11}$$

The in-phase and quadrature components of the envelope at the sampling instant $t_0 = 0$ are accordingly

$$\begin{aligned} \bar{R}(0) &= \sum_{n=-\infty}^{\infty} A_n \overline{P}(n\tau) \cos n\pi/2, \\ \bar{Q}(0) &= \sum_{n=-\infty}^{\infty} A_n \overline{P}(n\tau) \sin n\pi/2. \end{aligned} \tag{14.12}$$

Pulses in even positions, i.e., A_0, A_2, A_4, etc., will thus contribute an in-phase but no quadrature component while pulses in odd positions

A_1, A_3, A_5, etc., will contribute a quadrature but no in-phase component. It will be recognized from Fig. 42 that this is the same condition as encountered in vestigial sideband transmission with pulses in the latter case transmitted at intervals $\tau = \pi/2\omega_s = 1/4 f_s$.

To realize the above condition with pulses applied to a band-pass filter, it is necessary that in (14.10)

$$\omega_m n\tau = n\pi(\tfrac{1}{2} + N), \tag{14.13}$$

where N is an integer, or that

$$\tau = \frac{\pi(1 + 2N)}{2\omega_m} = \frac{1 + 2N}{4f_m}. \tag{14.14}$$

The interval between pulses must thus be an integral number of half-cycles plus one quarter cycle of the midband frequency f_m, as illustrated for a particular case in Fig. 48. When f_m is large in relation to the side-band frequency this condition can be achieved with substantially the same pulse spacing as with vestigial sideband transmission. To secure exactly the same rate of pulse transmission it is necessary that

$$\tau = 1/4 f_s,$$

which, in conjunction with (14.14), gives

$$N = \tfrac{1}{2}(f_m/f_s - 1). \tag{14.15}$$

Thus, if $f_m = 5f_s$, $N = 2$ and the interval τ between pulses as obtained from (14.14) is 1.25 cycles of f_m. If $f_m = 10f_s$, $N = 4.5$ and it is not possible to have exactly the same pulsing rate as with vestigial sideband transmission, since N must be an integer. It is then necessary to take $N = 4$ or 5. With $N = 4$ equation (14.14) gives $\tau = 9/40f_s$ and with $N = 5$, $\tau = 11/40f_s$. This compares with $\tau = 1/4f_s = 10/40f_s$ with vestigial sideband transmission, so that there is a minor difference in pulse intervals with the two methods.

15. DOUBLE VERSUS VESTIGIAL SIDEBAND SYSTEMS

From the preceding discussion it follows that, for the same bandwidth and margin against interference from characteristic distortion, a two-fold increase in transmission capacity can be approached with vestigial over double sideband transmission. This assumes that the carrier is transmitted at the proper level and that the phase characteristic is linear, or that otherwise homodyne detection is used to cancel the effect of the quadrature component.

For the same bandwidth, the same transmission capacity can be realized with a double sideband system employing four pulse amplitudes as with a vestigial sideband system with two pulse amplitudes. However, the latter type of system will have a greater tolerance to interference from characteristic distortion than the former. This follows when it is considered that in a quaternary system the maximum tolerable interference is $\frac{1}{6}$ the maximum pulse amplitude, as compared to $\frac{1}{2}$ the maximum pulse amplitude in a binary system. With $\mu = A_{\min}/A_{\max} = 0$, the quadrature component reduces the margin by a factor of 0.5, so that the maximum tolerable interference in relation to the maximum pulse

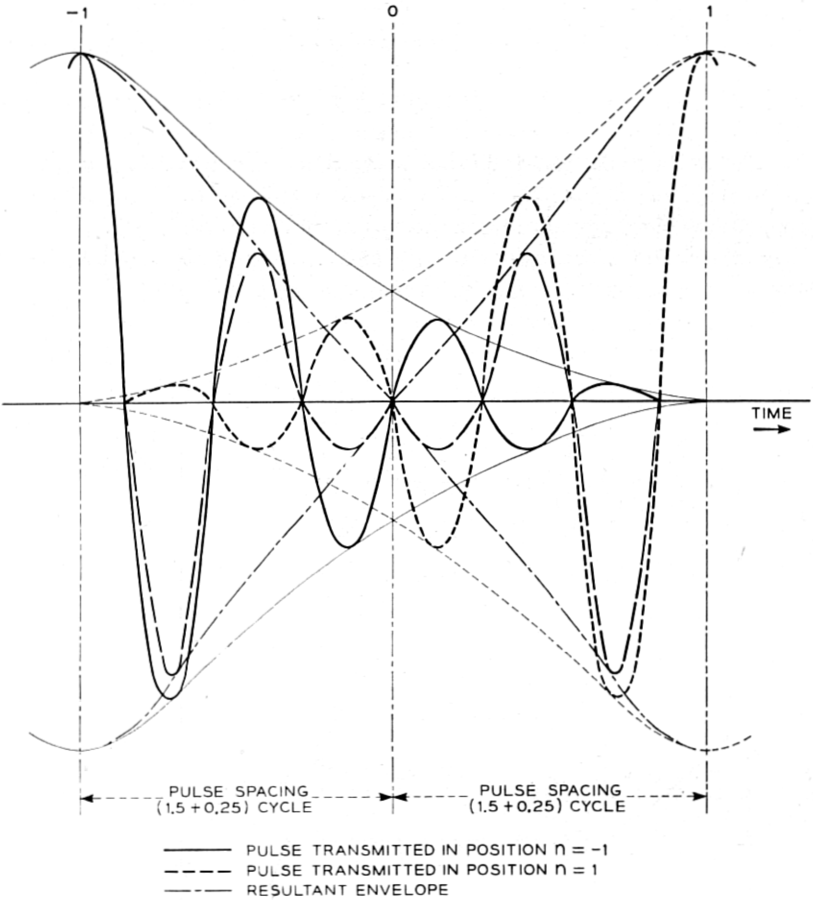

Fig. 48 — Impulse exitation of band-pass system with pulse spacing selected to provide equivalent of vestigial side-band transmission.

amplitude is ¼ as compared to ⅙ for a quaternary system. If the phase characteristic is linear and the carrier is transmitted at the optimum level, or if homodyne detection is used, the effect of the quadrature component is cancelled. The maximum tolerable interference is then ½ as compared with ⅙ for a quarternary double sideband system.

In the presence of phase distortion, a substantial advantage can also be realized with a binary vestigial system, which can be illustrated by considering the example in Section 14. For the optimum condition $\mu = 0.4$, the margin is reduced by a factor 0.4 and is thus 0.2. For a quaternary double sideband system the factor by which the margin is reduced is given by (13.07), with $q = 4$ and with

$$\frac{1}{P(0)} \sum_{n=1}^{\infty} | P(n\tau) | + | P(-n\tau) | = \frac{1}{R(0)} [\sum R^{+} + \sum R^{-}],$$

where $R(0) = 0.97$, $\sum R^{+} = 0.13$ and $\sum R^{-} = 0.09$, as in the example in Section 14. The reduction in margin thus obtained is $M/M_{\max} = 0.32$. Hence the maximum tolerable interference for a quarternary double sideband system is $0.32/6 = 0.053$ as compared with 0.20 for a binary vestigial sideband system under the optimum condition $\mu = 0.4$.

For the same transmission capacity and same number of pulse amplitudes, a substantial transmission advantage may be realized with vestigial over double sideband transmission in circuits with pronounced phase distortion, owing to the circumstance that a two-fold reduction in bandwidth with vestigial sideband transmission may afford a sub-

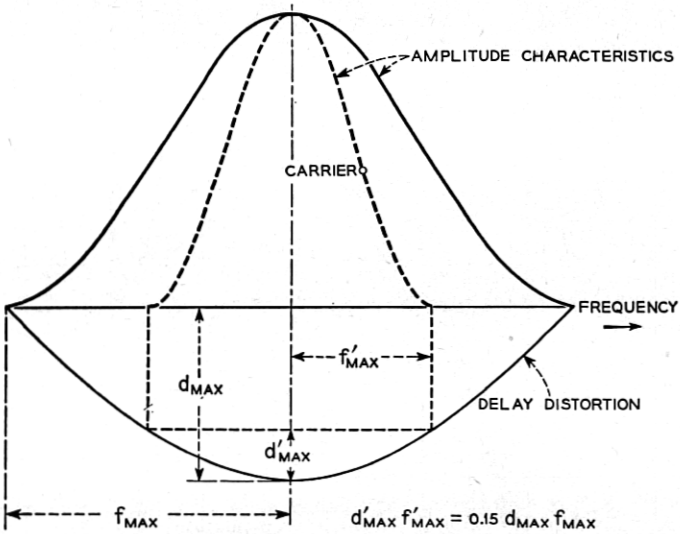

Fig. 49 — Comparison of double and vestigial side-band transmission in the presence of delay distortion.

stantial reduction in delay distortion over the transmission band. This is illustrated in Fig. 49, where a cosine variation in transmission delay is assumed. With a two-fold reduction in bandwidth, the product $d'_{max}f'_{max}$ for vestigial sideband transmission is about 15 per cent of the product $d_{max}f_{max}$ for double sideband transmission. Thus, with $d_{max}f_{max} = 8.3$, $d'_{max}f'_{max} = 1.25$, corresponding to $b = 5$ radians, as assumed in the example in Section 14. Vestigial sideband transmission is in this case feasible with an adequate margin, about 40 per cent of the maximum margin in the absence of phase distortion. Double sideband transmission would not be possible, as is evident from Fig. 43, since it would be necessary to have $d_{max}f_{max}$ less than 4, as compared with 8.3 in the above case.

The above discussion of vestigial vs double sideband transmission pertains to the effects of characteristic distortion rather than noise, and the relative complexity of terminal equipment was disregarded. Because of the simpler terminal equipment with double sideband transmission, this method is ordinarily used where bandwidth is not a primary consideration, as for example in providing a number of telegraph channels over a voice frequency circuit.

16. LIMITATION ON CHANNEL CAPACITY BY CHARACTERISTIC DISTORTION

For an idealized channel of bandwidth f_1 with a transmission-frequency characteristic as shown in Fig. 7, the transmission capacity in bits per second for a signal of average power P in the presence of random noise of average power N can with sufficiently complicated encoding methods approach the limiting value given by Shannon:[12]

$$C = f_1 \log_2 (1 + P/N). \tag{16.01}$$

The above expression also applies to certain other idealized channels with a linear phase characteristic, when f_1 is defined as in Fig. 10. In all of these cases the integral of the area under the amplitude characteristic, or the equivalent bandwidth, is f_1.

By way of comparison, for pulse code modulation systems the channel capacity is of the same basic form as (16.01), namely[13]:

$$C = f_1 \log_2 \left(1 + \frac{P}{KN}\right), \tag{16.02}$$

where $K \cong 8$. Thus about an 8-fold increase in signal power is required to attain the same channel capacity as with the idealized but impracticable encoding system underlying (16.01).

The above expressions give the limitation on channel capacity imposed by random noise. From the discussion in Sections 13 and 14 it follows that a limitation is placed on channel capacity by characteristic distortion, in the absence of noise. In idealized communication theory, characteristic distortion has been disregarded in determining channel capacity on the premise that unlike random noise it is predictable and can therefore be corrected, at least in principle. In actual systems, however, complete elimination though possible in principle cannot be accomplished in practice. The resultant limitation on transmission capacity may be as important as that imposed by the maximum signal power that can actually be provided to override noise.

In the following it will be assumed that correction of amplitude and phase deviations is made by equalization, so that the amplitude and phase characteristics are as assumed for an ideal channel, except for small fine structure residual deviations as illustrated in Fig. 30. These small fine structure deviations may be regarded as of random nature in the sense that they differ among channels and cannot be predicted, although for a given system they would remain fixed in the absence of temperature variations or changes in amplifiers with age.

From equation (13.12) it follows that the maximum number of pulse amplitudes or quantizing levels as limited by characteristic distortion is obtained from the relation

$$\frac{1}{q-1} = k\underline{U}\underline{A}/A_{\max}, \tag{16.03}$$

or

$$q = 1 + \frac{1}{k\underline{U}} A_{\max}/\underline{A}. \tag{16.04}$$

In the absence of characteristic distortion, the maximum number of pulse amplitudes as limited by an rms noise amplitude $\underline{A}_n$ or a peak noise amplitude $k\underline{A}_n$ is obtained from the following relation for a balanced pulse system.

$$\frac{A_{\max}}{q-1} = k\underline{A}_n, \tag{16.05}$$

or

$$q = 1 + \frac{\underline{A}}{k\underline{A}_n} A_{\max}/\underline{A}. \tag{16.06}$$

Comparison of (16.04) and (16.06) shows the following equivalence between intersymbol interference and noise from the standpoint of limitation on the permissible number of pulse amplitudes

$$\underline{U} = \underline{A}_n/\underline{A}, \tag{16.07}$$

or

$$\underline{U}^2 = D = N/P. \tag{16.08}$$

This means that random characteristic distortion has the same effect as a random noise power $N = DP$, where D is a distortion factor.

In view of the above equivalence, the channel capacity of a PCM system in the presence of random characteristic distortion, but without noise, as obtained by substitution of (16.08) in (16.02) becomes

$$C = f_1 \log_2 \left(1 + \frac{1}{KD}\right). \tag{16.09}$$

With random interference from both characteristic distortion and noise, the interfering powers add directly, so that for a PCM system

$$C = f_1 \log_2 \left(1 + \frac{1}{K(D + N/P)}\right). \tag{16.10}$$

The equivalence (16.08) was established above on the basis of discrete pulse amplitudes, but it is independent of q and would thus apply also for continuous signals. On this basis it would apply for any method of modulation or of encoding signals and the maximum channel capacity as given by (16.01) would be modified to

$$C = f_1 \log_2 \left(1 + \frac{1}{D + N/P}\right), \tag{16.11}$$

It follows from the above that for any modulation method the tolerable distortion factor is directly related to the average signal-to-noise ratio. Thus two modulation methods which are equivalent from the standpoint of signal-to-noise ratio are also equivalent from the standpoint of tolerable rms distortion, provided faithful reproduction of the transmitted signal is required, as assumed here.

From (8.14) the following relation is obtained between the distortion factor $D = \underline{U}^2$ and small rms deviations $\underline{a}$ (nepers) and $\underline{b}$ (radians) in the amplitude and phase characteristics

$$D = \underline{a}^2 + \underline{b}^2. \tag{16.12}$$

In order that characteristic distortion may be disregarded in comparison with noise, it is necessary that $D \ll N/P$ or

$$\underline{a}^2 + \underline{b}^2 \ll N/P. \tag{16.13}$$

For example, in communication systems employing the same bandwidth as the original signal, such as a pulse amplitude modulation system, a representative signal-to-noise ratio would be about 40 db, or $N/P = 10^{-4}$. In order that characteristic distortion may be disregarded in this case, it would be necessary for both $\underline{a}$ and $\underline{b}$ to be substantially less than 10^{-2} nepers and radians respectively. This would correspond to an rms gain deviation over the transmission band well below 0.08 db and an rms deviation from a linear phase characteristic well below 0.6 degrees. Since these tolerances are difficult to realize in actual systems, at least for wire circuits, characteristic distortion rather than noise may impose a limitation on channel capacity of systems employing about the same bandwidth as the original signal.

In accordance with (16.01), the bandwidth can in principle be halved without change in channel capacity if the signal-to-noise ratio is squared, i.e., if $N/P = 10^{-8}$ rather than 10^{-4} in the previous example. The tolerable rms amplitude and phase deviations would then be

$$\underline{a}^2 + \underline{b}^2 \ll 10^{-8}.$$

Thus both a and b would have to be substantially smaller than about 10^{-4}, which would preclude a substantial bandwidth saving in practical systems from the standpoint of characteristic distortion, even if it were feasible from the standpoint of signal power required to override noise.

The above considerations apply when faithful reproduction of the transmitted signal is required, as for example in data transmission. In speech transmission considerable distortion can be tolerated, a circumstance which permits appreciable phase distortion in the usual frequency division system without noticeable impairment of intelligibility, but which cannot be taken advantage of in time division pulse systems because of the resultant intersymbol interference. The characteristics of speech sounds also permit a reduction in the bandwidth of the original transmitted signal, by such devices as vocoders or frequency compandors, without excessive impairment of intelligibility.

ACKNOWLEDGMENTS

This paper is based on studies in connection with the application of pulse systems to wire circuits, suggested by M. L. Almquist and J. T. Dixon and carried out under their direction. Some of these studies were made on behalf of the Signal Corps, under contract W-36-039-SC-38115. The Signal Corps Engineering Laboratories have consented to the publication of results obtained in these studies. The writer had the benefit in these studies of discussions with C. B. Feldman and W. R. Bennett, and of some of their memoranda on pulse modulation systems. He is also thankful for comments and advice from F. B. Llewellyn and others in connection with preparation of the paper.

REFERENCES

See Part I.