# Graphical Analysis of Pulses on Lines

by B. L. Hart,* B.Sc., M.I.E.R.E.

**When transistor-transistor logic elements are connected together to form a digital system, spurious pulses can be introduced as a result of reflection effects on the interconnecting wires. The standard reflection chart approach is not appropriate to the analysis of waveshapes on transmission lines with non-linear terminations. The shape of the reflected pulses and, hence, the trouble they are likely to cause can be calculated by the graphical technique described using a new idea of a 'sliding load-line'.**

The conventional approach to the treatment of pulse reflections on uniform lossless transmission lines is based on the assumption of linear source and load impedance. The concept of reflection coefficient and the use of a suitable 'reflection chart' reduces many line problems to simple arithmetic calculations. In the practical construction of high-speed pulse generators capable of delivering waveforms with nanosecond and sub-nanosecond edges, precision coaxial cable (e.g. RG213U) and coaxial terminations (e.g. BNC, GR-874, etc.) are employed and the agreement between waveforms observed experimentally and predicted theoretically with the aid of a reflection chart is precise to a degree which is acceptable in the majority of applications likely to be encountered in practice. There are, however, some instances—specifically, those associated with the use of transmission lines with transistor-transistor logic—where the source and load impedances are distinctly non-linear. In these cases a reflection chart approach would be of limited use. Furthermore, a purely mathematical solution would present formidable difficulties.

However, a graphical approach, which avoids the computational problems, yields rapid results and enables us to see simply the effect of changes in system parameters. It is the background to the graphical technique which is explored in some detail in this article. First, though, it is necessary to review our basic ideas concerning pulses on lines.

## Line pulse fundamentals

Consider the simple arrangement shown in Fig. 1, in which a uniform, lossless, transmission line having characteristic impedance $Z_o$ and terminated by an impedance $Z_T$ is connected, at time $t = 0$, by a perfect switch to a battery in series with an impedance $Z_G$. (Lest it be thought that we are

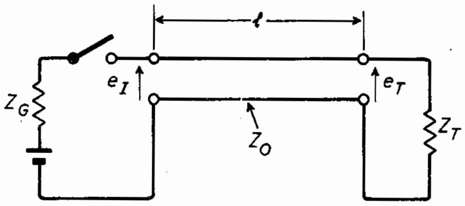



*Fig. 1. Transmission line with voltage step function drive*

already completely divorced from reality it must be borne in mind that a good practical approximation to a perfect switch is a specially-constructed mercury-wetted relay in a coaxial housing: the rise time of pulse edges using such a component is typically <2ns.)

A straightforward, though tedious, mathematical approach (see for example ref. 1) to the problem of finding the time function of current and voltage at any point, X, on the line is to solve the differential equations describing line behaviour in terms of its intrinsic properties for specified boundary conditions, viz. nature of driving function, particular values of $Z_G, Z_T$.

A practical engineering approach is different. The starting point is the appreciation that the appropriate differential equations are linear; thus, any two possible solutions can be a combined and the result is also an equally-valid solution. A physical interpretation of this is the possibility for the simultaneous existence of two waves travelling in opposite directions along the line. The voltage (or current) at a given $x,t$ is given by the algebraic sum of these two waves, the amplitudes of which are calculable for specified $Z_G, Z_T$. For each wave the voltage is related to the current by the impedance $Z_o$.

Thus, if in Fig. 2(a) the voltage and current components for a forward wave, i.e. travelling to the right in the direction of increasing $x$, are $e_f, i_f$ respectively, then

$$e_f/i_f = Z_o$$

Similarly, for a reverse wave (i.e. one travelling from right to left in the direction of decreasing $x$) characterized by voltage $e_r$ and current $i_r$ (Fig. 2(b))

$$e_r/i_r = Z_o$$

Now for any further quantitative treatment we must decide on, and adhere to, a consistent sign convention, and a long established one is this. A positive line voltage is one for which the 'top' conductor in Fig. 2 is positive with respect to the 'bottom', and a positive current is one flowing from left to right in the top conductor (and hence right to left in the bottom). For top and bottom read inner and outer respectively for a coaxial line system; a parallel line system is used here for each of drawing. On this basis $e_f, i_f$ are both positive. Write $e_f = e_+$, $i_f = i_+$ where '+' denotes the forward wave. Then the expression for the forward wave becomes

$$(e_+/i_+) = Z_o$$

In describing the reverse wave by a '−' subscript we have $e_r = e_-$ but now $i_- = -i_r$. Thus the expression for the reverse wave becomes

$$(e_-/i_-) = -Z_o$$

The emphasis on signs at this stage pays dividends later. The reason for the minus sign in this equation is not always well explained in textbooks and it is essential that it be clearly understood for future discussion.

Let us return to our specific problem:

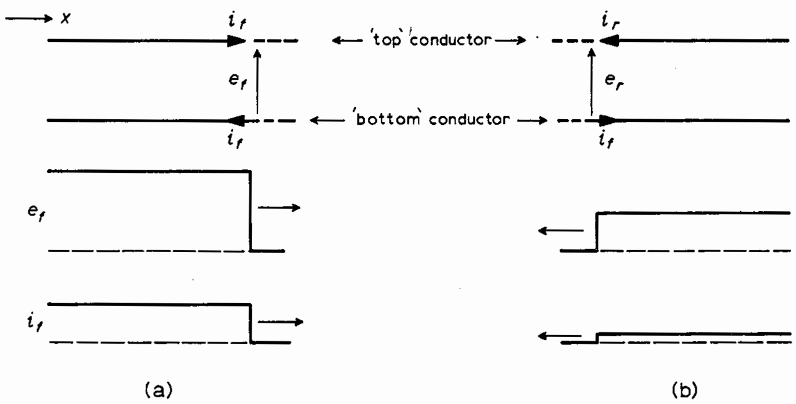



*Fig. 2. Sign conventions for (a) forward wave, and (b) reverse wave.*

*North-East London Polytechnic.

when the switch in Fig. 1 is closed at $t = 0$ the line appears as an impedance $Z_o$ irrespective of the nature of $Z_T$. (It is for this reason that $Z_o$ is sometimes referred to as 'surge impedance'.) For the low-loss cable assumed, this means a pure resistance, typically 50Ω. A forward wave thus starts down the line. No reverse wave is yet possible: the forward wavefront has got to discover first what exists at the other end of the line. Fig. 3 thus shows the simple equivalent circuit for calculating $e_+$. Clearly,

$$e_+ = e_I(0+) = Z_oE/(Z_o+Z_G)$$

and this voltage waveform travels towards $Z_T$ at a velocity $v$ dependent on the properties of the line and typically 0.3m/ns. As the wave progresses the electric and magnetic fields are in step and the energy is shared equally between the magnetic and electric fields. The current $i_f$ at $x$ charges up the line between $x$ and $(x+\delta x)$ to $e_+$ in a time $\delta t = (\delta x/v)$.

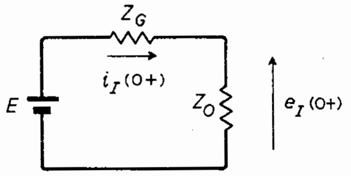



*Fig. 3. Equivalent circuit at line input at $t = 0+$.*

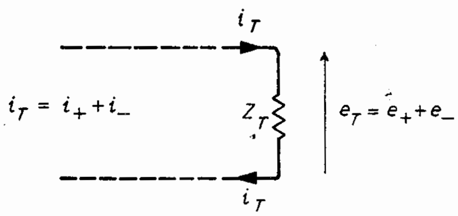



*Fig. 4. Circuit conditions at termination at $t = t_d$.*

At the termination all the incident energy in the forward wave is absorbed if the relationship that the line establishes between $e_+$ and $i_+$ is the same as the termination demands between the voltage across it and current in it at $t = t_d$, i.e. $e_T(t_d), i_T(t_d)$ respectively. For this condition

$$\frac{e_+}{i_+} = Z_o = \frac{e_T}{i_T} = Z_T$$

If $Z_T \neq Z_o$ all the energy in the forward wave cannot be absorbed at the termination. The law of conservation of energy thus demands a reflection phenomenon which is characteristic of all physical systems in which all the incident energy cannot be accepted. Thus light is reflected from the boundary between two media with different optical properties. Actually the optical analogy is a good one and it is profitable to regard the termination as a partially silvered mirror—but this goes beyond the scope of the present treatment.

The reflection of energy causes a reverse wave, characterized by $e_-$ and $i_-$ which must be such that, at the termination (see Fig. 4)

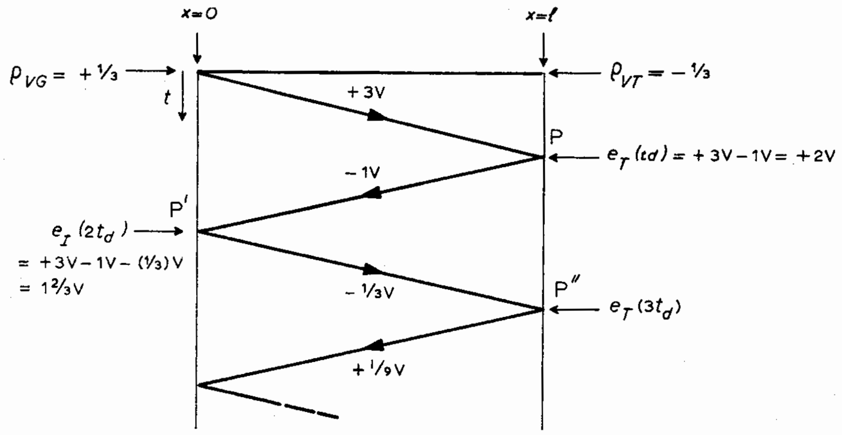



*Fig. 5. Reflection chart solution to problem of a mismatched line having linear source and terminating resistances.*

$$e_T = e_+ + e_-$$
$$i_T = i_+ + i_-$$

The plus sign is dictated by our sign convention. From these

$$\frac{e_T}{i_T} = \frac{(e_+ + e_-)}{(i_+ + i_-)}$$

$$\frac{e_T}{i_T} = \frac{(1+\rho_{VT})}{(1-\rho_{VT})} Z_o$$

where, by definition, $\rho_{VT} = (e_-/e_+)$ = voltage reflection coefficient at the termination. Rearranging this last equation gives

$$\rho_{VT} = \frac{(Z_T - Z_o)}{(Z_T + Z_o)} \qquad (1)$$

If, as is often the case, $Z_T \neq f(i_T)$, then $\rho_{VT}$ = constant. As it travels along the line $e_-$ causes the line to assume a voltage corresponding to that at the termination. When $e_-$ reaches the battery or source end there is complete absorption of energy in the $e_-$ wave if $Z_G = Z_o$. For $Z_G \neq Z_o$ a reflection occurs characterized by a voltage reflection coefficient, $\rho_{VG}$, where by analogy with equation 1

$$\rho_{VG} = \frac{(Z_G - Z_o)}{(Z_G + Z_o)} \qquad (2)$$

Again, if and only if $Z_G \neq f(i)$, then $\rho_{VG}$ = constant.

To illustrate the simplicity of a typical practical calculation assume $Z_G = R_G = 100\Omega$; $Z_o = R_o = 50\Omega$; $Z_T = R_T = 25\Omega$; $E = +9V$. From equations 1 and 2 $\rho_{VG} = (+1/3)$; $\rho_{VT} = (-1/3)$. Furthermore, $e_I(0+) = e_+ = \{(9 \times 150)/50\}V = 3V$.

The line behaviour can be represented by a reflection chart (Fig. 5) which is a convenient two-dimensional plot aiding and summarizing our calculations. By convention, distance along the line increases horizontally to the right on this chart and time vertically downwards. From the chart we can find the line conditions at an instant 'frozen' in time or the time function of voltage (or current) at any specified point. The motion down the line of the initial +3V step is indicated by the line OP: the magnitude of the slope of OP is the wave velocity, $v$. The progress of the −1V reflected wave is shown by the line PP′ sloping equally in the opposite direction.

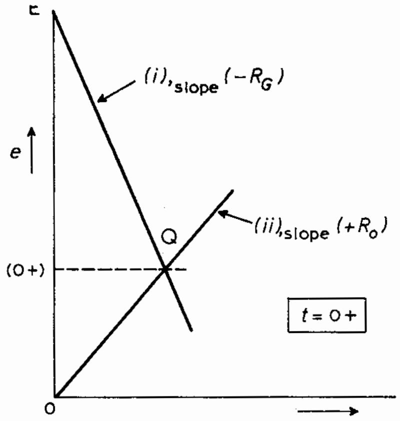



*Fig. 6. Graphical representation of Fig. 3. (Abscissa should be labelled i.)*

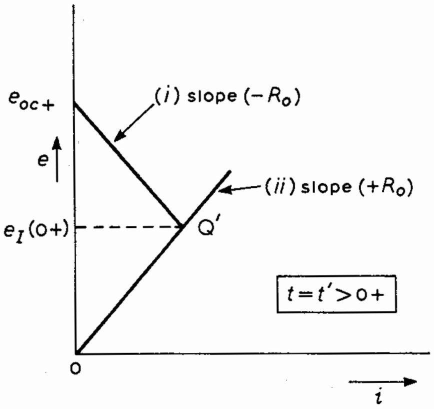



*Fig. 7. Construction giving line behaviour at point X reached by forward wavefront at $t'$ where $0+ < t' < t_d$.*

No fundamental importance attaches to the signs of the slopes: they have the polarities shown through the arbitrary (but convenient) choice of $t$ increasing vertically downwards. The manner of finding $e_I(t)$, $e_T(t)$ is clear from Fig. 5: we just sum the existing voltage, the incident wave voltage and the reflected wave voltage. If all $Z_G, Z_T$ were linear then what we have discussed is, arguably, all that one need to know in the majority of practical circuits in which line pulse reflections occurred.

The fundamental implication of non-linearity in source and termination impedances is that $\rho_{VG}, \rho_{VT}$ are $f(i)$ and are thus not constant: thus a simple reflection chart approach—with all its appealing elegance—is not possible. How then to deal with a practical non-linear termination such as a semiconductor diode? We will see that a graphical approach is appropriate. To develop this approach and in so doing to smooth the transition from familiar to less familiar concepts we will re-solve graphically the problem just considered. This will enable us to compare and check our analysis as we proceed.

## Graphical approach: the 'sliding Thévenin source' concept

For purely resistive impedances, i.e. $Z_G = R_G$, $Z_o = R_o$, $Z_T = R_T$, Fig. 6—a graphical interpretation of Fig. 3—is a 'snapshot' of the initial conditions at the input to the line. (We have chosen to plot $e = f(i)$ but the alternative choice $i = f(e)$ is equally valid.) In Fig. 6 curves (*i*) and (*ii*) are both straight lines: (*i*) represents the characteristic of the battery $E$ and source resistance $R_G$ while (*ii*) gives the instantaneous relationship between $e_I(t)$ and $i_I(t)$ at $t = 0$; (*ii*) is, of course, a straight line of slope $+R_o$ passing through the origin.

The intersection point $Q$, gives $e_I(0+) = ER_o/(R_o+R_G)$. Suppose we now wish to represent line behaviour at a point X, situated at a distance $x$ along the line, at a time $t'(=x/v)$ by an instantaneous time plot similar to Fig. 6. Such a plot is shown, correspondingly labelled, in Fig. 7. The dynamic resistance seen looking in either direction from the selected point is the characteristic resistance $R_o$. Thus curve (*ii*) represents, as before, the line to the right of X while (*i*) with a slope $-R_o$ (rather than $-R_G$) now represents the resistance looking backwards towards the battery. The point of intersection, $Q'$, of (*i*) and (*ii*) is the line voltage: but this is equal to $e_I(0+)$ because in travelling down the line the wave $e_+$ causes successive points to assume this value. It is thus evident that the apparent battery voltage seen from X is $e_{oc+} = 2e_I(0+)$.

Clearly, as far as points to the right of X are concerned we can imagine the line to be cut at X and the section to the left to be replaced by a Thévenin source circuit which 'slides' along the line. This process is illustrated in Fig. 8: the equivalent source components in the chain-dotted rectangle move to the right at the velocity $v$ of the $e_+$ wavefront. If we 'stack-up', one behind the other, a series of plots such as Figs. 6 and 7, we obtain the three-dimensional $e,i,t$ plot in Fig. 9, and this provides a good visualization of wavefront progress. Thus the projection on a plane drawn perpendicular to the $i$ axis and passing through $Q'$ gives $e = f(t)$, while the projection on a plane drawn perpendicular to the $e$ axis yields $i = f(t)$.

The conditions obtaining at the termination at $t = t_d$ are found by considering the $e,i$ curve there—see Fig. 10(a): $e_T(t_d)$ can be calculated from this or the associated circuit diagram in Fig. 10(b). Obviously,

$$e_T(t_d) = e_{oc+}R_T/(R_T+R_o)$$

$$\text{or, } e_T(t_d) = 2e_I(0+)R_T/(R_T+R_o)$$

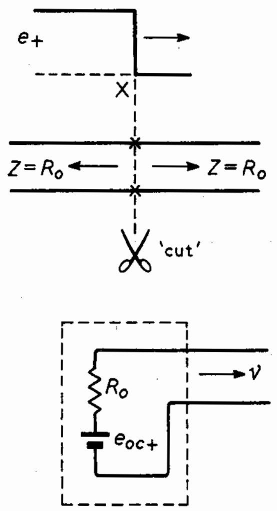



*Fig. 8. Illustrating the 'sliding Thévenin source' concept for forward wave.*

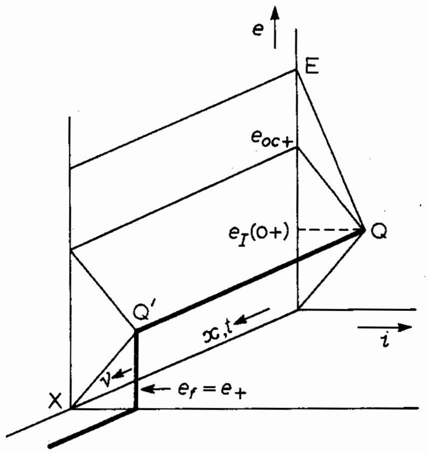



*Fig. 9. An e,i,t plot of line behaviour for 0+ < t < t_d.*

Using the numerical data for our example,

$$e_I(0+) = \{(9 \times 50/150)\}\text{V} = 3\text{V}$$

$$e_T(t_d) = \{(2 \times 3 \times 25)/75\}\text{V} = 2\text{V}$$

The value of $e_T(t_d)$ agrees, of course, with that indicated in Fig. 5. Now $R_T \neq R_o$ so there is a reflected wave $e_-$ which proceeds from the termination to the battery and causes successive points on the line to assume the voltage $e_T(t_d)$ as it reaches them. The line voltage $e$ at any point due to $e_-$ is given by

$$e = iR_o + e_{oc-}$$

This equation is represented in Fig. 11 by a line of slope $+R_o$ passing through a point, $Q''$, having co-ordinates $e_T(t_d), e_T(t_d)/R_T$. The plus sign for $R_o$ arises through our choice of sign for $i$, discussed earlier: we wish to draw the characteristic of the reflected wave on the same diagram as we have used for the forward wave in which the positive direction of current flow is taken to the right in the top conductor. To the left of the $e_-$ wavefront the line appears as a battery of magnitude $e_{oc+}$ in series with a resistance $R_o$. Fig. 11 should be compared with Fig. 7: Fig. 12 shows a sliding source equivalent representation of $e_-$ and should be compared with Fig. 8. We can, if we wish, draw for $e_-$ a diagram similar to Fig. 9,

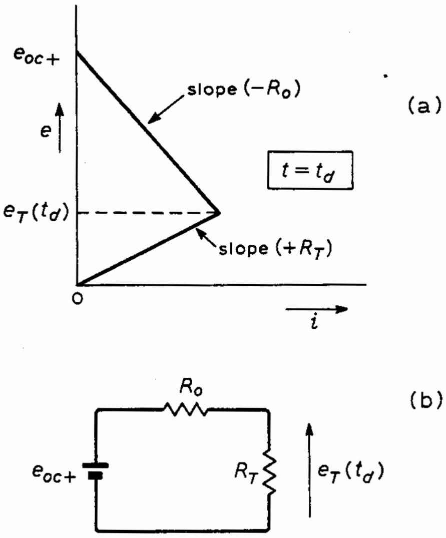



*Fig. 10. Line conditions at termination at $t = t_d$ (a) graphical construction and (b) equivalent circuit.*

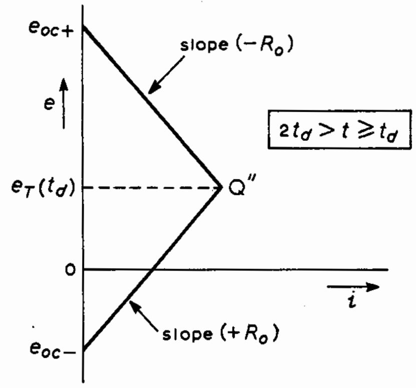



*Fig. 11. Counterpart of Fig. 7 for reverse wave for $t_d < t < 2t_d$.*

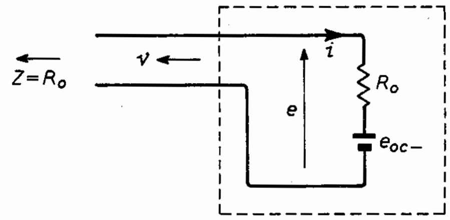



*Fig. 12. 'Sliding Thévenin source' representation of reverse wave (c.f. Fig. 8).*

but this has not been done as no new principle is involved.

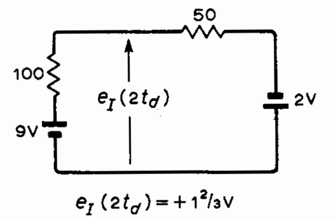
At the battery, Figs 13(a), 13(b) apply when the reflected wave arrives. Thus $e_I(2t_d) = \{9-(11 \times 100/150)\} = 1.67\text{V}$, as predicted by the chart of Fig. 5. The graphical construction can now be repeated for further forward and reverse waves. The intersections of successive lines of slope $+R_o$ with the source characteristic (in this case a battery) yield, respectively, $e_I(0+ \leqslant t < 2t_d)$, $e_I(2t_d \leqslant t < 4t_d)$, etc., while the intersections of the lines of slope $-R_o$ with the termination characteristic give, respectively, $e_T(t_d \leqslant t < 3t_d)$, $e_T(3t_d \leqslant t < 5t_d)$, etc. Theoretically, reflections occur until $t = \infty$ at which time $e_I(\infty) = e_T(\infty) = 1.8\text{V}$; this represents the intersection of the d.c.

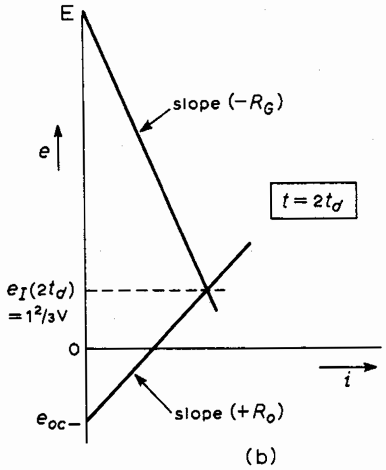
output characteristic of the resistance at the end of the, supposedly lossless, cable.

The lines with slope $\pm R_o$, used to describe the forward and reverse waves, are sometimes called Bergeron lines in honour of the Frenchman who appears to have been the first to use this graphical technique for describing the motion of waves. To describe the terminal behaviour of the line up till, say, $t = 2t_d$ we can superpose the individual time pictures to get on a single $e,i$ graph the composite time picture shown in Fig. 14(a). The terminal voltage waveform plots are readily obtained as in Fig. 14(b).

Though it offers some physical insight, the graphical technique would not normally be used for linear source and termination resistances since the reflection chart method is so quick and easy to apply. However, graphical analysis comes into its own with non-linear source and/or load resistances, as will be seen in the sections which follow.

*Fig. 13. (a) Diagram for finding $e_I(2t_d)$. (b) Graphical equivalent of (a).*



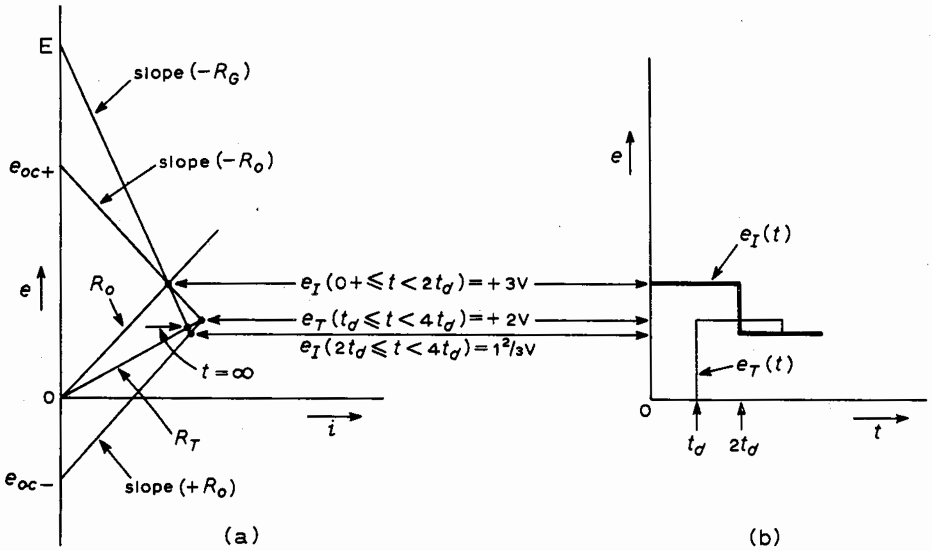




*Fig. 14(a) Composite time picture for $e_I(t), e_T(t)$ for $0+ \leqslant t \leqslant 2t_d$ and $t = \infty$. (b) Waveforms deduced from (a).*

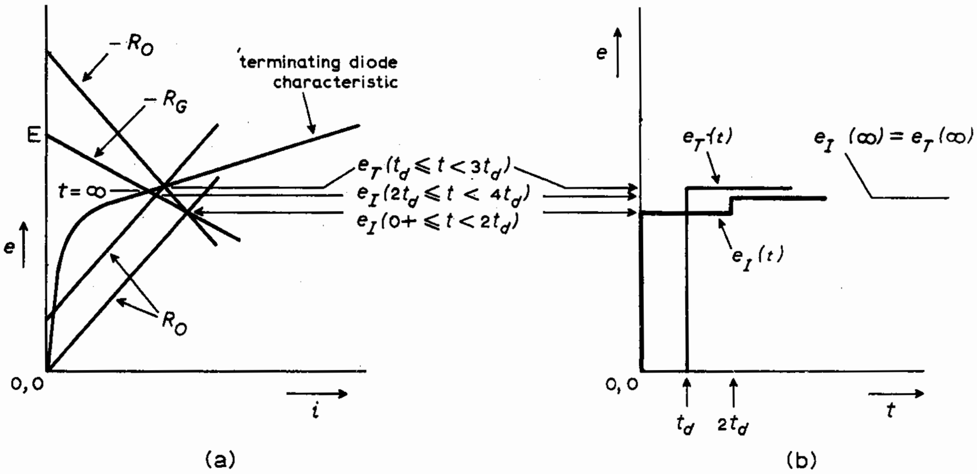



*Fig. 15(a) Composite time picture when terminating component is a semiconductor diode. (b) Waveforms for (a).*

## Graphical approach: non-linear termination

Suppose we replace the impedance $Z_T$ in Fig. 1 by a semiconductor diode with its anode connected to the top conductor and its cathode to the bottom. None of the theory behind the graphical construction so far discussed is altered—we do not change the mechanism of wave propagation on the line—only the boundary conditions are different. Thus, Fig. 15(a) shows a composite picture of terminal behaviour for $0+ \leqslant t < 2t_d$, for an arbitrary diode. Reflections occur and continue till the point $e_T(\infty)$ is reached. The photographs in Fig. 16 show the reflection behaviour in a practical case. Fig. 16(a) shows the open-circuit output voltage of a tunnel-diode pulse generator (H.P. type 213B). Figs 16(b), 16(c) respectively show $e_I(t)$, $e_T(t)$, when the generator is used to drive a length of precision 10-ns, 50-Ω delay cable terminated by a Schottky barrier diode (H.P. type 5082-2301).

Depending on the conditions of the problem it is possible to choose a diode which correctly terminates the line. This is easily seen graphically. The use of a diode to suppress overshoots has been known for many years; as far as the author is aware, no

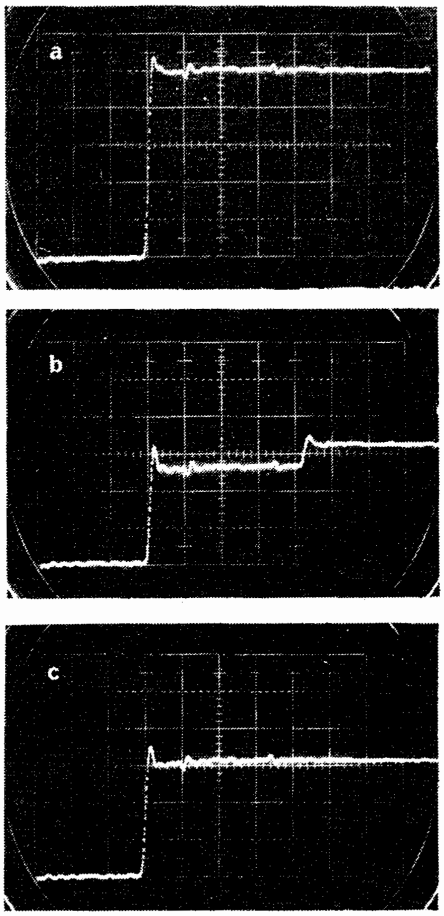



*Fig. 16. Experimentally observed waveforms using matched source impedance and Schottky diode as terminating component. Horizontal scale: 5ns/div., vertical scale: 100mV/div. (a) Open-circuit output voltage of pulse generator, (b) $e_I(t)$, (c) $e_T(t)$.*

exact mathematical solution of the problem has yet been worked out.

### Reflections with t.t.l. interconnections

For the preservation of pulse edges and the reduction in crosstalk effects it is common practice to interconnect high-speed digital integrated-circuit logic elements with some form of transmission line; this usually means strip-line ($R_o \approx 100\Omega$) on a printed circuit board and twisted-pair between board assemblies. Now, by their very nature, switching devices and circuits have pronounced non-linear impedances associated with them.

Thus the problem reduces to investigating reflection effects which occur and seeing if we can live with them. This will be accomplished if any reflections cause no voltage overstress in the driving and driven gates and no logic malfunction, through spurious triggering, of other gates to which they are connected.

The d.c. characteristics of a t.t.l. gate are shown in Fig. 17. (The interested reader is referred to ref. 2 for a good discussion of their origin). The positive logic convention is assumed, i.e. the more positive of the two discrete voltage levels in the system is taken as '1'. A single continuous curve—(*i*) in Fig. 17(a)—gives $e_I = f(i_I)$. For each value of $e_I$ there is a single value of $e_o$, but we are only interested in the *e,i* characteristics corresponding to $e_o \equiv$ '0' and $e_o \equiv$ '1'. These are shown respectively as curves (*ii*) and (*iii*) in Fig. 17(b), 17(c). Bearing in mind our *e,i* sign conventions we now superpose (*i*), (*ii*), (*iii*) to obtain the t.t.l. composite d.c. characteristics (Fig. 19) relevant to the two cascaded gates, A and B in Fig. 18, connected via a length of transmission line.

We consider only the 1 → 0 transition at the line input; the case of the 0 → 1 transition follows by analogous treatment. Thus, the output of A is initially 1 and has been so long enough for any past line transients, which may have occurred, to have died away: then $e_I = e_T = e_a$ and this corresponds to '*a*', the point of intersection of (*i*) and (*iii*). The dynamic representation of the line at $t = 0$, when $e_I$ switches to the 0 condition, is simply a resistance $R_o$ in series with a battery $e_a$; this is given by a straight line of slope $+R_o$ passing through '*a*'. Clearly, $e_I(0+) = e_I(0+ \leqslant t < 2t_d) = e_b$, where '*b*' is the intersection of this load line, extended back, with (*ii*). The sliding-source circuit for the wavefront which starts down the line is a battery $e_{oc+}$ in series with a resistance $R_o$. This is depicted graphically by a line with slope $-R_o$ passing through '*b*'. Where this intersects (*i*) gives $e_T(t_d) = e_T(t_d \leqslant t < 3t_d) = e_c$. We can continue to draw our Bergeron lines and thus determine $e_I(t)$, $e_T(t)$ for whatever time scale we choose. Incidentally, we have enough information to plot $e_x(t)$, $x$ being our general point on the line should this be required. Plots of $e_I(t), e_T(t)$ for $0+ \leqslant t \leqslant 2t_d$ are simply derived and are given in Fig. 20.

The first negative excursion in $e_I(t)$ can, if large enough, lead to a subsequent overshoot in $e_T$ sufficient to switch B. Various methods of reducing the likelihood of this happening have been described. They include control of the input characteristic[3] designated (*i*) in circuit design and the use of some form[4] of built-in clamping diode.

### Validity of graphical approach

The main assumptions of the graphical technique are; a well specified, effectively lossless, transmission path; transmission line pulses with step edges; terminating components which are completely specified by their d.c. characteristics. The inevitable departure from these ideal requirements, in practice, does not invalidate the approach. The twisted-pair is fairly reproducible in its characteristics and the lengths usually employed are unlikely to give any significant loss. Provided the pulse transition times are small compared with the electrical length of the line the effect of finite rise times and the presence of stray capacitance at terminations is to cause rounding of waveforms without having a serious effect on the amplitude of reflected pulses.

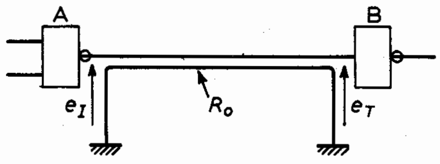



*Fig. 18. Interconnected t.t.l. gates.*

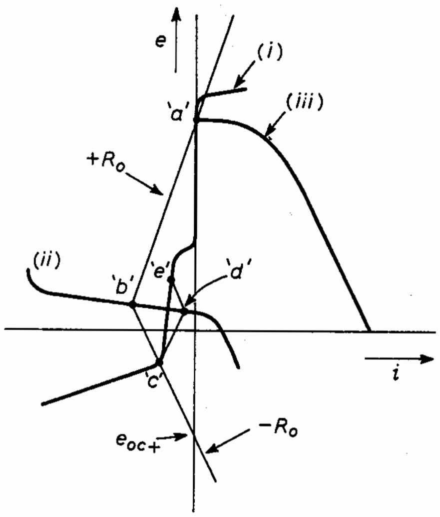



*Fig. 19. Composite d.c. characteristics of t.t.l. gate with 'Bergeron' lines.*

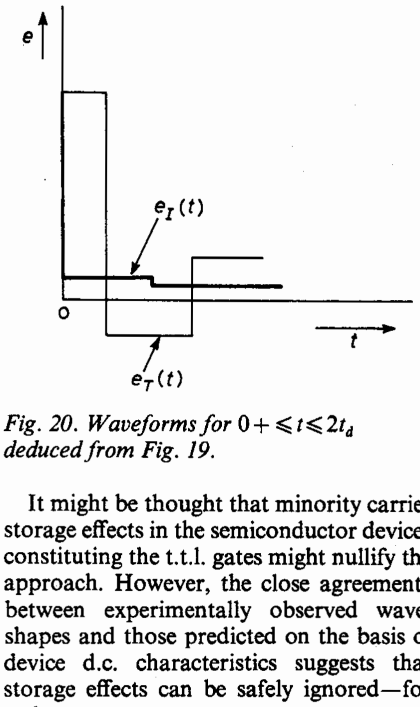



*Fig. 20. Waveforms for $0+ \leqslant t \leqslant 2t_d$ deduced from Fig. 19.*

It might be thought that minority carrier storage effects in the semiconductor devices constituting the t.t.l. gates might nullify the approach. However, the close agreement[5] between experimentally observed waveshapes and those predicted on the basis of device d.c. characteristics suggests that storage effects can be safely ignored—for t.t.l. at any rate.

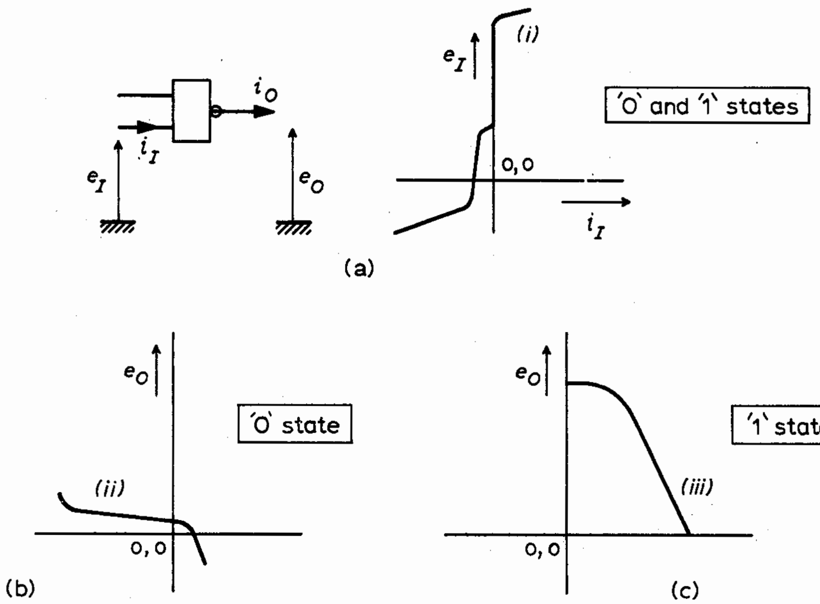



*Fig. 17(a) T.T.L. gate showing sign conventions and input characteristic. (b) '0' level output characteristic. (c) '1' level output characteristic.*

### References

1. S. Goldman, "Transformation calculus and electrical transients", Prentice Hall 1950, Chapter 10.
2. S. Garrett, "Integrated-circuit digital logic families. II—TTL devices", *I.E.E.E. Spectrum*, 1970, pp. 63–71.
3. G. O. Crowther, "Reflection phenomena when TTL gates are connected to long lines", *Electronic Equipment News*, Jan. 1970.
4. I.T.T. Semiconductors, "MIC 9000 Series TTL with input clamping diodes". A. Nguyen-Hall, "Solving the Ringing Problems of TTL Integrated Circuits", *Design Electronics*, Feb. 1969, pp. 42–5.
5. M. Abdel Latif and M. J. O. Stutt, "Simple graphical method to determine line reflections between high-speed logic integrated circuits", *Electronic Letters*, 1968, vol. 4, pp. 496–7.

# Diodes damp line reflections without overloading logic

As line terminations in digital-logic systems, diodes keep pulses almost as sharp as resistors would, yet they dissipate hardly any power

by E.E. Davidson and R.D. Lane, *IBM System Products Division, Hopewell Junction, N.Y.*

☐ In a digital-logic system, diode-terminated interconnections are a much better choice than resistively terminated lines. The diodes suppress multiple line voltage reflections almost as effectively as a resistor. Yet unlike a resistor they dissipate very little power and run no risk of overloading the logic circuitry.

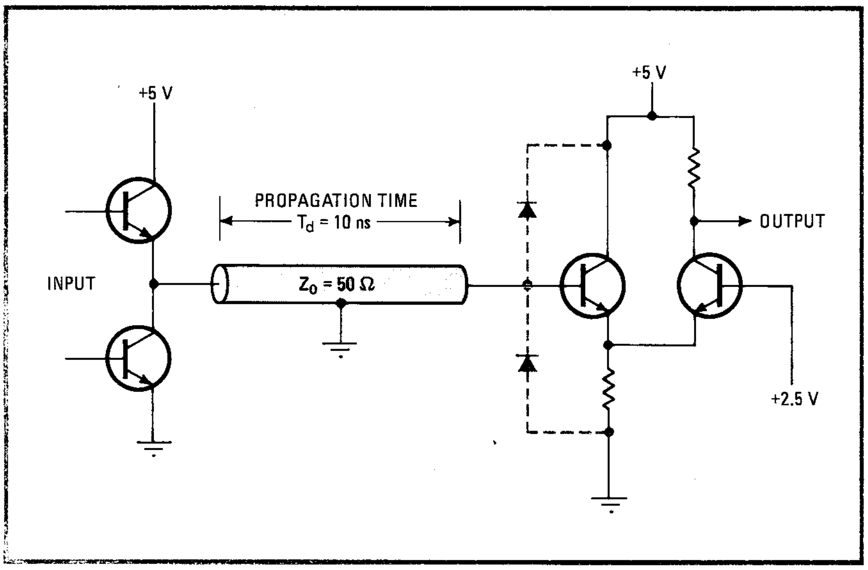
Multiple line reflections are a problem in a digital-logic system with high signal frequencies and long communication paths, because the interconnections behave much like transmission lines. If these lines are left unterminated, a signal pulse is reflected so often that stable switching can occur only after a long settling time. Moreover, the voltage doubling at the far end of such a line may exceed the breakdown values of the devices located there. Terminating the line with its characteristic resistance is the obvious and standard solution, yet it can be shown to be less satisfactory than using a pair of diodes. Therefore industry is turning to diode terminations (see "The diode story").

A typical logic interconnection circuit is shown in Fig. 1. The circuit arrangement consists of a push-pull driver, a transmission line, and a receiver/current switch. The input impedance of the switch is much greater than the characteristic impedance $Z_0$ of the line, so in the absence of the terminating diodes the line would see an open circuit. If the output impedance of the driver is assumed to be 7 ohms, the equivalent circuit for the logic stage, in which the signal voltage and source resistance ($R_S$) replace the driver, would be the one that is shown in Fig. 2(a).

A signal voltage ($E_S$) of a given level entering this line would eventually charge it to the same voltage level, but only after the line had oscillated considerably above and below that level. With an $E_S$ of 5 volts, for example, the voltage at the near end of the line ($E_n$) at time zero works out to be:

$$E_n = E_S Z_0/(R_S + Z_0) = 5 \times 50/(7+50) = 4.4\ V \qquad (1)$$

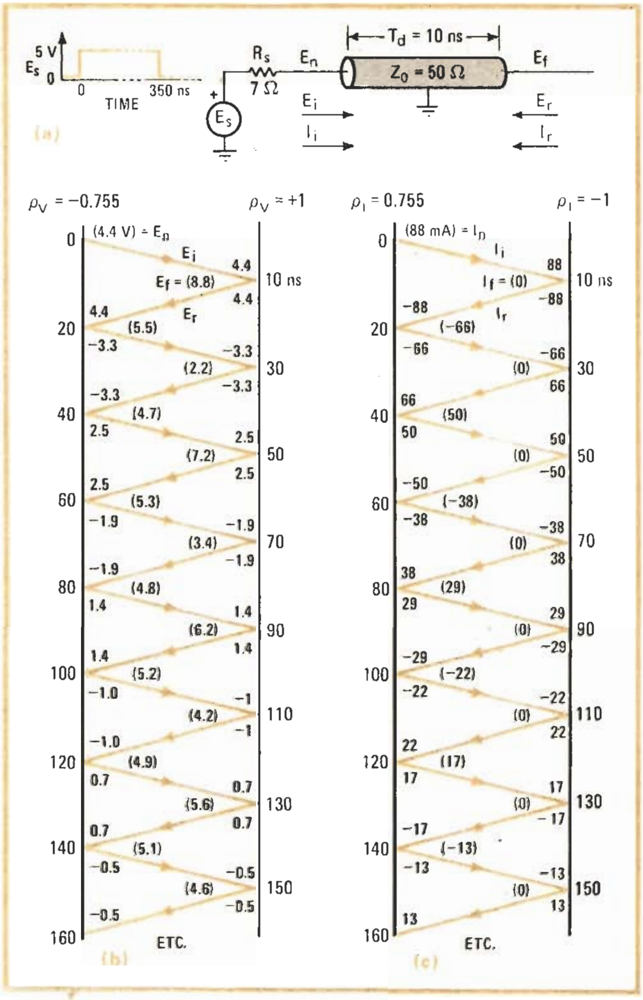
This incident voltage waveform $E_i$ travels down the line until it reaches the open-circuited far end. There, the boundary condition requires zero current flow, so a reflected voltage waveform ($E_r$) of equal value cancels the incident current and travels back up the line to the near



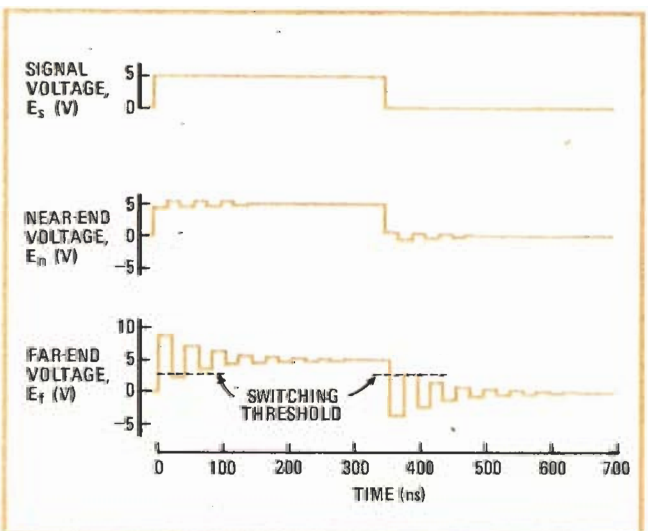

**1. Interconnection.** This logic circuit uses a transmission-line interconnection down which push-pull driver sends voltage pulses to a receiver/current-switch. Reflections on the line upset switching performance and may damage transistors in both driver and receiver, but terminating diodes (dotted color lines) suppress the reflections.

## The diode story

Resistive terminations are standard elements in the logic networks that the computer industry uses at present. But the advantages of diode terminators have not escaped notice by designers elsewhere. For example, Motorola's MECL System Design Handbook suggests use of diode terminations on circuit interconnections for its 10,000-series ECL. Fairchild's 10-k series includes the F10014 active terminator, a 16-pin dual in-line package that contains 14 "bilateral clamps" with the external characteristics of diode pairs. Moreover, an integrated logic circuit that's now being considered by IBM includes integral terminating diodes.—**D.J.B.**




**2. Unterminated.** For a logic circuit with an unterminated interconnection line (a), near-end voltage $E_N$ and far-end voltage $E_F$ change every 20 nanoseconds as incident and reflected voltages travel along line, as shown in voltage-reflection diagram (b). Current reflections are diagrammed in (c).




**3. End results.** A 5-V pulse with 350-ns duration and 50% duty cycle produces these voltage waveforms at near and far ends of unterminated transmission line. Note large voltage excursions and long settling time at far end. Overshoots may cause false switching.

end. In general, the reflected waveform obeys a relationship of the form:

$$E_r = \rho_V E_i \qquad (2)$$

where the voltage-reflection coefficient $\rho_V$ is given by:

$$\rho_V = (R_T - Z_0)/(R_T + Z_0) \qquad (3)$$

and $R_T$ is the termination resistance. For the open line $R_T$ is infinite, so $\rho_V$ is equal to unity.

The voltage-reflection diagram in Fig. 2(b) shows that the initial incident voltage of 4.4 V reaches the far end of the line after 10 nanoseconds and is fully reflected. The total voltage at the far end, $E_f$, is zero for the first 10 ns and then becomes 8.8 V with the appearance of the incident and reflected voltages.

The 4.4-V reflected wave reaches (is incident upon) the near end after 20 ns. It is re-reflected with a $\rho_V$ of (7 − 50)/(7+50) or −0.755, so a re-reflected voltage of −3.3 V is sent back down the line. The total near-end voltage is thus equal to the initial 4.4 V plus the sum of the +4.4-V incident and the −3.3-V re-reflected voltages. (Note that the reflected far-end waveform becomes the incident near-end waveform). The new near-end voltage is therefore 5.5 V.

This reflection process continues, as shown in Fig. 2(b), until the reflections die out. The line is then fully charged to the 5-V signal voltage.

Figure 2(c) diagrams the current reflections that accompany the voltage reflections of Fig. 2(b). The current-reflection coefficient $\rho_I$ is the negative of the voltage-reflection coefficient:

$$I_r/I_i = \rho_I = -\rho_V \qquad (4)$$

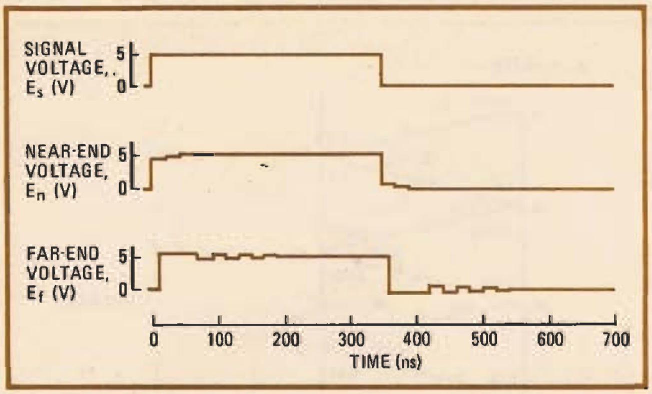
The incident and reflected currents ($I_i$ and $I_r$) are found by dividing the incident and reflected voltage ($E_i$ and $E_r$) by the characteristic impedance of the line:

$$I_i = E_i/Z_0;\ I_r = E_r/Z_0 \qquad (5)$$

When the signal transition is negative instead of positive, the positive information of Fig. 2(b) and (c) can be used to calculate the negative transition reflection information. Since the reflection coefficients remain the same, the incident and reflected values need only have their signs changed. Also, the starting values are no longer zero, so the near-end and far-end quantities for Fig. 2(b) and (c) have to be shifted by the magnitude of the line voltage and current at the time that the negative signal transition begins.

This procedure was used for generating the complete near-end and far-end wave forms that are shown in Fig. 3 for an unterminated line. In this figure, the positive and negative signal transition values of near-end and far-end voltages from Fig. 2(b) are plotted on a time axis. The resultant wave forms show a near-end voltage that is fairly well behaved (because the source impedance is low enough to make the voltage source seem not too far from the ideal) but a far-end voltage that exhibits quite large excursions.

These fluctuations can damage both the components and the operation of a logic system. For instance, because the signal is doubled the first time it reaches the far end, the voltage on the base of the receiver input




**5. Happy endings.** Voltages at near and far ends of diode-terminated line vary only slightly though for same 5-V signal as in Fig. 3. Diodes alter reflection mechanism to produce smoother pulses and stable switching, with no more power dissipation than for unterminated line. Resistive termination would eliminate reflections totally, but dissipate 33 times as much power.

stage with terminating diodes is shown in Fig. 4(a). The diodes are the Schottky variety with a forward voltage drop of 0.6 V. The upper diode handles positive-going transitions, and the lower diode handles negative-going transitions.

If a 5-V input is applied to the circuit, the voltage-divider effect at the near end again impresses 4.4 V onto the line, and this voltage travels to the far end. There it tries to double, because the line appears instantaneously unterminated, but fails, because the upper diode turns on and clamps the line as soon as the far end voltage exceeds 5.6 V. This 5.6 V now generates a 1.2-V reflected voltage, to satisfy the voltage boundary condition, which requires the sum of the incident and reflected voltages to equal the change in the far-end voltage (5.6 V). The corresponding reflection coefficient may be evaluated from Eq. 2. It may also be given by:

$$\rho_V = (V_C/E_i) - 1 \qquad (6)$$

where $V_C$ represents the change in the far-end voltage as determined by the clamping action of the diode when the diode is forwarded-biased. Here:

$$\rho_V = (5.6/4.4) - 1 = 0.273$$

As shown in the reflection diagram of Fig. 4(b), the 1.2-V far-end reflection returns to the near end after 20 ns. The near-end re-reflection is governed by Eq. 3, so the new voltage reaching the far end at 30 ns becomes -0.755 × 1.2 V, or -0.905 V. Since the upper terminator diode is on, the far end acts as though the line were short-circuited, and a full negative reflection occurs. The line continues to behave as if shorted until the diode turns off.

The corresponding current diagram in Fig. 4(c) illustrates the launching of a current wave of 88 milliamperes (4.4 $V$ ÷ 50 ohms) onto the line. When this current reaches the far end, it is modified by a reflection coefficient that is the negative of the corresponding $\rho_V$ at 10 ns. The reflections continue as if the line were shorted until there is no net positive current at the far end of the line. When this occurs, the diode stops conducting, and the line reverts to open-circuited behavior. In this case, the line becomes an open circuit after 70 ns.

The net line current at the far end must be zero when the diode is off. As a result, the current reflection when the diode turns off must dissipate any residual line current ($I_x$). This boundary condition leads to an equation for the current-reflection coefficient at this point that is analogous to Eq. 6 and is given by:

$$\rho_{if} = -(I_x/I_{if}) - 1 \qquad (7)$$

where $I_{if}$ is the last current value incident upon the far end before the diode turns off. Here $I_x$ is 0.4 A and $I_{if}$ is −10.7 A, so $\rho_{if}$ is −0.963. The voltage-reflection coefficient at the instant that the diode stops conducting is the negative of the value given by Eq. 7.

(For example in Fig. 4(c), the value of $\rho_{if}$ calculated from Eq. 7 at 70 ns happens to be very close to the open-circuit value to which the line subsequently reverts. But this kind of coincidence seldom occurs.)

After the diode ceases to conduct, both the voltage and current reflections at the far end are governed by the reflection coefficients for an open line. As illustrated in Fig. 4(b) and (c) these reflections continue until the current dies out and the line attains a level of 5 V throughout its length.

Non-ideal behavior of the diodes causes changes in the clamping voltage as the line current is made to vary by the current reflections. As a result, there are actually small deviations from the theoretical short-circuit behavior exhibited by the line in the example during the time the diode was on. But if the diode forward-voltage changes are small compared to the voltages on the line, the results are close to those predicted by Fig. 4(b) and (c). If the voltage changes are significant, Eq. 6 can be used to calculate each successive value for the reflection coefficient when the diode is on.

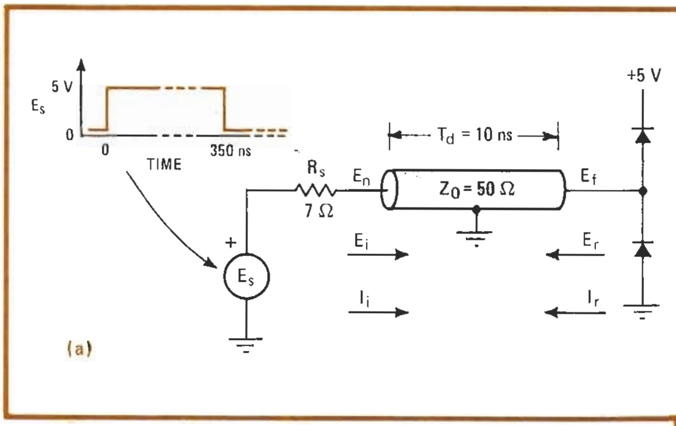
The marked improvement that is made by diode termination becomes very obvious when the information contained in Fig. 4(b) is transformed into a near-end and far-end voltage waveform diagram for positive and negative signal transitions (Fig. 5). The negative reflection information is obtained from Fig. 4(b) by following the procedure discussed previously; for negative-going transitions, the bottom diode in Fig. 4(a) takes over the role that the top diode played for positive-going transitions.

The voltage waveforms in Fig. 5 show that the terminating diodes have eliminated the large voltage excursions at the far end, so that the far-end voltage reaches the full signal level after one line-delay time. The near end steps up gradually to +5 V with no significant overshoot, and does so even when the source resistance is higher. Finally, the effects of the reflections die out much faster than in the unterminated case.

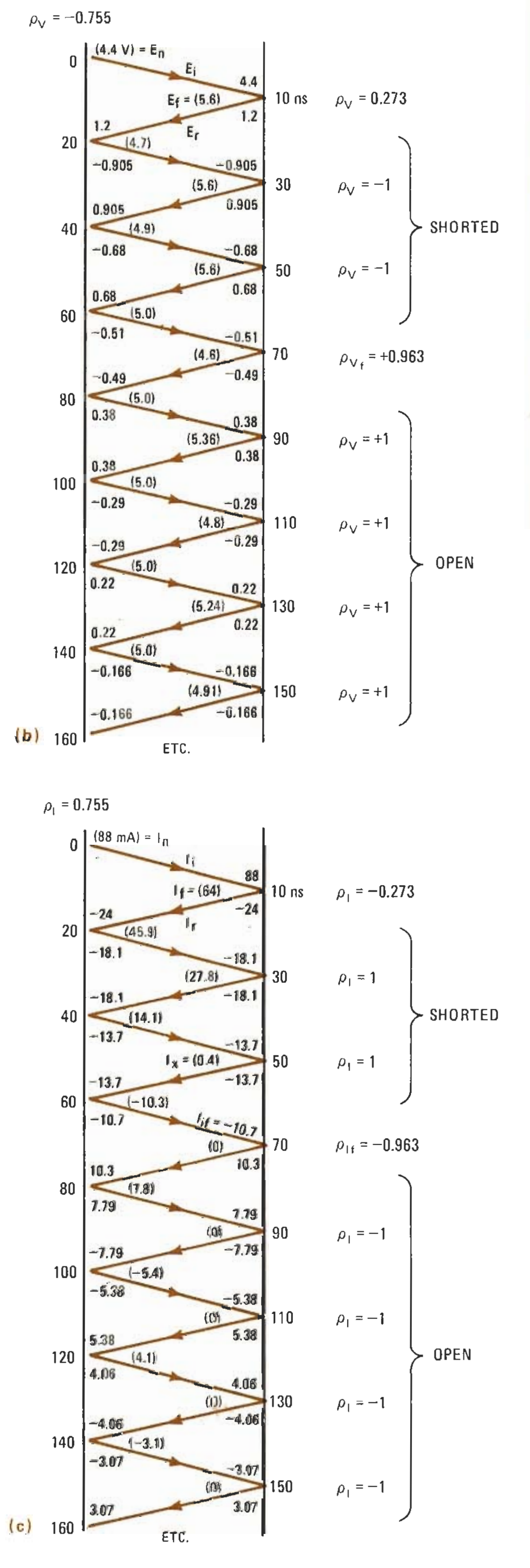
### The matter of power

Energy must be delivered to the line to charge it up to the signal voltage level. A diode-terminated line needs much less than a resistor-terminated line and in fact is as economical in this respect as an unterminated line.

In the case of an open-circuited line, energy is stored on the capacitance of the line. In order to calculate the power needed to fully charge a line, the transmission




**4. Diode-terminated.** A logic circuit in which interconnection line has diode-pair termination (a) has the voltage reflections diagrammed in (b) and the current reflections diagrammed in (c). When conducting, a diode short-circuits far end of line. When nonconducting, it leaves line open-circuited.




transistor in Fig. 1 is momentarily large enough to forward-bias the base-to-collector junction. The resulting current flow could destroy the transistor. One round trip later, the far-end voltage has dropped to 2.2 V. If the receiver threshold is at the halfway point of the input signal voltage (2.5 V), then the far-end voltage is below the threshold, and false switching may occur. Two round trips later, the far end is at 3.4 V. During this interval, the receiver does not tolerate noise nearly as well as when the line is at its quiescent level of 5 V. Finally, when the first negative-going transition reaches the far end, the input to the receiver falls well below ground. Such a condition might easily stress the base-emitter junction of the receiver input transistor in the reverse direction and once again damage the transistor.

Damage could also occur to the transistors in the output stage of the driver of Fig. 1. Related studies show that when the source impedance is higher than the 7 ohms discussed here, reflections at the near end may produce voltages that exceed the collector breakdown limits of the driver transistors. Furthermore, a negative-going reflection at the near end may forward-bias the isolation junction associated with the collector of the pull-down transistor, producing parasitics or even destroying the junction if substrate current limitations are inadequate.

## How a diode-terminated line differs

Terminating resistors are the standard way to prevent transmission line reflections, but the amount of power they dissipate severely limits logic packing density. An alternative termination technique is to connect a pair of diodes as shown by the dotted lines in Fig. 1.

When diodes terminate the transmission line, they squelch the voltage overshoots and reflections at the far end almost completely and, more surprisingly, they also greatly diminish the voltage reflections at the near end. The net effect is that the voltage wave forms hardly oscillate at all and, in fact, approach those of a resistively terminated line while dissipating as little power as an unterminated line.

The equivalent circuit for the logic interconnection line is replaced by its equivalent capacitance ($C_L$):

$$C_L = T_d/Z_0 \quad (8)$$

where $T_d$ is the propagation delay of the line. (The replacement is justified because both the line and the lumped capacitance store the same final energy.) The equivalent circuit of Fig. 6(a) can then be used to determine the power delivered to the line while a finite signal voltage is being applied. This value for the power delivered to an open-circuited line ($P_{OC}$) is given by:

$$P_{OC} = \frac{1}{T}\int_0^T E_S[(E_S/R_S)exp(-t/R_SC_L)]dt \quad (9)$$

where the term in brackets is the current that flows during the charging interval and T is the time period. The integration in Eq. 9 yields:

$$P_{OC} = (E_S^2C_L/T)[1 - exp(-T/R_SC_L)] \quad (10)$$

After the charge-up process is complete, i.e., when T is much larger than $R_SC_L$, the exponential term becomes zero and Eq. 10 reduces to:

$$P_{OC} = E_S^2C_L/T \quad (11)$$

Equation 11 represents the total power delivered by a power supply to a charged-up open-circuited transmission line during a period of time that exceeds the charging time. Half of this power is dissipated when the line is charged, the other half when the line is discharged. Note that the open-circuit power is not a function of the driving resistance, $R_S$.

For example, assume that the open-circuited line of Fig. 2(a) is fully charged and fully discharged during one cycle time of 700 ns. From Eq. 8, the line's capacitance is 200 picofarads. Substituting the appropriate numbers into Eq. 11 yields a value of 7 milliwatts for the power dissipated by the line.

When the transmission line is terminated with a resistance $R_T$, a steady-state current ($I_{SS}$) flows. Now energy is no longer stored only in the line's capacitance, but also stored in its inductance ($L_L$), given by:

$$L_L = T_dZ_0 \quad (12)$$

By following the procedure used for deriving Eq. 11 and referring to the equivalent circuit of the terminated line, as shown in Fig. 6(b), a general expression for the total power delivered to a terminated transmission line ($P_T$) can be written as:

$$P_T = I_{SS}^2R_SF + E_{SS}^2F/R_T + C_LE_{SS}^2/T + L_LI_{SS}^2/T \quad (13)$$

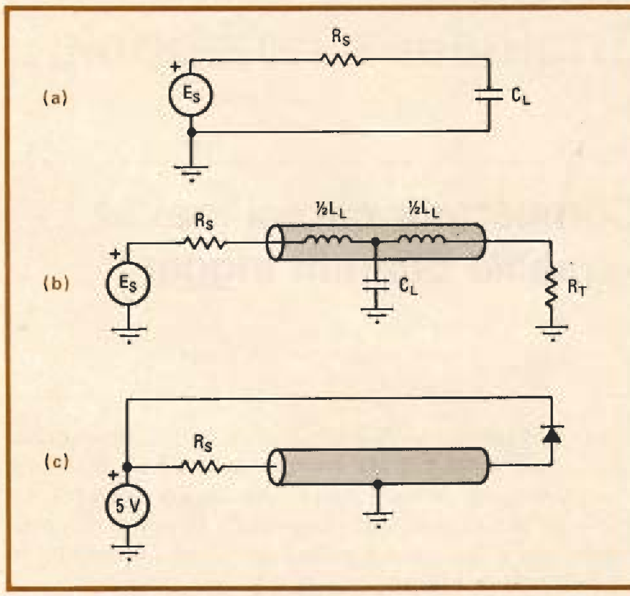
where F is the fraction of time T that the signal is applied, $I_{SS}$ is the steady-state current, and $E_{SS}$ is the steady state voltage on the line. The value of $E_{SS}$ is determined by the voltage divider:

$$E_{SS} = E_SR_T/(R_T + R_S) \quad (14)$$

From Fig. 6(b) the expression for the steady-state current may be seen to be:

$$I_{SS} = E_{SS}/R_T \quad (15)$$

The first term in Eq. 13 accounts for the dc power dissipated in the source resistance, the second term for the dc power dissipated in the load or termination resistance. The third term represents the power used to charge the capacitance of the line, the fourth the power that charges the inductance of the line. As in the case of the open-circuit power of Eq. 11, the total power of Eq. 13 represents the power dissipated in charging and discharging the line during a period T.




**6. Equivalent charging circuits.** For calculation of power required to charge up unterminated transmission line, line is replaced by capacitor in equivalent circuit (a). For similar calculation for terminated transmission line (b), line is represented by both inductance and capacitance; $R_T$ is terminating resistor. Diode-terminator equivalent circuit (c) shows diode dissipates no additional power.

In one of the above examples, the unterminated power for a 700-ns interval for the circuit of Fig. 2(a) was given as 7 milliwatts. Equation 13 shows that the terminated power for the same interval is 232 mW when the line is terminated in its characteristic impedance of 50 ohms and operates at a 50% duty factor. The terminated line dissipates 33 times as much power as the unterminated line; this disparity accounts for the desirability of using diode terminators.

Although the diode terminator consists of active devices that develop terminal voltages when current passes through them, they do not add to the total power associated with driving the line. To understand why this is so, consider Fig. 6(c), the effective circuit for the positive signal transition. Any current that flows through the diode emanates from 5 V and returns to 5 V. In other words, the power dissipated by the diode does not require additional power to be delivered from the power supply. Some of the power that would be dissipated in the source resistance for the unterminated case is instead dissipated in the diode for the diode-terminated case.

Since the total energy delivered to the diode-terminated line is required to charge the capacitance of the line, the power for the diode-terminated line is the same as that for the equivalent unterminated line as given by Eq. 11. In other words, a diode terminator dissipates no more power than an open line. □

turns of the screw cap for disconnect or connect, and that it shows some reflections when installed in a 50-ohm line. UHF connectors have similar voltage ratings to those of the BNC type and are not recommended for use above 300 MHz. The UHF connector will fit larger coaxial cables, up to the size of RG-8, that BNCs will not. This connector has been quite popular in the high-power and VHF communications equipment field. It is wise to have a few BNC to VHF adaptors available, especially if communication equipment is frequently worked on.

Occasionally when dealing with UHF and higher frequencies, it is necessary to connect to the larger cables such as RG-8 and a better connector than either the BNC or UHF is required. The type N is often used in such instances. The type N connector comes in both 50- and 75-ohm models and is noted for its extremely low reflections. Type N connectors are difficult to assemble properly and command a high price. Frequently BNC to N adaptors are used; however, such adaptors are not completely reflection free and should be kept out of exacting situations.

The type F connector is used in the 75-ohm television RF transmission line service. This connector makes use of the solid center conductor of the coaxial cable for the center conductor of the male connector. The main advantage of the type F connector is its low cost. Only instruments

**FIG. 11—COMMON COAXIAL CONNECTORS. The BNC type (left), the UHF type (middle), and the type N (right).**

**TABLE I Characteristics of Commonly used Coaxial Cables for Instrumentation**

| Type | Impedance ohms | Capacitance-per-foot (pF) | Voltage rating | Center conductor |
|---|---|---|---|---|
| RG-58/U | 53.5 | 28.5 | 1900 | #20 Solid |
| RG-58A/U | 50 | 29.5 | 1900 | #21 Stranded |
| RG-58C/U | 50 | 29.5 | 1900 | #21 Stranded |
| RG-59/U | 73 | 21.0 | 2300 | #22 Solid |
| RG-59A/U | 75 | 20.5 | 2300 | #23 Stranded |
| RG-59B/U | 75 | 20.5 | 2300 | #23 Stranded |
| RG-62/U | 93 | 13.5 | 750 | #22 Solid |
| RG-62A/U | 93 | 13.5 | 750 | #22 Solid |

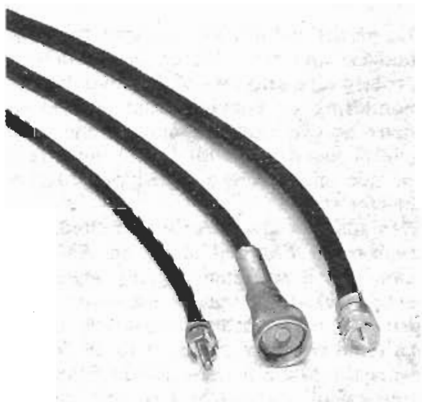


**FIG. 12—COMMON SHIELDED CABLE CONNECTORS. The phono connector (left), the single button microphone connector (middle), and the type F connector (right).**

specifically designed for television RF use are found with the type F connector.

Three other coaxial connectors are occasionally found, usually on shielded cables rather than transmission lines. The single button microphone connector was chosen for a few instruments (it can usually be neatly replaced by a BNC jack, in the same hole) and at this time has little to commend it. Although never really popular due to its inexpensive construction, the phonoplug and jack make a reasonably good connection for small coaxial cables. The phono connector has been used with reasonable success to 500 MHz, and has voltage ratings that will permit use to a hundred volts. The two- or three-conductor phone plug is sometimes found in audio oriented equipment and it is also used as a test lead connector on some VTVM's. Figures 11 and 12 show a few of these connectors.

## Coaxial cables

Three different coaxial cables are in common use for interconnections in instrumentation systems. These are RG-58 (the most common), a 50-ohm cable; RG-59, a 75-ohm cable most frequently found in television systems; and RG-62, a 93-ohm cable of much lower capacitance per unit length than the other two. The essential characteristics of these three cables are listed in Table 1.

Although the majority of coaxial cable usage for instrumentation work is done with the three previously mentioned coaxial cables, there are applications utilizing other cables. Specialized applications include low-cost microphone cable, where neither capacitance nor characteristic impedance are important; two center conductors inside a shield, where differential measurements are to be made; and double shielded coaxial cables, where the 80 dB shielding of a single braid is not sufficient and the 120 dB shielding of a double braid is required.

*(continued next month)*

# ELECTRONICS —it's easy!

# PART 38

## TRANSMISSION LINE THEORY

ELECTRONIC systems consist of basic analogue and digital subsystems interconnected to provide the required overall input-output relationships. It is important for the various subsystems to be interfaced correctly if they are to perform as intended. But with this condition satisfied, one cannot just assume that subsystems merely connect together without need to consider any other parameters in the interconnection process.

In practice the individual circuit assemblies may be geographically apart — such as the remote control of off-shore oil wells by a shore-based computer, the recording of test data from a missile, the control of banking accounts by a central computer centre or the sensors of a refinery which connect to the central control room. Each of these required some form of telemetry system.

When making connections it is also important, especially when noise sources are present that will interfere with the signal, to ensure that the signal is transferred from stage to stage without significant noise pick-up or signal degradation.

### TRANSMISSION LINKS

Several different transmission methods exist in which the signal is confined — open wires, coaxial cables and waveguides, optical fibres etc. Alternatively, information can be transferred via open radiation paths — radio, optical or acoustic links. The required signal bandwidth is one of the primary factors deciding which method is used. In radiation methods it is often necessary to use a carrier frequency higher than the signal bandwidth dictates because low frequency carriers will not radiate as well for the same amount of transmitted power.

**Confined Signal Links:** The simplest links are formed using an open-wire circuit (supported on insulators) or a multicore cable (such as is used in local telephone distribution).

Although apparently trivial, lines may, in fact, be an important part of the system. They are not as simple as they first appear because they have a frequency response that must be adequate for the signal bandwidth to be transmitted. Open-wire lines would not normally be used beyond 10MHz. Above that coaxial cables are needed — these are useful to about 5000MHz.

When current flow in a conducting line, magnetic and electric fields are set up around the wires. Figure 1 shows these plotted for the various kinds of cable. Open configurations radiate energy, the amount increasing with the frequency of the signal. A line is, in reality, a distributed inductance and capacitance component which also has losses due to the resistance of the wire and the resistance to ground. Figure 2 shows how lines can be considered as a lumped-element equivalent circuit which can be analysed more easily. Depending upon the factors that are negligible for a particular case the equivalent can be reduced to simpler circuits — see Fig. 3. For example, at very low frequencies (less than say 100 kHz) a medium length line may be represented by the series resistance of the cable shunted by the capacitance of the line. Typical

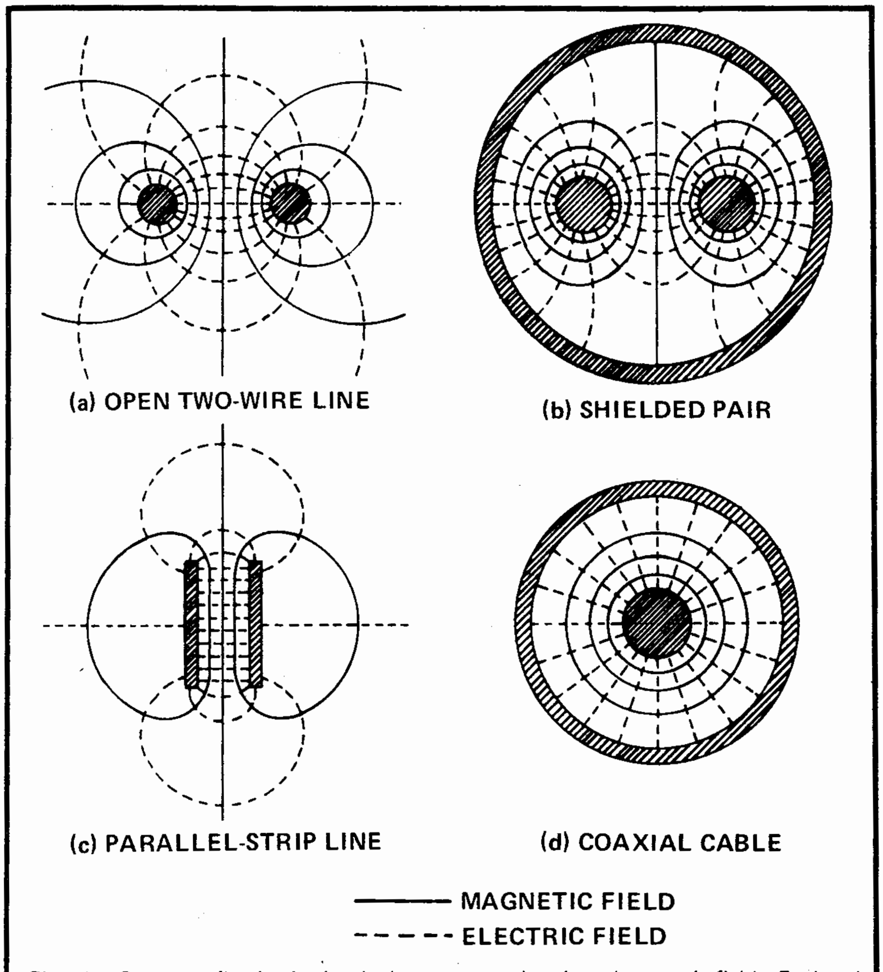



*Fig. 1. Currents flowing in signal wires generate electric and magnetic fields. Enclosed configurations can be used at higher frequencies because these fields are contained.*

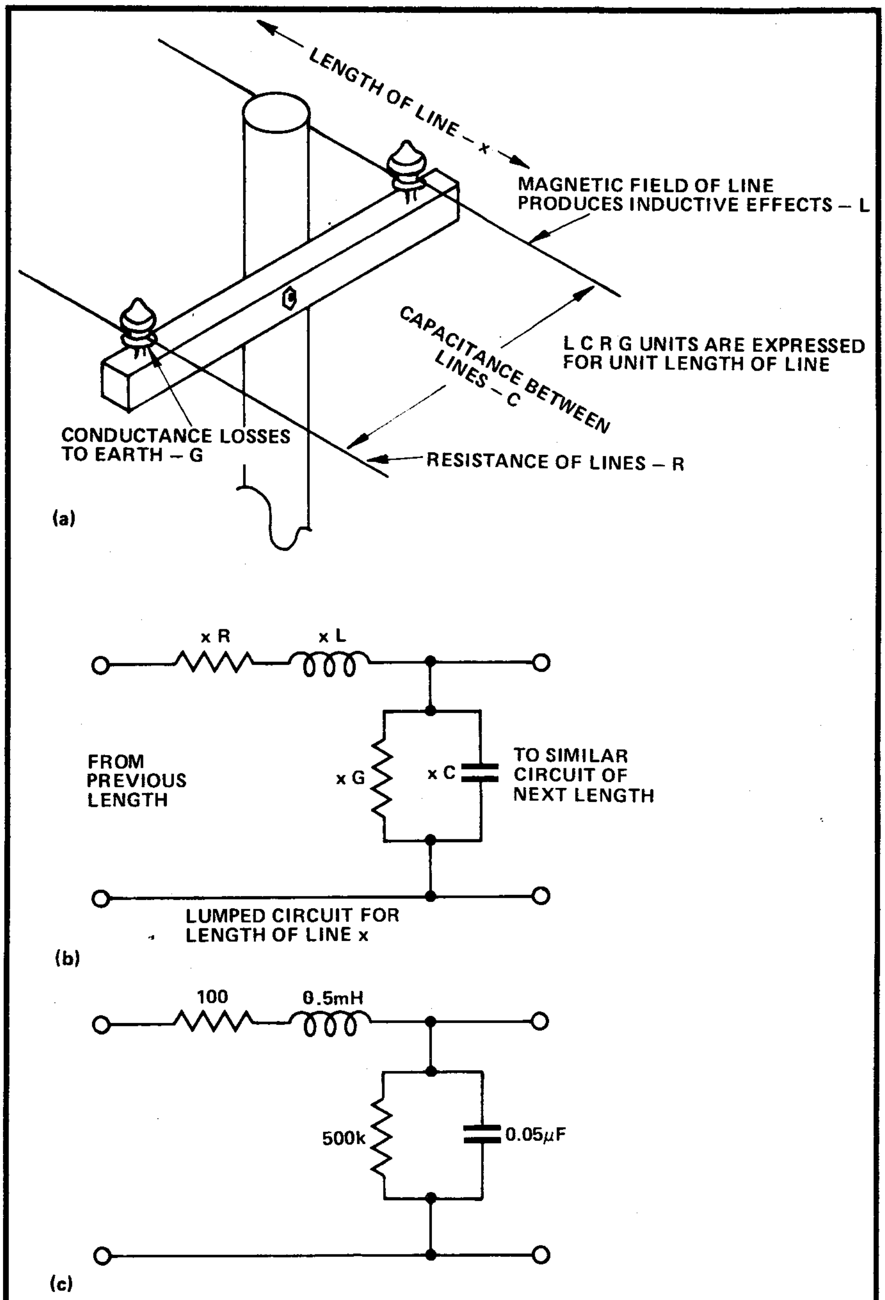



*Fig. 2. Transmission links are systems in which R, L and C are distributed uniformly over the length. For convenience we can consider the line as being composed of cascaded lumped-equivalent elements.*
*(a) A length of low frequency telephone line.*
*(b) Approximate lumped-element equivalent.*
*(c) Representation values for 1 km. of medium-size telephone line with earth return. (Actual constants vary widely depending upon design of line).*

cables may have a resistance of around 0.05 ohm per metre and a capacitance of 100 pF per metre. Hence a long length of shielded or open cable could provide a considerable shunting effect that attenuates and phase shifts the signal.

## APPLY OHMS LAW

When connecting high output-impedance sensors to lines, as little as one metre of cable may be sufficient to markedly attenuate the signal. It's a matter of applying Ohms law to the suitable equivalent circuit.

Because of the reactive effects of the cable the higher frequency signals transmitted will be degraded more than the low frequencies — for example, square waves become rounded as well as attenuated. The high-frequency performance of the line may be improved by "loading" it with inductors placed at regular intervals. The inductance value is chosen to tune out the inherent capacitive reactance at the upper frequency where response begins to fall off, a method that extends the bandwidth some way beyond the inherent, unloaded upper limit. This is used, for example, to broaden the bandwidth of submarine cables.

## CO-AXIAL

The coaxial cable, shown in Fig. 4, by virtue of the surrounding external shield (Fig. 1) acting as the second wire, has no external field and, therefore, does not radiate energy. Because of this a well designed coaxial cable will pass from dc to microwave frequencies — that is, such a cable can have a bandwidth of about 5000 MHz. Coaxial cable is, therefore, potentially able to transfer much more information than open wires. It does however need a common earth connection (asymmetric) and can't be used in a balanced mode (see later). The bandwidth of practical coaxial cables is limited by resistive and dielectric losses. In practice waveguides are generally used at frequencies above 1000 MHz or so.

## WAVE GUIDES

Waveguides consist of precise pipework — they look as if they had been made by a precision plumber! Waveguides carry travelling electromagnetic waves of very high frequency and behave vaguely in the same way that pipes carry water. They cannot however be used for low frequency transmission.

The cross-sectional area of a waveguide is inversely proportional to the design frequency. As a general rule of thumb guide the upper frequency limit of a waveguide is where the wavelength of the signal becomes one quarter of the guide aperture — millimetre wavelength signals (50 GHz or so) being the practical upper limit.

## OPTICAL FIBRES

Beyond this, a still wider bandwidth is obtainable using optical fibre transmission elements which will pass radiation in the visible lignt region) $10^{14}$ Hz to $10^{15}$ Hz). At our current state of technology, however, scientists have only been able to detect the frequencies of far infra-red signals (around $10^{11}$ Hz). We cannot, as yet, monitor individual cycles of light with electronic detectors.

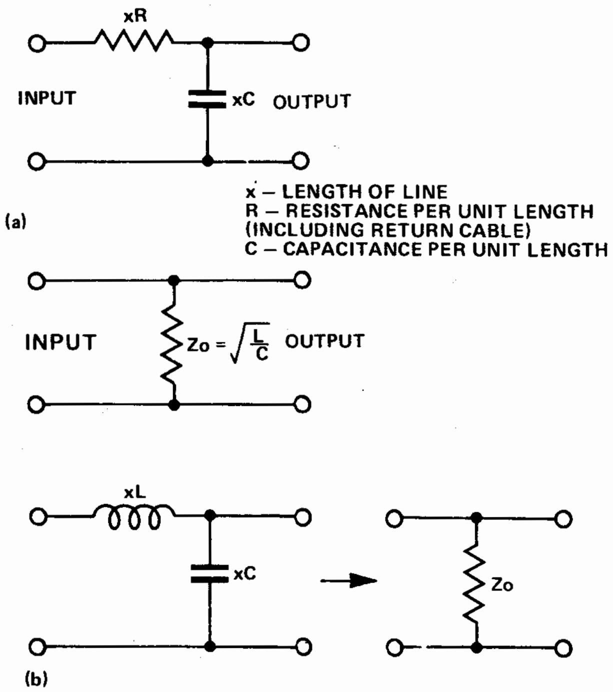



*Fig. 3. In certain practical cases the lumped equivalent reduces to simpler situations situations.*
*(a) Low frequency (negligible L assumption) short line in which only C and R are dominant. The R, C values are found from maker's data.*
*(b) High frequency lossless line (negligible R and G assumptions). The input and output impedances of the line are equal and constant regardless of length.*

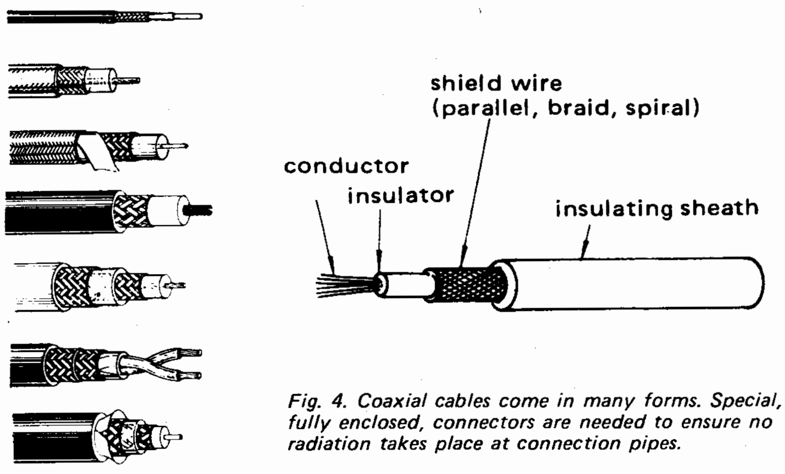



*Fig. 4. Coaxial cables come in many forms. Special, fully enclosed, connectors are needed to ensure no radiation takes place at connection pipes.*

## LUMPED LINES

When the losses of the line are insignificant (G = 0, R = 0, in Fig. 2b) the lumped-equivalent of the transmission lines reduces to L in series and C shunting, as shown in Fig. 3b. The nett result is, rather surprisingly, that the line exhibits only resistance of a fixed value when looking into the ends. This is called the characteristic impedance, Zo, for which Zo = (inductance per unit-length/capacitance per unit length) ½. The line appears to be purely resistive and the Zo value is decided by the design of the line or cable, not by its length! Examples are 600 ohm telephone lines, 75 ohm colour TV coaxial feeder cable. This means, in practice, that we can interconnect units on the basis of matching all connections to the Zo of the cable without having to worry about the cable length. If this rule is observed, no high-frequency energy will be reflected at the termination to change the information being transmitted . (The need for correct matching was also mentioned in the previous discussion about filters). However, if the line is very long matching must still be applied to obtain maximum transfer, but account must now be taken of losses. For example a typical 75 ohm coaxial cable will have losses of the order of 2 to 5 dB per one hundred metres.

**Radiation Links:** Electrical signals fed into open wires radiate energy out into the surrounding medium. As well as this radiated energy there also exists a "near field" that remains established, storing energy. This is the field we associate with, say, an electromagnet. As the frequency rises the ratio of radiated energy to stored energy increases. For this reason we are able to build efficient radio systems provided the frequency is kept above 100 kHz or so. Lower frequencies can be used as transmission systems but the power input needs rise enormously for the same distance radiated in free space. (The Omega navigation system uses extremely powerful VLF signals because of their ability to penetrate deep into the waters of the ocean). Beyond the gigahertz frequency region, circuitry becomes impracticable with current technology.

Even though the radiated energy must be at a very high frequency to operate efficiently we may not necessarily need to use the bandwidth available on the carrier, modulation techniques are used to super-impose a relatively narrow bandwidth signal on to the carrier. It might be thought that optical and infra-red links use extremely high carrier frequencies (330000 GHz for red light) but in these applications the carrier is not modulated on an individual cycle basis but rather as a variation of a continuous dc link. Fig. 6 is a modern link designed to transmit television plus speech commands — a bandwidth of 7.5MHz. Acoustic links using soundwave propagation operate with frequencies as low as 10Hz to well above the 10MHz region. These can be modulated on the individual cycle basis.

**Skin Effect:** The alternating magnetic field produced around a wire has the effect of causing the current flowing in the wire to flow at a greater density in the outer region of the wire. The higher the frequency the more pronounced this so-called skin-effect. At the very high frequencies so little current flows in the centre of the cable that the centre is often omitted completely, thus a tube is used as a conductor. For example, at 1 MHZ the majority of the current flows in a copper cable to a deoth of only 60 $\mu$m whereas at 60 Hz the distance would be 8.6 mm depth. This also means that the effectivle resistance of a wire rises significantly with frequency — by factors of 100.

**Process Industry Telemetry Links:** Process plants such as oil refineries, paper mills, brick kilns, power stations and aluminium refining plants are monitored by using hundreds of sensors connected to the control-room area via instrumentation links. These are invariably wired using shielded wire or coaxial cable. Because of the extreme electrical noise level of

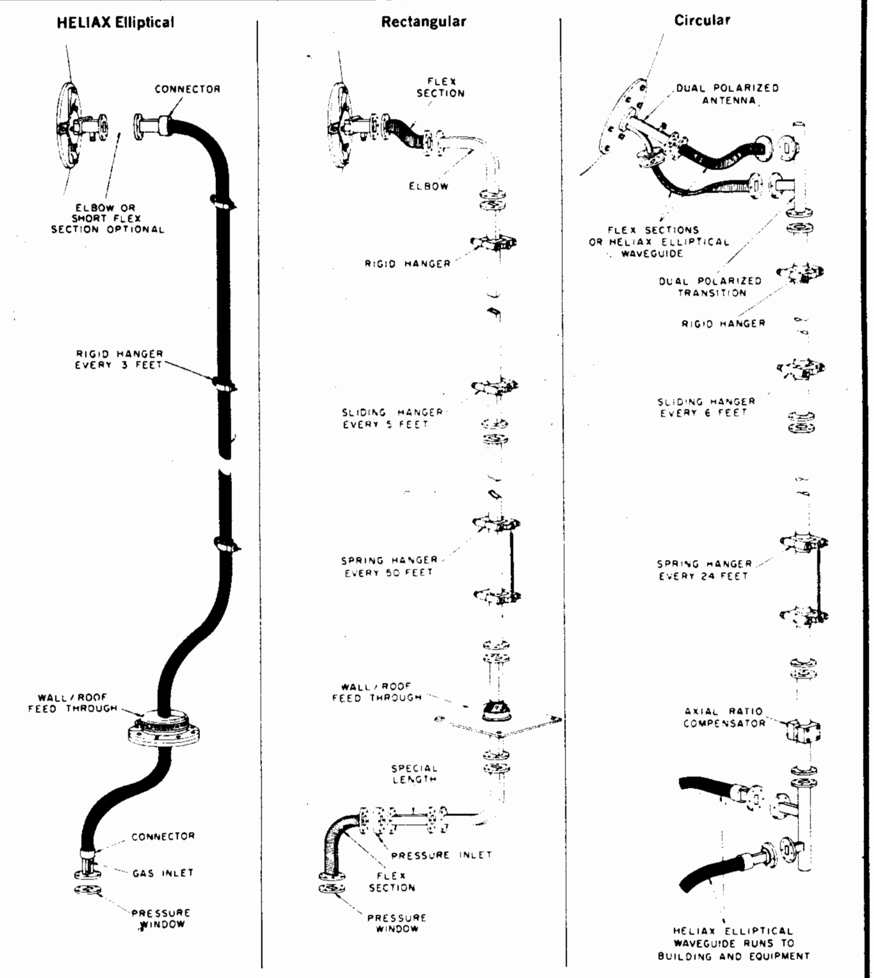



*Fig. 5. Waveguides transmit electro-magnetic energy via travelling fields that are conveyed through the pipework.*

such plants and low output signal level of the senors these links could pick up significant noise thus degrading the senor information. Over the years process instrument suppliers have standardised the design of the control systems, and their installation and noise pick-up by the cable has been avoided by several methods.

The first strategy is to superimpose the information signal on to a standing current or voltage thus raising the wanted signal level above expected noise levels. The two systems commonly used transmit the signal range of the data through 4-20 mA DC or 10-50 mV DC systems. An 0-20 mA system is also common. Current transmission has the advantage that the circuit is of low impedance — a few ohms — which reduces the level of induced noise power. Figure 7 is an example of these practices — Honeywell's arrangements used to test the temperature and pressure of natural gas wells in the Leman Field of the North Sea.

## SAFETY PRECAUTIONS

Often the sensor has to be placed at a location where an explosion could result from a spark or excessive overheating of a malfunctioning sensor circuit. The most obvious way of overcoming this is to place the whole unit in an explosion-proof enclosure. This, however, has disadvantages the cost is high, and testing and maintenance difficult due to the need to shut off the power when the enclosure is opened.

The alternative, more modern, method is known as intrinsic safety. As inflammables require a specific level of energy to ignite them, explosion can be prevented by ensuring that the sensor stage cannot, under any conditions, provide enough ignition energy. No enclosures are needed and the circuit can be maintained whilst it is operating. Originally the concept was implemented by ensuring the senor circuitry could not draw, or produce via storage, more than a specified power level. This level was found by experiment in a test rig set up for the situation involved.

The more recent idea is to use "safety barriers." At the exit from the declared hazardous area, the cables terminate into a zener-diode and attenuator arrangement which ensures that the current and voltage entering the area are limited to safe values. Figure 8 shows the circuit of a zener barrier. Another safety device uses a solid-state closely-coupled electro-optic link which provides DC electrical isolation between its input and output, the information being transferred from a light-emitting-diode mounted next to a silicon photo-diode detector. These ensure that overvoltage or induced earth-loop currents cannot enter the isolated hazardous area.

## MEDICAL MATTERS

In electro-medical instrumentation, safety precautions of another kind are vital to ensure the senor does not act as a pathway for a dangerous level of electric-current into the patient. At 240 VAC the human body's resistance, hand to hand is around 2000 ohms — 100 mA will flow. If totally connected (as by a conducting fluid) the resistance reduces to 200 ohm — 1 A will flow. About 75mA through the body will produce heart fibrillation; only 150 $\mu$A, through the heart itself, is needed to produce this effect. A person can usually hold (with the fingers) and release as much as a 10mA, 240 VAC current — beyond that the muscles become paralysed. Skin moisture largely decides the hand to hand resistance. When dry it will be (at 240 V) 2500 ohms and moist 1000 ohms. *Thus a hand-to-hand 240 V encounter will provide a shock at least double the fibrillation level!*

The instrumentation must, where the metal parts are earthed, be wired with the active, neutral and earthing wires connected correctly. Double-insulated systems avoid this problem. Earth-leakage balanced — core breakers are worth using. These defect minute difference currents in the active and neutral, tripping a breaker if they rise above milliamperes.

The sensor attached to the patient must not be capable of providing a lethal level of energy by means of feedback from the instrumentation.

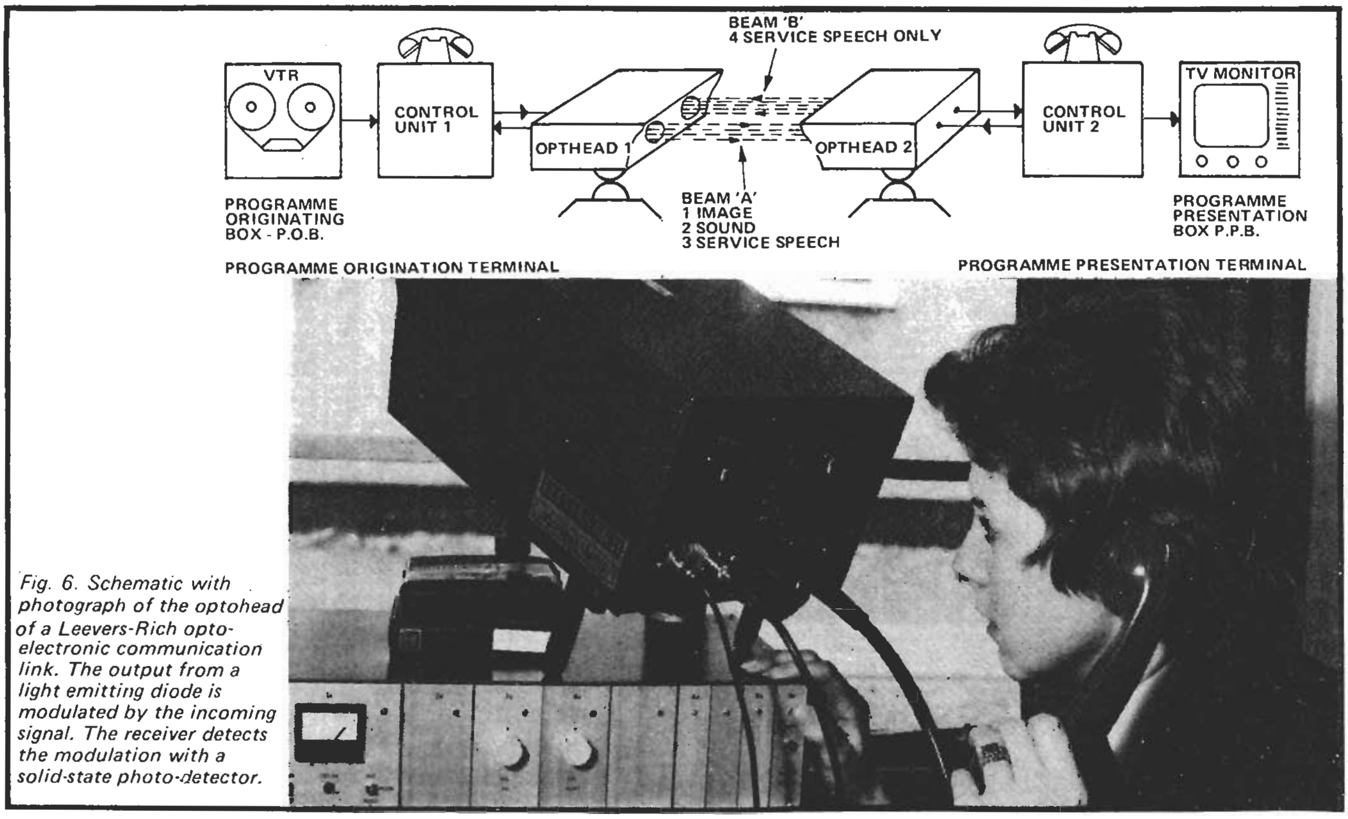



*Fig. 6. Schematic with photograph of the optohead of a Leevers-Rich opto-electronic communication link. The output from a light emitting diode is modulated by the incoming signal. The receiver detects the modulation with a solid-state photo-detector.*

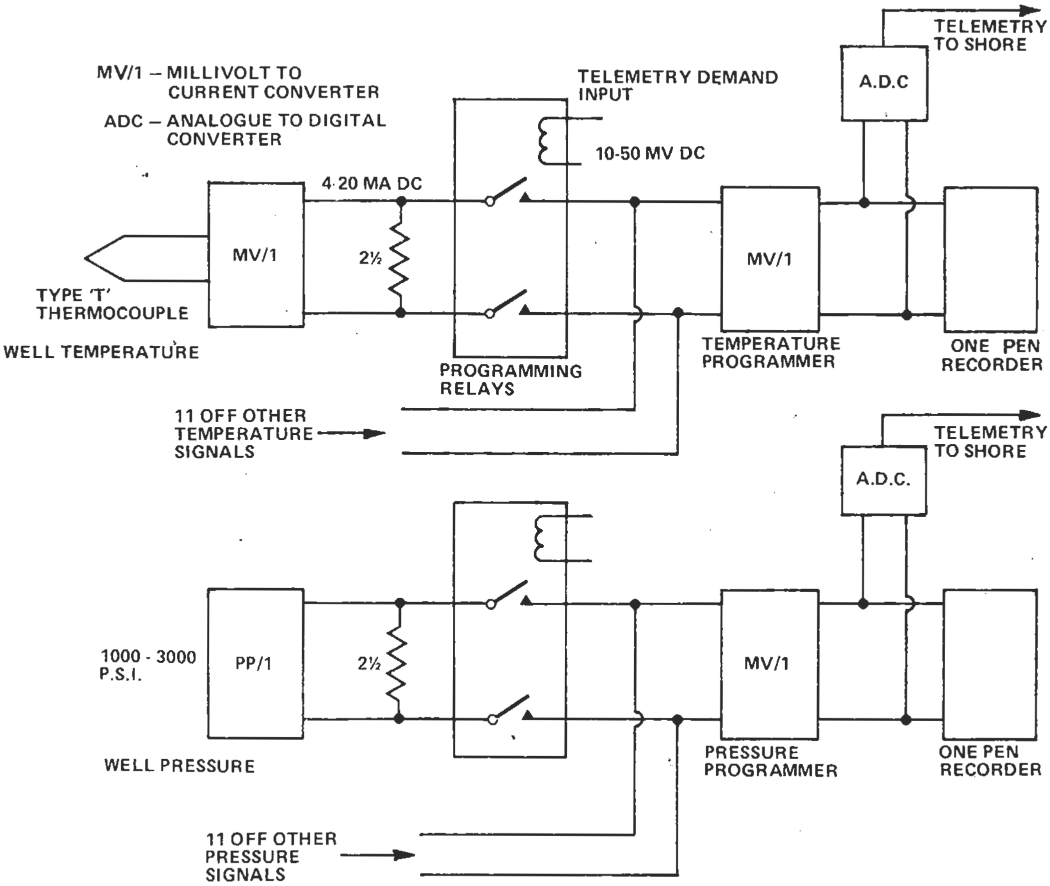
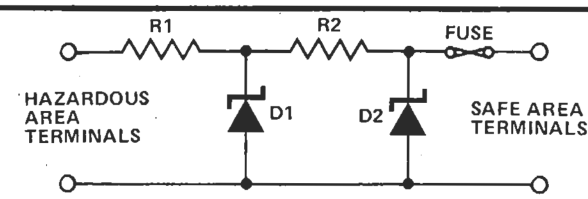



*Fig. 7. Process measurement link arrangements used between the oil-well and the off-shore platform. The A/D units send the data to a shore-based computer by digital telemetry links.*

*Fig. 8. Circuit used in a Zener barrier.*



# How Computers Detect and Correct Transmission Errors

***Parity checking and redundancy are two of the devices used in error detection.***

BY JEROME MAY

ALMOST everyone has had some experience in dealing with a computer mixup. One classic story goes like this: a computer, doing one of the earliest payroll jobs, put a 1 where a zero should have been, and the next week the janitor picked up his paycheck for exactly $1,000,147.38!!

Systems engineers and serious microprocessor hobbyists are aware of the problems that random noise can cause on data lines. Noise can cause the "three" that was sent to show up as a "seven."

Because "error" can be treated as a random event, having equal likelihood of occurring in any given datum, it becomes possible to apply techniques of *information theory, probability,* and *statistics* toward designing systems that are resistant to this type of error.

A simplifying assumption that will be useful is that, in a data word consisting of *n* bits, only *one* bit will be in error. The treatment of multiple-error-detecting systems uses methods similar to those discussed, but the treatment becomes extremely mathematical and complicated and is beyond the scope of an introductory article.

| DECIMAL | BINARY | BCD | EXCESS-3 | GRAY CODE | DECIMAL |
|---|---|---|---|---|---|
| 0 | 0000 | 0000 | 0011 | 0000 | 0 |
| 1 | 0001 | 0001 | 0100 | 0001 | 1 |
| 2 | 0010 | 0010 | 0101 | 0011 | 2 |
| 3 | 0011 | 0011 | 0110 | 0010 | 3 |
| 4 | 0100 | 0100 | 0111 | 0110 | 4 |
| 5 | 0101 | 0101 | 1000 | 0111 | 5 |
| 6 | 0110 | 0110 | 1001 | 0101 | 6 |
| 7 | 0111 | 0111 | 1010 | 0100 | 7 |
| 8 | 1000 | 1000 | 1011 | 1100 | 8 |
| 9 | 1001 | 1001 | 1100 | 1101 | 9 |
| 10 | 1010 | xx | xx | 1111 | 10 |
| 11 | 1011 | xx | xx | 1110 | 11 |
| 12 | 1100 | xx | xx | 1010 | 12 |
| 13 | 1101 | xx | xx | 1011 | 13 |
| 14 | 1110 | xx | xx | 1001 | 14 |
| 15 | 1111 | xx | xx | 1000 | 15 |

*Fig. 1. Five different 4-bit codes for the decimals 0 through 15.*

**Redundancy.** An error can be detected by redundancy, which is the inclusion of extra information with each data transmission. This extra information helps the receiver to decide if the data it has received has been altered in transmission.

Obviously, the simplest error detecting system, conceptually, would be the transmission of each data unit twice. Thus, if the first transmission does not match the second transmission, the receiver can signal that an error has occurred and to please repeat the data.

"This this has has some some obvious obvious disadvantages disadvantages,, and and the the search search goes goes on on for for better better methods methods.."

Restating the problem, the decimal

numbers 0 through 15 have exactly 16 representations in binary, as shown in Fig. 1. Since all possible combinations of 0 and 1 in four positions are used, there is no way to detect datum error because all combinations are equally likely; there is no room for redundancy.

**Binary-Coded Decimal.** Suppose only the decimal digits, 0 through 9, are to be transmitted. From these 10 digits any positive decimal integer can be constructed. The 10 binary representations for these digits are known as *binary-coded decimal*, or BCD.

In BCD the codes for the numbers 10 through 15 decimal are not used and, if they show up at a receiver, they can be detected as "illegal" by checking the received datum with a "legal word" list. By selecting BCD over straight binary, a designer makes it possible to detect one type of error.

**Excess-Three Code.** By making a simple change to the code, more information—redundancy—can be built right in. The *excess-three* (X-3) code is obtained by adding binary 0011 to the BCD codes for the decimal digits 0 through 9. Figure 1 shows that X-3 does not allow the codes 0000 or 1111. Thus, every legal X-3 word contains at least *one* 0 and 1, providing another bit of information—that the data channel is active and transmitting.

Another property that makes it interesting for error-checking is that X-3 is a *self-complementary* code. That is, if each 0 of a legal X-3 word is changed to a 1, and each 1 to 0, the process generates the 9's complement of the word. This feature makes error-checking in X-3 easier and statistically more reliable than BCD codes. In a BCD error-checking algorithm, for example, a minimum of *six* comparisons and table look-ups must be made before an error can possibly be detected.

BCD is a *weighted positional* code; in which the position of each bit in the data word determines its value. X-3 is not a weighted positional code. By counting 1's and 0's in each position of the X-3 representation of the digits 0 through 9, it can be seen that for a given X-3 word, each position has a 50% probability of being either 0 or 1. This eliminates any statistical bias in the code itself.

In an X-3 error-checking algorithm, with judicious use of the "complement" function (a very fast and very easy operation in most microprocessors), the algorithm can detect a bad code in six comparisons but only *two* memory look-ups. In some cases, the error checking can be done 100% *faster* by use of the X-3 code for data transmission.

As a further bonus, the X-3 code makes keeping track of carries and borrows in addition and subtraction of coded decimal digits significantly simpler than in the straight BCD code.

**Gray Code.** Turning to a completely different four-bit code, the *Gray code* finds an application in many analog-to-digital data-transmission systems. The Gray code's most significant feature is that, going from one number to another with a difference of only one, only a single bit changes in the Gray code. In a system where the analog signal is expected to change slowly with respect to the sample frequency (say, for example, temperature inside a house is being monitored and encoded), a change in more than one digit position would automatically signal an error to the receiver.

Some information—more redundancy—about the data itself is reflected in the code itself. The odd decimal digits (1, 3, 5 etc.) have Gray-code equivalents that contain an *odd* number of 0's and 1's. This extra information is designed right into the code system itself, taking advantage of its most likely application.

So far, discussion has been restricted to four-bit codes. This has been done keeping in mind the 8-bit data bus structures of microprocessors such as the 8008, 8080, or 6502. With more complicated codes, however, it is possible to add an extra bit to the data word to contain extra information about the data.

**Parity.** One of the more widely used complicated coding systems is called *parity checking.* Parity checking simply counts the number of 0's or 1's in a data word and assigns a value to an extra *parity bit*, depending on the result (Fig. 2). Thus a seven-bit code such as ASCII (American Standard Code for Information Interchange) might be transmitted in an eight-bit format, with one parity bit.

The *odd parity* system adds a 1 to a data word so that it always has an *odd* number of 1's in it. *Even* parity adds a 1 to cause an even number of 1's. Odd parity has a slight advantage, similar to excess-three, because every code word has at least one 1 or 0 in it, providing a verification of data-channel operation.

Note that with the parity system, if *two* errors occur in the same data word, parity check will not detect the error! But it will detect a three-bit error, or an error in any odd number of different bits.

| DECIMAL | BINARY | PARITY |
|---|---|---|
| 0 | 0000 | 1 |
| 1 | 0001 | 0 |
| 2 | 0010 | 0 |
| 3 | 0011 | 1 |
| 4 | 0100 | 0 |
| 5 | 0101 | 1 |
| 6 | 0110 | 1 |
| 7 | 0111 | 0 |
| 8 | 1000 | 0 |
| 9 | 1001 | 1 |
| 10 | 1010 | 1 |
| 11 | 1011 | 0 |
| 12 | 1100 | 1 |
| 13 | 1101 | 0 |
| 14 | 1110 | 0 |
| 15 | 1111 | 1 |

*Fig. 2. Binary codes and odd-parity digits for decimal numbers 0 through 15.*

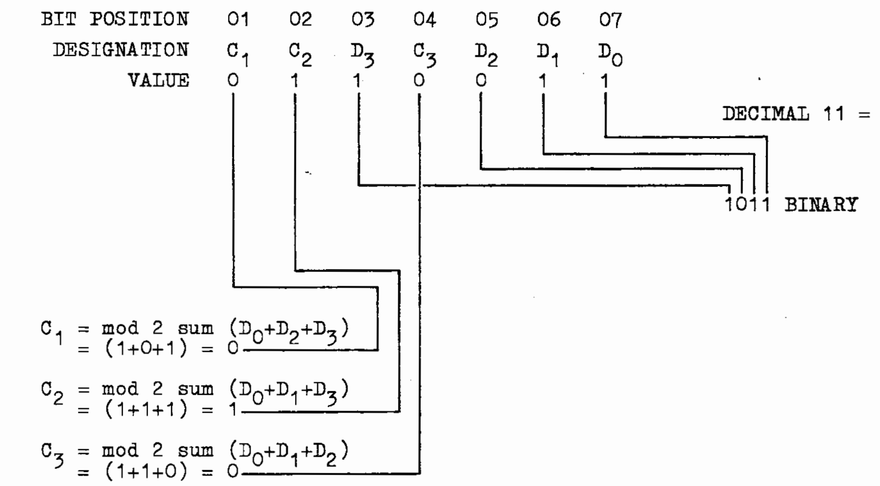



*Fig. 3A. How Hamming-code check digits are generated.*

| | 01 | 02 | 03 | 04 | 05 | 06 | 07 |
|---|---|---|---|---|---|---|---|
| | $C_1$ | $C_2$ | $D_3$ | $C_3$ | $D_2$ | $D_1$ | $D_0$ |
| RECEIVED VALUE | 0 | 1 | 1 | 0 | 0 | 1 | 1 |

$C_1'$ = mod 2 sum $(C_1 + D_0 + D_2 + D_3)$ = (0+1+0+1) = 0

$C_2'$ = mod 2 sum $(C_2 + D_0 + D_1 + D_3)$ = (1+1+1+1) = 0    $(C_3', C_2', C_1') = (0,0,0)$

$C_3'$ = mod 2 sum $(C_3 + D_0 + D_1 + D_2)$ = (0+1+1+0) = 0    NO ERROR

*Fig. 3B. Check digits show that word was transmitted correctly.*

| | 01 | 02 | 03 | 04 | 05 | 06 | 07 |
|---|---|---|---|---|---|---|---|
| | $C_1$ | $C_2$ | $D_3$ | $C_3$ | $D_2$ | $D_1$ | $D_0$ |
| RECEIVED VALUE | 0 | 1 | 1 | 0 | 0 | 0 | 1 |

$C_1'$ = mod 2 sum $(C_1+D_0+D_2+D_3)$ = (0+1+0+1) = 0

$C_2'$ = mod 2 sum $(C_2+D_0+D_1+D_3)$ = (1+1+0+1) = 1    $C_3', C_2', C_1' = (1,1,0)$

$C_3'$ = mod 2 sum $(C_3+D_0+D_1+D_2)$ = (0+1+0+0) = 1    ERROR IN BIT 6

*Fig. 3C. The check digits show the error is in bit 6.*

| | 01 | 02 | 03 | 04 | 05 | 06 | 07 |
|---|---|---|---|---|---|---|---|
| | $C_1$ | $C_2$ | $D_3$ | $C_3$ | $D_2$ | $D_1$ | $D_0$ |
| RECEIVED VALUE | 0 | 0 | 1 | 0 | 0 | 1 | 1 |

$C_1'$ = mod 2 sum $(C_1+D_0+D_2+D_3)$ = (0+1+0+1) = 0

$C_2'$ = mod 2 sum $(C_2+D_0+D_1+D_3)$ = (0+1+1+1) = 1    $(C_3', C_2', C_1') = (0,1,0)$

$C_3'$ = mod 2 sum $(C_3+D_0+D_1+D_2)$ = (0+1+1+0) = 0    ERROR IN BIT 2

*Fig. 3D. The check digits show an error in bit 2, proof that even the error-protect digits are safe from error.*

**Multiple Errors.** So far, the code structures and error-checking systems have provided enough redundancy to provide for some manner of single-error checking. But all fail pretty badly at detecting more than one error per datum.

Is single-error detection enough? Statistics (and the binomial theorem) say that for an error rate of one in 1000 transmitted bits, the odds of having two erroneous bits in a 5-bit word are one in 100,000. So, the single-error approximation seems a realistic, although simplistic, choice.

The amount of redundancy in an error-correcting code must be much higher than in a mere error-detecting code. Mathematicians, using such esoteric items as group theory, vector spaces and cyclic codes, have come up with a whole flock of codes that contain information about themselves, but most of them are best applicable when large bit-strings are processed.

**Hamming Code.** One of the simplest error-correcting codes is called the *Hamming code*, after its developer. The Hamming system for a four-bit datum, for example, generates three check bits, and the data must be transmitted in a particular manner (that is, the check and data bits interspersed in a particular manner) in order for the decoding and verification to correctly occur.

Suppose the decimal number 11 is to be transmitted. The binary representation is arranged in the indicated positions of the data word, in Fig. 3A and the check digits are generated by following the rules shown. The *modulo 2 sum* is the remainder (0 or 1) after adding a string of binary digits and disregarding any carry operations. The check digits go into the positions shown.

The resulting seven-bit word is now transmitted. The error-checking and results for a word transmitted with no error is shown in Fig. 3B.

Figure 3C shows the seven-bit word received with an error in data digit D1 in bit position 6. Note that, when the verification digits ($C_3'$, $C_2'$, $C_1'$) are arranged as a three-bit binary number, they point right to the erroneous bit—the number 110 in binary is 6 decimal.

Figure 3D shows the seven-bit word received with an error in check digit $C_2$. The verification digits now form the binary number 010, bit position 2, where the error is!

Again, note that if two errors occur, it is impossible to completely reconstruct the data word with this given scheme. Multiple-error-correction codes do exist, however. ◇