

Chapter 8: Fast Electronics and Transient Behavior on TEM Lines

8.1 Propagation and reflection of transient signals on TEM transmission lines

8.1.1 Lossless transmission lines

The speed of computation and signal processing is limited by the time required for charges to move within and between devices, and by the time required for signals to propagate between elements. If the devices partially reflect incoming signals there can be additional delays while the resulting reverberations fade. Finally, signals may distort as they propagate, smearing pulse shapes and arrival times. These three sources of delay, i.e., propagation plus reverberation, device response times, and signal distortion are discussed in Sections 8.1, 8.2, and 8.3, respectively. These same issues apply to any system combining transmission lines and circuits, such as integrated analog or digital circuits, printed circuit boards, interconnections between circuits or antennas, and electrical power lines.

Transmission lines are usually paired parallel conductors that convey signals between devices. They are fundamental to every electronic system, from integrated circuits to large systems. Section 7.1.2 derived from Maxwell's equations the behavior of transverse electromagnetic (TEM) waves propagating between parallel plate conductors, and Section 7.1.3 showed that the same equations also govern any structure, even a dissipative one, for which the cross-section is constant along its length and that has at least two perfectly conducting elements between which the exciting voltage is applied. Using differential RLC circuit elements, this section below derives the same transmission-line behavior in a form that can readily be extended to transmission lines with resistive wires, as discussed later in Section 8.3.1. Since resistive wires introduce longitudinal electric fields, such lines are no longer pure TEM lines.

Equations (7.1.10) and (7.1.11) characterized the voltage $v(t,z)$ and current $i(t,z)$ on TEM structures with inductance L [H m^{-1}] and capacitance C [F m^{-1}] as:

$$dv/dz = -L di/dt \quad (8.1.1)$$

$$di/dz = -C dv/dt \quad (8.1.2)$$

These expressions were combined to yield the *wave equation* (7.1.14) for lossless TEM lines:

$$\left(d^2/dz^2 - LC d^2/dt^2 \right) v(z,t) = 0 \quad (\text{TEM wave equation}) \quad (8.1.3)$$

One general solution to this wave equation is (7.1.16):

$$v(z,t) = v_+(z-ct) + v_-(z+ct) \quad (\text{TEM voltage}) \quad (8.1.4)$$

which corresponds to the superposition of forward and backward propagating waves moving at velocity $c = (LC)^{-0.5} = (\mu\epsilon)^{-0.5}$. The current $i(t,z)$ corresponding to (8.1.4) follows from substitution of (8.1.4) into (8.1.1) or (8.1.2), and differentiation followed by integration:

$$i(z,t) = Y_0 [v_+(z-ct) - v_-(z+ct)] \quad (\text{TEM current}) \quad (8.1.5)$$

Y_0 is the *characteristic admittance* of the line, and the reciprocal of the *characteristic impedance* Z_0 :

$$Z_0 = Y_0^{-1} = (L/C)^{0.5} \text{ [Ohms]} \quad (\text{characteristic impedance of lossless TEM line}) \quad (8.1.6)$$

The value of Y_0 follows directly from the steps above.

A more intuitive way to derive these equations utilizes an equivalent *distributed circuit* for the transmission line composed of an infinite number of differential elements with series inductance and parallel capacitance, as illustrated in Figure 8.1.1(a). This model is easily extended to non-TEM lines with resistive wires.

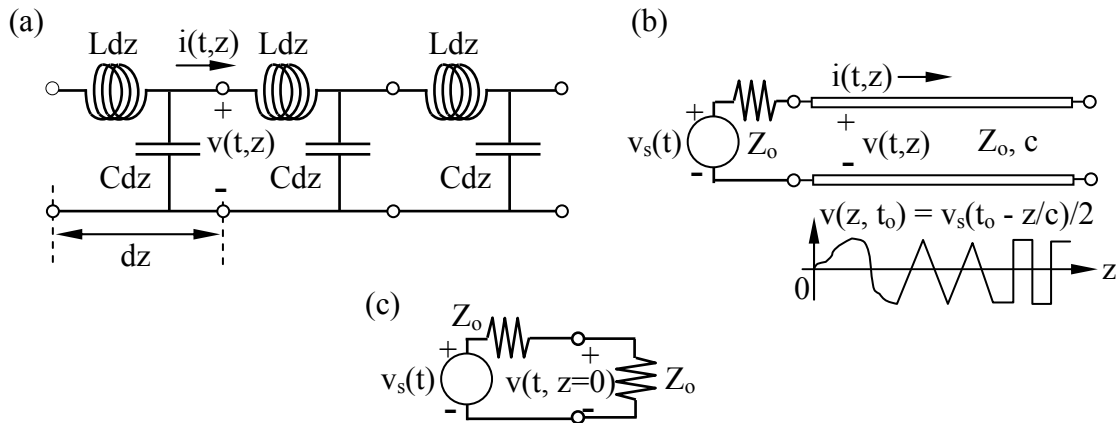


Figure 8.1.1 Distributed circuit model for lossless TEM transmission lines.

The inductance L [Henries m^{-1}] of the two conductors arises from the magnetic energy stored per meter of length, and produces a voltage drop dv across each incremental length dz of wire which is proportional to the time derivative of current through it⁴⁰:

$$dv = -L dz (di/dt) \quad (8.1.7)$$

Any current increase di across the distance dz , defined as $di = i(t, z+dz) - i(t,z)$, would be supplied from charge stored in C [$F m^{-1}$]:

⁴⁰ An alternate equivalent circuit would have a second inductor in the lower branch equivalent to that in the upper branch; both would have value $Ldz/2$, and $v(t,z)$ and $i(t,z)$ would remain the same.

$$di = -C dz (dv/dt) \quad (8.1.8)$$

These two equations for dv and di are equivalent to (8.1.1) and (8.1.2), respectively, and lead to the same wave equation and general solutions derived in Section 7.1.2 and summarized above, where arbitrary waveforms propagate down TEM lines in both directions and superimpose to produce the total $v(z,t)$ and $i(t,z)$.

Two equivalent solutions exist for this wave equation: (8.1.4) and (8.1.9):

$$v(z, t) = f_+ (t - z/c) + f_- (t + z/c) \quad (8.1.9)$$

The validity of (8.1.9) is easily shown by substitution into the wave equation (8.1.3), where again $c = (LC)^{-0.5}$. This alternate form is useful when relating line signals to sources or loads for which z is constant, as illustrated below. The first form (8.1.4) in terms of $(z - ct)$ is more convenient when t is constant and z varies.

Waves can be launched on TEM lines as suggested in Figure 8.1.1(b). The line is driven by the Thevenin equivalent source $v_s(t)$ in series with the source resistance Z_o , which is matched to the transmission line in this case. Equations (8.1.4) and (8.1.5) say that if there is no negative traveling wave, then the ratio of the voltage to current for the forward wave on the line must equal $Z_o = Y_o^{-1}$. The equivalent circuit for this TEM line is therefore simply a resistor of value Z_o , as suggested in Figure 8.1.1(c). If the source resistance is also Z_o , then only half the source voltage $v_s(t)$ appears across the TEM line terminals at $z = 0$. Therefore the voltages at the left terminals ($z = 0$) and on the line $v(t,z)$ are:

$$v(t, z = 0) = v_s(t)/2 = v_+(t, z = 0) \quad (8.1.10)$$

$$v(t, z) = v_+(t - z/c) = v_s(t - z/c)/2 \quad (\text{transmitted signal}) \quad (8.1.11)$$

where we have used the solution form of (8.1.9). The propagating wave in Figure 8.1.1(b) has half the amplitude of the Thevenin source $v_s(t)$ because the source was matched to the line so as to maximize the power transmitted from the given voltage $v_s(t)$. Note that (8.1.11) is the same as (8.1.10) except that z/c was subtracted from each. Equality is preserved if all arguments in an equation are shifted the same amount.

If the Thevenin source resistance were R , then the voltage-divider equation would yield the terminal and propagating voltage $v(t,z)$:

$$v(t, z) = v_s(t - z/c) [Z_o / (R + Z_o)] \quad (8.1.12)$$

This more general expression reduces to (8.1.10) when $R = Z_o$ and $z = 0$.

Example 8.1A

A certain integrated circuit with $\mu = \mu_0$ propagates signals at velocity $c/2$, and its TEM wires exhibit $Z_0 = 100$ ohms. What are ϵ , L , and C for these TEM lines?

Solution: $c = (\mu_0\epsilon_0)^{-0.5}$, and $v = c/2 = (\mu_0\epsilon)^{-0.5}$; so $\epsilon = 4\epsilon_0$. Since $v = (LC)^{-0.5}$ and $Z_0 = (L/C)^{0.5}$,
 $L = Z_0/v = 200/c = 6.67 \times 10^{-7}$ [Hy], and $C = 1/vZ_0 = 1/200c = 1.67 \times 10^{-11}$ [F].

8.1.2 Reflections at transmission line junctions

If a transmission line connecting a source to a load is sufficiently short, then the effects of the line on reflections can be modeled by simply replacing it with a small lumped capacitor across the source terminals representing the capacitance between the wires, and a resistor in series with an inductor and the load, representing the resistance and inductance of the wires. If, however, the line length D is such that the propagation time $\tau_{\text{line}} = D/c$ is a non-trivial fraction of the shortest time constant of the load τ_{load} , then we should use transmission line models governed by the wave equation (e.g., 8.1.3). That is, the TEM wave equation should be used unless the line length D is:

$$D \ll c\tau_{\text{load}} \quad (8.1.13)$$

For larger values of D the propagation delays become important and a transmission line model must be used, as explained in Section 8.1.1. Section 8.1.1 also explained how signals are launched and propagate on TEM lines, and how the Thevenin equivalent circuit (8.1.6) for a passive transmission line as seen by the source is simply a resistor $Z_0 = (L/C)^{0.5}$. This *characteristic impedance* Z_0 of the transmission line is the ratio of the forward voltage $v_+(t,z)$ to the associated current $i_+(z,t)$. TEM signals are partially transmitted and partially reflected at each junction they encounter, where these junctions may be the intended load or simply places where the impedance Z_0 of the transmission line changes. Sometimes multiple transmission lines meet at such junctions.

Section 7.2.2 (7.2.7) derived the *reflection coefficient* Γ for an arbitrary TEM wave $v_+(t,z)$ reflected by a load resistance R at z , where the normalized impedance of the load is $R_n = R/Z_0$:

$$v_-(t,z) = \Gamma v_+(t,z) \quad (8.1.14)$$

$$\Gamma = (R_n - 1)/(R_n + 1) \quad (8.1.15)$$

$$R_n \equiv R/Z_0 \quad (8.1.16)$$

It is important to distinguish the difference between Γ for purely resistive loads, which is real, and $\underline{\Gamma}(\omega)$, which is complex and applies to any complex load impedance \underline{Z}_L . Here R and Γ are real.

Consider the example illustrated in Figure 8.1.2(a), where a TEM line is characterized by impedance Z_0 and phase velocity c . The line is D meters long, open circuit at the right-hand end, and driven by a unit-step⁴¹ voltage $u(t)$. The equivalent circuit at the source end of the line is illustrated in (b), which is simply a voltage divider that places $v_s(t)/2$ volts across the line. But the voltage across the line equals the sum of the forward and backward moving waves, where a passive line at rest has no backward wave. Therefore the forward wave here at $z = 0+$ is simply $u(t)/2$, and the result is a voltage of 0.5 volts that moves down the line at velocity c , as illustrated in Figure 8.1.2(c) for $t = t_0$. The associated current $i(z, t)$ is plotted in (d) for $t = t_0$, and is proportional to the voltage.

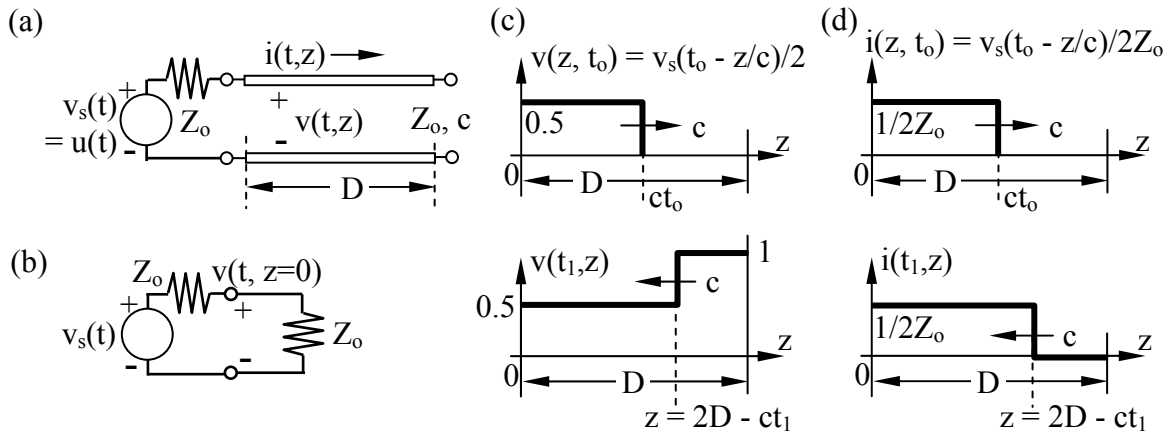


Figure 8.1.2 Step-function transients on a lossless transmission line.

Once the transient reaches the right-hand end, boundary conditions must again be satisfied, so there is a reflected voltage wave having $v_-(t, z=D) = \Gamma v_+(t, z=D)$, where $\Gamma = +1$, as given by (8.1.15) for $R_n \rightarrow \infty$. The total voltage on the line (8.1.9) is the sum of the forward and backward waves, each of value 0.5 volts, as illustrated in Figure 8.1.2(c) for $t = t_1 > D/c$. At t_1 the reflected voltage step is propagating leftward toward the source. The current at $t = t_1$ is plotted in Figure 8.1.2(d).

Although these voltage and current transients are most easily represented and understood graphically, they can also be derived and represented algebraically. For example, $v(t, z=0) = u(t)/2$ here, and therefore for $t < D/c$ we have $v(t, z) = v_+(t - z/c) = u(t - z/c)/2$. Note that if we translate an argument on one side of an equation, we must impose the same translation on the other; thus $v(t, 0) \rightarrow v(t - z/c)$ forces $u(t, 0) \rightarrow u(t - z/c)$. Once the wave reflects from the open circuit we have $v(z, t) = v_+(t - z/c) + v_-(t + z/c)$. At $z = D$ for $t < 3D/c$ the boundary condition at the open circuit requires $\Gamma = +1$, so $v_-(t + D/c) = v_+(t - D/c) = u(t - D/c)/2$. From $v_-(t + D/c)$ we can find the more general expression $v_-(t + z/c)$ simply by operating on their arguments: $v_-(t + z/c) = v_-(t + D/c - D/c + z/c) = u(t - 2D/c + z/c)$. The total voltage for $t < 2D/c$ is the sum of these forward and backward waves: $v(t, z) = [u(t - z/c) + u(t - 2D/c + z/c)]/2$. The same approach can represent line currents and also more complex examples.

⁴¹ We use the notation $u(t)$ to represent a *unit-step function* that is zero for $t < 0$, and unity for $t \geq 0$. A *unit impulse* is represented by $\delta(t)$, which is zero for all $|t| > \epsilon$ in the limit where $\epsilon \rightarrow 0$, and the integral of $\delta(t) \equiv 0$. $\int \delta(t) dt = u(t)$.

When the reflected wave arrives back at the source, $\Gamma = 0$ because this source is matched to the transmission line. In this special case there are no further reflections. Steady state is therefore one volt on the line everywhere, with $v_+ = v_- = 0.5$ in perpetuity. The total line current is the difference between the forward and backward wave (8.1.5), as plotted in Figure 8.1.2(d) for t_1 . The steady-state current is therefore zero. These steady state values correspond to $\omega \rightarrow 0$ and $\lambda \rightarrow \infty$, so the line is then much shorter than any wavelength of interest and can be considered static. We can easily see that an open-circuit line connected to a voltage source via any impedance at all will eventually assume the same voltage as the source, and the current will be zero, as it is here.

If the line were short-circuited at the right-hand end, then $\Gamma = -1$ and the voltage $v(z)$ at t_1 would resemble that of the current in Figure 8.1.2(d), with the values 0.5 and 0 volts, while the current $i(z)$ at t_1 would resemble that of the voltage in (c), with the values $0.5/Z_0$ and $1/Z_0$. The steady state values for voltage and current in this short-circuit case are zero and $1/Z_0$, respectively.

If the first transmission line were connected to a second passive infinite line of impedance Z_b , as illustrated in Figure 8.1.3(a), then the same computations would yield $v(t,z)$ and $i(t,z)$ on the first transmission line, where $R_n = Z_b/Z_0$. The solution on the second line follows from the boundary conditions: $v(t)$ and $i(t)$ are both continuous across the boundary. The resulting waveforms $v(t_1,z)$ and $i(t_1,z)$ at time $D/c < t_1 < 2D/c$ are plotted in Figure 8.1.3(b) for the case $R_n = 0.5$, so $\Gamma = -1/3$. In this case the current is increased by the reflection while the voltage is diminished. Independent of the incident waveform, the fraction of the incident power that is reflected is $(v_-/v_+)^2 = \Gamma^2$, where the reflection coefficient Γ is given by (8.1.15); the transmitted fraction is $1 - \Gamma^2$.

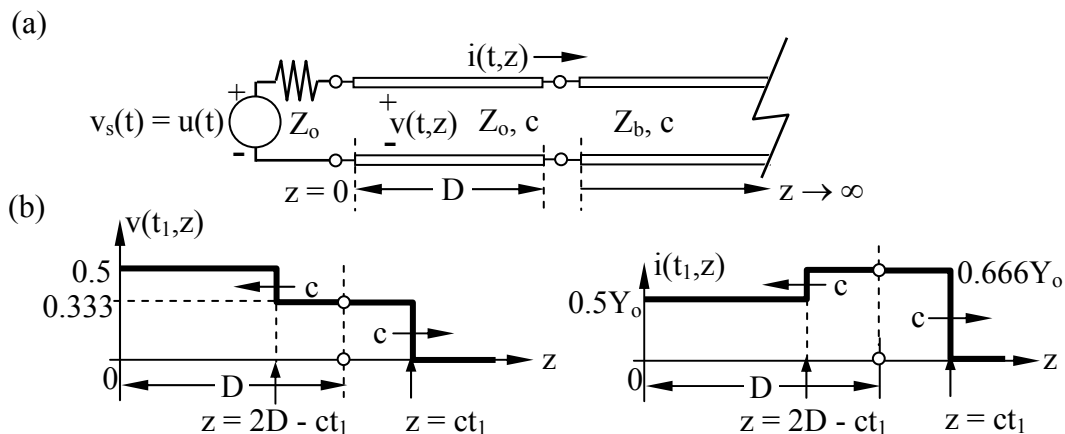


Figure 8.1.3 Step function incident upon a mismatched TEM line.

The principal consequence of this reflection phenomenon is that the voltage across a device may not be what was intended if there is an impedance mismatch between the TEM line and the device. This is an issue only when the line is sufficiently long that line delays are non-negligible

compared to circuit time constants (8.1.13). The analysis above is for linear resistive loads, but most loads are non-linear or reactive, and their treatment is discussed in Section 8.1.4.

8.1.3 Multiple reflections and reverberations

The reflected waves illustrated in Figures 8.1.2 and 8.1.3 eventually impact the source and may be reflected yet again. Since superposition applies if the sources and loads are linear, the contributions from each reflection can be separately determined and then added to yield the total voltage and current. That is, the reflected $v_-(t,z)$ will yield its own reflection at the source, and the fate of this reflection can be followed independently of the original forward wave. As usual when analyzing linear circuits, all sources are set to zero when determining the contribution of an independent source such as $v_-(t,z)$.

This paradigm is illustrated in Figure 8.1.4, which involves a unit-step current source driving an open-circuited TEM line that is characterized by Z_0 , c , and length D . Figures 8.1.4(a), (b), and (c) illustrate the circuit, the voltage at t_1 , and the current at t_1 , respectively, where $t_1 = D/2c$. The reflection coefficient $\Gamma = 1$ (8.1.15) for the open circuit at $z = D$, so the incident Z_0 -volt step is reflected positively, and the total voltage where they superimpose is $2Z_0$ volts, as illustrated in (d) for $t_2 = 1.5D/c$. The current at this moment is $Y_0[v_+(t - z/c) - v_-(t + z/c)]$, which is zero where the forward and reverse waves overlap, as illustrated in (e). When $v_-(t + z/c)$ is reflected from the left-hand end it sees $\Gamma = +1$ because, when using superposition, we consider the current source to be zero, corresponding to an open circuit. Thus an additional Z_0 volts, associated with v_{+2} , adds to v_{+1} and v_{-1} to yield a total of $3Z_0$ volts, as illustrated in (f) at t_3 ; the notation v_{+i} refers to the i th forward wave v_+ . This process continues indefinitely, with the voltage continuing to increase by Z_0 volts every D/c seconds until something breaks down. Voltage breakdowns are expected when current sources feed open circuits; the finite rate of voltage increase is related to the total capacitance of the TEM line.

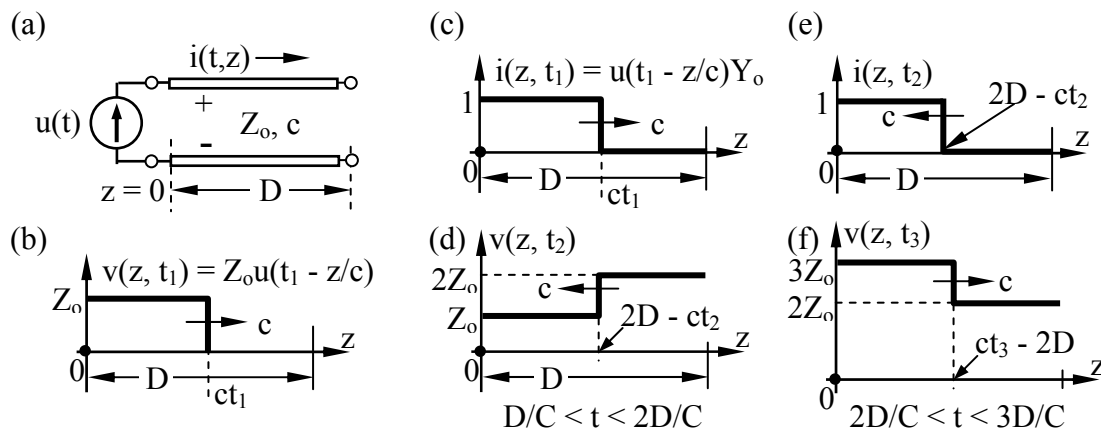


Figure 8.1.4 Transients for a current source driving an open-circuited TEM line.

The behavior of the current $i(t,z)$ is interesting too. Figure 8.1.4(c) illustrates how the one-ampere current from the current source propagates down the line at velocity c , and (e) shows how the “message” that the line is open-circuited is returned: the current is returning to zero.

When this left-moving wave is reflected at the left-hand end the current is again forced to be one ampere by the current source. Thus the current distribution (c) also applies to (f) at t_3 . This oscillation between one and zero amperes continues indefinitely, much like an unresolved argument between two people, each end of the line forcing the current to satisfy its own boundary conditions while that message propagates back and forth at velocity c .

Example 8.1B

A unit step voltage source $u(t)$ with no source resistance drives a short-circuited air-filled TEM line of length D and characteristic impedance $Z_0 = 1$ ohm. What current $i(t)$ flows through the short circuit at the end of this TEM line?

Solution: The unit step will propagate down the line, be reflected at the short circuit at $z = D$ where the reflection coefficient $\Gamma_D = -1$, and travel back to the voltage source at $z = 0$, which this transient sees as a short circuit ($\Delta v = 0$), also having $\Gamma_S = -1$. So, after a round-trip delay of $2D/c$, the voltage everywhere on the line is zero, after which a new step voltage travels down the line and superimposes on the first step voltage, thus adding a second step to the current $i(t, z=D)$. This process continues indefinitely as $i(t)$ steps in 1-ampere increments every $2D/C$ seconds monotonically toward infinity, which is the expected current when a voltage source is short-circuited. The effect of the line is simply to slow this result as the current and stored magnetic energy on the line build up. More precisely, $v(t, z=0) \equiv u(t) = v_{+1}(t, z=0)$. $v_{+1}(t, z=D) = u(t - D/c)$, so the TEM line presents an equivalent circuit at $z = D$ having Thevenin voltage $v_{Th} = 2v_{+1}(t, D) = 2u(t - D/c)$, and Thevenin impedance Z_0 ; this yields $i(t) = 2u(t - D/c)/Z_0$ for $t < 3D/c$. Therefore $v_{-1}(t, D) = \Gamma_D v_{+1}(t, D) = -u(t - D/c)$, so $v_{+2}(t, 0) = \Gamma_S v_{-1}(t, 0) = u(t - 2D/c)$. At $z = D$ this second step increases the Thevenin voltage by $2u(t - 3D/c)$ and increases the current by $2u(t - 3D/c)/Z_0$, where $Z_0 = 1$ ohm. Therefore $i(t) = \sum_{n=0}^{\infty} 2u(t - [2n+1]D/c)$.

8.1.4 Reflections by mnemonic or non-linear loads

Most junctions involve mnemonic⁴² or non-linear loads, where *mnemonic loads* are capacitors, inductors, or other energy storage devices that have characteristics depending on the past. *Non-linear loads* include diodes, transistors, and voltage- or current-dependent capacitors and inductors. In either case the response to arbitrary waveforms cannot be determined by the simple methods described in the previous section. However by simply replacing the transmission line by its equivalent circuit, the voltage and current can generally be easily found, first at the junction and then on the transmission line.

The equivalent circuit for an unexcited transmission line is simply a resistor of value Z_0 because the ratio $\Delta v/\Delta i$ for any excitation is always Z_0 . Determining the voltage across this Z_0 is generally straightforward even if the source driving the line contains capacitors, inductors,

⁴² Mnemonic means “involving memory”.

diodes, or similar devices. The forward-propagating wave voltage is simply the terminal voltage, as demonstrated in Figures 8.1.2–4.

The Thevenin equivalent circuit for an energized TEM line has a Thevenin voltage source V_{Th} in series with the Thevenin impedance of the line: $Z_{Th} = Z_o$. Note that the equivalent impedance for a TEM line is exactly Z_o , regardless of any loads on the line. The influence of the load at the far end of the line is manifest only in reflected waves that may propagate from it toward the observer, as discussed in the previous section.

The Thevenin equivalent voltage of any linear system is simply its open-circuit voltage. The open-circuit voltage of a transmission line is twice the amplitude of any incident voltage waveform because the reflection coefficient Γ for an open circuit is +1, which doubles the incidence voltage at the junction position z_J :

$$V_{Th}(t, z_J) = v_+(t, z_J) + v_-(t, z_J) = 2v_+(t, z_J) \quad (8.1.17)$$

The procedure for analyzing a TEM line terminated by any load at $z = z_J$ is then to: 1) solve for the wave $v_+(t - z/c)$ traveling toward the load of interest, 2) set $V_{Th} = 2v_+(t - z_J/c)$ and $Z_{Th} = Z_o$, 3) solve for the terminal voltage $v(t, z_J)$, 4) solve for $v_-(t, z_J)$, and 5) find $v_-(t + z/c)$, where we define z as increasing toward the load:

$$v_-(t, z_J) = v(t, z_J) - v_+(t, z_J) \equiv v_-(t + [z_J/c]) \quad (8.1.18)$$

$$v_-(t + z/c) = v_-(t + [(z - z_J)/c], z_J) \quad (\text{wave reflected by load}) \quad (8.1.19)$$

Equation (8.1.19) says $v_-(t + z/c)$ is simply the $v_-(t, z_J)$ given by (8.1.18), but delayed by $(z_J - z)/c$.

This procedure is best demonstrated by a simple example. Figure 8.1.5(a) illustrates a TEM line driven by a matched unit step voltage source and terminated with a capacitor C . This voltage step, reduced by a factor of two by the voltage divider, propagates toward the capacitor at velocity c , as illustrated in (b). The capacitor sees the Thevenin equivalent circuit illustrated in (c); it consists of Z_o in series with a Thevenin voltage source that is twice v_+ , where $v_+(t, D)$ is a 0.5-volt step delayed by the propagation time D/c . Therefore $V_{Th} = u(t - D/c)$, as illustrated in Figure 8.1.5(c) and (d). The solution to the circuit problem of (c) is the junction voltage $v_J(t)$ plotted in (e); it rises exponentially toward its 1-volt asymptote with a time constant $\tau = Z_o C$ seconds.

To solve for $v_-(t, z_J)$ we subtract $v_+(t, z_J)$ from $v_J(t)$, as shown in (8.1.18) and illustrated in (f); this then yields $v_-(t + z/c)$ using (8.1.19). The total voltage $v(t_1, z)$ on the line at time $D/c < t_1 < 2D/c$ is plotted in (g) and is the sum of $v_+(t - z/c)$, which is 0.5 volts, and $v_-(t + z/c)$. The corresponding current $i(t_1, z)$ is plotted in (h) and equals Y_o times the difference between the forward and reverse voltage waves, as given by (8.1.5). When $v_-(t, z)$ arrives at the source, it can be treated just as such waves were treated in Section 8.1.3. In this case the source is matched, so there are no further reflections.

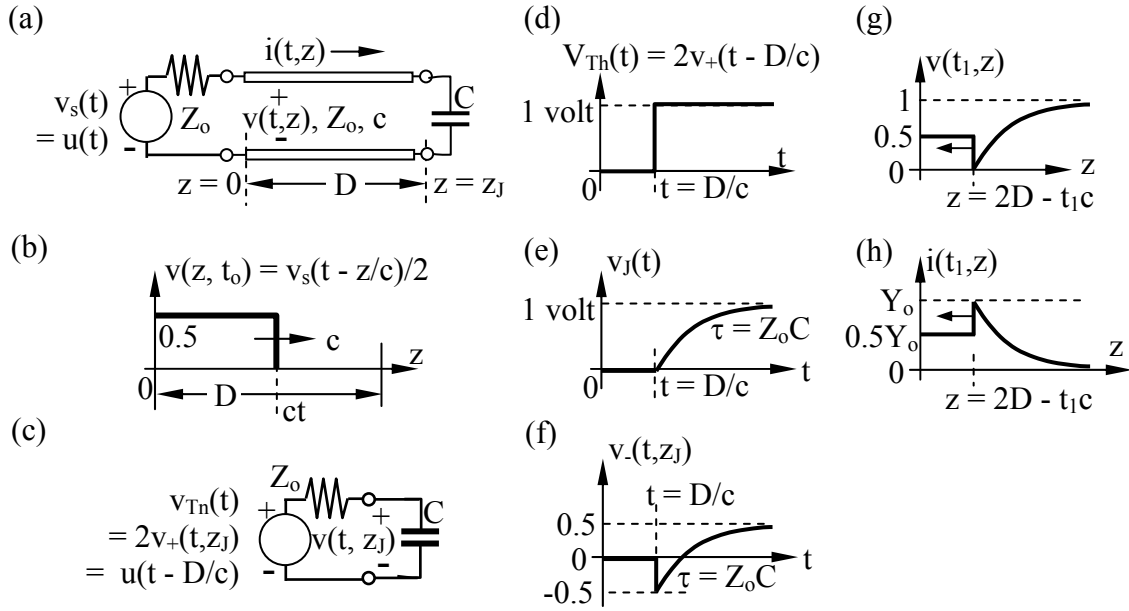


Figure 8.1.5 Transient voltages and currents on a capacitor-terminated TEM line.

Most digital circuits are non-linear, so this same technique is often used to determine the waveforms on longer TEM lines. Consider the circuit and ramp-pulse voltage source illustrated in Figure 8.1.6(a) and (b).

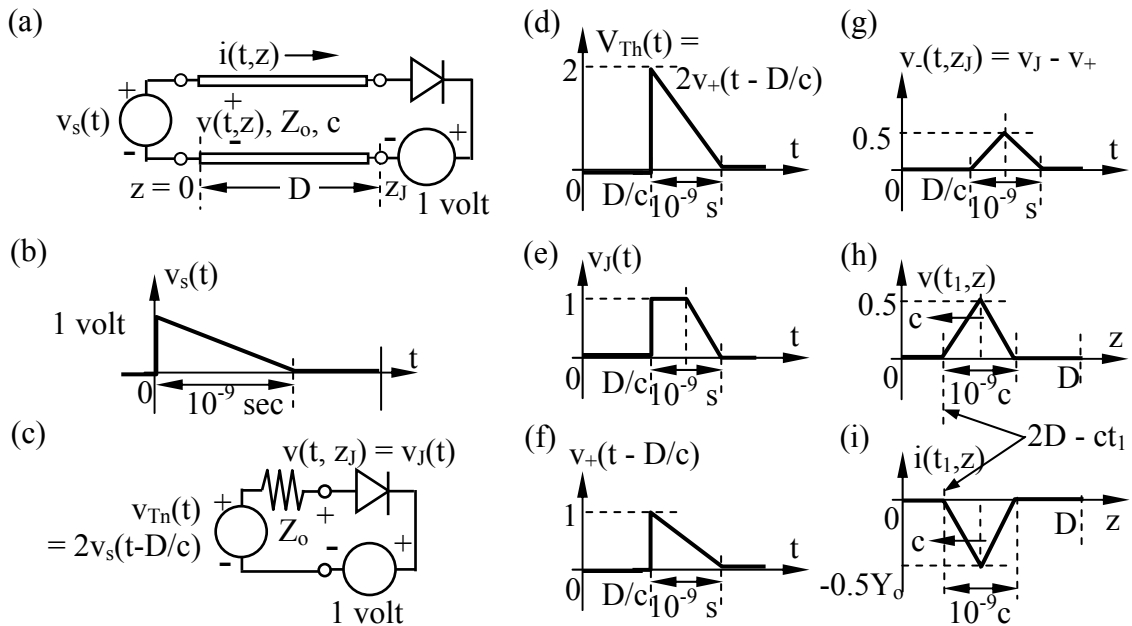


Figure 8.1.6 Transient TEM waveforms produced by reflection from a non-linear load.

In this case there is no source resistance (an arbitrary choice), so the full value of the source voltage appears across the TEM line. Part (c) shows the equivalent circuit of the transmission

line driving the load, which consists of a back-biased diode. The Thevenin voltage $V_{Th}(t, z_J) = 2v_+(t, z_J)$ is plotted in (d), the resulting junction voltage $v_J(t)$ is plotted in (e), $v_+(t, z_J)$ is plotted in (f), and $v_-(t, z_J) = v_J(t, z_J) - v_+(t, z_J)$ is plotted in (g). The line voltage and currents at $D/c < t_1 < 2D/c$ are plotted in (h) and (i), respectively. Note that these reflected waveforms do not resemble the incident waveform.

Example 8.1C

If the circuit illustrated in Figure 8.1.5(a) were terminated by L instead of C, what would be $v(t, D)$, $v_-(t, D)$, and $v(t, 0)$?

Solution: Figures 8.1.5(a–d) still apply, except that L replaces C. The one-volt Thevenin step voltage at $z = D$ in series with the Thevenin line impedance Z_0 yields a voltage $v(t, D)$ across the inductor of $u(t - D/c)e^{-tL/R}$. Therefore $v_-(t, D) = v(t, D) - v_+(t, D) = u(t - D/c)e^{-(t - D/c)L/R} - 0.5u(t - D/c)$ and, since there are no further reflections at the matched load at $z = 0$, it follows that $v(t, 0) = 0.5u(t) + u(t - 2D/c)e^{-(t - 2D/c)L/R} - 0.5u(t - 2D/c)$.

8.1.5 Initial conditions and transient creation

Often transmission lines have an initial voltage and current that is interrupted in some way, producing transients. For example, a charged TEM line at rest may have a switch thrown at one end that suddenly connects it to a load, or disconnects it; such a switch could be located in the middle of a line too, either in series or parallel. The solution method has two main steps: 1) determine $v_+(t, z)$ and $v_-(t, z)$ at $t = 0$, before the change occurs, and 2) solve for the subsequent behavior of the forward and backward moving waves for the given network configuration.

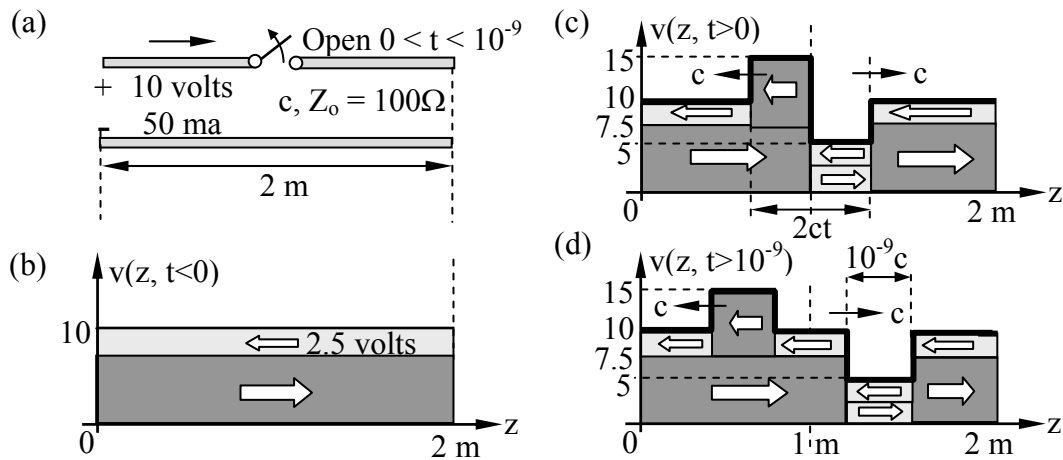


Figure 8.1.7 Transients induced by momentarily open-circuited active TEM line.

The simple example of Figure 8.1.7 illustrates the method. Assume an air-filled 2-meter long 100-ohm TEM line is feeding a 200-ohm load R with $I = 50$ milliamperes, when suddenly at $t = 0$ the line is open-circuited at $z = 1$ meters for 10^{-9} seconds, after which it returns to normal. What are the voltage and current on the line as a result of this temporary event?

Using the method suggested above, we first solve for the forward and backward waves prior to $t = 0$; the current $i(t < 0, z)$ is given as $I = 50$ milliamperes, and the voltage $v(t < 0, z) = IR$ is 0.05×200 ohms = 10 volts. Note that in steady state Z_o does not affect $v(t < 0, z)$. We know from (8.1.4) and (8.1.5) that:

$$v(z,t) = v_+(z - ct) + v_-(z + ct) \quad (8.1.20)$$

$$i(z,t) = Y_o [v_+(z - ct) - v_-(z + ct)] \quad (8.1.21)$$

Solving these two equations for v_+ and v_- yields:

$$v_+(z - ct) = [v(z,t) + Z_o i(z,t)]/2 = [10 + 5]/2 = 7.5 \text{ volts} \quad (8.1.22)$$

$$v_-(z + ct) = [v(z,t) - Z_o i(z,t)]/2 = [10 - 5]/2 = 2.5 \text{ volts} \quad (8.1.23)$$

These two voltages are shown in Figure 8.1.7(b), 7.5 volts for the forward wave and 2.5 volts for the reflected wave; this is consistent with the given 50-ma current.

When the switch opens at $t = 0$ for 10^{-9} seconds, it momentarily interrupts both v_+ and v_- , which see an open circuit at the switch and $\Gamma = +1$. Therefore in (c) we see 7.5 volts reflected back to the left from the switch, and 2.5 volts reflected back toward the right. At distances closer to the switch than ct [m] we therefore see 15 volts to the left and 5 volts to the right; this zone is propagating outward at velocity c . When the switch closes again, these mid-line reflections cease and the voltages and currents return to normal as the two transient pulses of 15 and 5 volts continue to propagate toward the two ends of the line, as shown in (d), where they might be reflected further.

The currents associated with Figure 8.1.7(d) can easily be surmised using (8.1.21). The effects of the switch are only felt for that brief 10^{-9} -second interval, and otherwise the current on the line is the original 50 ma. In the brief interval when the switch was open the current was forced to zero, and so zero-current pulses of duration 10^{-9} seconds propagate away from the switch in both directions.

Example 8.1D

A 100-ohm air-filled TEM line of length D is feeding 1 ampere to a 50-ohm load when it is momentarily short-circuited in its middle for a time $T < D/2c$. What are $v_+(z - ct)$ and $v_-(z + ct)$ prior to the short circuit, and during it?

Solution: For $t < 0$, $\Gamma = v_-(D - ct)/v_+(D + ct) = (Z_n - 1)/(Z_n + 1) = -0.5/1.5 = -1/3$ where $Z_n = 50/100$. Since the line voltage $v(z,t)$ equals the current i times the load resistance ($v = 50$ volts), it follows that $v_+ + v_- = 2v_+/3 = 50$, and therefore $v_+ = v_+(z - ct) = 75$ volts, and $v_-(z + ct) = -25$. During the short circuit the voltage within a distance $d = ct$ of the short is altered. On the source side the short circuit reflects $v_- = -v_+ = -75$,

so the total voltage ($v_+ + v_-$) within ct meters of the short circuit is zero, and on the load side $v_+ = -v_- = 25$ is reflected, so the total voltage is again zero. The currents left and right of the short are different, however, because the original $v_+ \neq v_-$, and $i_+ = v_+/Z_0$. Therefore, on the source side near the short circuit, $i = (v_+ - \Gamma v_-)/Z_0 = 2v_+/Z_0 = 2 \times 75/50 = 3$ [A]. On the load side near the short circuit, $i = -2 \times 25/50 = -1$ [A].

8.2 *Limits posed by devices and wires*

8.2.1 Introduction to device models

Most devices combine conducting elements with semiconductors, insulators, and air in a complex structure that stores, switches, or transforms energy at rates limited by characteristic time constants governed by Maxwell's equations and kinematics. For example, in *vacuum tubes* electrons are boiled into vacuum by a hot negatively charged *cathode* with small fractions of an electron volt of energy⁴³. Such tubes switch state only so fast as the free electrons can cross the vacuum toward the positively charged *anode*, and only as fast as permitted by the RL or RC circuit time constants that control the voltages accelerating or retarding the electrons. The same physical limits also apply to most semiconductor devices, as suggested in Section 8.2.4, although sometimes quantum effects introduce non-classical behavior, as illustrated in Section 12.3.1 for laser devices.

Design of vacuum tubes for use above ~ 100 MHz was difficult because high voltages and very small dimensions were required to shorten the electron transit time to fractions of a radio frequency (RF) cycle. The wires connecting the cathode, anode, and any grids to external circuits also contributed inductance that limited speed. Trade-offs were required. For example, as the cathode-anode gap was diminished to shorten electron transit times, the capacitance C between the cathode and anode increased together with delays associated with their RC time constant τ_{RC} . Exactly the same physical issues of gap length, capacitance, and τ_{RC} arise in most semiconductor devices. The kinematics of electrons in vacuum was discussed in Sections 5.1.2–3, and the behavior of simple RL and RC circuits was discussed in Section 3.5.1.

8.2.2 Semiconductor device models

One simple example illustrates typical sources of lag in semiconductor devices. Both pnp and npn *transistors* are composed of p-n junctions that contribute device-related delay. Field-effect transistors exhibit related lags. Figures 8.2.1(a) and (b) present a DC p-n junction i-v characteristic and a circuit model that exhibits approximately correct delay characteristics for the case where low-loss metal wires are used for interconnecting devices.

⁴³ A thermal energy E_0 of one electron volt ($\cong 1.6 \times 10^{-19}$ [J]) corresponds to a temperature T of $\sim 11,600$ K, where $E_0 = kT$ and k is Boltzmann's constant: $k \cong 1.38 \times 10^{-23}$ [J/°K]. Thus a red-hot cathode at ~ 1000 K would boil off free electrons with thermal energies of ~ 0.1 e.v.

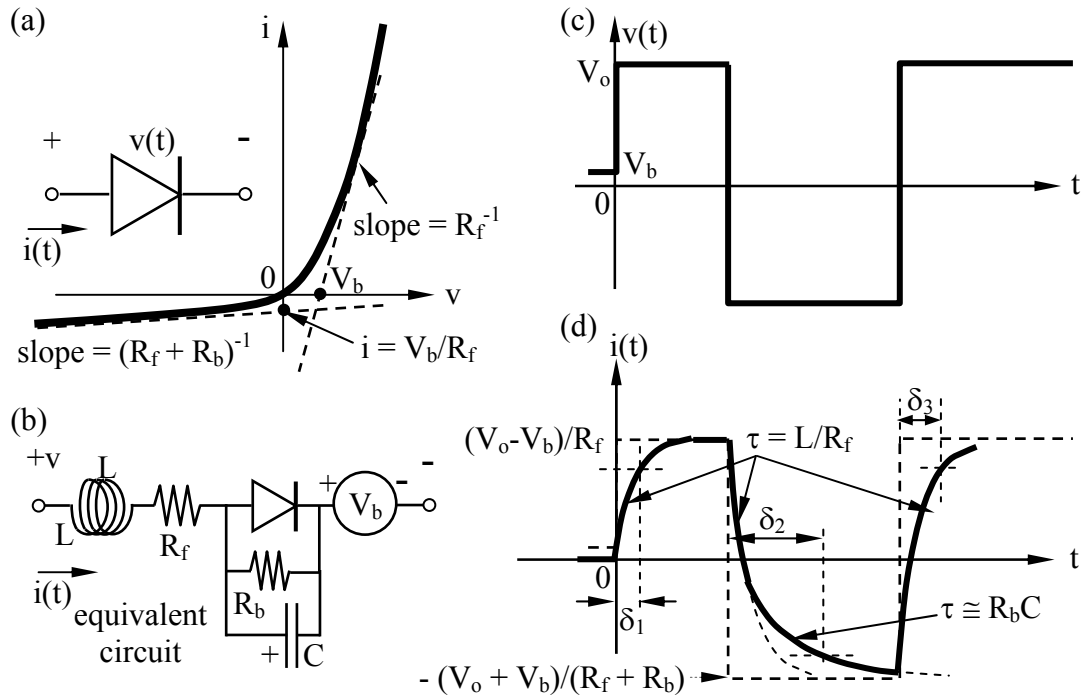


Figure 8.2.1 Circuit model for switching delays at p-n junctions.

When the ideal diode is forward biased the forward-bias resistance R_f determines the slope of the i - v characteristic. When it is back-biased beyond $\sim V_b$ the ideal diode becomes an open circuit so the junction capacitance C becomes important and the back-bias resistance $R_f + R_b \cong R_b$ determines the slope. C arises because of the charge-free depletion region that exists in back-biased diodes, an explanation of which is given in Section 8.2.4. C decreases as the back-bias voltage increases because the gap width d increases and $C \cong \epsilon A/d$ (3.1.10). The bias voltage V_b in the equivalent circuit is related to the band gap between the valence and conduction bands in the semiconductor, and is ~ 1 volt for silicon (see Section 8.2.4 for more discussion). The inductance L arises primarily from wires leading externally, and is discussed further in Section 8.2.3 for printed circuits, and estimated in Section 3.3.2 for isolated wires (3.3.17).

Figure 8.2.1(c) represents a typical test voltage across a p-n junction applied as suggested in Figure 8.2.1(a). It begins at that bias voltage $v(t) = V_b$ (~ 1 volt in silicon), for which the equivalent circuit in Figure 8.2.1(b) conducts no current and the capacitor is discharged. At $t = 0_+$, $v(t) \rightarrow V_o$ and the current $i(t)$ increases toward $(V_o - V_b)/R_f$ with zero incremental resistance offered by the diode and voltage source, so the time constant τ is L/R_f seconds (3.5.10), as illustrated in Figure 8.2.1(d). The capacitor remains uncharged. Section 3.5.1 discusses time constants for simple circuits. As a result of diversion of energy into the inductor, the current i does not reach levels sufficient to trigger the next circuit element until $t \cong \delta_1$, which is the lag time.

This time constant δ_1 can be easily estimated. For example, a p-n junction might be attached to wires of length $D = 0.001$ [m] and radius $r_o = 10^{-6}$ [m], and have a forward-biased resistance R_f of ~ 1 ohm. In this case (3.3.17) yields the wire inductance:

$$L \cong (\mu_o D / 16\pi) \ln(D/r_o) \cong 1.7 \times 10^{-10} \text{ [Henries]} \quad (8.2.1)$$

Thus $\delta_1 \cong \tau = L/R = 1.7 \times 10^{-10}$ seconds, so the diode might handle a maximum frequency of $\sim R/2\pi L$ Hz, or ~ 1 GHz. More conservatively the diode might be used at clock frequencies below ~ 0.2 GHz. Modern computers employ shorter wires and smaller R_f in order to work faster. The circuit model in Figure 8.2.1(b) does not include the capacitance between the wire and the substrate because it is negligibly small here relative to the effects of L .

When the test voltage $v(t)$ then goes negative, the ideal diode in Figure 8.2.1(b) continues to conduct until the current through the inductor decays to zero with the same L/R_f time constant. The current then begins charging C (as the depletion layer is cleared of charge) with a time constant $\sim R_b C$ that delays the current response for a total of $\sim \delta_2$ seconds. Note that for illustrative purposes the current scale for negative $i(t)$ in (d) has been expanded by a very large factor (R_b/R_f) relative to the scale for positive $i(t)$.

When the test voltage then returns to V_o from its strong negative value, it must first discharge C (re-populate the depletion layer with charge) before the ideal diode in Figure 8.2.1(c) closes, introducing a time constant of $\sim R_f C$ that we can estimate. If the capacitance C corresponds to a depletion layer of thickness $d \cong 10^{-7}$, area $A \cong 10^{-11}$ [m²], and permittivity $\epsilon \cong \sim 10\epsilon_o$, then (3.1.10) yields $C = \epsilon A/d \cong 10 \times 8.8 \times 10^{-12} \times 10^{-11}/10^{-7} \cong 9 \times 10^{-15}$ [F]. This yields $R_f C \cong 10^{-14}$ [s] $\ll L/R_f$, so L/R_f would dominate the entire transition, resulting in a total lag of $\delta_3 \cong \delta_1$ seconds. In reality $i(t)$ in this RLC circuit would ring at $\omega \cong (LC)^{-0.5}$ radians per second as $i(t)$ and the ringing decay toward the asymptote $i \cong (V_o - V_b)/R_f$.

In most bi-polar transistor circuits using metal wires it is L/R_f that controls the maximum clock speed for the system, which is limited by the slowest junction and the most inductive wires in the entire integrated circuit. In MOS integrated circuits, however, the resistivity of the polysilicon or diffusion layers used for conductors is sufficiently high that the wire inductance is often no longer controlling, as discussed in Section 8.3.1. Wire inductance is most easily reduced by using shorter wider wires, which also reduces wire resistance. Longer paths can be accommodated by using matched TEM lines, as discussed in Section 8.1.

8.2.3 Quasistatic wire models

The lag time for the p-n junction of Figure 8.2.1 was dominated by L/R_f , where L originated in the wires connected to the junction. The effects of depletion layer capacitance C were negligible in comparison for the assumed device parameters. In this section we examine the effects of wire capacitance and cross-section in limiting clock or signal frequencies.

In most integrated circuits the wires are planar and deposited on top of an insulating layer located over a conducting ground plane, as suggested in Figure 8.2.2.

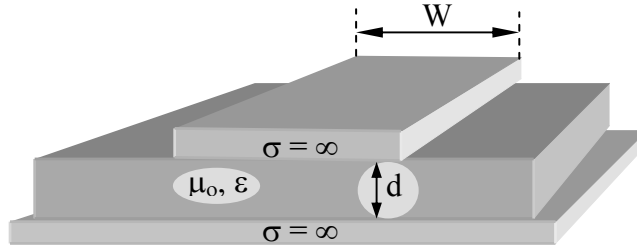


Figure 8.2.2 Idealized model for printed or integrated circuit wire.

The capacitance and inductance per unit length are C' and L' , respectively, which follow from (3.1.10) and (3.2.6) under the assumption that fringing fields are negligible:

$$C' = \epsilon W/d \quad [\text{F m}^{-1}] \quad (8.2.2)$$

$$L' = \mu d/W \quad [\text{H m}^{-1}] \quad (8.2.3)$$

$$L'C' = \mu\epsilon \quad (8.2.4)$$

Printed circuit wires with width $W \cong 1 \text{ mm}$ and length $D \cong 3 \text{ cm}$ printed over dielectrics with thickness $d \cong 1\text{-mm}$ and permittivity $\epsilon \cong 4\epsilon_0$ would add capacitance C and inductance L to the connected device, where:

$$C = \epsilon WD/d = 9 \times 8.8 \times 10^{-12} \times 10^{-3} \times 0.03/10^{-3} = 2.4 \times 10^{-12} \quad [\text{F}] \quad (8.2.5)$$

$$L = \mu dD/W = 1.2 \times 10^{-6} \times 10^{-3} \times 0.03/10^{-3} = 3.6 \times 10^{-8} \quad [\text{H}] \quad (8.2.6)$$

These values combine with nominal one-ohm forward-bias resistances of p-n junctions to yield the time constants $L/R = 3.6 \times 10^{-8}$ seconds, and $RC = 2.4 \times 10^{-12}$ seconds. Again the limit is posed by inductance. Such printed circuit boards would be limited to frequencies $f \leq \sim 1/2\pi\tau \cong 4 \text{ MHz}$.

The numbers cited here are not nearly so important as the notion that interconnections can strongly limit frequencies of operation and circuit utility. The quasistatic analysis above is valid because the physical dimensions here are much smaller than the shortest wavelength (at $f = 4 \text{ MHz}$): $\lambda = c/f \cong 700 \text{ meters}$ in air or $\sim 230 \text{ meters}$ in a dielectric with $\epsilon = 9\epsilon_0$.

We have previously ignored wire resistance in comparison to the nominal one-ohm resistance of forward-biased p-n junctions. If the printed wires above are $d = 10 \text{ microns}$ thick and have the conductivity σ of copper or aluminum, then their resistance is:

$$R = D/dW\sigma = 0.03/(10^{-5} \times 10^{-3} \times 5 \times 10^7) = 0.06 \quad [\text{ohms}] \quad (8.2.7)$$

Since some semiconductor devices have forward resistances much less than this, wires are sometimes made thicker or wider to compensate. Wider printed wires also have lower inductance [see (8.2.6)]. Width and thickness are particularly important for power supply wires, which often carry large currents.

If (8.2.7) is modified to represent wires on integrated circuits where the dimensions in microns are length $D = 100$, thickness $d = 0.1$, and width $W = 1$, then $R = 20$ ohms and far exceeds typical forward p-n junction resistances. Limiting D to 20 microns while increasing d to 0.4 and W to 2 microns would lower R to 0.5 ohms. The resistivities of polysilicon and diffusion layers often used for conductors can be 1000 times larger, posing even greater challenges. Clearly wire resistance is another major constraint for IC circuit design as higher operating frequencies are sought.

Example 8.2A

A certain integrated circuit device having forward resistance $R_f = 0.1$ ohms is fed by a polysilicon conductor that is 0.2 microns wide and thick, 2 microns long, and supported 0.1 micron above the ground plane by a dielectric having $\epsilon = 10\epsilon_0$. The conductivity of the polysilicon wire is $\sim 5 \times 10^4 \text{ S m}^{-1}$. What limits the switching time constant τ of this device?

Solution: R , L , and C for the conductor can be found from (8.2.7), (8.2.6), and (8.2.5), respectively. $R = D/dW\sigma = 2 \times 10^{-6} / [(0.2 \times 10^{-7})^2 5 \times 10^4] = 1000$, $L \cong \mu d D / W = 1.26 \times 10^{-6} \times 10^{-7} \times 2 \times 10^{-6} / (0.2 \times 10^{-6}) = 1.26 \times 10^{-12} \text{ [Hy]}$. $C = \epsilon W D / d = 8.85 \times 10^{-11} \times 0.2 \times 10^{-6} \times 2 \times 10^{-6} / 10^{-7} = 3.54 \times 10^{-16} \text{ [F]}$. $RC \cong 3.54 \times 10^{-13} \text{ [s]}$, $L/R = 1.26 \times 10^{-15} \text{ [s]}$, and $(LC)^{0.5} = 2.11 \times 10^{-14} \text{ [seconds/radian]}$, so RC limits the switching time. If metal substituted for polysilicon, then LC would pose the limit here.

8.2.4 Semiconductors and idealized p-n junctions

Among the most commonly used *semiconductors* are silicon (Si), germanium (Ge), gallium arsenide (GAs), and indium phosphide (InP). Semiconductors at low temperatures are insulators since all electrons are trapped in the immediate vicinity of their host atoms. The periodic atomic spacing of crystalline semiconductors permits electrons of sufficient energy to propagate freely without scattering, however. Diodes and transistors therefore exhibit conductivities that depend on the applied voltages and resulting electron energy distributions. The response times of these devices are determined by electron kinematics and the response times of the circuits and structures determining voltages and field strengths within the device.

The quantum mechanical explanation of such electron movement invokes the wave nature of electrons, which is governed by the Schroedinger wave equation (not explained here, although it is similar to the electromagnetic wave equation). The consequence is that semiconductors can be characterized by an *energy diagram* that shows possible electron energy states as a function of position in the z direction, as illustrated in Figure 8.2.3(a). At low temperatures all electrons occupy energy states in the lower *valence band*, corresponding to bound orbits around atoms. However a second *conduction band* of possible energy states occupied by freely moving electrons exists at higher energies separated from the valence band by an *energy gap* E_g that

varies with material, but is ~ 1 e.v. for silicon. For example, a photon with energy $hf = E_p > E_g$ can excite a bound electron in the valence band to a higher energy state in the conduction band where that electron can move freely and conduct electricity. In fact this photo-excitation mechanism is often used in semiconductor photo detectors to measure the intensity of light.

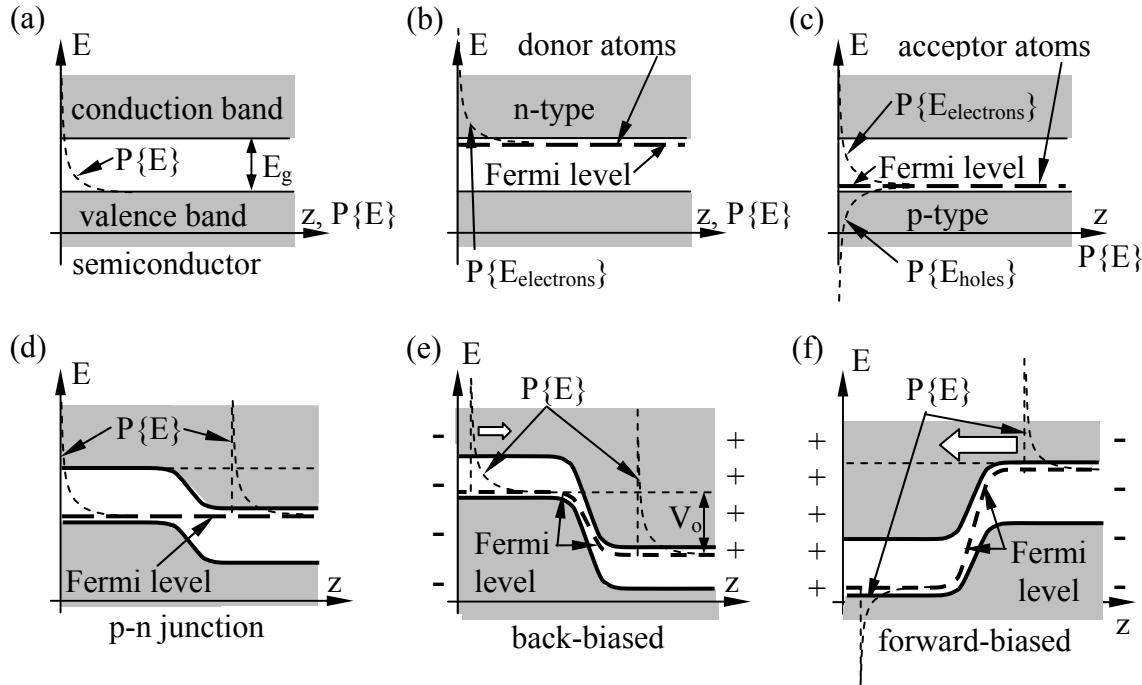


Figure 8.2.3 Energy diagrams for p and n semiconductors and p-n junctions.

The probability that an unbound electron in thermal equilibrium at temperature T has energy E is governed by the *Boltzmann distribution* $P\{E\} = e^{-E/kT}/kT$, where the *Boltzmann constant* $k = 1.38 \times 10^{-23}$ and one electron volt is 1.6×10^{-19} [J]. Therefore *thermal excitation* will randomly place a few free electrons in the conduction band since $P\{E > E_g\} > 0$. However this provides only extremely limited conductivity because a gap of ~ 1 e.v. corresponds to a temperature of $E/k \cong 11,600\text{K}$, much greater than room temperature.⁴⁴

To boost the conductivity of semiconductors a small fraction of doping atoms are added that either easily release one electron (called *donor atoms*), or that easily capture an extra electron (*acceptor atoms*). These atoms assume energy levels that are just below the conduction band edge (donor atoms) or just above the valence band edge (acceptor atoms), as illustrated in Figure 8.2.3(b) and (c), respectively. These energy gaps are quite small so a significant fraction of the donor and acceptor atoms are typically ionized at room temperature.

The probability that an electron has sufficient energy to leap a gap E_a is the integral of the Boltzmann probability distribution from E_a to $E \rightarrow \infty$, and E_a is sufficiently small that this integral can approach unity for some dopant atoms. The base energy for the Boltzmann distribution is

⁴⁴ The Fermi level of an undoped semiconductor is midway between the valence and conduction bands, so \sim one-half electron volt is actually sufficient to produce excitation, although far more electrons exist in the valence band itself.

the *Fermi level*; electrons fill the available energy levels starting with the lowest and ending with the highest being (on average) at the Fermi level. The Fermi level usually lies very close to the energy level associated with the donor or acceptor atoms, as illustrated in Figure 8.2.3(b–f). Since holes are positively charged, their exponential Boltzmann distribution appears inverted on the energy diagram, as illustrated in (c).

For each ionized donor atom there is an electron in the conduction band contributing to conductivity. For each negatively ionized acceptor atom there is a vacated positively charged “hole” left behind. An adjacent electron can easily jump to this hole, effectively moving the hole location to the space vacated by the jumping electron; in this fashion holes can migrate rapidly and provide nearly the same conductivity as electrons in the conduction band. Semiconductors doped with donor atoms so that free electrons dominate the conductivity are *n-type semiconductors* (negative carriers dominate), while holes dominate the conductivity of *p-type semiconductors* doped with acceptors (positive carriers dominate). Some semiconductors are doped to produce both types of carriers. The conductivity of homogeneous semiconductors is proportional to the number of charge carriers, which is controlled primarily by doping density and temperature.

If a p-n junction is short-circuited, the Fermi level is the same throughout as shown in Figure 8.2.3(d). Therefore the tails of the Boltzmann distributions on both sides of the junction are based at the same energy, so there is no net flow of current through the circuit.

When the junction is back-biased by V_0 volts as illustrated in (e), the Fermi level is depressed correspondingly. The dominant current comes from the tail of the Boltzmann distribution of electrons in the conduction band of the p-type semiconductor; these few electrons will be pulled to the positive terminal and are indicated by the small white arrow. Some holes thermally created in the n-type valence band may also contribute slightly. Because the carriers come from the high tail of the Boltzmann thermal distribution, the reverse current in a p-n junction is strongly dependent upon temperature and can be used as a thermometer; it is not very dependent upon voltage once the voltage is sufficiently negative. When a p-n junction is back-biased, the electric field pulls back most low-energy free electrons into the n-type semiconductor, and pulls the holes into the p-type semiconductor, leaving a carrier-free layer, called a charge-depletion region, that acts like a capacitor. Larger values of V_0 yield larger gaps and smaller capacitance.

When the junction is forward-biased by V_0 volts, as illustrated in (f), the charge-free layer disappears and the current flow is dominated by the much greater fraction of electrons excited into the conduction band in the n-type semiconductor because almost all of them will be pulled by the applied electric field across the junction before recombining with a positive ion. This flow of electrons (opposite to current flow) is indicated by the larger white arrow, and is proportional to that fraction of $P\{E\}$ that lies beyond the small energy gap separating the Fermi level and the lower edge of the conduction band. Holes in the valence band can also contribute significantly to this current. The integral over energy of the exponential probability distribution $P\{E\}$ above threshold $E_g - v$ is another exponential for $0 < v < E_g$, which is proportional to the population of conducting electrons, and which approximates the $i(v)$ relation for a p-n junction illustrated in Figure 8.2.1(a) for $v > 0$.

Transistors are semiconductor devices configured so that the number of carriers (electrons plus holes) available in a junction to support conductivity is controlled 1) by the number injected into the junction by a p-n interface biased so as to inject the desired number (e.g., as is done in p-n-p or n-p-n transistors), or 2) by the carriers present that have not been pulled to one side or trapped by electric fields (e.g., field-effect transistors). In general, small bias currents and voltages can thereby control the current flowing across much larger voltage gaps with power amplification factors of 100 or more. Although the range of device designs is very large, most can be understood semi-classically as suggested above, without the full quantum mechanical descriptions needed for precise characterization.

The response time of p-n junctions and transistors is usually determined by either the RC, RL, or LC time constants that limit the rise and fall times of voltages and currents applied to the device terminals, or by the field relaxation time ϵ/σ (4.3.3) of the semiconductor material within the device itself. In extremely fast devices the response time sometimes is $\tau \cong D/v$, where D is the junction dimension [m] of interest, and v is the velocity of light ($v = [\mu\epsilon]^{-0.5}$) or of the transiting electrons ($v = \int a dt$, $f = ma = eE$).

Although these physical models for semiconductor junctions are relatively primitive, they do approximately explain most phenomena.

Example 8.2B

What are the approximate temperature dependences of the currents flowing in forward- and reverse-biased p-n junctions?

Solution: If the bias voltage exceeds the gap voltage, and kT is large compared to the energy gap between the donor level and the conduction band, then essentially all donors are ionized and further temperature changes have little effect on forward biased p-n junctions; see Figure 8.2.3(f). The carrier concentration in reverse-biased diodes is proportional to $\int_{E_0+E_g}^{\infty} e^{-E/kT} dE$, and therefore to T [Figure 8.2.3(e)].

8.3 Distortions due to loss and dispersion

8.3.1 Lossy transmission lines

In most electronic systems transmission line loss is a concern because business strategy generally dictates reducing wire diameters and costs until such issues arise. For example, the polysilicon often used for conductors in integrated silicon devices has noticeable resistance.

The *TEM circuit model* of Figure 8.3.1 incorporates two types of loss. The series resistance R per meter arises from the finite conductivity of the wires, while the parallel conductance G per meter arises from leakage currents flowing between the wires through the medium separating them.

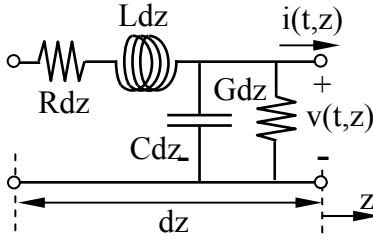


Figure 8.3.1 Distributed circuit model for lossy TEM transmission lines.

When these lossy elements are included, we obtain the *telegraphers' equations*:

$$dv/dz = -Ri - L di/dt \quad (\text{telegraphers' equation}) \quad (8.3.1)$$

$$di/dz = -Gv - C dv/dt \quad (\text{telegraphers' equation}) \quad (8.3.2)$$

If the wires are resistive, then current flowing through them introduces longitudinal electric fields E_z , violating the TEM assumption: $E_z = H_z = 0$. Since rigorous solution of Maxwell's equations for the non-TEM case is challenging, the telegraphers' equations are often used instead if the loss is modest. The same problem does not arise with G because it does not violate the TEM assumption, as shown in Section 7.1.3. Since propagation in such *lossy TEM lines* is frequency dependent, the telegraphers' equations (8.3.1–2) and their solutions are generally expressed using complex notation⁴⁵:

$$d\underline{V}(z)/dz = -(R + j\omega L)\underline{I}(z) \quad (\text{telegraphers' equation}) \quad (8.3.3)$$

$$d\underline{I}(z)/dz = -(G + j\omega C)\underline{V}(z) \quad (\text{telegraphers' equation}) \quad (8.3.4)$$

Differentiating (8.3.3) with respect to z , and substituting (8.3.4) for $d\underline{I}(z)/dz$ yields the wave equation for lossy TEM lines, where the sign of k^2 is chosen so that k is real, consistent with the lossless solutions discussed earlier in Section 7.1.2:

$$d^2\underline{V}(z)/dz^2 = (R + j\omega L)(G + j\omega C)\underline{V}(z) = -\underline{k}^2\underline{V}(z) \quad (\text{wave equation}) \quad (8.3.5)$$

$$\underline{k} = [-(R + j\omega L)(G + j\omega C)]^{0.5} = k' - jk'' \quad (\text{TEM propagation constant}) \quad (8.3.6)$$

Since the second derivative of $\underline{V}(z)$ equals a constant times itself, it must be expressible as the sum of exponentials that have this property:

$$\underline{V}(z) = \underline{V}_+ e^{-jkz} + \underline{V}_- e^{+jkz} \quad (\text{TEM voltage solution}) \quad (8.3.7)$$

⁴⁵ Complex notation is discussed in Section 2.3.2 and Appendix B. In general, $v(t) = \text{Re}\{\underline{V}e^{j\omega t}\}$, where $\text{Re}\{\bullet e^{j\omega t}\}$ is omitted from equations.

Differentiating (8.3.7) with respect to z and substituting the result in (8.3.3) yields both $\underline{I}(z)$ and \underline{Y}_o :

$$\underline{I}(z) = \underline{Y}_o (\underline{V}_+ e^{-jkz} - \underline{V}_- e^{+jkz}) \quad (\text{TEM current solution}) \quad (8.3.8)$$

$$\underline{Z}_o = \frac{1}{\underline{Y}_o} = \sqrt{\frac{R + j\omega L}{G + j\omega C}} \quad (\text{characteristic impedance}) \quad (8.3.9)$$

When $R = G = 0$, (8.3.9) reduces to the well known result $Z_o = (L/C)^{0.5}$.

Thus two new properties emerge when TEM lines are dissipative: 1) because \underline{k} is complex and a non-linear function of frequency, waves are attenuated and dispersed as they propagate in a frequency-dependent manner, and 2) \underline{Z}_o is complex and frequency dependent. Both k' and k'' (8.3.6) are functions of frequency, so signals propagating on lossy lines change shape, partly because different frequency components propagate and decay differently. The resulting attenuation and dispersion are discussed in Sections 8.3.1 and 8.3.2, respectively. Reflections are affected at junctions by losses, and also are attenuated with distance so the impedance of a lossy line $\underline{Z}(z) \rightarrow \underline{Z}_o$ regardless of load as $\underline{V}_-(z)$ becomes negligible. Reflections by junctions involving lossy lines are simply analyzed by replacing Z_o by a complex impedance \underline{Z}_o in the expressions developed in Section 7.2 for lossless lines.

Waves propagating only in the $+z$ direction obey (8.3.7), which becomes:

$$\underline{V}(z) = \underline{V}_+ e^{-jkz} = \underline{V}_+ e^{-jk'z} e^{-k''z} \quad (\text{decaying propagating wave}) \quad (8.3.10)$$

One combination of R , L , C , and G is particularly interesting because it results in zero dispersion and a frequency-independent decay that does not distort waveforms. We may discover this combination by evaluating \underline{k} using (8.3.6):

$$\underline{k} = [-(R + j\omega L)(G + j\omega C)]^{0.5} = \omega \{LC [1 - j(R/\omega L)] [1 - j(G/\omega C)]\}^{0.5} \quad (8.3.11)$$

It follows from (8.3.11) that if $R/L = G/C$, then the phase velocity ($v_p = \omega/k' = [LC]^{-0.5}$) and the decay rate ($k'' = R[C/L]^{0.5}$) are both frequency independent:

$$\underline{k} = (LC)^{0.5} (\omega - jR/L) = k' - jk'' \quad (\text{distortionless line}) \quad (8.3.12)$$

The ability to avoid signal distortion due to frequency-dependent absorption was first exploited by telephone companies who added small inductors periodically in series with their longer phone lines in order to reduce R/L so that it balanced G/C ; the result was called a

distortionless line, and the coils are called *Pupin coils* after their inventor⁴⁶. The consequences of dispersion are explored in Section 8.3.2.

Another limit is sometimes of interest when the effects of R dominate those of ωL . This occurs, for example, in resistive polysilicon or diffusion lines in integrated circuits, which may be approximately modeled by eliminating L and G from Figure 8.3.1. Then \underline{k} (8.3.11) becomes:

$$\underline{k} \cong (-j\omega RC)^{0.5} = (\omega RC/2)^{0.5} - j(\omega RC/2)^{0.5} = k' - jk'' \quad (8.3.13)$$

The square root of $-j$ was chosen to correspond to a decaying wave rather than to exponential growth. The phase and group velocities for this line are the same:

$$v_p = \omega/k' = (2\omega/RC)^{0.5} \text{ [m s}^{-1}\text{]} \quad (8.3.14)$$

$$v_g = (\partial k'/\partial \omega)^{-1} = 2(\omega/RC)^{0.5} \text{ [m s}^{-1}\text{]} \quad (8.3.15)$$

Although it is not easy to relate these frequency-dependent velocities to delays in digital circuits, they demonstrate that such delays exist and express their dependence on R and C. That is, larger line time constants RC lower pulse velocities and increase delays. Such lines are best used when they are short compared to the shortest wavelength of interest, $D < \lambda = v_p/f_{\max} = 2\pi(2/rc\omega_{\max})^{0.5}$. In polysilicon lines $\lambda_{\min} \cong 1 \text{ mm}$ for $\omega_{\max} = 10^{10}$. The response to arbitrary waveform excitation can be computed by: 1) Fourier transforming the signal, 2) propagating each frequency component as dictated by (8.3.13), and then 3) reconstructing the signal at the new location with an inverse Fourier transform. Typical values for R and C in metal, *polysilicon*, and diffusion lines are presented in Table 8.3.1, and correspond to velocities much less than c. The costs of these three options for forming conductors are unequal and must also be considered when designing fast integrated circuits.

Table 8.3.1 Resistance and capacitance per meter for typical integrated circuit lines.

Parameter	Metal	Polysilicon	Diffusion
R [$\text{M}\Omega \text{ m}^{-1}$]	0.06	50	50
C [nF m^{-1}]	0.1	0.2	1

Provided R is not so large compared to ωL that the TEM approximation is invalid because of strong longitudinal electric fields, then the power dissipated is:

$$P_d = (R|I|^2 + G|V|^2)/2 \text{ [W m}^{-1}\text{]} \quad (8.3.16)$$

⁴⁶ Pupin coils had to be inserted at least every $\lambda/10$ meters in order to avoid additional distortions, but the shortest λ for telephone voice signals is $\sim c/f = 3 \times 10^8/3000 = 100 \text{ km}$.

8.3.2 Dispersive transmission lines

Different frequency components propagate at different velocities on *dispersive transmission lines*. The nature and consequences of dispersion are discussed further in Section 9.5.2. Consider first a square-wave computer clock pulse at F Hz propagating along a dispersive TEM line. The Fourier transform of this signal has its fundamental at F Hz, with odd harmonics at $3F$, $5F$, etc., each of which has its own phase velocity, as suggested in Figure 8.3.2.

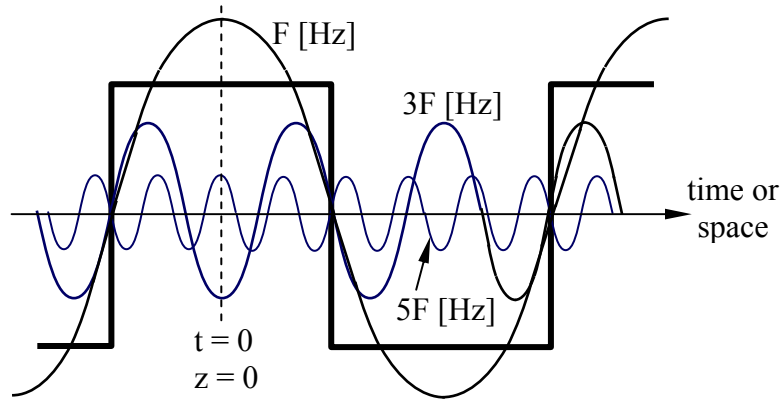


Figure 8.3.2 Square wave and its constituent sinusoids.

Significant pulse distortion occurs if a strong harmonic is shifted as much as $\sim 90^\circ$ relative to the fundamental. To determine the relative phase shift between fundamental and harmonic we can first multiply the difference in phase velocity at F and $3F$, e.g., $v_{pF} - v_{p3F}$, by the propagation time T of interest. This yields the spatial offset between these two harmonics, which we might limit to $\lambda/4$ for $3F$. That is, we might expect significant distortion over a transmission line of length $D = v_p T$ meters if:

$$(v_{pF} - v_{p3F})T > \sim \lambda_{3F}/4 = v_{p3F}/(4 \times 3F) \quad (8.3.17)$$

There is a similar limit to the propagation distance of narrowband pulse signals before *waveform distortion* becomes unacceptable. Digital communications systems commonly use narrowband pulses $s(t)$ for both wireless and cable signaling. For example, the square wave in Figure 8.3.2 could also represent the amplitude envelope $A(t) = \sum_i a_i \cos \omega_i t$ of an underlying sinusoid $\cos \omega_0 t$, where $\omega_0 \gg \omega_{i>0}$ and together they occupy a narrow bandwidth. That is:

$$s(t) = (\cos \omega_0 t) \sum_{i=1} a_i \cos \omega_i t = 0.5 \sum_i a_i \left\{ \cos [(\omega_0 + \omega_i)t] + \cos [(\omega_0 - \omega_i)t] \right\} \quad (8.3.18)$$

Since each frequency $\omega_0 \pm \omega_i$ propagates at a slightly different phase velocity, a narrowband pulse will also distort when a strong harmonic is $\sim \lambda/4$ out of phase relative to the original wave envelope, which is much larger than $\lambda = 2\pi c/\omega_0$ for narrowband signals. Some applications are more sensitive to dispersive distortion than others; for example, distorted digital signals can be

generally be regenerated distortion free, while analog signals require inverse distortion, which is often uneconomic.

Distortion of narrowband signals is usually computed in terms of the *group velocity* v_g , which is the velocity of propagation for the waveform envelope and equals the velocity of energy or information, which can never exceed c , the velocity of light in vacuum. The sine wave that characterizes the average frequency of a narrowband pulse propagates at the *phase velocity* v_p , which can be greater or less than c . Narrowband pulse signals (e.g., digitally modulated sinusoids) distort when the accumulated difference Δ in the envelope propagation distances between the high- and low-frequency end of the signal spectrum differs by more than a small fraction of the minimum pulse width W [m] (e.g., the length of a zero or one). Since the difference in group velocity across the bandwidth B [Hz] is $(\partial v_g / \partial f)B$ [m/s], and the pulse travel time is D/v_g , where D is propagation distance, it follows that the difference in envelope propagation distance across the band is:

$$\Delta = \frac{\partial v_g}{\partial f} \frac{BD}{v_g} \text{ [m]} \quad (8.3.19)$$

Since the minimum pulse width W is $\sim v_g/B$ [m], the requirement that $D \ll W$ implies that the maximum distortion-free propagation distance D is:

$$D \ll \left(\frac{v_g}{B} \right)^2 \left(\frac{\partial v_g}{\partial f} \right)^{-1} \quad (8.3.20)$$

Group and phase velocity are discussed further in Section 9.5.2 and their effect on distortion is explored in Section 12.2.2.

Example 8.3A

Typical 50-ohm coaxial cables for home distribution of television and internet signals have series resistance $R \cong 0.02(f_{\text{MHz}})^{0.5}$ ohms m^{-1} . Assume $\epsilon = 4\epsilon_0$, $\mu = \mu_0$. How far can signals propagate before attenuating 60 dB?

Solution: Since conductivity $G \cong 0$, (8.3.11) says $\underline{k} = \omega[LC(1 - jR/\omega L)]^{0.5} = k' + jk''$. The imaginary part of \underline{k} corresponds to exponential decay. For $R \ll \omega L$, $\underline{k} \cong \omega(LC)^{0.5}(1 - jR/\omega L)^{0.5}$, so $k'' = -(\omega RC)^{0.5}$. To find C we note the phase velocity $v = (\mu_0 4\epsilon_0)^{-0.5} = (LC)^{-0.5} = c/2 \cong 1.5 \times 10^8$ [m s^{-1}], and $Z_0 = 50 = (L/C)^{0.5}$. Therefore $C = 2/cZ_0 = 2/(3 \times 10^8 \times 50) \cong 1.33 \times 10^{-10}$ [F]. Thus at 100 MHz, $k'' = -(\omega RC)^{0.5} = -(2\pi 10^8 \times 0.2 \times 1.33 \times 10^{-10})^{0.5} \cong -0.017$. Since power decays as $e^{-2k''z}$, 60 dB corresponds to $e^{-2k''z} = 10^{-6}$, so $z = -\ln(10^{-6})/2k'' = 406$ meters. At 100 MHz the approximation $R \ll \omega L$ is quite valid. As coaxial cable systems boost data rates and their maximum frequency above 100-200 MHz, the increased attenuation requires amplifiers at intervals so short as to motivate switching to optical fibers that can propagate signals hundred of kilometers without amplification.

MIT OpenCourseWare
<http://ocw.mit.edu>

6.013 Electromagnetics and Applications
Spring 2009

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.