Samuel H. Russ

Signal Integrity

Applied Electromagnetics and Professional Practice

Second Edition

Springer

ENTROPIC EXHIBIT 2011
DISH v. Entropic
IPR2024-00373

Signal Integrity

Samuel H. Russ

Signal Integrity

Applied Electromagnetics and Professional Practice

Second Edition

Springer

Samuel H. Russ
University of South Alabama
Mobile, AL, USA

ISBN 978-3-030-86926-7 ISBN 978-3-030-86927-4 (eBook)
https://doi.org/10.1007/978-3-030-86927-4

© Springer Nature Switzerland AG 2016, 2022
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Contents

1 The Basics – Charge, Energy, Time, and Distance 1
Background and Objectives 1
Getting Back to Basics 1
Consequences of the Basics 3
Distance, Time, Speed, and c 4
The Effect of Dielectrics 6
Rise Time and Bandlimit 7
Knee Frequency 8
Lumped Versus Distributed 10
Combining Rise Times 12
Appendix 14

2 Circuit Boards 17
Background and Objectives 17
What Is a Circuit Board? 17
How Are Circuit Boards Made? 19
Components and Component Packages 22
How Are Circuit Boards Used to Make a Product? 24
Appendix 27

3 Gates, Packaging, and Boards: Properties and Modeling 29
Background and Objectives 29
What Limits What Is Possible? 29
Power and Heat Dissipation 31
Summarizing the Effects 34
Simplest Gate Model 34
Electrical Modeling 35
The Modeling Process 37
The Limits of Modeling 40
Appendix 41

4 Circuit Elements: Resistance, Capacitance, and Inductance 43
Background and Objectives 43
Reviewing the Review 43
What Is Resistance? 44
What Is Capacitance? 48
Finding Capacitance 50
What Is Inductance? 53
Calculating Inductance 57
Inductance and Return Current 58
Inductance and the Skin Effect 59
Appendix 61

5 Ground Bounce and Ringing 65
Background and Objectives 65
The Role of Inductance 65
What Is Ground Bounce? 66
Results of Ground Bounce 69
Minimizing Ground Bounce 72
What Is Ringing? 73
Results of Ringing 75
Minimizing Ringing 77
So Where Is Ringing Seen – And Not Seen – Today? 77
Appendix 78

6 Distributed Analysis: Transmission Lines and Z_0 81
Background and Objectives 81
Where Can You See a Transmission Line? 81
Transmission Line: The View from the Inside 82
Estimating Z_0 in Common Situations 86
Approximate Model of a Transmission Line 87
Appendix 88

7 Transmission Lines: Reflections and Termination 91
Background and Objectives 91
The Left Hand and the Right Hand: Reflection and Transmission 91
Terminating Impedance 96
Departures from the Ideal 99
Fixing It When It's Broke 101
Appendix 102

8 Lossy Transmission Lines 107
Background and Objectives 107
What Is Attenuation and How Is It Measured? 107
Frequency-Dependent Attenuation 109
Dielectric Loss 109

How Does Dielectric Loss Affect ε? 110
How Does Dielectric Loss Affect Signal Propagation? 112
Appendix 116

9 Differential Signaling 117
Background and Objectives 117
What Is Differential Signaling and How Does It Help? 117
What Is Z_{diff}? 119
Obstacles: Clock Jitter and Common-Mode Signals 121
Termination Strategies Revisited 123
Appendix 125

10 Crosstalk 127
Background and Objectives 127
What Is Crosstalk? 127
Near-End Versus Far-End Crosstalk 129
Estimating and Reducing Crosstalk 133
Diagnosing Crosstalk 134
Appendix 135

11 Power Distribution Network: Frequency Domain Analysis 139
Background and Objectives 139
The Power Distribution Network 139
Power Supplies and Resonance 141
Design Strategy: The Big Picture 144
Design Strategy: The Details 145
Design Strategy: The Role of Simulation 149
Selecting Bypass Capacitors 151
Appendix 153

12 EMI/EMC: Design and Susceptibility 155
Background and Objectives 155
EMI/EMC 155
Circuit-Board Design 157
Engineering Notebook – Mounting Holes, Cable Shields, and Grounding 159
Chassis 161
Cabling 163
Legal Certification: Standards and Test Requirements 164
Engineering Notebook – Get Down Here Right Away 165
Appendix 166

13 Electrostatic Discharge 167
Background and Objectives 167
What Is Electrostatic Discharge? 167
Where Can ESD Occur? 168

ESD and Lightning Standards and Testing 169
Components to Manage ESD 170
Design Techniques to Manage ESD 173
Appendix 173

14 Clocks, Jitter, and Eye Diagrams 175
Background and Objectives 175
Clock Jitter and Clock Skew 175
Sources of Jitter 178
Clock Sources 178
Layout to Reduce Clock Jitter 180
Engineer's Notebook – The Crystal Layout 181
Eye Diagrams 182
Appendix 185

15 Circuit-Board Design Process 187
Background and Objectives 187
How Are Circuit Boards Designed? 187
Component Selection 188
Engineer's Notebook – More About Capacitors 188
Schematic 191
Bill of Materials 193
Layout 193
Design Files 196
The Board Is Dead. Now What? 196
Appendix 197

16 Circuit-Board Attacks and Security 199
Background and Objectives 199
Circuit Board Recap 199
First Indication of Trouble 199
Review of Circuit-Board Design, Fabrication, and Assembly 200
Example Attacks 201
Review of Processor-Based Design 207
Taxonomy of Attacks 209
Conclusion 212
Appendix 212
References 213

17 Testing, Debugging, DFX, and Quality Management 215
Background and Objectives 215
Testing a Board: The Oscilloscope 215
Qualifying a Design for Mass Production 217
Transitioning to Mass Production 218
Appendix 223

18 Commercial and Legal Implications, Project Management, and Risk Mitigation 225
Background and Objectives 225
Legal Aspects of Design 225
Commercial Aspects of Design 227
Engineering Notebook – Predicting What Comes Next 229
Project Management 230
Risk Mitigation 231
Engineering Careers 232
Appendix 233

Index 235

Chapter 6
Distributed Analysis: Transmission Lines and Z_0

Background and Objectives

This chapter introduces distributed analysis, the art of analyzing circuits and structures that are so physically large that the voltage on wires is not necessarily the same at each point. The fact that the structure is so large means that parts of the circuit are "unaware" of what other parts are doing, and the results are, to say the least, counter-intuitive. Wires turn into resistors and excesses in current cause voltage waves to crash back and forth. When you finish this chapter, you should be able to:

- Explain what a transmission line is
- Explain the derivation of characteristic impedance (Z_0) and the physical consequences of its existence
- Understand the approximate model of a transmission line

Where Can You See a Transmission Line?

Maxwell's equations predicted a spooky world where energy could radiate outward from a point, move at a certain, fixed speed, and cause voltages to appear some distance away. Heinrich Hertz built one of the first systems to exploit this strangeness, creating the radio.

Albert Einstein used Maxwell's notion of the speed of light to construct a model of the universe that is relativistic. In effect, Einstein's theory outlines a universe where nothing is actually "simultaneous." A logic gate driving an output cannot "see" what happens at the other end instantaneously. For small wires or for slowly changing voltages, the net effect is nearly instantaneous, and this is the familiar "lumped" model of circuits. We are entering the domain of electrical engineering where this is no longer true – changes to circuits do not occur instantaneously. Like the old saw "the left hand does not know what the right hand is doing," a distributed

© Springer Nature Switzerland AG 2022

S. H. Russ, *Signal Integrity*, https://doi.org/10.1007/978-3-030-86927-4_6

circuit is one where the source of a changing voltage (or current) does not "know" what is happening at the destination. So now there is not only "spooky action at a distance," but there is also the fact that the results of the action at the distance are not known until later.

Getting back to radio... A radio uses a transmission line. In the case of a radio, the transmission line is the air between the two antennas (transmitting and receiving). A cable or DSL modem uses a transmission line to send and receive data, with the transmission line being, respectively, the coaxial cable and twisted-pair telephone line.

What do these examples have in common? First, the signal energy is propagating in a dielectric, an electrical insulator. Second, there is a relatively low loss of signal in the dielectric. This means that the signal tends to arrive intact at the other end, which makes it much easier to receive.

There are some differences. For example, the radio signal gets weaker farther from the antenna, but only because the energy is spreading out in space (i.e., the dielectric is getting larger as the signal moves away from the antenna). In the case of twisted-pair or coaxial cable, the signal energy is confined to the structure and so gets weaker at a much slower rate.

So how do transmission lines work? To understand that, consider a small piece of coaxial cable...

Transmission Line: The View from the Inside

Let's start with 1 m of ordinary RG-11 coaxial cable, commonly used for cable television or to carry satellite signals from the dish to the satellite set-top box. The term "coaxial" refers to the fact that it has a small center conductor surrounded by a cylinder of insulating material, which is in turn surrounded by a cylindrically shaped conductor. The two conductors are coaxial (get it?) and so they make a capacitor. If you measured the capacitance of 1 m of RG-11, it turns out to be 53.1 pF. If you measure its inductance, it is 298.7 nH. (This information is posted in datasheets for the cable.)

It is worth noting that if you double the length to two meters, both the capacitance and inductance go up by a factor of two. So we can say that RG-11 has 53.1 pF/m of capacitance and 298.7 nH/m of inductance. Stated symbolically, $C_L = 53.1$ pF/m and $L_L = 298.7$ nH/m.

What happens when one launches a very fast digital signal, say a 50 ps rising edge, down a long piece of RG-11 cable?

First, before the rising edge is launched, the voltage of the center conductor is less than the voltage of the outer conductor. Because of capacitance, the difference in voltage results in an E field between them.

Second, the rising edge "flips" the E field because, after the rising edge passes, the voltage of the center conductor is now higher. To be specific, and this is very important, *it is the reversal of the E field that travels at nearly the speed of light.*

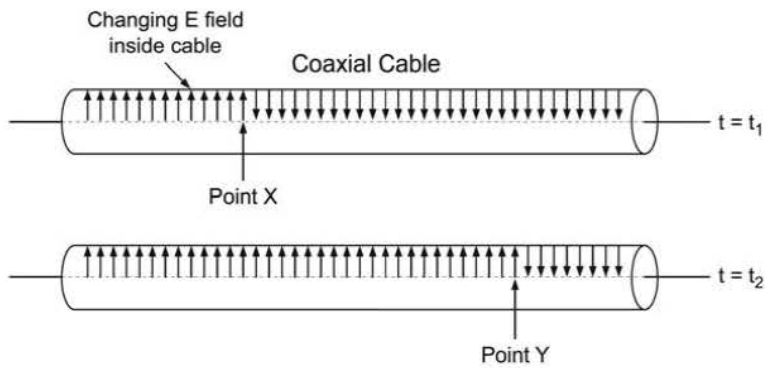
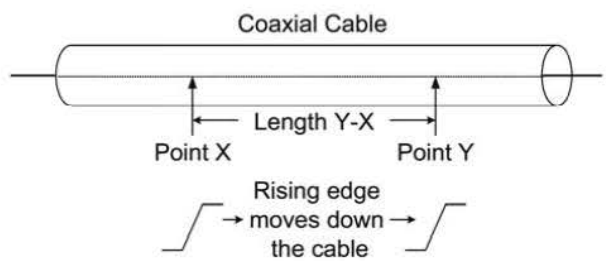


Fig. 6.1 Motion of a rising edge down a transmission line. Note that the edge passes point X at time t_1 and point Y at time t_2

Fig. 6.2 Derivation of Z_0



(To review, if the permittivity is higher than that of free space, light moves more slowly, and the E change in the E field moves at a slower speed.) It is exactly and specifically *not* the electrons that are moving that fast.

This is illustrated in Fig. 6.1.

The rising edge (more accurately, the changing E field) has to move from point X on the cable to point Y on the cable. The E field can only change if real electrical charge is applied to each region of the cable. To be clear, the E field can only change if the electrical charges on that region of the capacitor (the one formed between the two conductors at the point of the rising edge) are reversed, and the charge can only be reversed by applying electric current to redistribute the charges. Charge is applied at a certain rate, which of course is known as current. Since the changing E field is constantly moving down the cable, more current has to be put into the cable in order for the changing E field to continue down the line. One can then calculate the amount of current needed to enable the rising edge to race down the line. (Also see Fig. 6.2 and follow along.)

First, how much electric charge is needed to charge the piece of cable between point X and point Y? That is a simple calculation: charge is the product of the capacitance and the voltage. The capacitance is the length between points X and Y multiplied by the capacitance per unit length. The voltage is the voltage swing from a logic 0 to a logic 1, which is denoted "V".

$$Q = (Y - X)C_L V \tag{6.1}$$

Second, how long does it take the wave to move from point X to point Y? It is simple once one recalls Maxwell's formula for velocity, $1/\sqrt{\mu\varepsilon}$. The speed of propagation of the wave down the line is $1/\sqrt{C_L L_L}$ and so the time interval is the distance (e.g., in meters) divided by the speed (e.g., in meters per second).

$$t = (Y - X)\sqrt{C_L L_L} \tag{6.2}$$

(By the way, notice that $\sqrt{C_L L_L}$ is the *reciprocal* of the speed, so if speed is measured in cm/ps, the reciprocal is in ps/cm. Recall that this is a handy way to measure speed because it lets one easily calculate the delay given the length.)

So what is the current?

$$I = \frac{Q}{t} = \frac{(Y - X)C_L V}{(Y - X)\sqrt{C_L L_L}} = V\sqrt{\frac{C_L}{L_L}} \tag{6.3}$$

So the amount of current needed to create a rising edge of voltage difference V traveling at a velocity of $1/\sqrt{C_L L_L}$ is *linearly proportional to the voltage difference*. A linear relationship between current and voltage is a resistor. Rearranging Eq. 6.3, we get

$$V = I\sqrt{\frac{L_L}{C_L}} \equiv IZ_0 \tag{6.4}$$

and so

$$\sqrt{\frac{L_L}{C_L}} \equiv Z_0 \tag{6.5}$$

Again, not to put too fine a point on it, Z_0 is *the ratio of the desired voltage swing to the current needed to create that swing on a distributed line with capacitance per length C_L and inductance per length L_L*. In other words, it is the ratio of voltage swing to current on a distributed line.

The last part (distributed line) is important and easily overlooked. If you look back at Chap. 1, the distributed/lumped boundary involved the length of a rising edge relative to the length of the wire carrying the rising edge. The significance here is that Z_0 is only relevant while the rising edge is racing down the line. When viewed over a longer period of time, the wire turns back into a lumped wire with zero resistance.

Z_0 is a very easily misunderstood concept, so notice the ingredients. First, as just explained, the signal must be operating in a distributed regime (meaning that the length of the connection must be greater than the length of the signal transition).

Second, the signal must propagate in a medium with the capacitance per length and inductance per length as designated. Third, the capacitance and inductance dictate the speed at which the signal travels, which in turn dictates the rate at which electric charge must be pumped into the line. In other words, it dictates the current (rate of electric charge) that is needed. Fourth, electric charge is being pumped into the line in order to charge the capacitance per unit length, and it must continue to be pumped in if the signal is to continue down the line. The rate of charge is equivalent to current. Fifth, notice how units of distance cancel out.

To the gate that is driving a fast-rising edge down the line, the long piece of coaxial cable *looks like a real resistor of value* Z_0 *to ground*. The "wire" has turned into a "resistor". The resistance is purely real – not imaginary – which is highly counter-intuitive. Even though the wave propagation is due to inductance and capacitance, the current that is being sourced is charging a physical capacitance and so involves the transfer of actual electric charge. It is a real impedance, and Z_0 *is purely real*.

Example 6.1

What is the characteristic impedance of RG-11?

$Z_0 = \sqrt{\frac{298.7n}{53.1p}} = \sqrt{\frac{298700}{53.1}} = \sqrt{5625} = 75\ \Omega$. This is why RG-11 is referred to as "75-Ω cable". (Notice the units conversion needed on the inductance in order to perform division.)

As if this was not strange enough, consider this: what, exactly, is moving down the line and where is it located?

The "what exactly": According to Maxwell's equations, the "thing" that is moving down the line at speed $1/\sqrt{C_L L_L}$ is an electromagnetic wave, a changing E field. It is specifically NOT the electrons themselves – they actually move at a much slower pace. Just like waves can ripple across an ocean with the water remaining basically motionless, electromagnetic waves can move much, much faster than electrons.

The "where": Remember that waves in a transmission line move *in the dielectric*. (If you are hazy on this, remember how your cell phone works!) In our RG-11 example, the rising edge (again, the changing E field) is NOT located in the center conductor; it is located in the dielectric between the two conductors.

On a circuit board, this change becomes crucial to the thought process that accompanies good design. The rising edge is *located in the fiberglass between the signal and the ground plane*. Any imperfection in the fiberglass *or* the ground plane will distort the signal. At high frequencies (more accurately, at a wire length greater than $\ell/6$), the signal is no longer in the wire – it is in the *combination* of the wire, the insulator, and the ground plane. And so any perturbation to this tangled web of signal propagation (any change to the geometry of the wire, insulator, or ground plane) results in signal distortion.

There is one aspect of this newly interconnected world that must be emphasized forcefully. The signal and its return current are carried in the dielectric, and so the signal trace and ground plane are now *equally responsible* for carrying the signal.

"Ground" (i.e., the ground plane) is no longer a DC signal reference or part of the power infrastructure; it is the second conductor of a matched pair that is needed to carry the signal.

Once logic gate has launched a fast-rising edge down a long piece of RG-11 coaxial cable, how long does it take to reach the end? One can look back at Eq. 6.2. Defining T_0 as the one-way transmission time and l as the length of the cable, we have

$$T_0 = l\sqrt{C_L L_L} \tag{6.6}$$

Example 6.2
What is one-way transmission time down 2.5 meters of RG-11?

$T_0 = 2.5\sqrt{53.1p \times 298.7n} = 2.5 \times 3.98n = 9.96$ ns. Thc spccd of light in a vacuum is 33.3 ps/cm so light traverses 2.5 m of vacuum in $33.3 \times 250 = 8325$ ps= 8.325 ns. So electromagnetic waves travel about 20% slower in RG-11 than in free spacc.

Estimating Z_0 in Common Situations

Z_0 is a function of a transmission line's inductance and capacitance per unit length. To calculate Z_0 precisely for a transmission line usually requires sophisticated electromagnetic simulation, but fortunately there are some formulas that estimate Z_0 for commonly encountered circuit-board structures. These are approximations (actually, they are industry-standard approximations), but fortunately "approximately-correct" termination results in "nearly-perfect" signal transmission. It is a situation where being off by a few percent will usually work correctly.

The formulas were published by IPC, a standards-making body for the printed circuit board and surface-mount industry, and unfortunately are based on English units. First, for microstrip (which is a conductor on the top layer of a circuit board over a ground plane separated by a dielectric):

$$Z_0 = \frac{87}{\sqrt{\varepsilon_r + 1.41}} \ln\left(\frac{5.98h}{0.8w + t}\right) \tag{6.7}$$

$$T_p = 85\sqrt{0.475\varepsilon_r + 0.67} \tag{6.8}$$

Z_0 = Characteristic Impedance (ohms)
T_p = Delay in ps/in
w = conductor width (mils)
t = conductor thickness (mils)
h = spacing from conductor to plane (mils)

Second, for stripline, which is a conductor centered vertically between two planes and completely surrounded by a single, uniform dielectric:

$$Z_0 = \frac{60}{\sqrt{\varepsilon_r}} \ln\left(\frac{1.9b}{0.8w + t}\right) \tag{6.9}$$

$$T_p = 85\sqrt{\varepsilon_r} \tag{6.10}$$

Z_0 = Characteristic Impedance (ohms)
T_p = Delay in ps/in
w = conductor width (mils)
t = conductor thickness (mils)
b = spacing from plane to plane (mils)

Example 6.3
A line is laid out in microstrip topology using FR-4 fiberglass ($\varepsilon_r = 4.2$). The line is 5 mils wide and lies 5 mils above the ground plane. The trace is constructed from "1-ounce" copper which is 35 μm (1.4 mils) thick. What is the Z_0 and T_p of the trace?

$$Z_0 = \frac{87}{\sqrt{4.2 + 1.41}} \ln\left(\frac{5.98 \times 5}{0.8 \times 5 + 1.4}\right) = 36.7 \ln(5.53) = 62.8\ \Omega$$

$$T_p = 85\sqrt{0.475 \times 4.2 + 0.67} = 85\sqrt{2.665} = 138.8\ \text{ps/in}$$

5-mil-wide traces and 1-ounce copper are typical values for most modern circuit boards. The designer then chooses the thickness of the insulating layer between the signal layer and the ground layer to get the desired Z_0.

Approximate Model of a Transmission Line

There is another way to think about transmission lines, based on a model of how they work. Consider a pair of conductors in close proximity that travel over a long distance. There is a mutual capacitance between the conductors because they can be at different voltages and there is insulating material between them, the conductors have mutual inductance because they are physically separated and their fluxes do not completely cancel, and the conductors have series resistance. Also, because the insulating material is not a perfect insulator, there is a conductance between the two conductors. (This is called "dielectric loss" – more on this later.) The model is shown in Fig. 6.3.

To model a transmission line, it is divided into a very large number of tiny segments, and each segment is assigned some mutual capacitance, mutual inductance, resistance, and conductance. Using this model, all the behaviors (such as Z_0)

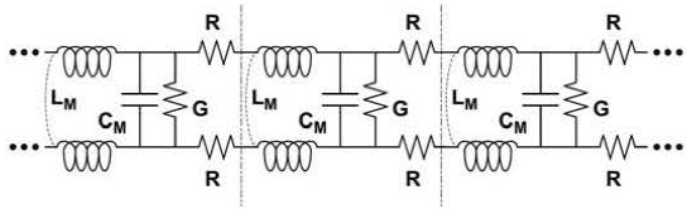


Fig. 6.3 Lumped model of a transmission line

emerge. This model does not help us understand the physical significance of Z_0 but it helps programmers simulate transmission lines.

By the way, this model leads to "the telegrapher's equation" because telegraph wires were among the first transmission lines ever developed.

It is helpful to describe the four terms (C, L, R, and G) as an amount per unit length. When this is done, it is common to add a subscript L to indicate that the unit is per length. Capacitance C, for example, becomes C_L and has units like pF/cm.

This model will come in handy in the chapter on dielectric loss and lossy transmission lines.

Appendix

1. Consider a transmission line.
   (a) What structures or features cause impedance discontinuities and therefore signal reflections? (Name three examples.)
   (b) Where is the signal located as it propagates?
   (c) What factors determine a transmission line's characteristic impedance?
2. A vendor has submitted a cable for you to evaluate for use in a high-speed signal environment. When measuring 10 inches of cable, you find that it has 64 nH of inductance and 11.4 pF of capacitance. What is the cable's characteristic impedance?
3. Consider a transmission line with impedance $Z_0 = 50\ \Omega$.
   (a) What is the phase angle between the voltage and the current? Stated differently, is the 50-Ω impedance seen at the source real, complex, or imaginary?
   (b) If the trace was moved farther from the ground plane, would you expect Z_0 to go up or down?
4. Consider a 5-mil-wide circuit trace located between two ground planes that are 10 mils apart. (That is, the two planes are 10 mils apart and the circuit trace is centered between them.) The dielectric is FR-4 ($\varepsilon_r = 4.2$). The trace is routed stripline, completely surrounded by the FR-4 and the two planes. The trace is 9 inches long and is carrying a signal with a 200 ps rise time.

(a) What is the knee frequency of the signal?
(b) What is the speed of propagation of a signal on the trace, in ps/in?
(c) How long must the trace be in order to be considered distributed?
(d) What is the Z_0 of the trace?
(e) What are the T_p and T_0 of the trace?

5. Consider a transmission line with impedance $Z_0 = 50\ \Omega$. The transmission line is a trace on a circuit board.

   (a) What is the phase angle between the voltage and the current? Stated differently, is the 50-Ω impedance seen at the source real, complex, or imaginary? Real
   (b) If the trace was moved farther from the ground plane, would you expect Z_0 to go up or down? Farther, more inductive, Z_0 goes up
   (c) If the trace was made narrower, would you expect Z_0 to go up or down? Narrower = more inductance and less capacitance. Z_0 goes up.

6. Consider a 5-mil-wide circuit trace located 10 mils over a ground plane. The dielectric is FR-4 ($\varepsilon_r = 4.2$). The trace is routed microstrip, on top of the FR-4 and ground plane. The trace is 25 cm long and is carrying a signal with a 200 ps rise time.

   (a) What is the knee frequency of the signal? 1/400 ps = 2.5 GHz
   (b) What is the Z_0 of the trace? (Use the formula from the slides.) I get 88.3 Ω. . .
   (c) What are the T_p and T_0 of the trace? T_p = 138.8 ps/in = 54.6 ps/cm. T_0 = 54.6 ps/cm × 25 cm = 1.366. ns.

Chapter 7
Transmission Lines: Reflections and Termination

Background and Objectives

This chapter introduces the concept of terminating a transmission line. Because of the "left hand does not know what the right hand is doing" property of transmission lines, changes in voltage can have unexpected effects at the source and load (and points in between) in the form of reflections. Managing these reflections requires a termination strategy. When you finish this chapter, you should be able to:

- Describe how reflections occur on improperly terminated lines
- Calculate and simulate simple transmission lines
- Calculate the correct impedance to terminate transmission lines
- Describe different termination strategies and their relative merits
- Explain how to diagnose and deal with common termination issues

The Left Hand and the Right Hand: Reflection and Transmission

Reviewing briefly, if a gate launches a very fast rising edge into RG-11 coaxial cable, the gate "sees" a resistance of 75 Ω to ground. The rising edge then takes 10 ns to get to the end of 2.5 m of cable.

What happens next is completely unexpected. There are two basic issues in play.

First, the 75-Ω impedance of the cable is in series with whatever resistance is at the source and at whatever resistance is at the other end. (We call the resistance at the other end the *load* or *load resistance*.)

Second, the source is isolated from the load by the relative slowness of the speed of light. It is as if the signal is being carried by carrier pigeons and takes days to reach the other end. The source does not "know" what is happening at the load, at least not at first. In our RG-11 example, it takes 10 ns for the edge to reach the load and then

© Springer Nature Switzerland AG 2022

S. H. Russ, *Signal Integrity*, https://doi.org/10.1007/978-3-030-86927-4_7

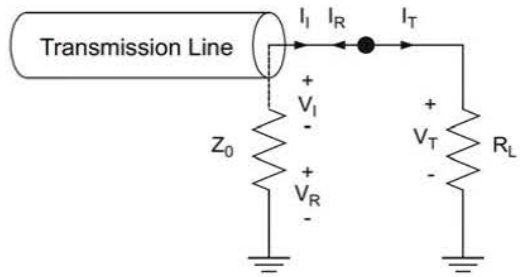


Fig. 7.1 Currents and voltages used to derive the reflection-coefficient formula. Note that the resistor of Z_0 to ground is how the transmission line appears to the load

another 10 ns for news of what happened to get back to the source. Because we are operating well above the critical length, the theory of relativity makes clear that there is no way for the source to have instantaneous knowledge of the load. Again, this is what makes the circuit distributed instead of lumped: the "left hand" (source) does not know what the "right hand" (load) is doing.

So what happens when the wave hits the load? It depends on the load.

If the impedance of the load is greater than the impedance of the line (i.e., if $Z_L > Z_0$) then the rising edge traveling down the line has too much current for the desired voltage swing at the load. The excess current crashes (more accurately, it *reflects*) back onto the line in the form of a positive reflection. The reflection is positive because there is an excess coming back off of the load.

If the impedance of the load is less than the impedance of the line (i.e., if $Z_L < Z_0$) then the rising edge traveling down the line does not have enough current for the desired voltage swing at the load. The shortage of current appears in the form of a negative reflection off of the load. One can imagine it as if the load is sending back to the source for more current.

So how can we model this? Referring to Fig. 7.1, an *incident* current and voltage emerge out of the line and strike the load. Part of the voltage and current are *reflected* back onto the line and part is *transmitted* onto the load. We will denote the incident wave with the subscript I, the reflected wave with R, and the transmitted wave with T. Our goal is to calculate the ratio of the reflected voltage (V_R) to the incident voltage (V_I).

The line has an impedance Z_0 and the load has an impedance R_L. We assume here that the load is purely resistive, although the math works out exactly the same if it is not. Note that we do not have to assume that Z_0 is purely real – it *is* purely real!

I_I is coming out of the line and I_R is going back onto the line, and so we know that $I_I = V_I/Z_0$ and $I_R = V_R/Z_0$. Likewise, $I_T = V_T/R_L$ because it is flowing through the load resistance. In other words, the resistances of the line and of the load constrain the respective ratios of current to voltage.

The voltage at the end of the line must equal the voltage across the load, since the end of the line and the load have the same node. We can apply lumped analysis here because the load is lumped, and the coaxial cable's Z_0 has the illusion of being a lumped resistance. In other words, this analysis is being applied at an instantaneous moment in time so that we can use lumped analysis. We know that

$$V_I + V_R = V_L \tag{7.1}$$

because the voltage of the load equals the voltage at the line (they are the same node) and the voltage of the line equals the sum of the incident and reflected voltages.

The incident current must equal the transmitted and reflected current. So we know that $I_I = I_R + I_L$. (If you study everything to this point, the sign for V_I looks backward relative to I_I, but that is because the incident wave is a source of current and the reflection and transmission are sinks of current.)

Substituting the three formulas for current into the current equation, we have

$$V_I/Z_0 = V_R/Z_0 + V_T/R_L \tag{7.2}$$

and so

$$V_I - V_R = \frac{Z_0 V_T}{R_L} \tag{7.3}$$

If we subtract and add Eqs. 7.1 and 7.3, we get, respectively,

$$2V_R = V_T\left(1 - \frac{Z_0}{R_L}\right) \tag{7.4}$$

$$2V_I = V_T\left(1 + \frac{Z_0}{R_L}\right) \tag{7.5}$$

So dividing Eq. 7.4 by Eq. 7.5, and canceling out, we get

$$\frac{V_R}{V_I} = \frac{\left(1 - \frac{Z_0}{R_L}\right)}{\left(1 + \frac{Z_0}{R_L}\right)} = \frac{R_L - Z_0}{R_L + Z_0} = \rho_L \tag{7.6}$$

So we define the ratio V_R/V_I as the *reflection coefficient at the load* and denote it ρ_L. It lets us calculate the reflected voltage as a function of the incident voltage.

To summarize the derivation, by expressing the incident, reflected, and transmitted waves and their respective impedances, we can "balance the books" and construct a model for what happens to the signal energy.

If $R_L > Z_0$, then $\rho_L > 0$ and there is a positive reflection (and as $R_L \to \infty$, $\rho_L \to 1$). Likewise, if $R_L < Z_0$, $\rho_L < 0$ and there is a negative reflection. If $R_L = Z_0$, $\rho_L = 0$ and there is no reflection. This is in line with the earlier statement that if $R_L > Z_0$, there is an excess of current and a positive reflection.

We can also define the ratio V_L/V_I as the *transmission coefficient at the load* and denote it T_L. Dividing Eq. 7.6 by V_I, we get

$$\frac{V_I + V_R}{V_I} = 1 + \rho_L = \frac{V_T}{V_I} = T_L \tag{7.7}$$

In other words, $1+ \rho_L = T_L$. T_L lets us calculate the transmitted voltage (the voltage appearing at the load) as a function of incident voltage.

Let's review what happens. When the waveform hits the "other end" of the transmission line, part of the waveform is reflected and part is transmitted. The transmitted part becomes visible as a voltage across the load. The reflected part goes back onto the transmission line and goes back to the source.

Sometime later (T_0 later, to be exact), the reflected wave hits the *source*. The source has a similar reflection and transmission coefficient, and, once again, part of the energy is transmitted, becoming a voltage visible at the source, and part is reflected, heading back to the load again.

The result is a pinball game of signal energy, with the size of the pinball a function of the reflection coefficients at the source and load.

So let's work an example to see what happens... (Example 7.1)

Example 7.1

Consider a long 50-Ωtransmission line on a circuit board. A fast-falling edge is being driven on the line by a source with a 10-Ω resistance. At the other end of the line, the receiving logic gate looks like a 5 MΩ resistance to ground. (The source and load resistances are typical for CMOS logic.) The voltage swing is 4 V, and both the source and load are quietly resting at +4 V when everything starts. Plot the source and load voltages in time units of T_0.

First we first need to know the source and load reflection and transmission coefficients. ρ_L = (5Meg-50)/(5Meg + 50) = 1. $T_L = 1 + \rho_L = 2$. ρ_S = (10–50)/(10 + 50) = −2/3. $T_S = 1 + \rho_S = +1/3$.

Then we need to consider what happens to the wave when it is launched. The source is a square wave from 4 V down to 0 V, but there is a resistor divider between the 10-Ω source resistance and the transmission line. As a result, only 50/(10 + 50) = 5/6 of the 4 V drop actually reaches the line. 4 V*5/6 = 3.333 Volts. So at $t = 0$, a −3.333-V wave is launched down the line.

At $t = T_0$, the −3.333 V wave hits the load. A wave of 1*–3.333 = −3.333 V is reflected onto the line and a wave of 2*–3.333 = −6.666 V is transmitted out of the load.

At $t = 2\ T_0$, the −3.333 V wave hits the source. A wave of −2/3*–3.333 = +2.222 V is reflected onto the line and a wave of +1/3*–3.333 = −1.111 V is transmitted out of the source.

At $t = 3\ T_0$, the +2.222 V wave hits the load. A wave of 1* + 2.222 = +2.222 V is reflected onto the line and a wave of 2*2.222 = 4.444 V is transmitted out of the load.

At $t = 4\ T_0$, the +2.222 V wave hits the source. A wave of −2/3*2.222 = −1.481 V is reflected onto the line and a wave of +1/3*2.222 = 0.741 V is transmitted out of the source.

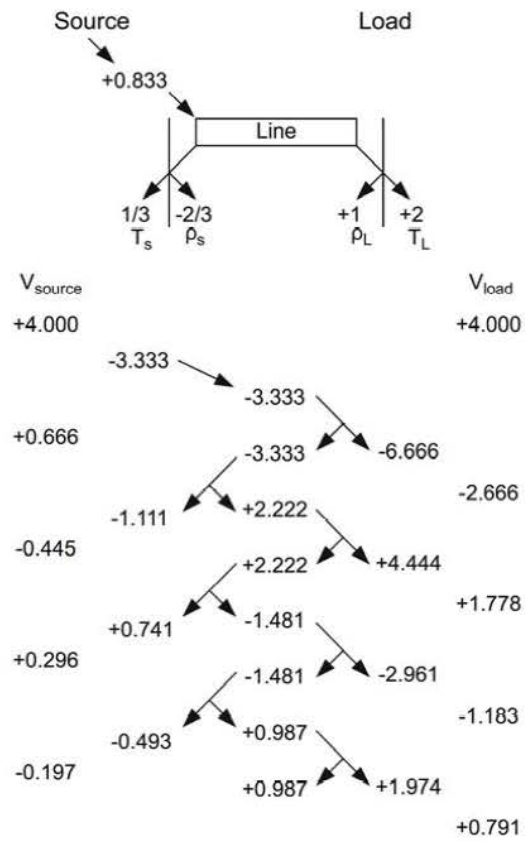


Fig. 7.2 Bounce diagram using reflection and transmission coefficients

At t = 5 T_0, the −1.481 V wave hits the load. A wave of 1*−1.481 = −1.481 V is reflected onto the line and a wave of 2*−1.481 = −2.961 V is transmitted out of the load.

And so on…

So tallying at the load, the initial voltage was +4 V. At time T_0, the voltage drops to 4–6.666 = −2.666 V. At time $3T_0$, the voltage rises to −2.666 + 4.444 = 1.778 V. At time $5T_0$, the voltage drops to 1.776–2.961 = −1.183 V, and so on.

All of this is summarized in the "bounce diagram" in Fig. 7.2.

So what happened in the example? The load voltage bounced up and down; it plunged to −2.7 V and then rose to +1.8 V and then dipped to −1.2 V, and so on. In fact, the load voltage looked *just like ringing*. However, the cause of the waveform has nothing whatsoever to do with inductance (except for the indirect connection between inductance and Z_0). The ringing-like behavior is caused by an *impedance mismatch*. A ringing-like reflection waveform is only possible if the source and load reflection coefficients have opposite signs, and is only significant if the product of the two is close to 1 (otherwise it damps out rapidly) (Fig. 7.3).

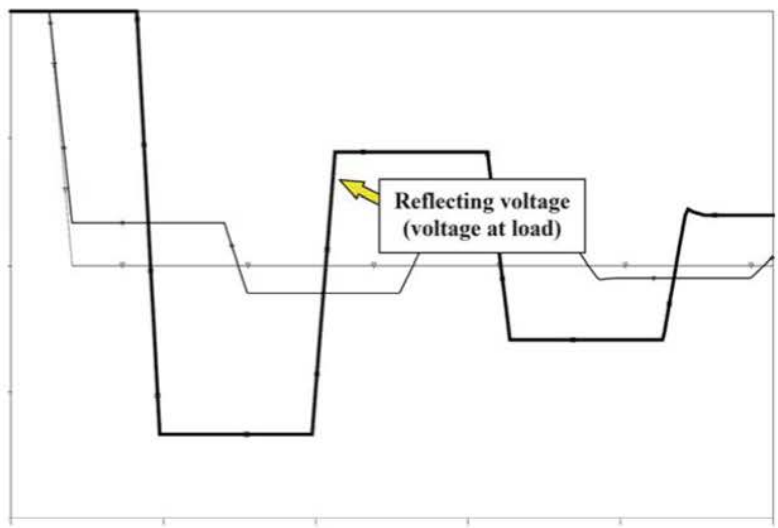


Fig. 7.3 SPICE simulation of the example. The reflecting voltage is the thick line, and it matches the voltages calculated in the bounce diagram

Why is the bouncing waveform bad? Look at the load voltage – first it plunges well below 0 V, which is a good thing because clearly, the logic gate will treat it like a logic "0." Next, however, it rises up to about 1.8 V or about 44% of a full logic swing. This is well beyond the noise margin, and the receiving gate might treat the peak as an extra logic transition. If the gate input is a clock, for example, it might treat the reflection as an extra clock edge.

Finally, what happens to a transmission line if the source stops switching as time approaches infinity? Over a sufficiently long period of time, the line becomes electrically short again, and the transmission line turns back into a wire with zero ohms series resistance. It returns to a lumped circuit. This also lines up with our intuition – that given enough time, the line acts like an ordinary wire and the entire wire drops to 0 V. This will affect the DC power consumption calculations, as we will see.

Terminating Impedance

We have seen that lines that are long (long in the electrical sense) take on two spooky properties, characteristic impedance and reflections. Is there a way to eliminate reflections altogether?

From the reflection-coefficient formula (Eq. 7.6) it is clear that if the load impedance exactly matches the line impedance, there is no reflection. This is called *load termination*. It is a parallel termination because the load is positioned as a shunt resistance from the line to ground.

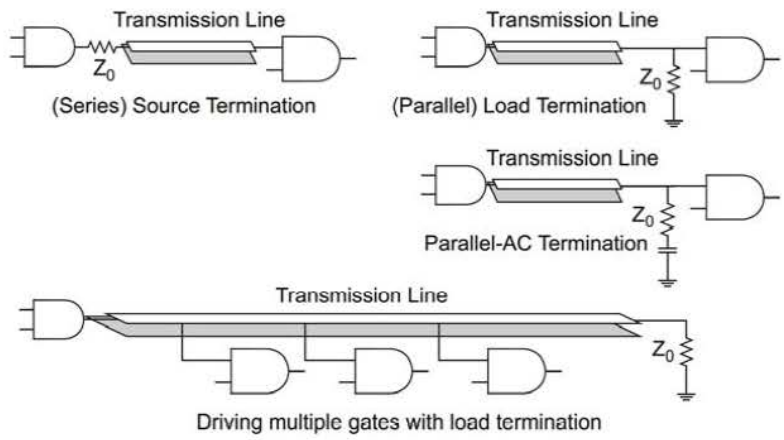


Fig. 7.4 Several termination strategies

Similarly, if the load is left open-circuited and the source impedance matches the line impedance, there is one big reflection off the load; it travels back to the source and then no more reflections. This is called *source termination*, and it (the source termination) is placed in series between the source (the driving logic gate) and the line. These are illustrated in Fig. 7.4.

These two possibilities are the two main types of termination, and they have some interesting trade-offs.

Parallel-load termination places a total resistance equal to the characteristic impedance in between the end of the line (the load end) and ground, in parallel with the logic input at the load. Typically, the source impedance is extremely low. The source reflection coefficient is close to -1, but the load reflection coefficient is close to 0 (and the load transmission coefficient is close to 1) so it does not matter. Because the resistor divider between the source resistor and the line is so lopsided, almost a full logic swing is launched into the line, and is transmitted to the load.

One disadvantage of load termination is DC power consumption. The source has to drive the load termination continuously in the steady state, since the load termination is in parallel with the logic input. One alternative to classic parallel termination, therefore, is parallel-AC termination in which a small capacitor is added in series with the resistor. This blocks the DC consumption but at AC, it still appears to be correctly terminated.

One advantage of load termination is that the load termination can be placed past the end of all of the logic inputs. It is commonly the case that one line drives multiple logic inputs, and the line can then continue past the last input and then be terminated. This is how DDR3 logic is terminated, for example, and it has the interesting (and important) advantage that the line is correctly terminated no matter how many RAM cards are inserted.

Series-source termination places a total resistance equal to the characteristic impedance in series between the source and the line, and the load is typically a

very high impedance, almost an open circuit. (Again, CMOS inputs typically look like open circuits with a few pF of capacitance.) If the source has a nonzero resistance (which is the normal case) then the external series resistor needs to be reduced in value to compensate. For example, if the source has 10 Ω resistance and the line has an impedance of 50 Ω, adding a 40-Ω series resistor will create a "perfect" termination.

One unusual property of source termination has to do with the voltage divider at the source. When the source switches, only ½ of the voltage swing is applied to the line; the other ½ of the swing is across the series resistance. A half-swing is what is transmitted down the transmission line. At the load, the reflection coefficient is almost 1 and the transmission coefficient is almost 2 (like Example 7.1). The half-swing times a transmission coefficient of 2 sends a full swing to the load. Meanwhile, another half-swing races back to the source, bringing the line up to a full logic swing.

At the load, the termination is perfect – it sees one full logic swing and no reflections. At any point on the transmission line other than the load, however, there is a half swing followed, later, by another half. This is bad if there are additional logic inputs mid-way down the line, so series termination does not work well if the line drives multiple inputs, unless special provisions are made.

So what are the advantages? First, the DC power consumption is zero. Once the load has been raised to a full logic level, the voltage across the series resistor is zero and there is no current draw. Second, no special termination is needed at the load, and so the line can be extended as long as needed. PCI uses this strategy, relying on the reflection off of the open circuit to obtain a full logic swing.

When the advantages and disadvantages are traded off, parallel-load termination is the more common of the two. The half-swing issue turns out to be a deal-breaker for most logic families. And with respect to DC consumption, most modern high-speed serial interfaces are "always on" (some sort of data is always being transmitted unless the interface is completely powered down). There is continuous AC consumption, a regime in which both source and load termination have the same power consumption. Most modern high-speed signal standards operate at very low voltages, which lets logic gates switch faster and also reduces power consumption.

So how is a 50-Ω line, for example, laid out correctly? From the logic output, the line is routed as close as possible to each of the inputs where the signal is being routed. By moving the line as close as possible, the "stub" that connects the line to the logic input is made as small as possible, which minimizes the load capacitance (the stub forms a capacitor with the ground plane) and keeps the stub as close as possible to a lumped element. If the stub is long enough, it becomes another transmission line, and the resulting reflections ruin the signal integrity.

After the last logic input, the line continues past the end and is then terminated with the resistor. This makes the last logic input look like a small, lumped capacitor and places a 50-Ω resistance at the end. The termination is purely resistive (no capacitor in parallel) and correctly matched to Z_0, which minimizes the reflection coefficient.

It turns out that this ideal situation cannot always be achieved, as we will see in the upcoming Engineer's Notebook, but before we read about it, we need to see what happens on less-than-ideal transmission lines.

Departures from the Ideal

A transmission line is a stable, easily fabricated structure that can carry signals well up into the 10's of GHz. The key to perfect signal integrity is for the transmission line structure to have a completely uniform cross-section down its entire length.

So what causes departures from this ideal? Many things – including slots in the ground plane, vias (passages from one layer to another), right-angle bends or turns, and additional logic inputs in the middle of the line – can cause changes to the uniformity.

Most obstructions, such as vias, bends, and additional inputs, cause extra capacitance. A shunt capacitance lowers the Z_0 at the point where it connects to the line, and so causes a small negative reflection. It can be modeled as a break between two transmission lines. At that point, the first line "sees" the second line in parallel with a capacitor to ground. The capacitor is in parallel and so lowers the net impedance seen by the first line. A drop in impedance results in a negative reflection coefficient.

Since the capacitor has a derivative relationship between current and voltage, the reflection is the first derivative of the incoming waveform. For example, if the waveform is a rising edge, the reflection is a negative pulse.

These negative pulses can propagate back toward the source and look like false logic transitions at inputs that are located nearer the source. Also, if load termination is used, the pulses can reflect off of the poorly terminated source and travel back out onto the line.

In the transmission direction, the combination of the capacitance and the Z_0 of the line actually creates an RC filter that slows the signal and increases the rise time. At the point the capacitor contacts the transmission line, the Z_0 of the upstream and downstream directions are in parallel, so the effective resistance "seen" by the capacitor is $Z_0/2$. The overall rise time adder is $2.2^* Z_0/2^*C = 1.1\ Z_0C$. If the signal passes multiple capacitive loads (a common case on a bus), then the rise time steadily grows as the signal moves down the line (and a stream of negative pulses is sent back to the source).

A SPICE simulation of a capacitive imperfection is shown below in Fig. 7.5. The circuit in the simulation used source termination, so the initial reflection of the falling edge off of the capacitor is absorbed at the source. However, in source termination, there is a nearly total reflection off of the load. The load reflection actually reflects off of the capacitor and comes back to the load. Since the sign of the capacitor's reflection coefficient is negative, the reflection of the falling edge is a rising spike. Since the sign is opposite, it can be mistaken by the load as an extra logic transition. The simulation also shows how the "reflection of the reflection" turns into a second-derivative double pulse. (The voltage at the load is drawn in the thick line in Fig. 7.5.)

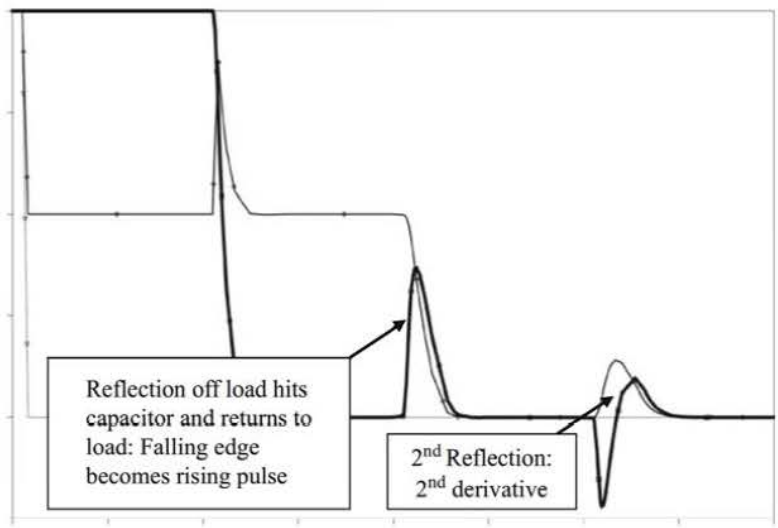


Fig. 7.5 Capacitor in middle of the line with source termination

To recap, the source sent a half-logic-swing waveform down the line and past the capacitor. The half-logic-swing reflects off of the load and heads back to the source. When the reflection hits the capacitor, there is a reflection back to the load, which is the first pulse seen in the simulation. Additional reflections also occur later.

If there is a slot or hole in the ground plane, the return current has to divert around it, and the effect is an increase in inductance at that point in the line. It looks, electrically, like a small series inductor is interrupting the transmission line. It raises Z_0 and so causes a positive first-derivative reflection. For example, a rising edge is reflected as a positive pulse. The positive pulses are not a problem (they overdrive the incoming signal and do not look like a false logic transition) but the pulses reflected off of the source are negative and can cause problems. Like their capacitive brethren, the inductive imperfections are low-pass filters in the forward direction with a rise time adder of 1.1 L/Z_0.

A SPICE simulation of an inductive imperfection is shown below in Fig. 7.6. This circuit used load termination, so the positive reflection off of the inductor travels back to the source, hits the source's negative reflection coefficient, and heads back down to the load. So the falling edge created a falling spike which, after reflecting off of the source, became a rising spike.

To recap, a falling edge was launched down the line. The edge reflected off of an inductor (and became a pulse in the process) and went back to the source. The source then reflected the pulse back out onto the line and back toward the load. This is the first pulse visible in the simulation.

These capacitive and inductive issues highlight the general process of obtaining good signal integrity – minimize vias, right-angle turns, and slots or cuts in the ground plane. One common way to reduce the effect of right-angle turns is to turn corners gradually; come circuit-board layout software will let you lay it out this way.

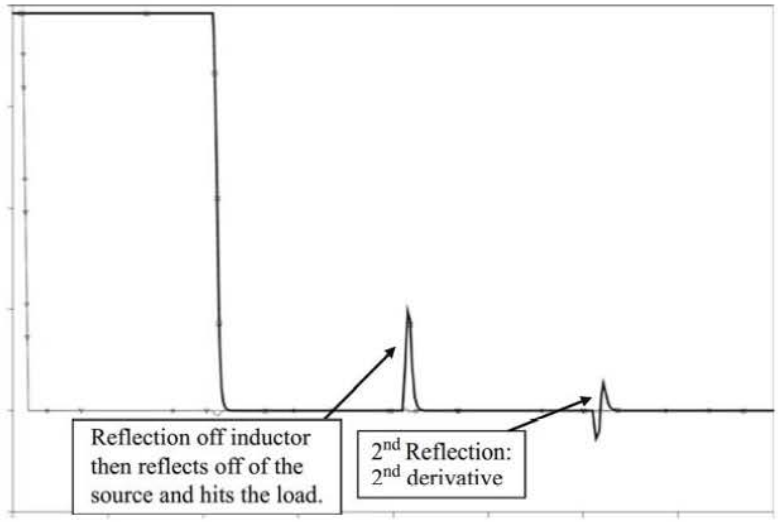


Fig. 7.6 SPICE simulation of an inductive imperfection

Fixing It When It's Broke

In most designs, termination tends to be straightforward. More problematic is trying to keep lines straight, on the same circuit-board layer, and avoiding interruptions in the power or ground plane. The problem arises because circuit boards are so crowded that often signals have to take detours. The straight line from chip A to chip B might be right under chip C, for example, or chip B might be on the opposite side of the board.

The "trick" is to keep track of the high-speed buses and pay special attention to their layout. We did not, and we got burned. . .

Engineer's Notebook: The High-Speed Bus Disaster

Two of the main chips in one of our designs were connected by a very high-speed parallel bus (16 bits wide, roughly 500 MHz speed). Parallel buses are automatically harder to design because, in order for all the bits to stay in sync, each bit must have the same length.

On our first version of the board, the bus was a mess, full of errors. Then we sat down and looked at the layout, and it was embarrassing. We had made one major rookie mistake and discovered a subtle issue that kept us from being able to follow the rules.

The bus had to be routed on the bottom side of the board. This turns out to be an important distinction because, as is typical in a four-layer-board design, the bottom side is routed over a power plane instead of over a ground plane. (You can add almost as many planes as you want; it just gets more expensive, and our designs were very cost-sensitive.) The power plane was a hodge-podge of voltages, including 1.0,

1.2, 1.8, and 3.3 V, to name a few. The bus was routed over all of these different DC voltages, and had to cross a slot in the plane every time. It was as if the signal had to drive across 10 different countries and stop and show a passport every time it crossed a border.

So the first fix was to carve out space underneath the signal routing and fill it with a dedicated 0 V ground plane all the way from under the first chip to under the second chip. The bus then had a single, uniform ground plane under it.

Then we looked at the termination. We had "followed the rules" and routed the signals under the chip, to the input pins, and then continued the signals back out from under the chip to terminating resistors. That is, we located the termination at the end of the line, past the inputs.

The problem was the signal routing underneath the chip. Having a wire go from the outside to the input and back out doubled the wiring that was needed, and the chip was so small and the wiring so crowded that the wiring was a tangled-up mess (to use the precise technical term). So we broke the rules.

As near the chip as we could place them, we added terminating resistors on the top side of the board. We then ran a via from each bus bit straight up to one side of the resistor and then ran from the other side of the resistor through a via to the top-side ground plane. In other words, we had nearly perfect termination wiring. Then we continued the bus on the bottom side under the chip. The wire lengths past the terminating resistors were pretty long, over an inch in some cases, but it actually resulted in much cleaner termination.

After all those changes, we had error-free data transmission.

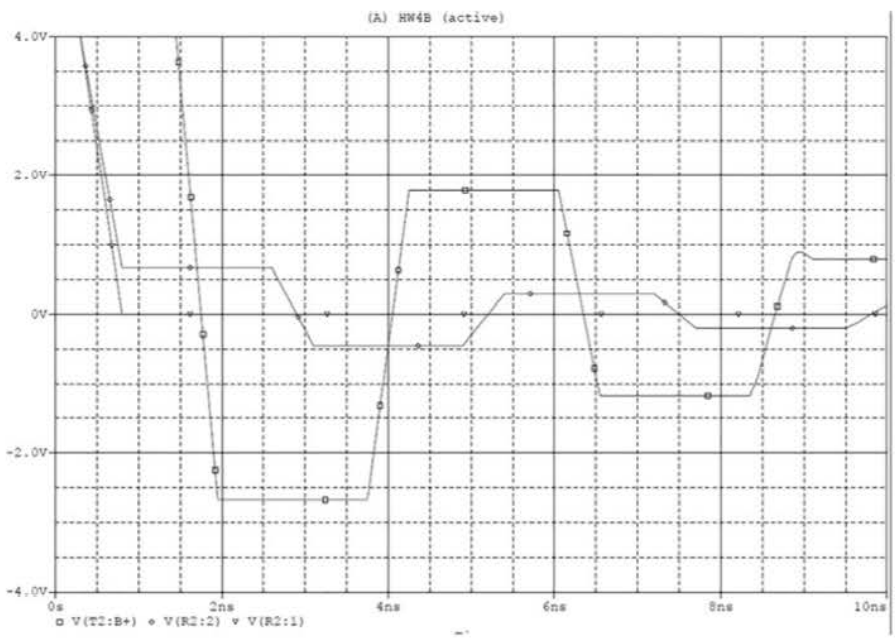
Appendix

1. Reconstruct the transmission-line Example 7.1 from the chapter. The example was 8″ of microstrip with $Z_0 = 50\ \Omega$ and $\varepsilon_r = 5$.
   (a) Compute the Tp (ns/ft), C_L (pF/ft), and T_0 (one-way propagation delay) of the line.
   (b) Run a Spice simulation of the line. The source has 10 Ω of impedance and the load has 5 MΩ of impedance. The source is a 4–0 V transition with a 500 ps fall time.
   (c) What is the undershoot and overshoot seen at the load.
      (i) At the first arrival of the falling edge
      (ii) At the arrival of the first reflection
   (d) Change the load termination to 50 Ω, and rerun the simulation.
      (i) What is the waveform seen at the load?
      (ii) What is the waveform seen at the source?

2. Reconstruct the transmission-line example (6.4) from the lecture (p. 29 from the PowerPoint). The example was 8″ of microstrip with $Z_0 = 50\ \Omega$ and $\varepsilon_r = 5$.

   (a) Compute the Tp (ns/ft), C_L (pF/ft), and T_0 (one-way propagation delay) of the line. The Tp formula is in the PowerPoint slides, and recall that $C_L = T_p/Z_0$ and that $T_0 = T_p \times$ length of line. Tp = 85*sqrt(0.475*5 + 0.67) = 148.3 ps/in = 1.779 ns/ft. CL = Tp/Z0 = 1779 ps/ft./50 = 35.6 pF/ft. T_0 = 1.779 ns/ft. * 8/12 feet = 1.186 ns.

   (b) Run a Spice simulation of the line. The source has 10 Ω of impedance and the load has 5 MΩ of impedance. The source is a 4–0 V transition with a 500 ps fall time.



   (c) What is the undershoot and overshoot seen at the load.

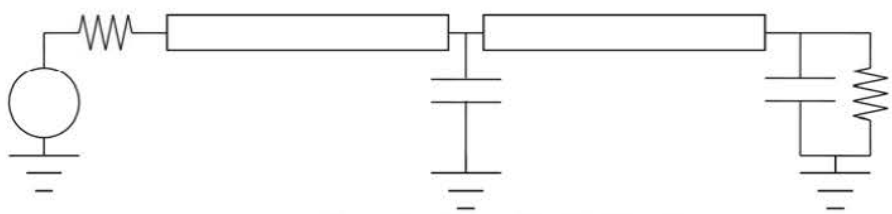
      (i) At the first arrival of the falling edge −2.666 V
      (ii) At the arrival of the first reflection +1.777 V

   (d) Change the load termination from 5 MΩ to 50 Ω, and rerun the simulation.

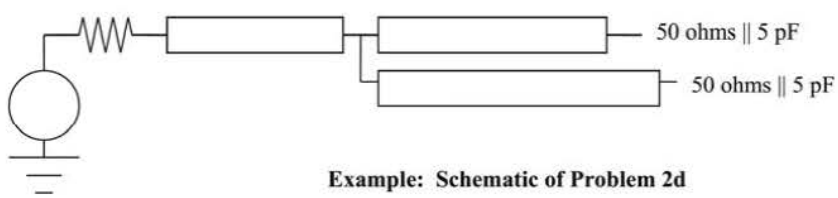
      (i) What is the waveform seen at the load? Falling edge
      (ii) What is the waveform seen at the source? Falling edge

3. Construct a SPICE simulation of a transmission line with $Z_0 = 50\ \Omega$ and $T_0 = 2$ ns. The source is a 1–0 V transition with a 200 ps fall time. The source has 5 Ω impedance, and the load impedance is 50 Ω in parallel with 5 pF capacitance. Then split the line in half (i.e., create 2 50-Ω transmission lines in series each with $T_0 = 1$ ns) and simulate the following issues in the middle of the line. Plot the voltage at the source, in the middle of the line, and at the load for each of the 4 cases.

   (a) Add a logic gate input in the middle of the line (6 pF total capacitance – 5 pF gate plus 1 pF via). The extra logic gate does not have a 50-Ω terminating impedance. (Why?) Simulate in SPICE.
   (b) Add a break in the signal-return plane (5 nH inductance) and simulate in SPICE.
   (c) Consider case b. Compute a value of capacitance to add before and after the break to compensate for the extra inductance. Add the capacitance and then simulate it again.
   (d) Add a third transmission line by placing it parallel to the second. That is, the source should be connected to one transmission line, and the other end of the line should be connected to two lines in parallel. The third transmission line should have $T_0 = 2$ ns and should be terminated with a logic gate (50 Ω in parallel with 5 pF).



Example: Schematic of Problem 2a



Example: Schematic of Problem 2d

4. You are helping a coworker debug a nonfunctioning board. One signal trace has a one-way propagation delay of $T_0 = 4$ ns. You and the coworker probe the signal source. 2 ns after the source creates a *rising* edge, there is a large *negative* spike visible on the oscilloscope.

   (a) What is the likeliest cause for the spike?
   (b) The signal trace is about 27 inches long. How far down the trace would you expect to find the cause of the spike?

Chapter 8
Lossy Transmission Lines

Background and Objectives

In Chaps. 6 and 7, the concepts of a transmission line, reflections, and terminations were unveiled. Transmission lines have the very desirable property that they carry signals nearly completely intact, and that is why they are widely used. However, if you go all the way back to the beginning of Chap. 6, one assumption was made: That is, in a coaxial cable or on a twisted-pair cable, there is no loss of signal energy. This turns out to be false, but the source of the loss is surprising. This chapter discusses dielectric loss, explains how it manifests itself, and explains how it affects signals on real circuit boards. When this chapter is finished, you should be able to:

- Understand the units of attenuation
- Contrast frequency-independent and frequency-dependent attenuation
- Identify the two sources of frequency-dependent attenuation
- Explain dielectric loss and its physical origins
- Explain how dielectric loss affects ε
- Show how to read dielectric loss off of a datasheet
- Describe how dielectric loss affects signal propagation

What Is Attenuation and How Is It Measured?

This chapter is ultimately about attenuation, which is a loss of intensity. In the context of signal integrity, attenuation is when the energy of a signal becomes weaker, usually over some distance or space.

One common source of attenuation is the propagation of a signal through a three-dimensional medium. Radio waves, for example, become weaker at the square of the distance. The waves are "spreading out" to fill space, spreading out over larger areas, and therefore the energy per unit area drops.

© Springer Nature Switzerland AG 2022

S. H. Russ, *Signal Integrity*, https://doi.org/10.1007/978-3-030-86927-4_8

If the signal energy is confined, as is the case with a transmission line, the signal energy should remain constant but, in fact, it also drops over time. Every "real" transmission line has attenuation.

But first... How is attenuation measured? Almost all attenuations are exponential. For example, over a distance of 1 m, a signal may be reduced in amplitude by ½. So if a signal has an amplitude of 8 V at the source it may be reduced to 4 V at 1 m, 2 V at 2 m, and so on. Over a distance of 1 m, the signal is reduced by a factor of 2, which corresponds to a 6 dB drop in amplitude. So the attenuation is 6 dB/1 m = 6 dB/m. This unit of measure (dB per distance) is a common measure of attenuation, and reflects the exponential nature of attenuation.

Example 8.1
A cable has an attenuation of 0.1 dB/m.

(a) What is the amplitude of a 5 V signal after traveling down 6 m of this cable?

Attenuation = 0.1 dB/m * 6 m = 0.6 dB.
0.6 dB of loss = $10^{-(0.6/20)}$ = 0.933.
Output = 5 * 0.933 = 4.665 V.

(b) At what distance does the signal have an amplitude of 4 V?

4 V/5 V = 0.8.
$10^{(-x/20)} = 0.8$; $-x/20 = -0.9691$; $x = 1.9382$ dB of loss.
Distance = 1.9382 dB/0.1 dB/m = 19.4 m.

The next issue to consider is whether the attenuation is constant with frequency or varies with frequency. This turns out to be a profound issue.

If the attenuation is constant with frequency, then the attenuation causes the signal to shrink linearly. Most importantly, the rise and fall times are not degraded and so the signal retains very sharp edges. While the receiver may need to use amplification, the basic process of receiving and decoding the signal is relatively simple because the shape of the signal has not been altered. The receiver will eventually reach a point where the amplifier is not enough, but this still leads to a robust system. For example, a Wi-Fi receiver may have a dynamic range (also known as link margin) of 100 dB – five orders of magnitude. (Actually, Wi-Fi is a bad example – it operates at a specific frequency and so all of the signal is attenuated the same amount, even if there is frequency-dependent attenuation present.)

If the attenuation varies with frequency, specifically if the attenuation is worse at high frequency, this leads to significant signal distortion. Portions of a signal with high-frequency content, such as rising and falling edges, are attenuated more than the lower-frequency counterparts. This manifests itself as a signal with significant rise time degradation, which is the same as saying a signal that has been low-pass filtered. On an oscilloscope, for example, the relatively square-ish rising and falling signals become much rounder, and the receiver has a much harder time sampling whether the input is 0 or 1.

The contrast is illustrated below in Fig. 8.1.

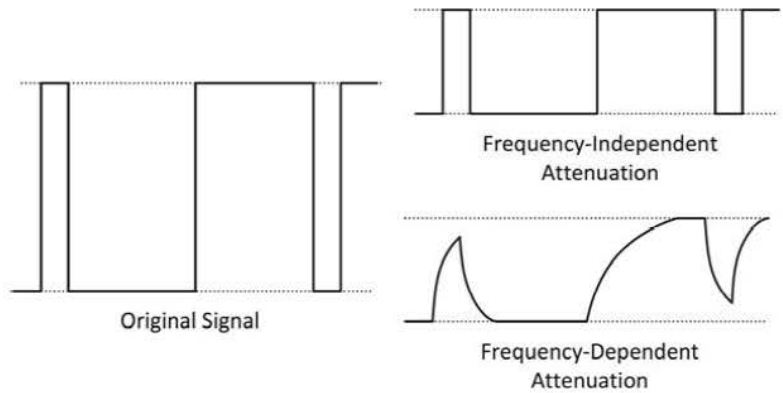


Fig. 8.1 Frequency-independent and frequency-dependent attenuation

Frequency-Dependent Attenuation

The first example of frequency-dependent attenuation is one we have already studied, the skin effect. Because of this effect (which is due to inductance, see Chap. 4), the conductor appears to get smaller and thinner as frequency goes up. It is as if the resistance of the conductor is going up. As you may recall from the discussion, the skin depth goes down (gets smaller) as the square root of frequency, and so the effect of skin depth is that the apparent resistance of a conductor goes up as the square root of frequency.

The frequency region below which the skin depth is greater than half the conductor thickness is called the "low loss" region because, in those frequency ranges, the signal can spread out over the entire conductor and the resistance is constant with frequency. The frequency region above is called the "skin effect" region, and the resistance goes up with the square root of frequency.

However, once the frequency becomes even higher, another effect appears.

Dielectric Loss

Think about this: If you put an empty glass bowl in a microwave oven and run the microwave for a few minutes, will it get hot? The answer is, of course, is "yes." But why? Microwave ovens work by emitting microwaves at a frequency that water can absorb (2.4 GHz). But there is no water in the glass. So how does it get hot?

Backing up for a moment, what is the significance of the bowl getting hot? It means that something in the bowl is absorbing real energy (not imaginary energy) from the microwaves in the oven. So what is it?

It turns out that molecules have microscopic regions of positive and negative charges, due basically to the fact that they are asymmetric. When an electric field is

applied, the molecules rotate to line up with the field. Even in a solid, the molecules rotate a small amount.

When an alternating electric field is applied, the molecules actually rock or twist back and forth. At microwave frequencies, the incident alternating E field begins to match the speed at which the molecules move. Even though the molecules twist really fast, if the E field is alternating faster, the molecules never stop moving. As the frequency goes up, the speed of rotation of the molecules goes up. The molecules undergo an actual, physical, mechanical motion that becomes faster at higher frequencies.

Conversely, the molecules require real energy to do this. Actual work (force times distance) is being done, and so real energy is being consumed. So what is another term for the microscopic vibration of molecules? Temperature! The result of this microscopic motion is that the molecules, and therefore the entire substance (e.g., the entire glass bowl), get hot.

Since transmitting microwaves through insulating materials is accompanied by the loss of real energy, the name for this phenomenon is *dielectric loss*.

How Does Dielectric Loss Affect ε?

So if one applies high frequencies (e.g., GHz) to a capacitor, two different phenomena are observed.

The first is that, for a fixed voltage, the current goes up with frequency, and this current is 90° out of phase with the voltage. This is the current we are all familiar with, a manifestation of $I = C\mathrm{d}V/\mathrm{d}t$.

The second is that, for a fixed voltage, the current also goes up with frequency, but this current is *in phase* with the voltage. This is the current due to dielectric loss. It is in phase with the voltage because it is real energy, and it is real energy because real work is being done and being dissipated as heat. It is proportional to frequency because higher frequencies twist the molecules twice as fast.

Pause for a second: Note that both currents are linearly proportional to the voltage V and to the frequency ω.

For the analysis, it is convenient to flip everything around and speak in terms of admittance. So instead of $V = IR$ we will use $I = VG$. The reason for this will become clear soon.

For an AC voltage V at frequency ω, we have

$$I = j\omega CV + \omega\widehat{C}V \tag{8.1}$$

The first term is the familiar one, the 90-degree-out-of-phase current. The second is due to this new phenomenon, dielectric loss. In Eq. 8.1, the real part of the current is being treated as if it is a new type of capacitance. This abstraction works because the real current behaves like a capacitance (conductance increasing with frequency,

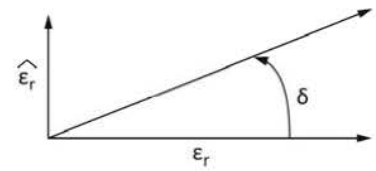


Fig. 8.2 Real and imaginary part of dielectric constant

current proportional to frequency). In fact, if one measures dielectric loss in a lab, it becomes apparent it is an intrinsic property of the material.

One can convert the "C" term above into two terms, and so

$$C \equiv \varepsilon_r C_0 \tag{8.2}$$

where C_0 is the *intrinsic capacitance*. The intrinsic capacitance is the capacitance that would be there if the dielectric were replaced with a vacuum. A very similar substitution can be made for the dielectric-loss term $\widehat{C}$, and so

$$I = j\omega\varepsilon_r C_0 V + \omega\widehat{\varepsilon}_r C_0 V = \omega C_0 V(\ j\varepsilon_r + \widehat{\varepsilon}_r) \tag{8.3}$$

If you look at the term in parentheses, it becomes clear that *the relative dielectric constant is now a complex number*! This is where it gets very confusing, so pay attention.

The convention is to express $\widehat{\varepsilon}_r$ in terms of its ratio to ε_r. In order to compute this ratio, one treats the combination of $\widehat{\varepsilon}_r$ and ε_r as a complex number, with ε_r being the real part and $\widehat{\varepsilon}_r$ as the imaginary part. So ε_r, which is associated with the *imaginary* component of the current, is treated as the *real part* of the dielectric constant. Conversely, $\widehat{\varepsilon}_r$, which is associated with the real component of the current, is treated as the imaginary part of the dielectric constant. It sounds backwards, but that is the convention. It dates back to the fact that the imaginary current/real dielectric constant association came along first, and the discovery of dielectric loss later.

One can observe the dielectric constant in the Argand plane (Fig. 8.2).

The phase angle of the now-complex dielectric constant is termed δ and so we have

$$\tan\delta = \widehat{\varepsilon}_r/\varepsilon_r \tag{8.4}$$

From (8.4), it can be seen that $\widehat{\varepsilon}_r$ can be expressed as $\widehat{\varepsilon}_r = \varepsilon_r \tan\delta$ and so

$$I = \omega C_0 V(\ j\varepsilon_r + \widehat{\varepsilon}_r) = \omega C_0 V(\ j\varepsilon_r + \varepsilon_r \tan\delta) = \omega\varepsilon_r C_o V(\ j + \tan\delta) \tag{8.5}$$

Since $C = \varepsilon_r C_0$,

$$I = \omega C V(\ j + \tan\delta) \tag{8.6}$$

The real part of I/V, therefore, is $\omega C \tan \delta$. Also note that the real part of I/V is conductance. (And, to recap, the imaginary part is the current one normally associates with a capacitor, 90 degrees out of phase with the voltage.)

To review, real dielectrics exhibit a phenomenon called dielectric loss, which causes the dielectric to absorb real energy. It is intrinsic to the material and can be expressed as a ratio relative to the ε_r of the material. The ratio is called $\tan \delta$, and can be thought of as the imaginary component of the dielectric constant of the material.

So how big is $\tan \delta$? For example, the $\tan \delta$ for FR-4 fiberglass (the most common circuit-board material) is roughly 0.02.

How do you know $\tan \delta$? It is normally listed on a datasheet or in reputable online sources. So dielectric loss is measured in the relatively obscure unit of $\tan \delta$, and this unit is readily available.

How Does Dielectric Loss Affect Signal Propagation?

The basic assumption of a transmission line was that the conductance of the insulator was zero and the resistance of the conductors was zero. If the conductance is nonzero, then the model breaks down, and we have to go back to the lumped model of a transmission line (see Fig. 6.3) to figure out what actually happens.

Before, in the approximate model of a transmission line, the assumption was that $R = 0$ and $G = 0$. The assumption was that the conductors were perfectly conductive and the insulator was perfectly insulating. Here, because of dielectric loss, we discover that $G > 0$. However, we will still assume that the resistance R is zero; this assumption turns out to be accurate because the resistance is much smaller than Z_0.

If one considers the circuit from the point of view of the conductance G (again, see Fig. 6.3), it is apparent that there is a resistor divider. One piece of the approximate model is reproduced in Fig. 8.3, to help understand the math.

The conductance G "sees" the series resistance of Z_0 on its input and output, and so the net series resistance is $Z_0/2$. The inductance, capacitance, and resistance R are shown as grayed out in the picture – we are neglecting them because we are considering the pure DC behavior of the circuit and, as noted above, R is approximately 0. If a voltage V_{in} is applied, then the voltage V_{out} is the resistor divider

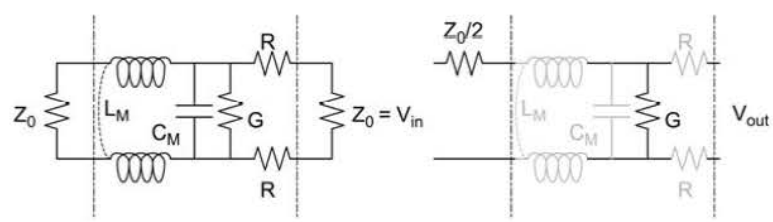


Fig. 8.3 One "piece" of a transmission line, with the other pieces replaced with Z_0

formed by the conductance G (actually, by the resistance $1/G$) and $Z_0/2$. Also, we can express G as a conductance per unit length G_L, which is the convention when we divide the transmission line into little pieces. So we have

$$V_{out} = V_{in}\left(\frac{\frac{1}{G_L}}{\frac{1}{G_L}+\frac{Z_0}{2}}\right) = V_{in}\left(\frac{1}{1+\frac{1}{2}G_L Z_0}\right) \tag{8.7}$$

We will need to convert this ratio to an exponential form. To make it solvable, we start by expressing the ratio as an exponent of e.

$$\frac{V_{out}}{V_{in}} = \frac{1}{1+\frac{1}{2}G_L Z_0} = e^{-x} \tag{8.8}$$

$$\ln\left(\frac{1}{1+\frac{1}{2}G_L Z_0}\right) = -x \tag{8.9}$$

$$\ln\left(1+\frac{1}{2}G_L Z_0\right) = x \tag{8.10}$$

For the next step, we can use the approximation $\ln(1 + a) \approx a$ because, in most practical cases, $G_L Z_0$ is a very small number and so $G_L {Z_0}^2$ is much less than $G_L Z_0$. So

$$x \approx \frac{1}{2}G_L Z_0 \tag{8.11}$$

and

$$\frac{V_{out}}{V_{in}} = e^{-x} = e^{-\frac{1}{2}G_L Z_0} \tag{8.12}$$

Converting this to dB is straightforward. For a loss b in dB, we can write

$$e^{-\frac{1}{2}G_L Z_0} = 10^{-(b/20)} \tag{8.13}$$

$$\log_{10}\left(e^{-\frac{1}{2}G_L Z_0}\right) = \log_{10}\left(10^{-(b/20)}\right) \tag{8.14}$$

$$-\frac{1}{2}G_L Z_0 \log_{10}(e) = -b/20 \tag{8.15}$$

$$10 G_L Z_0 \log_{10}(e) = b \tag{8.16}$$

Since $\log_{10}(e) = 0.434$,

$$b = 4.34 G_L Z_0 \tag{8.17}$$

where b is in units of dB/length. We can express both G_L and Z_0 in terms of their per-length parameters. Equation 8.6 above teaches that $I = \omega CV \tan\delta$. Since $G_L = I/V$ and C is C_L (capacitance per length), we have that $G_L = \omega C_L \tan\delta$. From Chap. 6, we have $Z_0 = \sqrt{\varepsilon_r}/_{cC_L}$. (And, yes, c is the speed of light.) So b, the loss in dB, can be written as

$$b = 4.34 G_L Z_0 = 4.34 \frac{\omega C_L \tan\delta\sqrt{\varepsilon_r}}{cC_L} = \frac{4.34\omega \tan\delta\sqrt{\varepsilon_r}}{c} \tag{8.18}$$

where b is the loss in dB/length. The units of length are whatever unit of length is used for c; all other measurements cancel out or are dimensionless.

Example 8.2
Consider FR-4 fiberglass with $\varepsilon_r = 4.2$ and $\tan\delta = 0.02$.

(a) Find the loss in dB/cm due to dielectric loss at a frequency of 10 GHz.

Recall that c, the speed of light, is 300,000 km/s $= 3 \times 10^{10}$ cm/s. (We need c to be in cm/s so that the final answer is in dB/cm.) So we have $b = (4.34 * 2 * \pi *$ 10 GHz $* 0.02 * \sqrt{4.2})/3 \times 10^{10}$ cm/s $= 0.372$ dB/cm.

(b) Over what length of conductor will a 10 GHz signal have 3 dB of loss?

3 dB/0.372 dB/cm = 8.05 cm.

The example shows that dielectric loss can significantly reduce GHz-range signals in FR-4 fiberglass.

Is there an easier way to understand the effect? First, recall that t_r is roughly equal to $0.35/f_{3dB}$. That is, the rise time of a circuit element is 0.35 divided by its 3 dB frequency (the frequency at which the element attenuates its input by 3 dB). Second, we can re-write (8.17) by multiplying the loss in dB/length by length (resulting in a loss in dB) and setting that loss equal to 3 (that is, 3 dB). So we have, for distance d,

$$\begin{aligned} 3 &= \frac{4.34\omega \tan\delta\sqrt{\varepsilon_r}}{c} d = \frac{4.34 \cdot 2\pi f_{3\ \mathrm{dB}} \tan\delta\sqrt{\varepsilon_r}}{c} d \\ &= \frac{27.26 f_{3\ \mathrm{dB}} \tan\delta\sqrt{\varepsilon_r}}{c} d \end{aligned} \tag{8.19}$$

Solving for $f_{3\ \mathrm{dB}}$

$$f_{3\ \mathrm{dB}} = \frac{3c}{27.26 \tan\delta\sqrt{\varepsilon_r} d} \tag{8.20}$$

Substituting (8.19) into the rise time formula, we have

$$t_r = \frac{0.35}{f_{3\ \mathrm{dB}}} = 0.35\frac{27.26\tan\delta\sqrt{\varepsilon_r}d}{3c} = \frac{3.18\tan\delta\sqrt{\varepsilon_r}d}{c} \tag{8.21}$$

Example 8.3

Continue Example 8.2 with FR-4 fiberglass with $\varepsilon_r = 4.2$ and $\tan\delta = 0.02$.

(a) Find the rise time adder of 1 cm of conductor routed over FR-4 due to dielectric loss.

Again, c is 3×10^{10} cm/s. So we have $t_r = 3.18 * 0.02 * \sqrt{4.2} * 1$ cm/3×10^{10} cm/s = 0.434 ps. This is a very important figure of merit: FR-4 has a rise time adder of 0.434 ps/cm due to dielectric loss. Stated differently, every cm of FR-4 is a low-pass filter with a rise time equivalent to a 0.434 ps rise time adder.

(b) A 1 Gigabit/s SATA signal with a rise time of 130 ps is routed over 10 cm of FR-4. What is its rise time at the load?

Total rise time adder = 0.434 ps/cm * 10 cm = 4.34 ps. $\sqrt{130^2 + 4.34^2} =$ 130.1 ps. This is not a significant issue.

(c) A 6 Gigabit/s SATA signal with a rise time of 20 ps is routed over 30 cm of FR-4. What is its rise time at the load?

Total rise time = 0.434 ps/cm * 30 cm = 13.02 ps. $\sqrt{20^2 + 13.02^2} = 23.9$ ps. Since this is a 20% increase in rise time, this is a significant problem.

(d) How much shorter does the trace need to be to keep the rise time under 22 ps?

$\sqrt{20^2 + x^2} = 22$ ps. Solving, we have $x = 9.17$ ps. 9.17 ps/0.434 ps/cm = 21.1 cm.

This example teaches two things.

First, the dielectric loss ultimately manifests itself as a rise time adder. The units can be a little confusing, but the dielectric loss in FR-4 creates a rise time of 0.434 ps per cm of trace length. The trace length in cm times 4.34 is the rise time adder in ps. This rise time must then be combined with the rise time of the signal (using the rise time adder formula) in order to calculate the rise time of the load.

Second, this example shows that, as signals move past 1 Gbit/s, FR-4 fiberglass is going to be harder to use. Because of the combination of dielectric loss and high-speed signaling, the era of FR-4 is nearing its end.

Appendix

1. A circuit board trace has an attenuation of 0.3 dB/m and is carrying a waveform that is 1.2 V peak-to-peak.
   (a) What is the amplitude of the signal after traveling down 30 cm of trace?
   (b) What attenuation would be needed to lose less than 10% of the peak-to-peak voltage at a distance of 30 cm?
2. Consider a PTFE ceramic circuit-board material (Rogers R3003™) with $\varepsilon_r = 3.0$ and $\tan\delta = 0.0013$.
   (a) Find the loss in dB/cm due to dielectric loss at a frequency of 7 GHz.
   (b) Over what length will a conductor with a 10 GHz signal have 3 dB loss?
   (c) How does this compare to the FR-4 fiberglass in Example 9.2?
3. Consider a material with $\varepsilon_r = 1.7$ and $\tan\delta = 0.01$.
   (a) Find the rise time adder of 1 cm of conductor routed over the material due to dielectric loss.
   (b) A microwave radio signal with a rise time of 15 ps is routed over 30 cm of the material. What is the rise time at the load?
   (c) What length does the trace need to be for a rise time of 17 ps?
4. Consider a circuit trace carrying a 3 Gbps SATA signal. It has a 250-mV logic swing. The trace is routed over a circuit board material that has a dielectric loss adder of 4 ps/cm.
   (a) If the trace is routed 10 cm on the board, what is the effective rise time of the dielectric loss? 4 ps/cm * 10 cm = 40 ps.
   (b) If the SATA signal has a rise time at the driver of 50 ps, what is the rise time of the signal at the other end of the trace? Sqrt($50^2 + 40^2$) = 64 ps.
   (c) How long can the signal be routed and keep the overall rise time (combination of driver rise time and rise time due to dielectric loss) below 60 ps? sqrt ($50^2 + x^2$) = 60 ps. $X = 33.2$ ps. 33.2 ps/4 ps/cm = 8.3 cm.

Chapter 9
Differential Signaling

Background and Objectives

In exploring transmission lines and signals, it is clear that signals that are close to each other have both capacitive and inductive coupling between them. It turns out that this is both good news and bad news. In this chapter, we will start with the good news and introduce differential signaling. In the next chapter, we will break the bad news and discuss crosstalk.

One common method for better signal integrity is to move to differential signaling. This seemingly simple change, using both a positive and negative waveform in a matched set, eliminates the need for an intact ground plane, takes care of return current, manages ground bounce, and generally improves signal integrity. When this chapter is finished, you should be able to:

- Explain how differential signaling improves signal integrity.
- Calculate Z_{diff}, the differential impedance.
- Explain how changes in mutual inductance and capacitance can affect Z_{diff}.
- Terminate cables or conductors that carry differential signals.
- Simulate differential signaling.
- Explain the effects of common-mode signals and clock jitter on differential signals.

What Is Differential Signaling and How Does It Help?

Up until now, a ground plane has been used as a return path. This has one advantage; you can lay out signals over a ground plane and never have to worry about laying out a return path. But there are disadvantages. For example, the ground plane has to be kept intact, or else the inductance (both self and mutual) will go up. The ground plane is used for pretty much everything on the board (e.g., for power distribution,

© Springer Nature Switzerland AG 2022

S. H. Russ, *Signal Integrity*, https://doi.org/10.1007/978-3-030-86927-4_9

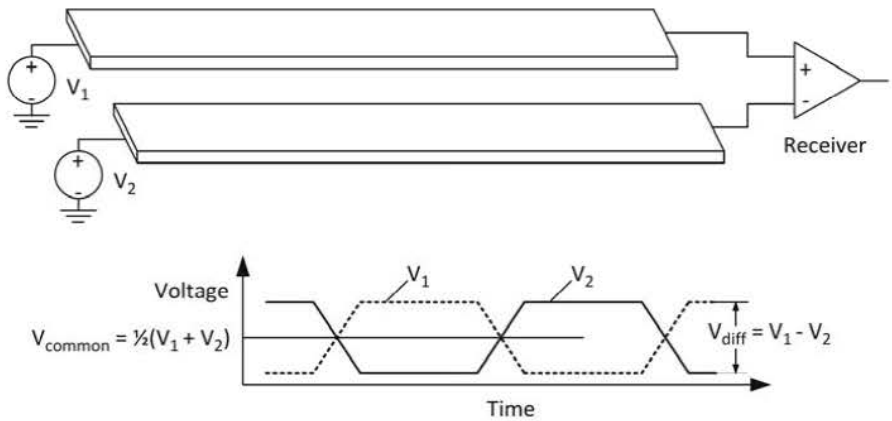


Fig. 9.1 Example of differential signaling. V_1 and V_2 are the differential signals, and their voltages over time are shown on the bottom

for return currents for unrelated signals, etc.). Is there an alternative to using a ground plane?

It turns out there is. In this alternative strategy, each signal is routed with its dual, so that when one switches high, the other signal switches low. This tends to cancel out ground bounce, and the receiver can operate off of the difference between the two signals. Most importantly, each signal now has a dedicated return-current path and so the signals stay "cleaner." It sounds strange, but by routing a signal's dual near the signal, the dual of the signal functions as a return path. Since the ground plane is no longer needed for signal propagation, the pair of signals is relatively robust if the surroundings change, such as if there is a slot in the ground plane. In fact, the ground plane can be done away with altogether.

This strategy of routing each signal with its own dual is called *differential signaling*. In fact, low-voltage differential signaling (LVDS) is taking over most modern interconnection standards. USB, PCI Express, SATA, and even the old Firewire standard all use differential signaling. It is the "trick" that enables the modern standards (like SATA) to run at Gbps over cheap cabling.

So what does it look like electrically? Consider a differential pair (sometimes informally called a "diff pair") using signals V_1 and V_2. The routing and voltage levels are illustrated below in Fig. 9.1.

The two signals are routed over a ground plane (or perhaps not – it turns out not to matter much) and go to a receiver. The receiver subtracts the two signals in order to extract the differential signal. The voltage graph in Fig. 9.1 bears some explanation. As V_1 rises, V_2 falls. Thus V_2 is always the opposite (or dual) of V_1.

V_1 is the voltage on line 1, and V_2 is the voltage on line 2. So we have

$$V_{\text{diff}} = V_1 - V_2 \tag{9.1}$$

$$V_{\text{common}} = (V_1+V_2)/2 \tag{9.2}$$

where V_{diff} is the differential voltage and V_{common} is the common voltage. If it is not already clear, V_{diff} is the voltage that carries information, and it is defined as the difference between the two voltages. A logic 1 may be defined as the case where $V_1 > V_2$, for example. The receiver can subtract the two signals (like at the input of an op-amp), and can thereby make a precise voltage measurement. Sometimes the common voltage is zero and sometimes it is some positive DC offset. The receiver uses the differential voltage almost exclusively, and for most practical purposes the common voltage can be ignored.

There are only two situations where the common voltage becomes relevant. First, if the common voltage becomes high enough (or low enough), it saturates the receiver. (Amplifiers do not behave normally if inputs are close to the power or ground voltages.) Second, common voltage can cause unwanted radiated emissions, as will be discussed in more detail below.

So the common voltage can usually be ignored, and this is why differential signaling is so robust. Since the receiver is subtracting the two signals, the system is tolerant to noise and other signal impairments. As long as the two signals are corrupted in the same way (e.g., receive the same noise pulse), the difference between them is still intelligible at the receiver.

To summarize, the recipe for a differential pair is quite simple. First, you need two lines that are identical in length, Z_0, and layout impairments. Second, it is preferable that the lines be close enough to obtain capacitive and inductive coupling.

What Is Z_{diff}?

The characteristic impedance of the differential pair is called Z_{diff} and it turns out to be easy to estimate.

Consider the case when V_1 is rising and V_2 is falling. V_1 is sourcing a current I into its line and V_2 is sinking and equal and opposite current I from its line. This is illustrated in Fig. 9.2.

The voltage swing is doubled, because the falling edge of V_2 is subtracted from the rising edge of V_1, and the two swings are of equal magnitude. That is, the total voltage swing of V_2–V_1 is double the swing of a single line. However, the current looks (to the differential source) like a single loop of current. So the voltage is doubled but the apparent current is not.

$$Z_{diff} = {}^{2V}/_{I} = 2Z_0 \tag{9.3}$$

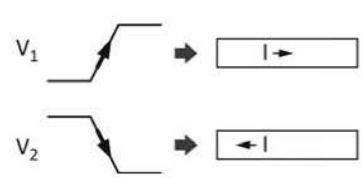


Fig. 9.2 Driving a signal into the differential pair

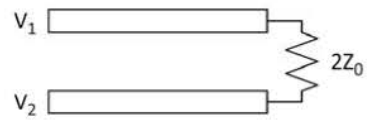


Fig. 9.3 Termination for differential signals. Note that each line has a characteristic impedance of Z_0

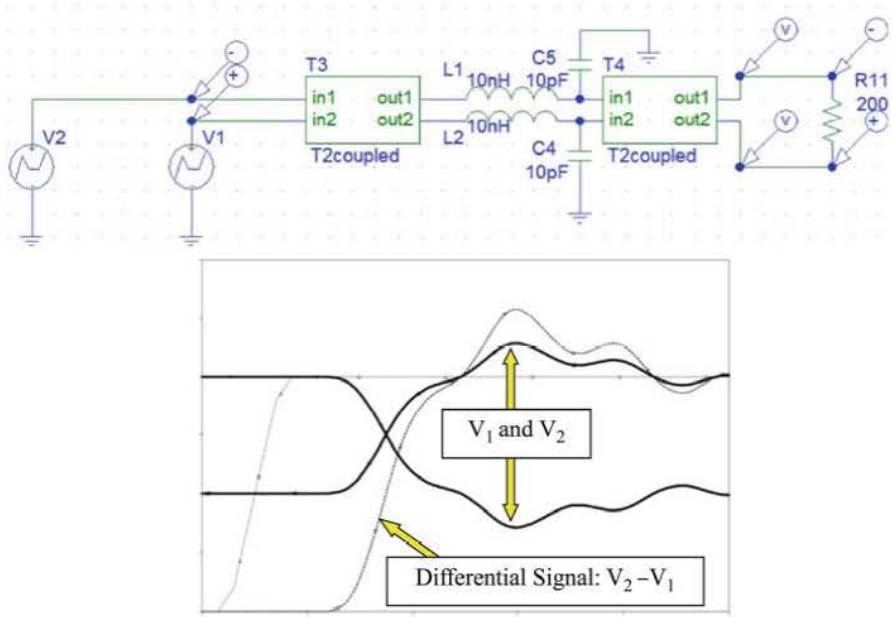


Fig. 9.4 Simulation of differential signal with inductive and capacitive impairments

So $Z_{diff} \approx 2Z_0$. This leads to the termination strategy for differential signaling (Fig. 9.3).

So how well does it work? One can simulate a case where there is a big inductance and big capacitance on both lines. This is shown in Fig. 9.4.

The results are well-behaved, even when one stops to consider that there is a lot of inductance and capacitance in the circuit. The differential signal (shown in the dotted line in Fig. 9.4) shows about a 50% overshoot, but then settles down to a normal level fairly soon. The signals still work correctly, since the receiver is subtracting the two signals, even though there is capacitance and inductance. This is a dramatic improvement over conventional signaling. This example works well because inductance and capacitance are symmetric – each line has the same inductance and capacitance.

What happens if the differential-voltage swing lines are brought closer? Consider the situation from the point of view of V_1. Its own signal is rising, for example, and the opposite signal is falling. V_1 has to charge its own line and, at the same time, combat the negative current introduced from V_2 due to capacitive coupling. In other words, it has to drive harder and source more current in order to obtain the same

voltage swing. The effect is to lower the apparent Z_{diff} "seen" by V_1 and V_2. At extremely close spacing, Z_{diff} can drop as much as 12%.

There are formulas that enable one to approximate Z_{diff} for practical cases such as microstrip and stripline. However, these formulas are very approximate, and so a field solver is often used if precision is needed.

What happens if the ground plane is moved farther from the two signal lines? As the ground plane moves farther, the coupling from the signals to the ground drops, and at some point, becomes negligible. At that point, the differential pair is carrying all of its own return current. This point is typically when the conductor-to-ground spacing is roughly double the conductor-to-conductor spacing. At that point, if it is not already clear, the ground can be removed altogether with little observable difference.

While removing the ground plane may not make much sense on a circuit board, since it always has a ground plane, this situation is common in cabling. There, the "ground" can be an extra, grounded line on the outside of the cable called a "shield." The shield can be removed with little impairment. This is what Ethernet uses – unshielded twisted pair (UTP). This is important because, as one might guess, the unshielded cable is cheaper than shielded.

So are there situations in which differential propagation is less than ideal? There are two common cases, and their origins are unexpected.

Obstacles: Clock Jitter and Common-Mode Signals

One of the more subtle requirements for differential signaling to work correctly is that the two signals (V_1 and V_2) have to be synchronized in time. Backing up, there can be several reasons why the two are not in sync. For example, one line may be longer than the other line. There may be some asymmetry, such as a test pad or via. As we have noted, both of these cases are actually mistakes – recall that the line carrying V_1 needs to be as identical to the line carrying V_2 as possible. A more common cause of nonsynchronized signals lies inside the chip that creates the signals. Consider the clock signals inside the chip that drive the circuits that create V_1 and V_2 – if there is any difference in the clocks (a situation called *clock skew*), then the two signals will exit the chip at slightly different times.

So, consider what happens if the two outputs (V_1 and V_2) are out of sync. One of the outputs changes before the other, and so the differential voltage acquires a hiccup. This is illustrated below in Fig. 9.5.

Figure 9.5 illustrates what happens, but it needs some explanation.

In the top case, the two voltages are in sync. The common voltage is a DC voltage (halfway between V_1 and V_2) and the differential voltage is double the swing of a single signal.

In the bottom case, V_2 is delayed relative to V_1.

Consider the common voltage – V_1 rises first and then V_2 falls. The average of the two voltages first rises (one voltage is rising while the other stays the same) and then

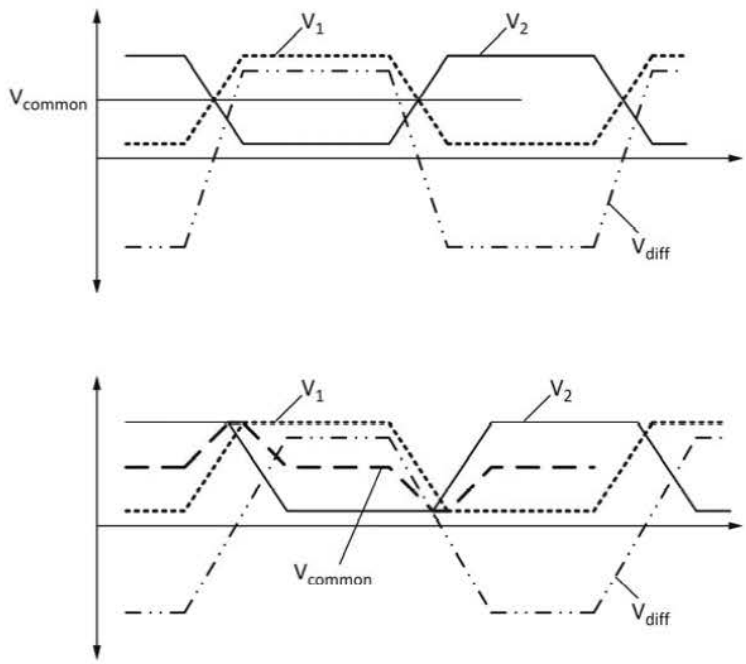


Fig. 9.5 Effects of clock skew on differential signals. (Top) Signal with no skew. (Bottom) Signal with V_2 slower than V_1

falls (the later voltage then falls). The common-mode voltage now has a small positive triangle-shaped "blip" in it, the width of which is roughly proportional to the skew (time difference) between the two differential voltages. Later, when V_1 falls, the "blip" is of opposite polarity because V_1 falls before V_2 rises.

Consider the differential voltage – When only one output switches, the differential voltage undergoes half of a voltage swing. Later, the other output switches and the remaining half-voltage-swing occurs. This creates a spread-out differential waveform, which looks like a slower-rise-time version of the original waveform. This is, in itself, not necessarily bad – it is the same signal but with a slightly slower rise and fall time.

So the result of a time delay between differential outputs is a train of impulses seen on the common-mode voltage and a slower rise and fall time seen on the differential voltage. The slower rise and fall time is likely negligible, but what does a train of impulses look like in the frequency domain? A train of impulses! The result is signal energy out to the clock frequency of the signal. A 1 Gbps SATA cable radiates out to 1 GHz.

And why is this a problem? First, the common-mode energy makes the cable act like a very large antenna. Second, there is no termination strategy for the common-mode signals. If you look back up to Fig. 9.3, it should be apparent that each of the lines, considered individually, is completely unterminated! So the

impulses tend to run up and down the cable like a terrified squirrel at a wedding reception.

There are basically four ways to deal with this situation. The first is to eliminate asymmetries and skew. This is what is known in the industry as a root-cause fix – it eliminates the source of the problem. The second is to add back the shield. The shield can potentially help (e.g., it is required by the SATA standard) but it must be extremely well-grounded or else the shield will reradiate. (If it is not properly grounded, the shield turns back into a hunk of metal and therefore a big antenna.) The third option is to add a ferrite bead to the cable. The ferrite bead has the property that only signals that cancel themselves out (signals that are accompanied by a return path) do not "see" the inductance of the bead. So stray signals do "see" this inductance and see a much higher impedance. In other words, this train of pulses is attenuated by the inductance of the ferrite bead. This is discussed in more depth in the chapter on electromagnetic interference (EMI), but suffice to say it makes the cable cost more. Keep in mind that if you only pass EMI testing by using a cable with a ferrite, you have to ship a cable with a ferrite with every unit you sell. The unit is only legal for sale if it is accompanied with whatever special cabling was needed to make it pass the test.

The fourth way is to modify the termination so as to terminate common-mode voltage transients.

Termination Strategies Revisited

Consider a simulation of a differential signal with significant clock skew, shown in Fig. 9.6.

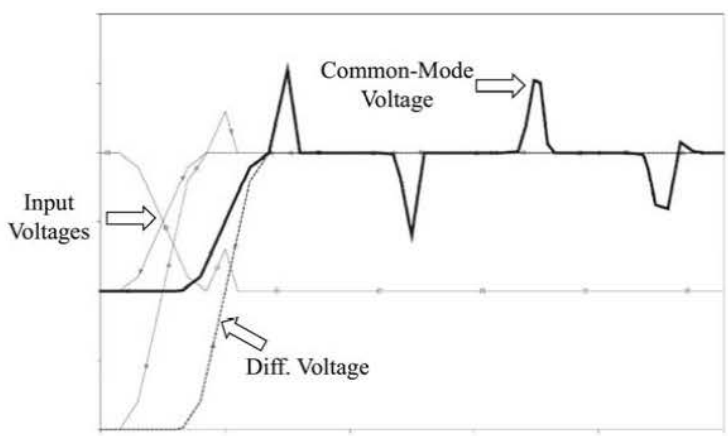


Fig. 9.6 Simulation of the case shown in Fig. 9.5

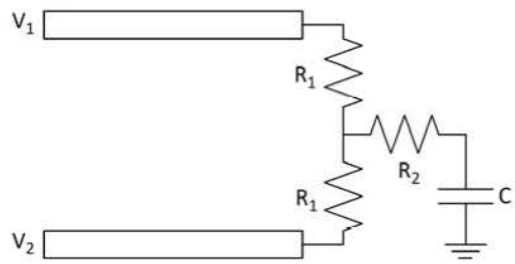


Fig. 9.7 Modified termination strategy

The simulation results are dramatic – the impulse train can clearly be seen. The results can be a little confusing. The input voltages shown in the figure are the two differential voltages V_1 and V_2. The differential voltage is $V_2 - V_1$, and is actually very well-behaved. In other words, this arrangement would function normally! The common-mode voltage is the voltage measured on one of the two differential lines, in this case V_2. This highlights what happens when each of the differential voltages is unterminated – a single transient results in a train of impulses due to reflections. Each of the single lines in the pair has the same train of impulses, and so the impulses subtract out when the differential voltage is calculated. However, they are present on the line and can cause radiated emissions.

So what is the fix? The solution is to add a common-mode termination path to the differential termination. In other words, one adds termination to each of the individual transmission lines. By adding a series resistor and capacitor at the halfway point, the differential signal still "sees" the same resistance, but each individual line sees half the differential resistance (i.e., it sees the correct single-signal terminating resistance) in series with the added resistor and capacitor. This is illustrated below in Fig. 9.7.

As shown in the figure, the differential termination has been split into two equal resistors labeled R_1 and the resistor R_2 and capacitor C have been added. For effective termination, there are some rules of thumb.

$$R_1 = 0.9Z_0 \tag{9.4}$$

$$R_2 = 0.1Z_0 \tag{9.5}$$

$$C = \frac{5t_r}{Z_0/2} = \frac{10t_r}{Z_0} \tag{9.6}$$

This if for the signal rise time t_r. If t_r is in ps, then C is in pF.

Example 9.1

Consider a differential pair with $Z_0 = 100\ \Omega$ (and therefore $Z_{diff} = 200\ \Omega$). The pair is driving a signal with a rise time of 800 ps.

(a) Calculate the resistors to terminate both differential and common-mode signals.
$R_1 = 0.9Z_0 = 90\ \Omega$. $R_2 = 0.1Z_0 = 10\ \Omega$. $C = (10 * 800\ \text{ps})/100\ \Omega = 80\ pF$..

(b) Simulate the result.

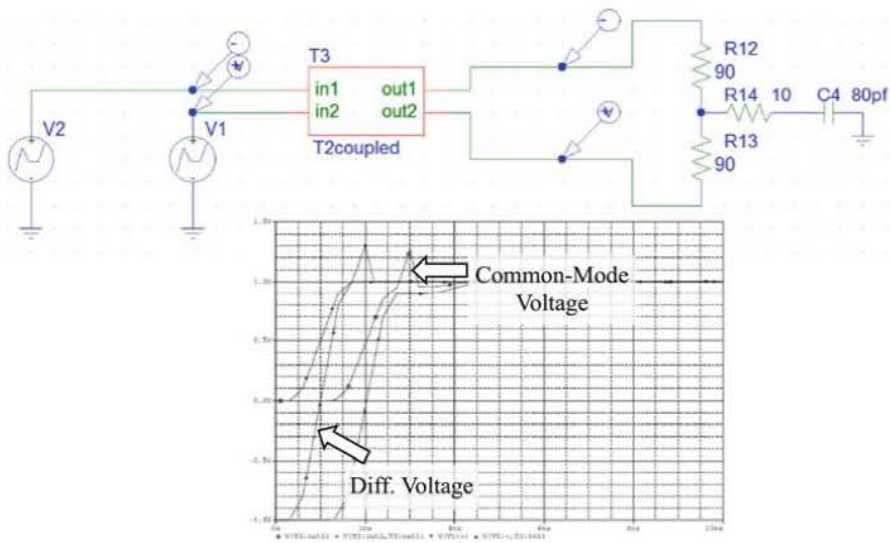


Fig. 9.8 Simulation of common-mode termination strategy

See Fig. 9.8 below.

As the simulation shows, the addition of the series resistor and capacitor has eliminated the train of impulses. (Contrast with Fig. 9.6.) There is only a single impulse because now the line is terminated, and the pulses do not reflect. However, there is still a single impulse, which shows that eliminating skew is still preferable to termination.

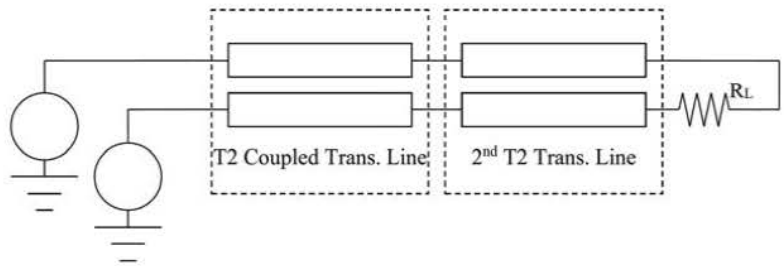
Appendix

Construct a SPICE simulation of two coupled dual transmission lines with LEN = 1, $L = 100$ nH, $C = 10$ pF, and LM = CM = 0. Connect the lines serially and connect a voltage source (800 ps rise time) at one end and a single terminating resistor across the outputs of the lines. (See the diagram below.)

1. Calculate the Z_0 of each line. For $R_S = 0$ (i.e., no source resistance) and $R_L = 2Z_0$, simulate the signal and show that the signal has "perfect" signal integrity.
2. Add a 2 nH inductor in series and a 5 pF shunt capacitor (to ground) to *one* of the signal lines at the midpoint between the two sets of lines. Which is affected more – the differential signal or the signal of a single line?
3. Restore the setup of the first problem (i.e., eliminate the L and C) and instead create 200 ps of clock skew between the two voltage sources. Do this by making

one of the differential signals switch 200 ps after the other one. What effect does it have on the differential signal? And on each of the individual signals?

4. Restore the setup of the first problem and instead change LM to 10 nH and CM to 5 pF. What effect does this have? How do you explain the result?



Example: Schematic of Problem 1 and Problem 4 Plot the voltage at each source, the differential voltage across the inputs (measured at the input to the first set of coupled transmission lines), and the differential voltage across the output (measured at the output of the second set of coupled transmission lines).

Consider a system that has a connector for a high-speed differential signal with a rise time of 200 ps. The system needs to terminate the signal correctly.

5. If the trace carrying the signal has an impedance of 75 Ω, what is the differential impedance needed for the signal? 150 Ω.
6. Calculate the R and C that needs to be added to the termination in order to terminate common-mode signals correctly. (See the example at the end of the lecture.) $R = 0.9 * 75 = 67.5$ Ωs. $R2 = 0.1 * 75 = 7.5$ Ω. $C = 10 * 200/75 = 26.7$ pF.

Index

A
AC signal, 4
Actual decoupling capacitors, 152
Added component mode, 210
Aluminum-electrolytic capacitors, 189
American Wire Gauge, 46
Amortization, 228
Attenuation, 107, 108

B
Bandlimit, 8, 9
Baseboard management controller (BMC), 208
B field, 2–4
Bill of materials cost (BOM) cost, 227
Bypass capacitor, 72, 73, 141, 144, 147
  selection, 151–153

C
Cabling, 163
Capacitance
  dielectric constant, 50
  fringing, 51
  IPC equation
    for microstrip capacitance, 52
    for stripline capacitance, 52
  parallel plate capacitor equation, 51
  relative dielectric constant, 50
Capacitor, 100, 142
  definition, 48
  model, 38
  properties, 49
  SPICE model, 39
Cell phones, 157
Ceramic capacitor, 189
Certification, 227
Charge, 83
Circuit board
  bare, 22
  CMOS, 34
  component packaging, 23, 24
  components, 22
  designing, 19–22
  electrical modeling, 35–37
  heat dissipation, 32
  layer stack, 19
  microstrip, 19
  pick-and-place machines, 24
  power and ground plane, 19
  power dissipation, 32
  purpose, 17
  selective-solder pallet, 26
  stack up of signals, 18
  stripline, 19
  surface-mount component, 22, 23
  surface-mount process, 25
  through-hole board, 26
  through-hole component, 23
  transistors
    finite drive strength, 30
    heat, 30
    interconnection between transistors, 30
    of nonzero size, 29
  wave solder machine, 25
Circuit-board attacks and security
  circuit board recap, 199
  design, fabrication and assembly, 200, 201
  example attacks, 201–205
  first indication of trouble, 199, 200

© Springer Nature Switzerland AG 2022

S. H. Russ, *Signal Integrity*, https://doi.org/10.1007/978-3-030-86927-4

Circuit-board attacks and security (*cont.*)
processor-based design, 207, 208
taxonomy of attacks, 209–212
Circuit-board design, 157–160, 194
component selection, 188–191
design files, 196
Circuit-board layout, 173
Circular magnetic-field loops, 53
Clock jitter
and common-mode signals, 121–123
Clocks and clock sources
clock jitter, 176, 177
layout reduction, 180, 181
sources of, 178
clock period, 176
clock signal, 175
clock skew, 176
crystal and amplification, 179
eye diagram, 182–184
phase-error measurement circuits and low-pass filters, 180
voltage-controlled oscillator, 179
CMOS integrated circuits, 168
Coaxial, 82
Combine rise times, 13
Common-mode termination strategy, 125
Common-mode voltage, 122
Complementary metal-oxide-semiconductor (CMOS), 34, 35
Conducted emissions, 155
Conductor, 44
Conversion cost, 228
Copper, 18–21, 27, 45
Cost risk, 232
Coupled-noise issues, 178
Craft ports, 208
Critical length, 11
Critical path, 231
Crosstalk, 127, 128
diagnosing, 134, 135
estimating and reducing crosstalk, 133, 134
near-end versus far-end crosstalk, 129–133
Crowbar circuit, 171, 172
Crowbar devices, 171
Crystal oscillator, 179
Current, 83
definition, 2
Cycle-to-cycle jitter, 177

D

DC-isolating circuit, 172
Debugging
mass production, design for, 217–223
Degraded eye diagram, 183
Delta-I noise, *see* Switching noise
Depreciation, 228
Design
commercial aspects of, 227–230
legal aspects of, 225–227
Design file attack, 209
Design files, 196
Design qualification testing, 217
Design verification testing, 217
Deterministic jitter, 185
DFX, 218, 219
Dielectric constant, 6, 30, 49, 50, 62
Dielectric loss, 109, 110
affect ε, 110–112
Dielectrics, 6
Differential signaling, 117, 118
clock jitter and common-mode signals, 121–123
termination strategies revisited, 123–125
Zdiff, 119–121
Differential voltage, 120, 122
Digital integrated circuit, 140
Dirty ground, 69
Dirty power, 69
Discharged charge, 67
Distance, 4, 5, 7
Distributed analysis
transmission line, 81–86
transmission line, approximate model of, 87, 88
Z_0, 86, 87
Double-pulse waveform, 67, 69–71
Drive strength, 30
Driving gate, 66, 67, 69, 72, 73

E

E field, 2–4, 6
Einstein's theory, 81
Electrical engineering, 1, 8
Electrical modeling, 35–37
Electromagnetic interference (EMI)/electromagnetic compatibility (EMC)
cabling, 163, 164
chassis, 161, 162
circuit-board design, 157–161
legal certification, 164–166
Electromagnetic pulse (EMP), 172
Electron tunneling, 29
Electrostatic discharge (ESD), 167, 168
components to, 170–172
design techniques to, 173

and lightning standards and testing, 169, 170
occurs in, 169
Energy dissipation, 171
Engineering careers, 232, 233
Equivalent series inductance (ESL), 38
Equivalent series resistance (ESR), 38
ESD-management device, 170
Ethernet, 121

F
Fall time, 7
Falling-edge waveform, 67
Faraday's Law, 56
Far-end crosstalk, 129–133
Federal Communications Commission (FCC), 164
Fiberglass circuit, 3
Filter capacitor, 144
Flowing solder, 25
Footprint, 21
Four-layer board, 18–20
FPGA board, 71
Frequency-dependent attenuation, 109
Frequency domain, 140, 141, 146, 147
Frequency-independent and frequency-dependent attenuation, 109
Fringing, 51

G
Gantt chart, 230, 231
Gauge, 46
Gaussian distribution, 68
Gerber-editing software, 202
Gerber file format, 196
Gross margin, 228
Gross profit, 228
Ground bounce
changing gate, 70
double-pulse, 71
equivalent circuit, 66
minimizing, 72, 73
pulse, 70
Ground loop, 163
Ground plane, 18–20, 27
Guard traces, 133

H
Half-logic-swing waveform, 100
HDMI connector, 221
Heat dissipation, 31, 32
Heatsink, 32, 33
Hertz, Heinrich, 81
Hi-pot tested, 226

I
IBIS models, 36
IC modeling
capacitor model, 38
ESL, 38
ESR, 38
impedance of a capacitor versus frequency, 37
limits of, 40
process, 37
SPICE model of capacitor, 39
SPICE simulation results, 39
SRF, 38
Impedance, 3
Inductance
calculation, 57, 58
current and magnetic field, 53
magnetic flux, 56
mutual inductance, 55–59, 62
relation of current, magnetic flux and, 54
self-inductance, 55–58, 62
Insulator, 3, 6, 15
Integrated circuits, 180
Intel Xeon processor, 30
Intentional radiation, 157
Intentional radiator, 156
Interconnection between transistors, 30
Intrinsic capacitance, 111

K
Knee frequency, 8, 9, 15, 162

L
Latch-up, 168
Layer stack, 19
Layout, 193, 209
Lightning immunity, 172
Lightning standards
ESD, 169, 170
Lightning strikc, 172
Lightning-protection device, 170
Load termination, 96
Logic gates, 29–31, 35
Log-log impedance graph, 144
Lossy transmission lines

Lossy transmission lines (*cont.*)
attenuation, 107, 108
dielectric loss, 109, 110
affect signal propagation, 112–115
affect ε, 110–112
frequency-dependent attenuation, 109
Low-pass filter, 12, 14
Low-voltage differential signaling (LVDS), 118
Lumped analysis, 65
Lumped model of a transmission line, 88
Lumped versus distributed circuit, 10–12

M
Magnetic field, 2
Magnetic flux, 53–57
Markings, 195
Mass production, design for, 217–223
Masters in business administration (MBA), 233
Maxwell's equations, 4, 81, 85
Memory cells, 140
Metal-oxide varistors (MOVs), 171
Microstrip, 19
Minizap, 169
Moating, 161
Modern integrated circuits, 179
Moving air, 32
Mutual inductance, 55–59, 62, 127

N
Nationally Recognized Test Lab (NRTL), 226
N-channel FET, 34
N-channel transistor, 34
Near-end crosstalk, 129–133
Nonrecurring expenditures (NREs), 228
Nonzero permeability, 55
Nuclear power industry, 227

O
Oscilloscope, 77, 215–217

P
Parallel C-series RL resonance, 143
Parallel plate capacitor equation, 51
Passive components, 22
P-channel FET, 34
P-channel transistor, 34
Period jitter, 177
Phase-locked loops (PLLs), 179
Photoresist, 20
Pick-and-place machines, 24
Piezoelectric effect, 179
Plated through-hole, 20
Polarity marking, 22, 195
Population option, 26
Power consumption, 140
Power dissipation, 31, 32
Power distribution network, 139–141
design strategy, 144–148
simulation role, 149–151
power supplies and resonance, 141–144
selecting bypass capacitors, 151, 152
Power plane, 18–20
Power supplies and resonance, 141–144
Power supply, 34, 35, 37
Power-supply impedance, 145
Power-supply network (PSN), 149
Pre-preg, 20
Printed circuit board (PCB), *see* Circuit board
Printed wiring board (PWB), *see* Circuit board
Processor-based design, 207, 208
Project management, 230, 231
Public switched telephone network (PSTN), 164

Q
Quality factor (Q), 73

R
Radiated emissions, 155
Recurring expenditures, 228
Reference design, 192
Reference designators, 194
Reflection coefficient at the load, 93
Reflow oven, 25
Relative dielectric constant, 49, 50
Relative permittivity, 6
Resistance
definition, 46
sheet resistance, 47
Resistivity, 45
Resistor, 4, 12, 191
Resonances, 142
Resonant frequency, 74, 76
Return current, 35, 58, 59
Reverse crosstalk, 131
Rework attack, 209
RF energy, 157
Right-hand rule, 53
Ringing
minimizing, 77

resonant frequency, 74
results of, 75, 76
seen and not seen on signals, 77
SPICE simulation, 74
Rise time, 7, 12–14, 69, 73
Rising edge, 7–12, 131
Risk mitigation, 231, 232
Routing, 194
RS-274X, 196

S
Sarbanes-Oxley Act (SOX), 229
Schedule risk, 232
Schematic attack, 209
Selective-solder pallet, 26, 27
Self-inductance, 55–58, 62
Self-resetting board, 71
Self-resonant frequency (SRF), 38
Semiconductors, 171
Sensitivity analysis, 40
Serial ATA (SATA), 2, 169
Series RLC resonance, 142
Series-source termination, 97
Sheet resistance, 47, 61
Signal integrity, 1, 17, 182
Signal propagation, 112–115
Silkscreen, 21
Silver, 18
Simplest gate model, 34, 35
Skin depth, 60, 62
Skin effect, 59, 60
Solder mask, 21
Source termination, 97
Spark gaps, 173
Speed of light, 4, 5, 7, 10–12, 14
Speed of propagation, 3, 6, 12
SPICE model, 36, 38
SPICE simulation, 100, 149
SPICE simulation of a capacitive imperfection, 99
SPICE simulation of an inductive imperfection, 101
Static gate, 70
Stripline, 19
Surface-mount components, 21
Surface-mount part, 22
Surface-mount process, 25
Susceptibility, 155
Switch, 34
Switching logic gates, 35
Switching noise, 72

T
Tantalum capacitor, 189
Taxonomy of attacks, 209–212
Technical risk, 231
Terminating impedance, 96–99
Termination strategies revisited, 123–125
Testing
oscilloscope, 215–217
Thermal coefficient, 32, 33
Through-hole board, 26
Through-hole component, 21, 22
Time, 4, 5, 7, 8, 12, 13
Tolerances, 40
Trace, 20
Transient voltage suppression (TVS) diodes, 171
Transmission, 91
Transmission coefficient at the load, 93
Transmission line, 11, 77, 81–86
approximate model of, 87, 88
departures from the ideal, 99, 100
fixing, 101, 102
reflection and transmission, 91–96
terminating impedance, 96–99

U
Uncorrelated jitter, 178
Underwriters Laboratory (UL), 226
Unintentional radiator, 156
Unshielded twisted pair (UTP), 121

V
Volt, 1
definition, 2
Voltage drop, 2
Voltage gain, 2
Voltage swing, 119

W
Wave solder machine, 25
Wire wrap, 17

X
X-acto®, 216

Z
Z_0, 83–87
Zdiff, 119–121