R. Landauer

Shock Waves in Nonlinear Transmission Lines and Their Effect on Parametric Amplification

Abstract: The propagation of a periodic signal on a transmission line with a nonlinearity in the distributed capacitance is examined. The signal is deformed during its propagation and electromagnetic shock waves are generated. It is pointed out that the shock wave will form in a distance which is short for any parametric amplification purposes. The subsequent growth of the shock and its decay, due to the inevitable dissipation associated with a shock, are analyzed assuming that the capacitance variations are small compared to the total capacitance. The propagation of a small deviation from a signal which is perfectly periodic in time is also examined, and it is shown that the small deviation may spread out in time but cannot be changed in its sign. This result was invoked in an earlier paper demonstrating the impossibility of parametric amplification on dispersionless nonlinear lines.

1. Introduction

An earlier paper[1] pointed out that a large-amplitude signal moving along a nonlinear transmission line was subject to deformation, since various portions of the signal move with different velocities. The present note is intended to describe this deformation in more detail, particularly with regard to the formation of shock waves and their effect on parametric amplification. The deformation process is of some interest in itself, as a method of harmonic generation and as a method by which one end of a pulse can be sharpened at the expense of the other. This paper will be entirely concerned with the behavior of traveling waves. A separate note[2] has pointed out that some interesting parametric amplification effects are obtainable with standing waves.

Our considerations on this subject were largely motivated by the possibility of constructing transmission lines with a ferroelectric material just above its Curie point. In a short temperature range, just above the temperature at which ferroelectric materials lose their spontaneous polarization, they are strongly nonlinear dielectrics.[3,4] The details of the theory developed below, however, are not tailored particularly to the ferroelectric case. First of all, ferroelectrics are not the only possible embodiment for a nonlinear electromagnetic medium. Secondly, ferroelectrics have special properties which would considerably complicate the analysis but probably would not affect the relevant phenomena very greatly.

Of the special properties of ferroelectrics that we have in mind we may list first the dispersion at low frequencies, which is due to electromechanical effects. At low frequencies, the lattice distorts mechanically as induced polarization is developed. At high frequencies the mechanical inertia of the lattice prevents this distortion and results in a different D-vs-E relationship. This dispersion does not show up in the dielectric constant measured at $D=E=O$. It does, however affect the nonlinear behavior.[3] If we assume that all the frequency components of the electrical signal are well above the important mechanical transverse resonance of the structure, and if furthermore the velocity of electromagnetic waves along the structure is high compared to that of sound waves, then this source of dispersion will be unimportant. This condition will certainly be well satisfied in the microwave range. A second property of ferroelectrics that will be ignored, but only in part, is a high-frequency dispersion. There will, in actuality, be deviations from a simple capacitive D-vs-E relationship at frequencies which are high enough so that the particles which move to establish the polarization are moved sufficiently fast to be subject to damping and/or inertial effects. This is presumably the source of the high loss factor measured near the Curie point in $BaTiO_3$ by Benedict and Durand.[5] We will not ignore the existence of this dispersion but will assume that it occurs at very high frequencies, so that it is relevant only in determining the width of electromagnetic shock waves. Our theory is thus based upon a simple, frequency-independent D-E relationship. It is unlikely that there will be any nonlinear reactance, electric or magnetic, which really meets such specifications. This theory must then be regarded as a limiting case, which can be approached in varying degrees by practical systems. The phenomena of wave distortion and shock wave formation will presumably show up, in a watered-down way, even in such systems as a linear transmission line, periodically loaded with nonlinear *p-n* junction capacitances.

It is interesting to note that the connection between the nonlinear electromagnetic problem and the older hydrodynamics problems was first pointed out by Salinger[6] in 1923. Salinger treats the deformation process much as we will and also recognizes the fact that there must be shock wave formation. He, however, requires a propagating shock to be a nondissipative phenomenon. This inconsistency leads to erroneous conclusions about the conditions under which shocks form, and completely eliminates the signal decay caused by the shock formation.

Our treatment is intended, at least in principle, to be readable without a previous knowledge of shock wave theory, but the arguments are likely to be much more transparent to the reader who is familiar with the general notions of shock wave treatments.

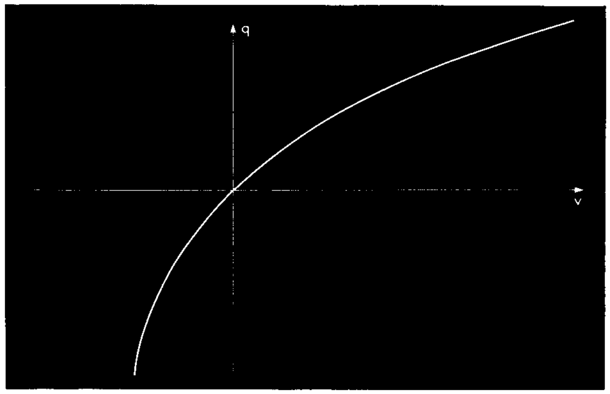


Figure 1 **Capacitive characteristic used in most of the discussion in this paper. The symbol q is the charge per unit length of line and v the voltage across the line. The capacitance is a monotonically decreasing function of the voltage.**

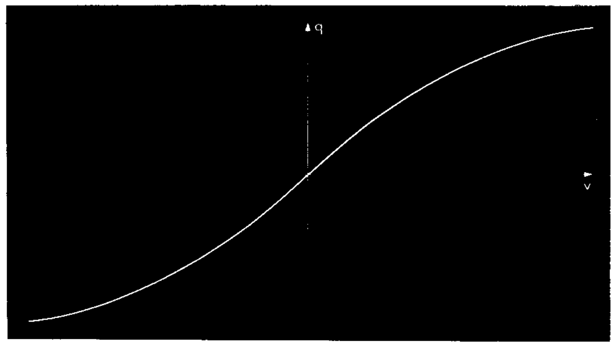


Figure 2 **Capacitive characteristic typical of an unbiased ferroelectric line. The capacitance has its maximum value in the absence of a signal.**

2. General shock wave behavior

In this section we will describe the general behavior of a propagating periodic pump signal on a nonlinear line. We will state results here, without much justification, leaving the detailed arguments to the subsequent sections. In this section as well as in all subsequent ones, the nonlinear characteristic we have in mind is sketched in Fig. 1. The relevant point is that $d^2q/dv^2 < 0$. The treatment of the case where $d^2q/dv^2 > 0$ is a trivial variation on the one we shall discuss, and does not require separate comments at each stage. Wherever the case $d^2q/dv^2 < 0$ leads to a sharpening of wave fronts, the case $d^2q/dv^2 > 0$ leads to a spreading out, and vice versa. The characteristic shown in Fig. 2, which is typical of an unbiased ferroelectric, leads to a somewhat more complicated analysis. For a signal which includes both positive and negative voltages, d^2q/dv^2 changes sign at $q = v = 0$. The modifications to our basic theory, necessary for the description of such a signal, are indicated in Section 7.

Figure 3 is a typical sketch of a propagating pump signal on a transmission line with the kind of capacitance characteristic shown in Fig. 1. The significant point in the analysis of the signal is that each value of voltage moves with its own characteristic value of velocity, $1/\sqrt{lc}$, where c is the differential capacitance dq/dv determined by that voltage. Near the beginning of the line the signal is still close to its initial sinusoidal form. Point A, however, moves faster than Point B. In the cycle which has progressed farther down the line this has caused Point C to close in on the preceding minimum, Point D, and to move away from the succeeding one, Point B. Eventually the wave form becomes infinitely steep, and a shock wave starts to form. FG denotes a shock which has not yet grown very large. The shock has a velocity which is intermediate between the velocities characteristic of its end points. The shock FG will, therefore, move faster than the point H, and as the shock catches up with portions of GH, the shock will grow. Similarly the wave between E and F will move faster than the shock and, as it catches up with the shock, the shock also grows in that direction. The shock reaches a maximum amplitude, somewhere near the stage shown at JK in Fig. 3. Subsequently it will diminish. The point I will move faster than the shock, and by the time it has caught up with the shock, the upper end of the shock will be reduced to the voltage of I. Similarly the shock JK moves faster than the point L, and as it catches up with L, the lower end of the shock will be raised the voltage of level L. Eventually all portions of the wave move into the shock, or are caught by the shock, and the shock deteriorates asymptotically to the vanishing point.

The general description we have given is, in part, the result of an approximation whose nature will be pointed out in the subsequent discussion.

3. Criterion for shock formation

If l is the inductance per unit length of our line, $c(v) = dq/dv$, the differential capacitance per unit length, and if i, q, and v denote respectively the transmission line current, charge per unit length and voltage, then the basic equations are

$$l \frac{\partial i}{\partial t} = -\frac{\partial v}{\partial z}, \tag{3.1}$$

$$\frac{\partial i}{\partial z} = -\frac{\partial q}{\partial t} = -c(v)\frac{\partial v}{\partial t}. \tag{3.2}$$

These equations have a set of particularly simple solutions, corresponding to the traveling waves for a strictly linear transmission line. These "simple waves"[8] have the property that in the (z, t) plane the lines of constant v are straight lines with a slope

$$\frac{dz}{dt} = \frac{1}{\sqrt{lc(v)}}. \tag{3.3}$$

Every portion of the wave has a velocity characteristic of the voltage at that portion of the wave. The correctness of these solutions, $v = v\left(z - \frac{1}{\sqrt{lc(v)}}t\right)$ can be verified by direct substitution in the wave equation, resulting from (3.1) and (3.2)

$$\frac{\partial^2 v}{\partial z^2} - l\frac{\partial}{\partial t}\left[c(v)\frac{\partial v}{\partial t}\right] = 0, \tag{3.4}$$

or else by the more physical method of reasoning, employed in the earlier paper.[1] The difference in velocities produces a distortion in the moving wave. Voltages associated with a low capacitance, $c(v)$, move fast and will tend to catch up with earlier slower portions of the wave which are associated with high values of $c(v)$. Figure 4 shows four successive stages of deformation in the history of a pulse moving to the right. The deformation, as shown, is a result of the nonlinearity displayed in Fig. 1. Initially, in Fig. 4a, the pulse is symmetrical. Subsequently Point B comes closer to A, and moves away from

Figure 3 **Voltage as a function of distance along the line, due to a periodic voltage applied at z=0. *FG* is a shock wave that has just formed. The shock wave *JL* is near its maximum amplitude. *MN* is a shock wave decreasing in amplitude.**

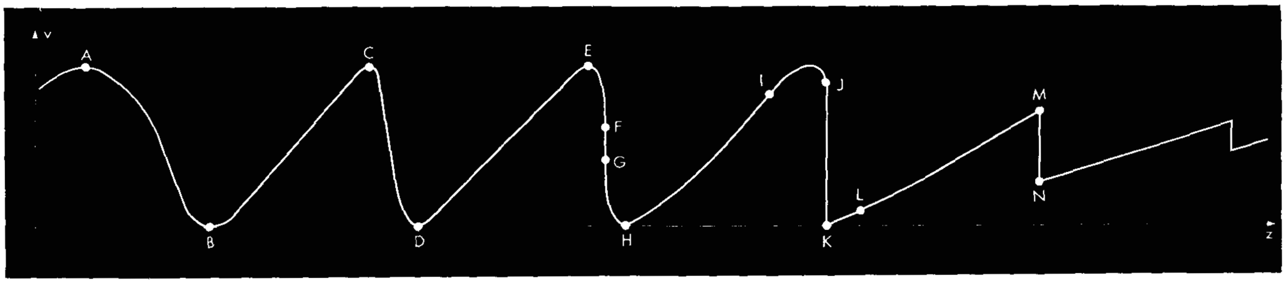
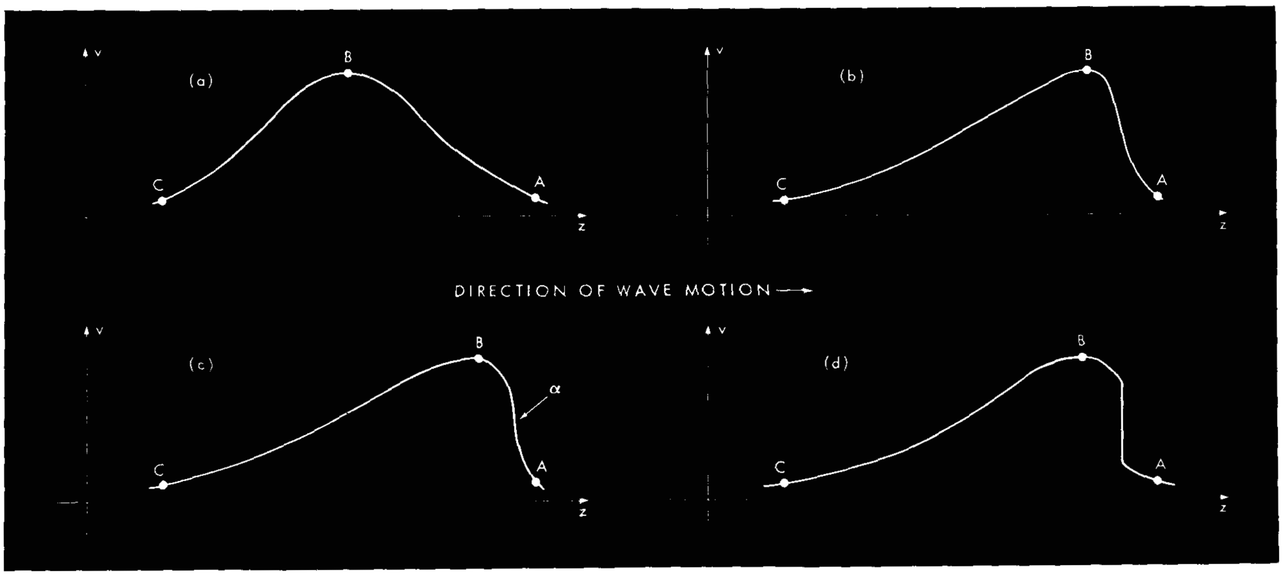


Figure 4 **Four successive stages in the deformation of a propagating pulse. ∂v/∂z becomes infinite at point α, which is the point where the shock formation starts.**



C. Eventually voltages launched at slightly different times, will catch up with each other, and $\partial v/\partial z$ becomes infinite, as at point α in Fig. 4c. Subsequent voltages, between α and B, will tend to catch up with this steep portion of the wave front, while the steep portion of the wave front catches up with voltages ahead of it, between α and A. This results in the growth of the shock wave, as shown in 4d. The motion and growth of the shock wave will be treated in subsequent sections. In this section we will be concerned with the first onset of the shock, as shown in Fig. 4c.

The propagation of a given value of v along the line is given by Eq. (3.3) or equivalently

$$z=(t-t_i)u(v)\,, \tag{3.5}$$

where $u=1/\sqrt{lc}$ and where t_i is the moment when the voltage v started down the line. The shock wave onset in Fig. 4c corresponds to

$$(\partial z/\partial t_i)_t=0\,. \tag{3.6}$$

Applying the condition of Eq. (3.6) to Eq. (3.5) yields

$$t-t_i=u(t_i)/u'(t_i)\,. \tag{3.7}$$

Let us now consider the special case relevant to the propagation of the pump signal in parametric amplification. Consider a pump signal which causes a sinusoidal capacitance variation at the input end of the line

$$c=c_0(1+\xi\sin\omega t_i)\,. \tag{3.8}$$

Then

$$u'(t_i)=\frac{d}{dt_i}\left(\frac{1}{\sqrt{lc(t_i)}}\right)=-\frac{1}{2\sqrt{lc}}\frac{1}{c}\frac{dc}{dt_i}\,. \tag{3.9}$$

From Eq. (3.8) we have

$$\frac{dc}{dt_i}=c_0\xi\omega\cos\omega t_i\,. \tag{3.10}$$

Substituting (3.10) in (3.9) yields

$$u'(t_i)=-\frac{u}{2}\xi\omega\cos\omega t_i/(1+\xi\sin\omega t_i)\,. \tag{3.11}$$

For analytical simplicity we will now specialize to the case of small ξ. This permits us to replace the denominator on the right-hand side of Eq. (3.11) by unity. Using this simplified form in turn in Eq. (3.7) gives

$$t-t_i=-2/(\xi\omega\cos\omega t_i)\,, \tag{3.12}$$

or

$$t=t_i-\frac{2}{\xi\omega\cos\omega t_i}\,. \tag{3.13}$$

Adjacent trajectories of the form (3.5) cross whenever (3.13) is satisfied, and there is a continuum of such situations.

We are interested in the earliest time, $t(t_i)$, permitted by (3.13). Differentiating gives

$$\frac{dt}{dt_i}=1-\frac{2}{\xi\omega\cos^2\omega t_i}\,\omega\sin\omega t_i=0\,, \tag{3.14}$$

which has its only real solution

$$\sin\omega t_i=(\sqrt{1+\xi^2}-1)/\xi\simeq\xi/2\,. \tag{3.15}$$

To first order in ξ therefore, $\cos\omega t_i=-1$, and $u=1\sqrt{lc_0}$ at the formation of the shock. The distance from the input end, at the time of the first shock formation is

$$z=u(t_i)(t-t_i)=\frac{1}{\sqrt{lc_0}}(t-t_i)\,, \tag{3.16}$$

where for $(t-t_i)$ we can use Eq. (3.12), resulting in

$$z=\frac{1}{\sqrt{lc_0}}\left(\frac{-2}{\xi\omega\cos\omega t_i}\right)=\frac{2}{\sqrt{lc_0}\,\xi\omega}\,. \tag{3.17}$$

The question now arises, how much parametric amplification is available in this distance? The most optimistic treatment for parametric transmission line amplification is that of Tien and Suhl.[9] According to their theory a signal can grow along the line as $e^{\alpha z}$ where $\alpha=\frac{1}{4}(\xi^2\beta_1\beta_2)^{1/2}$. Here ξ is the same nonlinearity parameter we introduced in Eq. (3.8). For the dispersionless line being presently considered the propagation constants β_1 and β_2 equal ω_1/u_0 and ω_2/u_0 respectively, where ω_1 is the frequency being amplified and ω_2 is the idling frequency. $u_0=1/\sqrt{lc_0}$ is the propagation velocity along the unpumped line. The maximum value of α occurs when ω_1 and ω_2 are approximately equal, and close to one-half of the pump frequency (if ω_1 and ω_2 are *exactly* equal to one-half the pump frequency, the amplification depends on the signal phase, and the Tien and Suhl treatment does not apply). This case gives $\alpha=\frac{1}{4}\xi\omega/u_0$. The gain obtainable in a length of line short enough to prevent the formation of shock waves must then be less than $e^{\alpha z}$, where the value of z given by Eq. (3.17) is relevant. The maximum gain then is:

$$\exp(\alpha z)=\exp\left(\frac{1}{4}\frac{\xi\omega}{u_0}\cdot\frac{2}{\sqrt{lc_0}\,\xi\omega}\right)=\exp(1/2)\,. \tag{3.18}$$

We see therefore that the deformation which leads to shock wave formation becomes effective before very much parametric amplification is available. Actually, of course, no parametric gain at all is available on a dispersionless line,[1] and the preceding exercise only shows that if we have two lines, one dispersionless, and one to which the Tien and Suhl theory is really applicable, and if the capacitance variation at the input is the same for both lines, then the dispersionless line will develop shock waves in a length of line in which the amplifying line cannot amplify by more than exp(1/2).

4. Shock wave motion

If we were to continue to apply Eqs. (3.1) and (3.2) to a simple wave, after it has reached the stage depicted in Fig. 4c, the solution would become multiple-valued, as shown in Fig. 5, since the trajectories of the type of Eq. (3.5) will overlap. This is clearly not a physically significant solution. On the other hand there is no reason for the wave front to become less steep again. The only way out of this dilemma is to assume that a discontinuity, as shown in Fig. 4d, will be formed, and that at this discontinuity the $q-v$ relationship of Fig. 1 becomes inapplicable. This is not unreasonable, since a very steep wave front implies very rapid charge changes, and if the charge motions are sufficiently rapid the *D-E* relationship must break down and show a dispersion. We therefore assume that in actuality some effect, such as a finite relaxation time for the ferroelectric when changing its polarization, will really prevent the wave front from ever actually achieving infinite slope, but that the relaxation time is short enough, so that the wave front can become very steep. The motion of such a steep wave front can be treated without taking into account the detailed behavior of the relaxation (or other dispersion mechanism). A treatment of the shock front, calculating its thickness as a function of the dispersion behavior, can be given[7] but is not relevant to our present purposes.

Consider the two relationships

$$l\frac{\partial i}{\partial t}=-\frac{\partial v}{\partial z}, \tag{4.1}$$

$$\frac{\partial q}{\partial t}=-\frac{\partial i}{\partial z}. \tag{4.2}$$

They can be assumed to be accurately valid, even in the region where q and v are changing very rapidly. Now consider two planes, such as 1 and 2 in Fig. 6. Plane 2 is assumed to be just ahead of the steep portion of the wave front, while Plane 1 is just behind it. Assume that the wave front has a velocity of motion u. Now let us integrate equations (1) between the two planes. This gives

$$l\frac{\partial}{\partial t}\int_{z_1}^{z_2} i dz=-[v(z_2)-v(z_1)], \tag{4.3}$$

$$\frac{\partial}{\partial t}\int_{z_1}^{z_2} q dz=-[i(z_2)-i(z_1)]. \tag{4.4}$$

Now if z_1 and z_2 are sufficiently close to the steep region, then the integrals involved in Eqs. (4.3) and (4.4) have a time derivative only because of the motion of the steep portion, and the slower changes due to the motion of the parts of the wave with moderate slope can be neglected. If the steep wave front moves with velocity u, then the integrals change because in a time dt a section of line udt in length and within the range of integration has values $i(z_2)$ and $q(z_2)$ replaced by $i(z_1)$ and $q(z_1)$. Eqs. (4.3) and (4.4) therefore become:

$$lu[i(1)-i(2)]=-[v(2)-v(1)], \tag{4.5}$$

$$u[q(1)-q(2)]=-[i(2)-i(1)]. \tag{4.6}$$

Eliminating $[i(2)-i(1)]$ from Eqs. (4.5) and (4.6) we find

$$u^2=\frac{1}{l}\frac{v(2)-v(1)}{q(2)-q(1)}=\frac{1}{lc_s}, \tag{4.7}$$

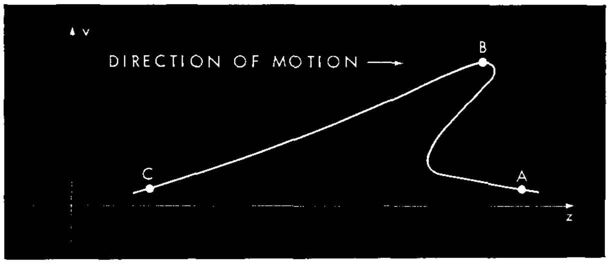


Figure 5 **Multivalued potential function, which results from invoking Eqs. (3.1) and (3.2) or Eq. (3.5) after the propagating wave reaches the stage shown in Fig. 4c.**

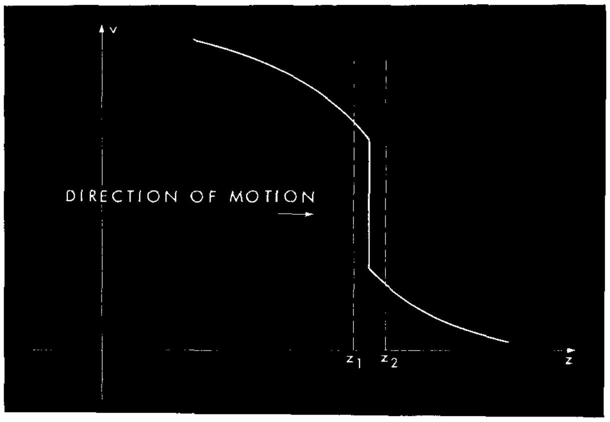


Figure 6 **Shock wave moving to the right. Lines z_1 and z_2 are positions of planes close to the shock.**

where c_s is the average capacitance between z_2 and z_1, i.e., the slope of the straight line, as shown in Fig. 7. We can immediately see from Fig. 7 that $u(2)<u<u(1)$. Therefore the parts of the wave following the shock move faster than the shock, and will catch up with it and enlarge the shock. Similarly the shock will catch up with the slower-moving wave ahead of it, and this will also enlarge the shock. Actually when a portion of a signal merges with the shock, reflections are generated. These are, however, weak if the capacitance variation between the end points of the shock is small compared to the capacitances. This is the case which we will consider hereafter. The motion of a shock is inevitably accompanied by the dissipation of energy, and the dissipation is essentially independent of how narrow the shock is, once the wave front has already become steep compared to the rest of the wave shape and is limited from further steepening by the dielectric relaxation time. This energy dissipation can be evaluated by taking the energy flow across the plane at z_1 subtracting from it the energy flow across the plane at z_2 and also subtracting the rate at

which the shock, in causing a change from State 2 to State 1, leaves stored energy in its wake. The result of this consideration is that a shock, in passing through a unit length of line, dissipates an amount of energy equal to the cross-hatched area of Fig. 7.

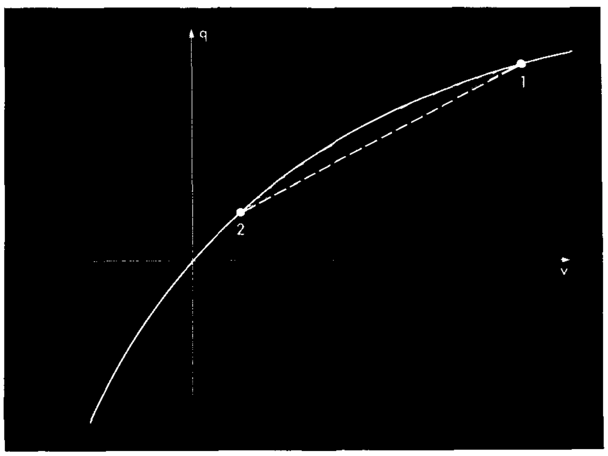
This dissipation actually causes the entropy of the final State "1" to be different from that of the initial State "2". Since Fig. 7 represents an adiabatic q-v characteristic, the two states involved cannot really be on the same curve, as drawn in Fig. 7. It can be shown, however, that for typical ferroelectric nonlinearities, the correction involved is negligible.[7]

5. Signal attenuation by shock waves

In this section we will be concerned with determining the behavior of the pump signal, once shock waves have formed and caused energy loss. In its complete generality this is a rather formidable problem, and we have been able to handle it only by confining ourselves to the case where the capacitance variation is a small fraction of the capacitance. In this case the reflections formed when part of the continuous wave merges with a shock can be neglected. As a consequence the continuous portions of the wave are unaffected by the shock formation until they actually merge with the shock. This method is taken from the literature on fluid flow problems,[10] but a brief independent justification of it is given in Appendix A.

The shock waves first form at the position given by Eq. (3.17). They will grow subsequently, reach a maximum, and then deteriorate, as shown in Fig. 3. We shall describe the growth and attenuation of the shock wave,

Figure 7 **Point 2 is the initial state of a line, just before being reached by a shock wave, and corresponds to the location z_2 in Fig. 6. Point 1 corresponds to z_1 and gives the charge and voltage after shock passage. The slope of the dotted straight line is c_s, the shock capacitance. The cross-hatched area between curve and straight line is the dissipation, per unit length of line, caused by the shock passage.**



using the approximation described above, which permits us to use trajectories as given by Eq. (3.5) to describe the propagation of those parts of the signal which have not yet merged with the shock. We shall also assume that at the input end Eq. (3.8) applies, with $\xi \ll 1$.

This problem becomes very simple if we assume that the shock has the velocity $u_0 = 1/\sqrt{lc_0}$ at all times. Is this reasonable? When the shock first forms, $u = u_0$, neglecting second-order terms in ξ^2; this was pointed out in Section 3. The shock forms at a voltage halfway between the maximum and minimum pump voltages. What happens in its subsequent growth? The voltage associated with a velocity $u_0+\delta$ and the voltage associated with the velocity $u_0-\delta$ will merge with the shock, at the same time (neglecting again, second-order effects in ξ) as long as the shock retains its velocity u_0. As long as the shock remains symmetrical about the halfway voltage, however, its velocity will stay at u_0, to second order in ξ. Hence the shock velocity can remain at u_0. The argument just given shows that there is a solution which has a shock velocity which remains at u_0 (to first order in ξ). Is this a stable equilibrium? If for some reason the shock should suddenly grow faster on the low-voltage side than on the high-voltage side, the shock velocity would be depressed below u_0. Thereafter the high-voltage wave form would catch up with the shock more rapidly than before, while the shock would catch the low-voltage wave form less rapidly. As a result the shock would be accelerated, showing that the solution with fixed velocity u_0 is, in fact, stable.

Equation (3.8) leads to

$$u = \frac{1}{\sqrt{lc}} = u_0\left(1 - \frac{\xi}{2}\sin \omega t_i\right), \tag{5.1}$$

and this variation is depicted in Fig. 8. Let us consider two points on this curve, corresponding to $\omega t_i = \pm\theta$. These two points will join the shock wave at the same time. If this joining occurs at a distance z from the initial end of the line, then we must have:

$$\frac{z}{u_0\left(1 - \frac{\xi}{2}\sin\theta\right)} = \frac{z}{u_0\left(1 + \frac{\xi}{2}\sin\theta\right)} + \frac{2\theta}{\omega}, \tag{5.2}$$

where the final term, $2\theta/\omega$, represents the time delay with which the faster speed is launched on the line, as compared to the slower speed. Eq. (5.2) yields

$$z = \frac{2u_0}{\xi\omega}\frac{\theta}{\sin\theta}. \tag{5.3}$$

The shock first forms for values of θ near zero. This corresponds to the minimum value of $z = 2u_0/\xi\omega$, in agreement with the earlier results of Eq. (3.17). The shock reaches its peak amplitude for $\theta = \pi/2$, giving $z = \pi u_0/\xi\omega$. Thus the shock develops to its maximum amplitude, in about half the additional distance that it took the shock to form in the first place. The shock decays when θ becomes greater than $\pi/2$. Its asymptotic decay corresponds to $\theta = \pi - \varepsilon$ where $\varepsilon \ll \pi$. This yields

$z = 2\pi u_0 / \xi\omega\varepsilon$, (5.4)

so that eventually the shock amplitude decreases inversely as the distance along the line.

6. Localized deviation

In this section we will be concerned with a signal which deviates from the periodic pump signal by a small amount. The deviation will be assumed highly localized in space and time. As has been pointed out in the earlier paper,[1] all other signals can be considered as combinations of such localized signals. The earlier paper considered deviations of the form shown in Fig. 9. Points P and Q follow their unperturbed trajectories. The intermediate points, for both the original signal and the perturbed signal, have intermediate velocities. The localized small signal will be compressed, or expanded depending on the relationship of $u(P)$ to $u(Q)$. The extra charge associated with the deviation must, however, be conserved.

The development of the signal outside of the range PQ is unaffected by the small deviation, as long as the range PQ has not merged with a shock. What happens after PQ merges with a shock? After this time the continuous portions of both the pure pump signal and of the perturbed signal consist of voltages whose trajectories were undisturbed by the small signal between P and Q. This, of course, assumes the approximation of Section 5, which neglects reflections. The two waves under consideration, the perturbed and the pure pump signal, must differ, however, since charge is conserved. To see how they can differ, even though their continuous portions are defined by the same trajectories, consider these trajectories in the immediate vicinity of a shock. As is apparent from Fig. 5, a shock forms after the continuous solution, predicted by the method of trajectories, becomes multivalued. The shock represents a transition from one branch of the multivalued function to another. The multivalued function, by itself, does not determine the position of the shock, except that it limits the shock to the region in which the function has three branches. Hence two solutions can correspond to the same trajectories as far as their continuous portions are concerned, but can differ in the position of the shock.

After PQ has merged with the shock, the two waves must then be as shown in Fig. 10. The extent to which the two shocks are separated is determined by the condition of charge conservation. We thus see that a localized charge remains localized, even after it merges with the shock. This is the basic fact which was already invoked in the earlier paper[1] to show that parametric transmission line amplification was impossible.

It is interesting to note, however, that in a very special way, a nonlinear transmission line can provide amplification. Consider a small signal of the form shown in Fig. 9. If it is situated on a line where $u(P) > u(Q)$, the small signal will be contracted in time as it travels down the line. Charge conservation then requires the local voltage disturbance v, due to the small signal, to be inversely proportional to the length of time T of the small signal.

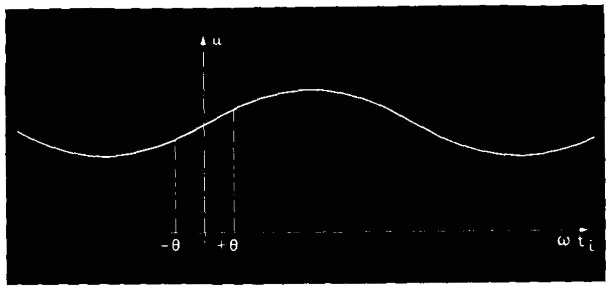


Figure 8 **Variation of wave velocity, as a function of time, at the initial end of the line. The portion of the wave launched at $\omega t_i = +\theta$ will merge with the shock at the same time as the portion of the wave launched at $\omega t_i = -\theta$.**

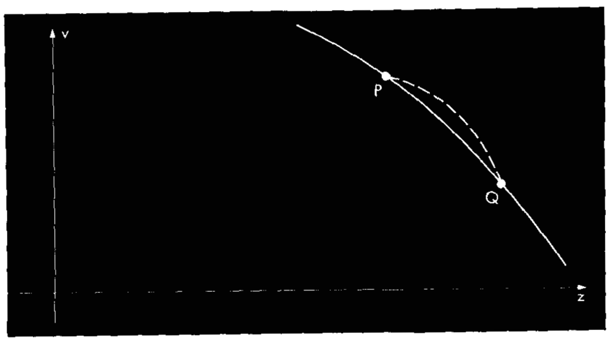


Figure 9 **The solid line is a portion of the periodic pump signal. The dotted line indicates a small, highly localized deviation.**

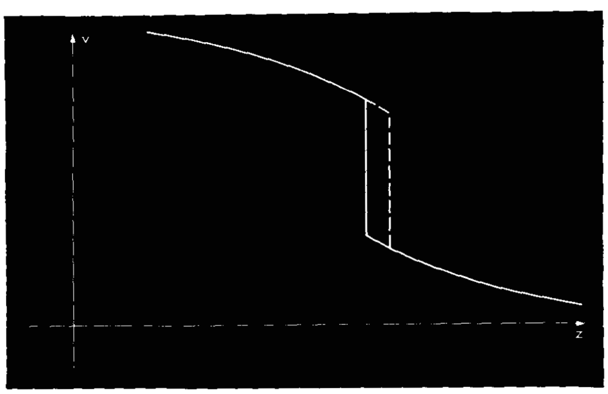


Figure 10 **The solid curve and the dotted curves represent two signals whose constant voltage trajectories in the (z, t) plane are identical, but whose shock positions differ.**

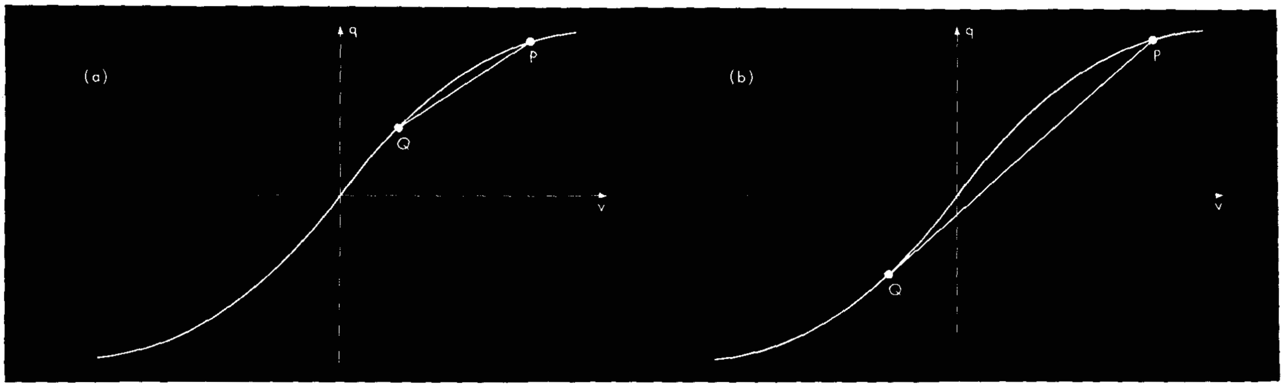
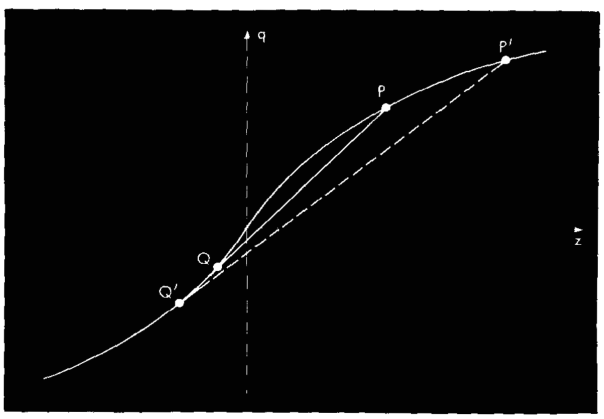


Figure 11 **a) Shows a symmetrical capacitance characteristic, with Q and *P* as the states, respectively, before and after shock passage. b) Shows the *same* shock after it has grown. Line *PQ is* tangent to the *q-v* curve at Q.**

Since the small-signal impedance level does not change as the signal propagates, the energy, which is proportional to $\int v^2 dt$, varies as $1/T$. Hence a small signal, riding on a portion of a wave which is contracting in time, will have its frequencies raised and its energy content raised proportionately. The simple blob shown in Fig. 9 is not essential to this argument—the small signal can be oscillatory. A small signal can, of course, also be attenuated by being placed on the line during a time when the small-signal capacitance at the input end is being increased. Quite aside from the energy changes involved, when a small signal travels down the nonlinear line together with a large signal, the small signal has its transit time controlled by the large signal. Such a line could therefore be used for frequency and phase modulation.

The earlier paper[1] dealt separately with the case in which the small-signal frequency was a multiple of exactly one-half the pump frequency. Consider the most important case, in which the signal frequency is exactly half the pump frequency. If suitably synchronized with the pump signal, then each positive half-cycle of the small signal charge can merge with one shock wave, and the succeeding negative half-cycle with the next shock wave. Thus each half wave of the small signal charge is compressed to a Dirac δ-function. This kind of compression was considered in the earlier paper, and it was pointed out that it corresponds to an increase in the fundamental component, by a factor $4/\pi$. In the earlier paper, however, this was put forth only as an upper limit for the possible gain, and it was not shown to be actually achievable.

Figure 12 **Line *PQ* represents a shock with a tangency at Q so that $c_s = c(Q)$. Subsequently a higher voltage *P′* catches up with the shock. This permits the shock to grow at its lower end to the new point of tangency *Q′*.**



At the other exceptional frequencies, the higher multiples of one-half the pump frequency, several half cycles of the signal will coalesce with one shock. If the number of half cycles is even, the signal will disappear completely. If the number of half cycles is odd, the signal will not disappear, but the component at the fundamental frequency will always be reduced.

In the preceding discussion of the behavior at the exceptional frequencies we have assumed a *q-v* characteristic and a pump signal wave form leading to one shock wave per pump cycle. One can conceive cases which lead to more than one shock per pump cycle. It can, however, be shown that even in these cases $4/\pi$ is a limit on the possible gain.

7. Symmetrical line

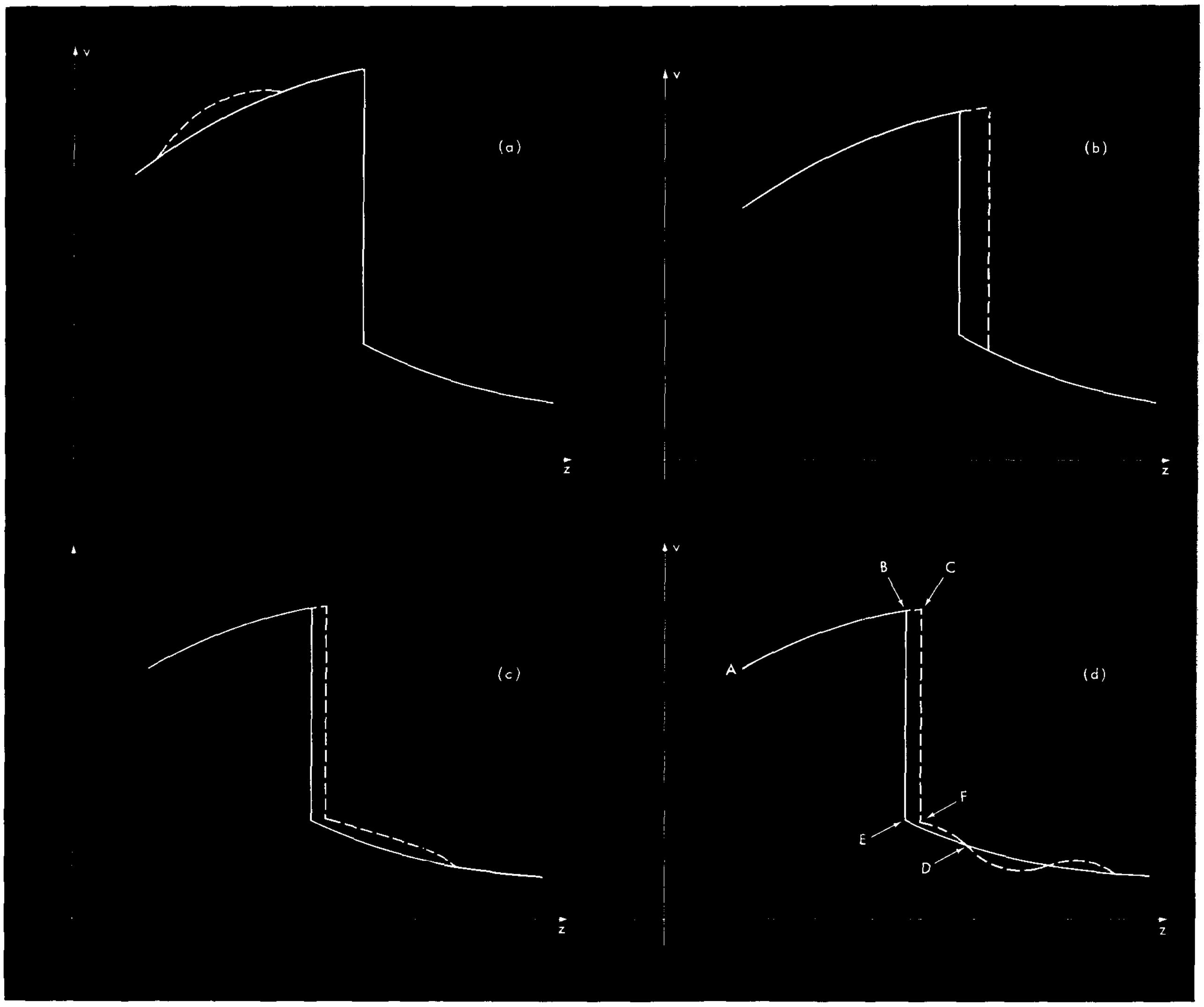
We will briefly indicate here the modifications involved if we are concerned with a *q-v* relationship of the type shown in Fig. 2. First of all, it is clear that a single cycle of the pump can take us through two high-capacitance and two low-capacitance portions of the wave, and that therefore each cycle of the pump signal will lead to two shock waves instead of one.

A more subtle difference arises after shock wave formation. Figure 11a shows the terminal points for a shock which has not yet developed very far. In its growth it will eventually reach the stage shown in Fig. 11b. Here P and Q are the terminal points of a shock. The straight line which connects P to Q is tangent to the curve at A. Hence the shock capacitance, c_s, equals the capacitance of the line immediately ahead of the shock and is larger than the line capacitance some distance ahead of the shock. Therefore the lower end of the shock will not continue to grow, if P remains fixed. If P is not yet the maximum voltage in the cycle, the shock can still grow at this end. This will decrease the shock capacitance and simultaneously permit the lower end of the shock to catch up with the continuous portions of the wave until the condition of tangency is restored. This sequence is shown in Fig. 12. Once P has reached the maximum voltage in the cycle, it must decrease subsequently, causing the shock velocity to be decreased. As a result some portions of the shock, at its lower end, are free to travel faster than the shock and will peel off, restoring the condition of tangency. The portions that peel off one shock will eventually reach the preceding shock and merge with it.

How does this affect the behavior of the small signal? Fig. 13a shows a pure pump signal, with a shock wave, and also shows the deviation due to a highly localized small signal. When the small signal has reached the shock, we can expect an immediate effect, as in Section 6 and as illustrated in Fig. 13b. Since the shock, in this case, generates a continuous signal ahead of itself we must expect at some later time a situation as shown in Fig. 13c, where the effects of the small signal are not localized to the immediate shock vicinity any more. What we wish to show specifically is that the situation must be as represented in Fig. 13c rather than in Fig. 13d.

Assume for the moment that Fig. 13d represents a possible situation. The portions AB and AC represent

Figure 13 **The solid line is the periodic pump signal, the dotted line indicates the history of a small deviation. Part a) is before the deviation merges with the shock; b) is just after the deviation catches up with the shock; c) is a still later state; and d) is not possible.**



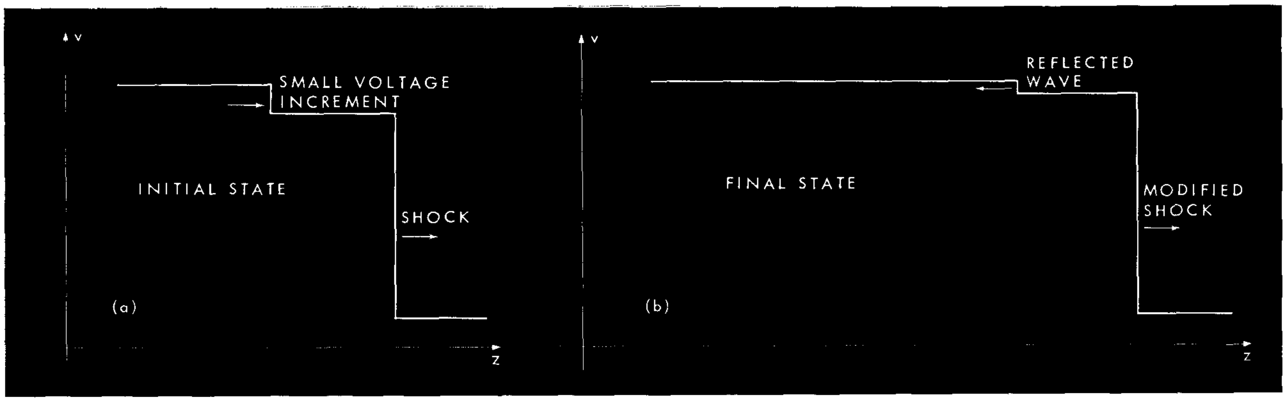


Figure 14 **In a) a small voltage increment is catching up with a shock. In b) it has caught up with the shock, produced a modified shock and a reflected wave.**

voltages whose trajectories of the form (3.5) are unaffected by the small signal. The trajectories all belong to the same branch of the same continuous solution of Eqs. (3.1) and (3.2). The difference between the two signals occurs only in the location of the shock, i.e., in the time and location where the signal departs from this branch. The two signals have the same value of voltage at D and therefore also the same velocity. The two signals thus have identical trajectories associated with the identical voltages at D. Since the two shock positions BE and CF are separated in space, the trajectory of D, followed backwards in time, must intersect the two shock positions at different times. The intersection of the trajectory of D with a shock, represents the moment when D "peeled off" the shock. This is the situation that was illustrated in Fig. 11b; at the moment when D "peels off," it must have the velocity of the shock. The velocity of the voltage at point D is, however, time-independent. Therefore when D peeled off the shock which subsequently becomes CF, it must have had the same velocity as when it peeled off the shock which subsequently becomes BE. Both shocks at the time of "peel off" have the same lower end voltage, namely the voltage later found at D. The two shocks can only have had the same velocity, i.e., the velocity of D, if they also had the same upper end (terminal) voltage, at their respective "peel off" moments. Hence the trajectory of D must intersect the trajectory of this upper shock end voltage in two points. The upper shock end corresponds, however, to a higher velocity than the shock and its lower end, and therefore the two trajectories can cross in only one point.

Since the situation shown in Fig. 13d is ruled out, the small signal retains its original sign, even though it does spread out ahead of the shock. It will retain its original sign, even when portions of it catch up with the preceding shock wave and "peel off" again from this preceding shock. It is this lack of sign change which was invoked in the earlier paper to show that parametric transmission line amplification did not exist.

Conclusion

The main conclusion has already been reached in the earlier paper,[1] i.e., parametric amplification cannot be achieved on transmission lines which are relatively dispersionless. For harmonic generation, for wave shaping, for intermittent amplification accompanied by a signal compression in time, or for the control of transit time through transmission systems these lines do have possibilities.

Appendix: Reflections can be ignored for weak shocks

Consider the situation shown in Fig. 14a, which is intended to depict a situation on a transmission line with a charge-voltage relationship as shown in Fig. 1. The large voltage drop is a shock wave, and a small voltage increment is catching up with it. After the two steps coalesce, the shock may be modified. No transmitted signal can appear ahead of the modified shock, since the velocity of the shock is higher than that of a small signal propagating on the line in its initial state. There can, however, be a reflected wave as shown in Fig. 14b. Let Z_{su} be the impedance of the upper end of the shock, that is, the ratio of change in shock voltage to change in current discontinuity as the initial state of the shock is kept unchanged, but the final state of the shock is changed slightly. Furthermore let Z' be the characteristic impedance of the line in its final state, after shock passage. Then the balances of currents and voltages that determine the strength of the reflected wave are exactly the same as for the problem in which a voltage wave on a line of impedance Z' is incident on a termination of impedance Z_{su}. Hence the ratio of reflected voltage to incident voltage is

$$\frac{v_r}{v_i}=\frac{Z_{su}-Z'}{Z_{su}+Z'}\,. \tag{A.1}$$

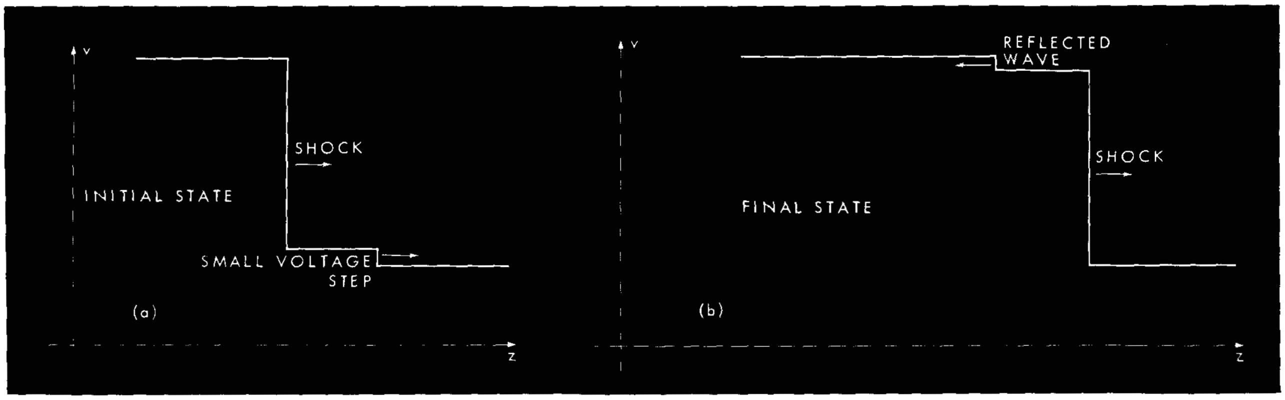


Figure 15 **Here a shock catches up with a small voltage increment, again causing a modified shock and a reflection.**

A straightforward auxiliary calculation shows

$$Z_{su} = \sqrt{\frac{l}{c_s}} \frac{2c_s}{c_s + c'}\,, \tag{A.2}$$

where c_s is as defined in Eq. (4.7) and c' is the capacitance associated with the final state. Somewhat similar considerations for the initial case of Fig. 15a, resulting in a final configuration shown in Fig. 15b, yield

$$\frac{v_r}{v_i} = \frac{Y'' - Y_{sl}}{Y_{su} + Y'}\,, \tag{A.3}$$

where the admittances are reciprocals of the previously defined impedances. Y'' characterizes the initial state of the line, ahead of the shock. Y_{sl} is an admittance defined for the shock when its initial state is changed and its final state kept fixed. The admittance Y_{sl} is given by:

$$Y_{sl} = \sqrt{\frac{c_s}{l}} \frac{c_s + c''}{2c_s}\,. \tag{A.4}$$

It is clear that v_r/v_i for both (A.1) and (A.3) is a fraction of the relative capacitance change associated with the shock wave. Hence if the latter is small, the reflections are small.

We could leave the argument at this point, if we were concerned with a fixed length of line. In most practical parametric amplifier schemes, however, as the strength of the nonlinearity is decreased, the length of time during which interaction occurs must be increased, which in the transmission line case means an increase in line length. As the line is increased, the number of shocks on it is increased and the number of reflections reaching a given point, but arising from different shocks, also increases. It is not clear, therefore, that as the line is made less nonlinear, but simultaneously longer, that the total of the reflections actually decreases and goes to zero as the linear case is approached.

It is clear, however, that as the line is made more linear, the wave following the shock takes longer to move into the shock. The reflections that are generated, therefore, are stretched out more, as the line becomes more linear. Within any one pump cycle on a mildly nonlinear line, therefore, the reflections due to shock waves ahead of it at best affect the dc level, rather than the detailed wave form, and its rate of distortion (unless the degree or sign of the nonlinearity is very critically dependent on dc bias level). It can, in fact, be shown that the approximate method of Section 6, which neglects reflections, can give somewhat incorrect values for the dc levels. For most purposes, however, this is not likely to be a serious error. The degree of error can generally be recognized, since Eqs. (3.1) and (3.2) require that the exact solution have the same time average value of v and i all along the line.

Acknowledgments

The author is indebted to L. H. Thomas of Watson Laboratory, who introduced him to the basic concepts concerning electromagnetic shock waves. A conversation with E. T. Jaynes of Stanford University also stimulated interest in this field.

References

1. R. Landauer, *J. Appl. Phys.*, **31,** 479 (1960).
2. R. Landauer, *Proc. IRE,* **48,** 1328 (1960).
3. M. E. Drougard, R. Landauer, and D. R. Young, *Phys. Rev.*, **98,** 1010 (1955).
4. S. Triebwasser, *IBM Journal,* **2,** 212 (1958).
5. T. S. Benedict and J. L. Durand, *Phys. Rev.*, **109,** 1091 (1958).
6. H. Salinger, *Arch. Elektrotech.*, **12,** 268 (1923).
7. R. Landauer and L. H. Thomas, *Bull. Amer. Phys. Soc.*, **4,** 424 (1959).
8. R. Courant and K. O. Friedrichs, *Supersonic Flow and Shock Waves,* Interscience Publishers, Inc., New York, 1948.
9. P. K. Tien and H. Suhl, *Proc. IRE,* **46,** 700 (1958).
10. Courant and Friedrichs, *op. cit.*, Sec. 74, p. 161.

Received October 7, 1959