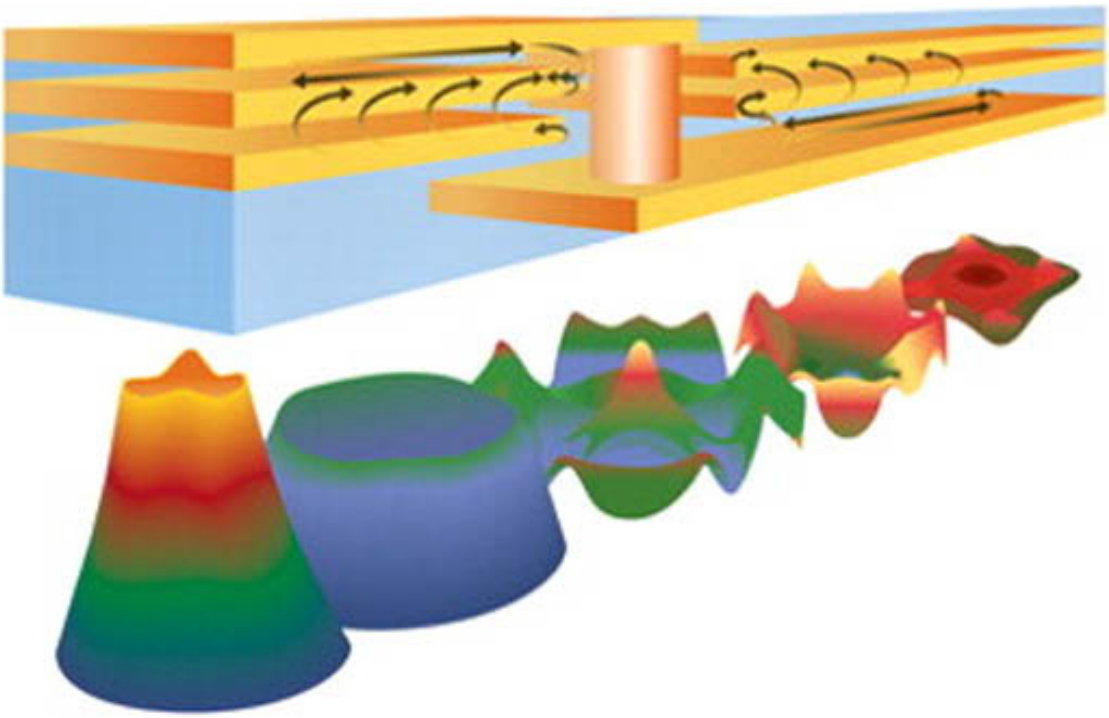
PRENTICE HALL

Signal and Power Integrity—SIMPLIFIED

Second Edition



Eric Bogatin

Prentice Hall Modern Semiconductor Design Series

Prentice Hall Signal Integrity Library

Signal and Power Integrity—

Simplified

Second Edition

Prentice Hall Modern Semiconductor Design Series

James R. Armstrong and F. Gail Gray
VHDL Design Representation and Synthesis

Mark Gordon Arnold
Verilog Digital Computer Design: Algorithms into Hardware

Jayaram Bhasker
A VHDL Primer, Third Edition

Mark D. Birnbaum
Essential Electronic Design Automation (EDA)

Eric Bogatin
Signal and Power Integrity—Simplified, Second Edition

Douglas Brooks
Signal Integrity Issues and Printed Circuit Board Design

Ken Coffman
Real World FPGA Design with Verilog

Alfred Crouch
Design-for-Test for Digital IC's and Embedded Core Systems

Dennis Derickson and Marcus Müller (Editors)
Digital Communications Test and Measurement

Greg Edlund
Timing Analysis and Simulation for Signal Integrity Engineers

Daniel Foty
MOSFET Modeling with SPICE: Principles and Practice

Tom Granberg
Handbook of Digital Techniques for High-Speed Design

Nigel Horspool and Peter Gorman
The ASIC Handbook

William K. Lam
Hardware Design Verification

Geoff Lawday, David Ireland, and Greg Edlund
A Signal Integrity Engineer's Companion

Mike Peng Li
Jitter, Noise, and Signal Integrity at High-Speed

Farzad Nekoogar and Faranak Nekoogar
From ASICs to SOCs: A Practical Approach

Farzad Nekoogar
Timing Verification of Application-Specific Integrated Circuits (ASICs)

Samir Palnitkar
Design Verification with e

David Pellerin and Scott Thibault
Practical FPGA Programming in C

Christopher T. Robertson
Printed Circuit Board Designer's Reference: Basics

Chris Rowen
Engineering the Complex SOC

Madhavan Swaminathan and A. Ege Engin
Power Integrity Modeling and Design for Semiconductors and Systems

Wayne Wolf
FPGA-Based System Design

Wayne Wolf
Modern VLSI Design, Fourth Edition: IP-Based Design

Signal and Power Integrity—

Simplified

Second Edition

Eric Bogatin

Upper Saddle River, NJ • Boston • Indianapolis • San Francisco
New York • Toronto • Montreal • London • Munich • Paris • Madrid
Capetown • Sydney • Tokyo • Singapore • Mexico City

Library of Congress Cataloging-in-Publication Data
Bogatin, Eric.
Signal and power integrity—simplified / Eric Bogatin. —2nd ed.
p. cm.
Rev. ed. of: Signal and power integrity—simplified / Eric Bogatin. c2004.
Includes bibliographical references and index.
ISBN-10: 0-13-234979-5 (hardcover : alk. paper)
ISBN-13: 978-0-13-234979-6 (hardcover : alk. paper)
1. Signal processing—Digital techniques. 2. Impedance (Electricity) 3. Signal integrity (Electronics) I. Bogatin, Eric. Signal integrity—simplified. II. Title.
TK5102.9.B58 2010
621.382'2—dc22 2009021261

ISBN-13: 978-0-13-234979-6
ISBN-10: 0-13-234979-5
Text printed in the United States on recycled paper at Edwards Brothers in Ann Arbor, Michigan.
First printing, July 2009

The dedication of a book is called "The Dedication" because it requires dedication by the author to complete the work. As every author will tell you, writing is a solitary, alone process, the opposite of social activity. It's easy to become engulfed in the writing and researching and exclude more and more of the real world. Successful authors are either not married or married to an understanding, supportive spouse who sees his or her role as providing a nurturing environment in which creative juices can ferment.

Susan, my wife, patiently put up with my solitary writing, giving me the space to put in the dedication to finish the first and second editions. At the same time, she was also my anchor to the real world and forced me to keep a healthy balance between alone work and social life. The second edition is as much due to her efforts as from mine, and as I get to write the words, I am dedicating the second edition to her.

This page intentionally left blank

CONTENTS

Preface to the Second Edition xv
Preface to the First Edition xvii
Acknowledgments xxiii
About the Author xxv

Chapter 1 Signal Integrity Is in Your Future 1
1.1 What Is Signal Integrity? 2
1.2 Signal Quality on a Single Net 5
1.3 Cross Talk 9
1.4 Rail-Collapse Noise 11
1.5 Electromagnetic Interference (EMI) 13
1.6 Two Important Signal Integrity Generalizations 16
1.7 Trends in Electronic Products 16
1.8 The Need for a New Design Methodology 22
1.9 A New Product Design Methodology 23
1.10 Simulations 24
1.11 Modeling and Models 27
1.12 Creating Circuit Models from Calculation 30
1.13 Three Types of Measurements 35
1.14 The Role of Measurements 38
1.15 The Bottom Line 39

Chapter 2 Time and Frequency Domains 43
2.1 The Time Domain 44
2.2 Sine Waves in the Frequency Domain 46
2.3 Shorter Time to a Solution in the Frequency Domain 48
2.4 Sine Wave Features 49
2.5 The Fourier Transform 51
2.6 The Spectrum of a Repetitive Signal 53

2.7 The Spectrum of an Ideal Square Wave 55
2.8 From the Frequency Domain to the Time Domain 57
2.9 Effect of Bandwidth on Rise Time 58
2.10 Bandwidth and Rise Time 62
2.11 What Does *Significant* Mean? 63
2.12 Bandwidth of Real Signals 67
2.13 Bandwidth and Clock Frequency 68
2.14 Bandwidth of a Measurement 70
2.15 Bandwidth of a Model 72
2.16 Bandwidth of an Interconnect 74
2.17 The Bottom Line 78

Chapter 3 Impedance and Electrical Models 81
3.1 Describing Signal-Integrity Solutions in Terms of Impedance 82
3.2 What Is Impedance? 84
3.3 Real Versus Ideal Circuit Elements 86
3.4 Impedance of an Ideal Resistor in the Time Domain 88
3.5 Impedance of an Ideal Capacitor in the Time Domain 89
3.6 Impedance of an Ideal Inductor in the Time Domain 92
3.7 Impedance in the Frequency Domain 94
3.8 Equivalent Electrical Circuit Models 99
3.9 Circuit Theory and SPICE 101
3.10 Introduction to Modeling 105
3.11 The Bottom Line 110

Chapter 4 The Physical Basis of Resistance 113
4.1 Translating Physical Design into Electrical Performance 114
4.2 The Only Good Approximation for the Resistance of Interconnects 115
4.3 Bulk Resistivity 118
4.4 Resistance per Length 119
4.5 Sheet Resistance 121
4.6 The Bottom Line 124

Chapter 5 The Physical Basis of Capacitance 127
5.1 Current Flow in Capacitors 128
5.2 The Capacitance of a Sphere 130
5.3 Parallel Plate Approximation 131
5.4 Dielectric Constant 133

5.5 Power and Ground Planes and Decoupling Capacitance 135
5.6 Capacitance per Length 138
5.7 2D Field Solvers 143
5.8 Effective Dielectric Constant 146
5.9 The Bottom Line 150

Chapter 6 The Physical Basis of Inductance 151
6.1 What Is Inductance? 151
6.2 Inductance Principle #1: There Are Circular Rings of Magnetic-Field Lines Around All Currents 152
6.3 Inductance Principle #2: Inductance Is the Number of Webers of Field Line Rings Around a Conductor per Amp of Current Through It 154
6.4 Self-Inductance and Mutual Inductance 156
6.5 Inductance Principle #3: When the Number of Field Line Rings Around a Conductor Changes, There Will Be a Voltage Induced Across the Ends of the Conductor 158
6.6 Partial Inductance 161
6.7 Effective, Total, or Net Inductance and Ground Bounce 167
6.8 Loop Self- and Mutual Inductance 173
6.9 The Power-Distribution Network (PDN) and Loop Inductance 178
6.10 Loop Inductance per Square of Planes 183
6.11 Loop Inductance of Planes and Via Contacts 184
6.12 Loop Inductance of Planes with a Field of Clearance Holes 187
6.13 Loop Mutual Inductance 188
6.14 Equivalent Inductance of Multiple Inductors 189
6.15 Summary of Inductance 191
6.16 Current Distributions and Skin Depth 193
6.17 High-Permeability Materials 202
6.18 Eddy Currents 204
6.19 The Bottom Line 206

Chapter 7 The Physical Basis of Transmission Lines 209
7.1 Forget the Word *Ground* 210
7.2 The Signal 211
7.3 Uniform Transmission Lines 212
7.4 The Speed of Electrons in Copper 214
7.5 The Speed of a Signal in a Transmission Line 215
7.6 Spatial Extent of the Leading Edge 219

7.7 "Be the Signal" 220
7.8 The Instantaneous Impedance of a Transmission Line 224
7.9 Characteristic Impedance and Controlled Impedance 227
7.10 Famous Characteristic Impedances 230
7.11 *The* Impedance of a Transmission Line 233
7.12 Driving a Transmission Line 238
7.13 Return Paths 241
7.14 When Return Paths Switch Reference Planes 244
7.15 A First-Order Model of a Transmission Line 257
7.16 Calculating Characteristic Impedance with Approximations 262
7.17 Calculating the Characteristic Impedance with a 2D Field Solver 265
7.18 An n-Section Lumped Circuit Model 270
7.19 Frequency Variation of the Characteristic Impedance 278
7.20 The Bottom Line 279

Chapter 8 Transmission Lines and Reflections 281

8.1 Reflections at Impedance Changes 282
8.2 Why Are There Reflections? 284
8.3 Reflections from Resistive Loads 288
8.4 Source Impedance 290
8.5 Bounce Diagrams 292
8.6 Simulating Reflected Waveforms 295
8.7 Measuring Reflections with a TDR 295
8.8 Transmission Lines and Unintentional Discontinuities 299
8.9 When to Terminate 301
8.10 The Most Common Termination Strategy for Point-to-Point Topology 304
8.11 Reflections from Short Series Transmission Lines 306
8.12 Reflections from Short-Stub Transmission Lines 309
8.13 Reflections from Capacitive End Terminations 311
8.14 Reflections from Capacitive Loads in the Middle of a Trace 314
8.15 Capacitive Delay Adders 317
8.16 Effects of Corners and Vias 319
8.17 Loaded Lines 325
8.18 Reflections from Inductive Discontinuities 327
8.19 Compensation 331
8.20 The Bottom Line 334

Chapter 9 Lossy Lines, Rise-Time Degradation, and Material Properties 337

9.1 Why Worry About Lossy Lines? 338
9.2 Losses in Transmission Lines 340
9.3 Sources of Loss: Conductor Resistance and Skin Depth 342
9.4 Sources of Loss: The Dielectric 346
9.5 Dissipation Factor 351
9.6 The Real Meaning of Dissipation Factor 354
9.7 Modeling Lossy Transmission Lines 360
9.8 Characteristic Impedance of a Lossy Transmission Line 367
9.9 Signal Velocity in a Lossy Transmission Line 369
9.10 Attenuation and the dB 371
9.11 Attenuation in Lossy Lines 376
9.12 Measured Properties of a Lossy Line in the Frequency Domain 385
9.13 The Bandwidth of an Interconnect 390
9.14 Time-Domain Behavior of Lossy Lines 397
9.15 Improving the Eye Diagram of a Transmission Line 400
9.16 Pre-emphasis and Equalization 402
9.17 The Bottom Line 403

Chapter 10 Cross Talk in Transmission Lines 405

10.1 Superposition 406
10.2 Origin of Coupling: Capacitance and Inductance 407
10.3 Cross Talk in Transmission Lines: NEXT and FEXT 409
10.4 Describing Cross Talk 411
10.5 The SPICE Capacitance Matrix 413
10.6 The Maxwell Capacitance Matrix and 2D Field Solvers 417
10.7 The Inductance Matrix 424
10.8 Cross Talk in Uniform Transmission Lines and Saturation Length 425
10.9 Capacitively Coupled Currents 431
10.10 Inductively Coupled Currents 435
10.11 Near-End Cross Talk 438
10.12 Far-End Cross Talk 441
10.13 Decreasing Far-End Cross Talk 448
10.14 Simulating Cross Talk 451
10.15 Guard Traces 457
10.16 Cross Talk and Dielectric Constant 464
10.17 Cross Talk and Timing 466

10.18 Switching Noise 469
10.19 Summary of Reducing Cross Talk 473
10.20 The Bottom Line 474

Chapter 11 Differential Pairs and Differential Impedance 475

11.1 Differential Signaling 476
11.2 A Differential Pair 480
11.3 Differential Impedance with No Coupling 482
11.4 The Impact from Coupling 486
11.5 Calculating Differential Impedance 493
11.6 The Return-Current Distribution in a Differential Pair 496
11.7 Odd and Even Modes 502
11.8 Differential Impedance and Odd-Mode Impedance 507
11.9 Common Impedance and Even-Mode Impedance 508
11.10 Differential and Common Signals and Odd- and Even-Mode Voltage Components 511
11.11 Velocity of Each Mode and Far-End Cross Talk 513
11.12 Ideal Coupled Transmission-Line Model or an Ideal Differential Pair 519
11.13 Measuring Even- and Odd-Mode Impedance 520
11.14 Terminating Differential and Common Signals 522
11.15 Conversion of Differential to Common Signals 529
11.16 EMI and Common Signals 534
11.17 Cross Talk in Differential Pairs 539
11.18 Crossing a Gap in the Return Path 542
11.19 To Tightly Couple or Not to Tightly Couple 544
11.20 Calculating Odd and Even Modes from Capacitance- and Inductance-Matrix Elements 546
11.21 The Characteristic Impedance Matrix 550
11.22 The Bottom Line 553

Chapter 12 S-Parameters for Signal Integrity Applications 555

12.1 S-Parameters, the New Universal Metric 555
12.2 What Are S-Parameters? 557
12.3 Basic S-Parameter Formalism 559
12.4 S-Parameter Matrix Elements 562
12.5 Simulating Return and Insertion Loss 567

12.6 A Transparent Interconnect 570
12.7 Changing the Port Impedance 573
12.8 The Phase of S21 for a Uniform 50-Ohm Transmission Line 574
12.9 The Magnitude of S21 for a Uniform Transmission Line 577
12.10 Coupling to Other Transmission Lines 582
12.11 Insertion Loss for Non-50-Ohm Transmission Lines 589
12.12 Data-Mining S-Parameters 594
12.13 Single-Ended and Differential S-Parameters 596
12.14 Differential Insertion Loss 601
12.15 The Mode Conversion Terms 605
12.16 Converting to Mixed-Mode S-Parameters 607
12.17 Time and Frequency Domains 609
12.18 The Bottom Line 613

Chapter 13 The Power Distribution Network (PDN) 615

13.1 The Problem 615
13.2 The Root Cause 618
13.3 The Most Important Design Guidelines for the PDN 620
13.4 Establishing the Target Impedance Is Hard 621
13.5 Every Product Has a Unique PDN Requirement 629
13.6 Engineering the PDN 631
13.7 The VRM 633
13.8 Simulating Impedance with SPICE 635
13.9 On-die Capacitance 637
13.10 The Package Barrier 639
13.11 The PDN with No Decoupling Capacitors 644
13.12 The MLCC Capacitor 646
13.13 The Equivalent Series Inductance 650
13.14 Approximating Loop Inductance 652
13.15 Optimizing the Mounting of Capacitors 661
13.16 Combining Capacitors in Parallel 667
13.17 Engineering a Reduced Parallel Resonant Peak by Adding More Capacitors 673
13.18 Selecting Capacitor Values 675
13.19 Estimating the Number of Capacitors Needed 681
13.20 How Much Does a nH Cost? 683
13.21 Quantity or Specific Values? 687

13.22 Sculpting the Impedance Profiles: The Frequency Domain Target Impedance (FDTI) Method 692
13.23 When Every pH Counts 699
13.24 Location, Location, Location 703
13.25 When Spreading Inductance Is the Limitation 707
13.26 The Chip View 710
13.27 Bringing It All Together 713
13.28 The Bottom Line 717

Appendix A 100 General Design Guidelines to Minimize Signal-Integrity Problems 719

Appendix B 100 Collected Rules of Thumb to Help Estimate Signal-Integrity Effects 727

Appendix C Selected References 739

Index 741

PREFACE TO THE SECOND EDITION

Since the publication of the first edition of *Signal Integrity—Simplified*, the principles of signal integrity haven't changed. What has changed, though, is the prolific use of high-speed serial links and the critical role power integrity now plays in the success or failure of new product introductions.

In addition to fleshing out more details and examples in many of the chapters, especially on differential pairs and losses, two new chapters have been added to this second edition to provide a strong foundation to meet the needs of today's engineers and designers.

This first new chapter—Chapter 12—provides a thorough introduction to the use of S-parameters in signal integrity applications. If you deal with any high-speed serial links, you will encounter S-parameters. Because they are written in the foreign language of the frequency domain, they are intimidating to the high-speed digital designer. Chapter 12, like all the chapters in this book, provides a solid foundation in understanding this formalism and enables all engineers to harness the great power of S-parameters.

Chapter 13, the second new chapter, is on power integrity. These issues increasingly fall in the lap of the design engineer. With higher speed applications, interconnects in the power distribution path affect not just power delivery, but also signals' return paths and passing an EMC certification test.

We start at the beginning and illustrate the role of the power distribution interconnects and how design and technology selection can make or break the performance of the power distribution network. The essential principles of plane

impedance, spreading inductance, decoupling capacitors, and the loop inductance of capacitors are introduced. This valuable insight helps feed the intuition of engineers enabling them to apply the power of their creativity to synthesize new designs. Hand in hand with the creation of a design is the analysis of its performance so that cost-performance trade-offs can be explored and the PDN impedance profile can be sculpted to perfection.

If you are new to signal integrity, this second edition of *Signal and Power Integrity—Simplified* provides your starting place to build a strong foundation and empowers you to get your new signal integrity designs right the first time, every time.

PREFACE TO THE FIRST EDITION

"Everything should be made as simple as possible, but not simpler."

Albert Einstein

Printed circuit-board and IC-package design used to be a field that involved expertise in layout, CAD, logic design, heat transfer, mechanical engineering, and reliability analysis. With modern digital electronic systems pushing beyond the 1-GHz barrier, packaging and board designers must now balance signal integrity and electrical performance with these other concerns.

Everyone who touches the physical design of a product has the potential of affecting the performance. All designers should understand how what they do will affect signal integrity or, at the very least, be able to talk with engineers who are responsible for the signal integrity.

The old design methodology of building prototypes, hoping they work, and then testing them to find out is no longer cost effective when time to market is as important as cost and performance. If signal integrity is not taken into account from the beginning, there is little hope a design will work the first time.

In our new "high-speed" world, where the packaging and interconnect are no longer electrically transparent to the signals, a new methodology for designing a product right the first time is needed. This new methodology is based on predictability. The first step is to use established design guidelines based on engineering discipline. The second step is to evaluate the expected performance by "putting in the numbers." This is what distinguishes engineering from guesswork. It takes

advantage of four important tools: rules of thumb, analytic approximations, numerical simulation tools, and measurements. With an efficient design and simulation process, many of the trade-offs between the expected performance and the ultimate cost can be evaluated early in the design cycle, where the time, risk, and cost savings will have the biggest impact. The way to solve signal integrity problems is to first understand their origin and then apply all the tools in our toolbox to find and verify the optimum solution.

The design process is an intuitive one. The source of inspiration for a new way of solving a problem is that mysterious world of imagination and creativity. An idea is generated and the analytical powers of our technical training take over to massage the idea into a practical solution. Though computer simulations are absolutely necessary for final verification of a solution, they only rarely aid in our intuitive understanding. Rather, it is an understanding of the mechanisms, principles and definitions, and exposure to the possibilities, that contribute to the creation of a solution. Arriving at that initial guess and knowing the places to look for solutions require understanding and imagination.

This book emphasizes the intuitive approach. It offers a framework for understanding the electrical properties of interconnects and materials that apply across the entire hierarchy from on-chip, through the packages, circuit boards, connectors, and cables.

Those struggling with the confusing and sometimes contradictory statements made in the trade press will use this book as their starting place. Those experienced in electrical design will use this book as the place to finally understand what the equations mean.

In this book, terms are introduced starting at the ground floor. For example, the impedance of a transmission line is the most fundamental electrical property of an interconnect. It describes what a signal will see electrically and how it will interact with the interconnects. For those new to signal integrity, most of the problems arise from confusion over three terms: the *characteristic* impedance, *the* impedance, and the *instantaneous* impedance a signal sees. This distinction is even important for experienced engineers. This book introduces the reader to each of these terms and their meanings, without complex mathematics.

New topics are introduced at a basic level; most are not covered in other signal integrity books at this level. These include partial inductance (as distinct from loop inductance), the origin of ground bounce and EMI, impedance, transmission line discontinuities, differential impedance, and attenuation in lossy lines affecting

the collapse of the eye diagram. These topics have become critically important for the new high-speed serial links.

In addition to understanding the basic principles, leveraging commercially available tools is critical for the practicing engineer who wants to find the best answer in the shortest time. Tools for solving signal integrity problems fall in two categories: analysis and characterization. Analysis is what we usually refer to as a calculation. Characterization is what we usually refer to as a measurement. The various tools, guidelines on when they should be used, and examples of their value are presented throughout the book.

There are three types of analysis tools: rules of thumb, analytic approximations, and numerical simulation. Each has a different balance between accuracy and effort to use. Each has a right and a wrong place for its appropriate use. And each tool is important and should be in the toolbox of every engineer.

Rules of thumb, such as "the self inductance of a wire is about 25 nH/inch," are important when having a quick answer NOW! is more important than having an accurate answer later. With very few exceptions, every equation used in signal integrity is either a definition or an approximation. Approximations are great for exploring design space and balancing design and performance trade-offs. However, without knowing how accurate a particular approximation really is, would you want to risk a $10,000 board-fabrication run and four weeks of your schedule based on an approximation?

When accuracy is important, for example, when signing off on a design, numerical simulation is the right tool to use. In the last five years, a whole new generation of tools has become available. These new tools have the powerful combination of being both easy to use and accurate. They can predict the characteristic impedance, cross talk, and differential impedance of any cross-section transmission line and simulate how a signal might be affected by any type of termination scheme. You don't have to be a Ph.D. to use this new generation of tools so there is no reason every engineer can't take advantage of them.

The quality of the simulation is only as good as the quality of the electrical description of the components (i.e., the equivalent circuit models). Engineers are taught about circuit models of gates that perform all the information processing, but rarely are the circuit models of the interconnects reviewed. Fifteen years ago, when interconnects looked transparent to the signals, all interconnects were considered as ideal wires—no impedance and no delay. When these terms were added, they were lumped together as "parasitics."

Today, in a high-speed digital system with a clock frequency above about 100 MHz, it is the real wires—the wire bonds, the package leads, the pins, the circuit board traces, the connectors, and the cabling—that create signal-integrity problems and can prevent products from working correctly the first time. Understanding these "analog" effects, designing for them, specifying correct values for them, and including them in the system simulations before the design is committed to hardware, can enable moving a more robust product to market more quickly.

This book provides the tools to enable all engineers and managers involved in chip packaging and board, connector and interconnect design, to understand how these passive elements affect the electrical performance of a system and how they can be incorporated in system simulation. It illustrates how to perform engineering estimates of important electrical parameters and evaluate technology trade-offs. Examples are selected from a wide variety of common systems, including on-chip interconnects, wire bonds, flip chip attach, multilayer circuit boards, DIPs, PGAs, BGAs, QFPs, MCM connectors, and cables.

While most textbooks emphasize theoretical derivation and mathematical rigor, this book emphasizes intuitive understanding, practical tools, and engineering discipline. We use the principles of electrical engineering and physics and apply them to the world of packaging and interconnects to establish a framework of understanding and a methodology of solving problems. The tools of time- and frequency-domain measurement, two- and three-dimensional field solvers, transmission-line simulations, circuit simulators, and analytical approximations are introduced to build verified equivalent circuit models for packages and interconnects.

There are two important questions that all designers should ask of any model they use: How accurate is it? And what is the bandwidth of the model? The answers to these questions can come only from measurements. Measurements play the very important role of risk reduction.

The three generic measurement instruments, the impedance analyzer, the vector-network analyzer (VNA) and the time-domain reflectometer (TDR) are introduced and the interpretation of their data explained. Examples of measurements from real interconnects such as IC packages, printed circuit boards, cables, and connectors are included throughout this book to illustrate the principles and, by example, the value of characterization tools.

This book has been designed for use by people of all levels of expertise and training: engineers, project managers, sales and marketing managers, technology developers, and scientists. We start out with an overview of why designing the

interconnects for high-speed digital systems is difficult and what major technical hurdles must be overcome to reach high-frequency operation.

We apply the tools of electrical engineering and physics to the problems of signal integrity in digital signals through the entire range of interconnects. The concept of equivalent circuit models is introduced to facilitate the quantified prediction of performance. The rest of the book describes how the circuit models of interconnects affect the electrical performance of the system in terms of the four families of noise problems: reflections, cross talk, rail collapse in the power distribution network, and EMI.

This book originated from a series of short courses and semester-long courses the author gave to packaging, circuit-board, and design engineers. It is oriented to all people who need to balance electrical performance with all other packaging and interconnect concerns in their system designs. This book provides the foundation to understand how the physical design world of geometries and material properties affects electrical performance.

If you remember nothing else about signal integrity, you should remember the following important general principles. These are summarized here and described in more detail throughout this book.

Top Ten Signal Integrity Principles

1. The key to efficient high-speed product design is to take advantage of analysis tools that enable accurate performance prediction. Use measurements as a way of validating the design process, reducing risk, and increasing confidence in the design tools.
2. The only way to separate myth from reality is to put in the numbers using rules of thumb, approximations, numerical simulation tools, or measurements. This is the essential element of engineering discipline.
3. Each interconnect is a transmission line with a signal and a return path, regardless of its length, shape, or signal rise time. A signal sees an instantaneous impedance at each step along its way down an interconnect. Signal quality is dramatically improved if the instantaneous impedance is constant, as in a transmission line with a uniform cross section.
4. Forget the word *ground*. More problems are created than solved by using this term. Every signal has a return path. Think *return path* and you will train

your intuition to look for and treat the return path as carefully as you treat the signal path.

5. Current flows through a capacitor whenever the voltage changes. For fast edges, even the air gap between the edge of a circuit board and a dangling wire can have a low impedance through the fringe field capacitance.
6. Inductance is fundamentally related to the number of rings of magnetic-field lines completely surrounding a current. If the number of rings of field lines ever changes, for whatever reason, a voltage will be created across the conductor. This is the origin of some reflection noise, cross talk, switching noise, ground bounce, rail collapse, and some EMI.
7. Ground bounce is the voltage created on the ground return conductor due to changing currents through the total inductance of the return path. It is the primary cause of switching noise and EMI.
8. The bandwidth of a signal is the highest sine-wave frequency component that is significant, compared to an equivalent frequency square wave. The bandwidth of a model is the highest sine-wave frequency at which the model still accurately predicts the actual performance of the interconnect. Never use a model in an application where the signal bandwidth is higher than the model's bandwidth.
9. Never forget, with few exceptions, every formula used in signal integrity is either a definition or an approximation. If accuracy is important, do not use an approximation.
10. The problem caused by lossy transmission lines is the rise-time degradation. The losses increase with frequency due to skin depth and dielectric losses. If the losses were constant with frequency, the rise time would not change and lossy lines would be only a minor inconvenience.
11. The most expensive rule is the one that delays the product ship.

ACKNOWLEDGMENTS

Many colleagues, friends, and students contributed to my understanding that went into the content for both the first and second editions. Literally thousands of engineers from Intel, Cisco, Motorola, Altera, Qualcomm, Raytheon, and other companies who attended my classes provided feedback on what explanations worked and what didn't work.

In the second edition, my reviewers, Greg Edlund, Tim Swettlen, and Larry Smith, provided excellent feedback. I learned a lot from these experts.

My publisher, Bernard Goodwin, was always patient and encouraging, even when I missed deadlines, and never complained when my science fiction novel was completed ahead of this second edition.

Thank you all for the wonderful support and encouragement.

This page intentionally left blank

ABOUT THE AUTHOR

Eric Bogatin received his B.S. in physics from MIT in 1976 and his M.S. and Ph.D. in physics from the University of Arizona in Tucson in 1980. For more than thirty years he has been active in the fields of signal integrity and interconnect design. He worked in senior engineering and management roles at AT&T Bell Labs, Raychem Corp., Sun Microsystems, and Interconnect Devices Inc.

Recognizing a need in the industry for quality technical training in signal integrity, he created Bogatin Enterprises, which has grown to be a world-leading provider of signal integrity training services. As a "Signal Integrity Evangelist," Eric turns complexity into practical design principles, leveraging commercially available analysis techniques and measurement tools.

Over the years, he has created a number of live classes and presented to more than five thousand engineers, worldwide. These include topics on transmission lines, differential pairs, switching noise, ground bounce, the power distribution network, and EMI. Some courses cover the use of TDR and VNA measurement techniques to characterize interconnects.

In addition, Eric is a prolific author with more than three hundred publications, many posted on his web site, www.beTheSignal.com, for free download. He regularly presents at DesignCon, PCB West, Mentor User Group Meetings, and the IPC's Designer Day annual event.

In the past ten years, he has written regular monthly columns for *Printed Circuit Design and Fabrication* magazine, Semiconductor International, Electronic Packaging and Production, Altera Corporation, and Mentor Graphics Corporation. He is on the editorial board of *Printed Circuit Design and Fabrication* magazine and *Microinterconnects Newsletter.*

While many engineers specialize in the electromagnetics aspects of signal integrity, Eric has equal experience in interconnect technologies. He has written four other books on packaging technologies and incorporates examples of the entire interconnect hierarchy in his materials, from on-chip, through packages, sockets, interposers, circuit boards, connectors, and cables.

In 2005, he recognized the increasing value of web-based, distance learning and created an online presence at www.beTheSignal.com. His was the first company to introduce a complete curriculum of web-based signal integrity content, including lectures, labs, and quizzes, available through a subscription plan for individuals or entire companies.

Signal and Power Integrity—

Simplified

Second Edition

In physical science the first essential step in the direction of learning any subject is to find principles of numerical reckoning and practicable methods for measuring some quality connected with it. I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely in your thoughts advanced to the state of Science, whatever the matter may be.

—Lord Kelvin (1824–1907)

CHAPTER 1

Signal Integrity Is in Your Future

"There are two kinds of designers, those with signal-integrity problems and those that will have them."

on a white board at a large systems company

Ironically, this is an era when not only are clock frequencies increasing and signal integrity problems getting more severe, but the time design teams have available to solve these problems and design new products is getting shorter. Product design teams have one chance to get a product to market; the product must work successfully the first time. If identifying and eliminating signal integrity problems isn't an active priority as early in the product design cycle as possible, chances are the product will not work.

TIP As clock frequencies increase, identifying and solving signal-integrity problems becomes critical. The successful companies will be those that master signal-integrity problems and implement an efficient design process to eliminate these problems. It is by incorporating new design rules, new technologies, and new analysis tools that higher performance designs can be implemented and meet ever-shrinking schedules.

In high-speed products, the physical and mechanical design can affect signal integrity. An example of how a simple 2-inch-long section of trace on a printed

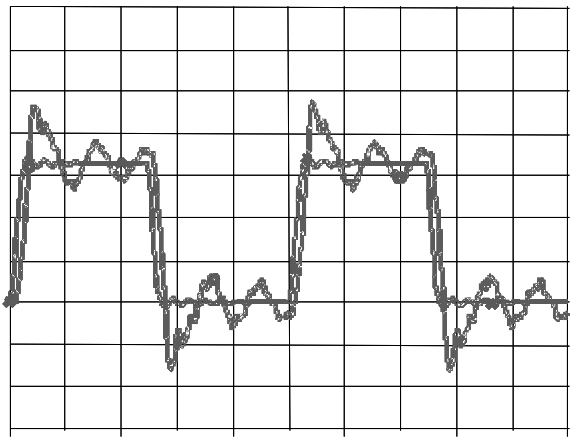


Figure 1-1 100-MHz clock waveform from a driver chip, when there is no connection (smooth plot) and when there is a two-inch length of PCB trace connected to the output (ringing). Scale is 1 v/div and 2 nsec/div, simulated with Mentor Graphics Hyperlynx.

circuit board (PCB) can affect the signal integrity from a typical driver is shown in Figure 1-1.

The design process is often a very intuitive and creative process. Feeding your engineering intuition about signal integrity is critically important to reaching an acceptable design as quickly as possible. All engineers who touch the product should have an understanding of how they influence the performance of the overall product. By understanding the fundamental principles of signal integrity at an intuitive and engineering level, every engineer involved in the design process can see the impact of their decisions on system performance. This book is about the fundamental principles needed to understand signal-integrity problems and their solutions. The engineering discipline required to deal with these problems is presented at an intuitive and a quantitative level.

1.1 What Is Signal Integrity?

In the good old days of 10-MHz clock frequencies, the chief design challenges in circuit boards or packages were how to route all the signals in a two-layer board and how to get packages that wouldn't crack during assembly. The electrical properties of the interconnects were not important because they didn't affect system performance. In this sense, we say that "the interconnects were transparent to the signals."

A device would output a signal with a rise time of roughly 10 nsec and a clock frequency of 10 MHz, for example, and the circuits would work with the crudest of interconnects. Prototypes fabricated with wire-wrapped boards worked as well as final products with printed circuit boards and engineering change wires.

But clock frequencies have increased and rise times of signals have decreased. For most electronic products, signal-integrity effects begin to be important at clock frequencies above about 100 MHz or rise times shorter than about 1 nsec. This is sometimes called the *high-frequency* or *high-speed regime*. These terms refer to products and systems where the interconnects are no longer transparent to the signals and, if you are not careful, one or more signal integrity problems arise.

Signal integrity refers, in its broadest sense, to all the problems that arise in high-speed products due to the interconnects. It is about how the electrical properties of the interconnects, interacting with the digital signal's voltage and current waveforms, can affect performance.

All of these problems fall into one of the following categories:

1. Timing

2. Noise

3. Electromagnetic interference (EMI)

Timing is a complicated field of study by itself. In one cycle of a clock, a certain number of operations must happen. This short amount of time must be divided up and allocated, in a budget, to all the various operations. For example, some time is allocated for gate switching, for propagating the signal to the output gate, for waiting for the clock to get to the next gate, and for waiting for the gate to read the data at the input. Though the interconnects affect the timing budget, timing is not covered in this book. We refer any interested readers to a number of other books listed in the reference section for more information on this topic. Instead, this book concentrates on the effects of the interconnects on the other generic high-speed problem, too much noise.

We hear about a lot of signal-integrity noise problems, such as ringing, ground bounce, reflections, near-end cross talk, switching noise, non-monotonicity, power bounce, attenuation, and capacitive loading. All of these relate to the electrical properties of the interconnects and how the electrical properties affect the waveforms of the digital signals.

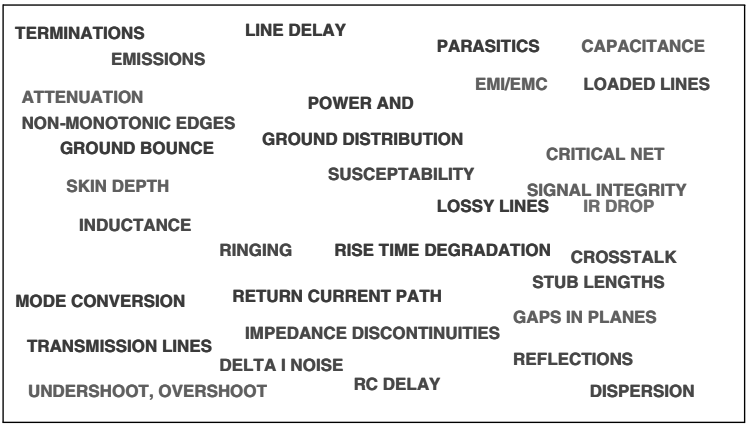


Figure 1-2 The list of signal-integrity effects seems like a random collection of terms, without any pattern.

It seems at first glance that there is an unlimited supply of new effects we have to take into account. This confusion is illustrated in Figure 1-2. Few digital-system designers or board designers are familiar with all of these terms as other than labels for the craters left in a previous product-design minefield. How are we to make sense of all these signal-integrity problems? Do we just keep a growing checklist and add to it periodically?

All the effects listed above, associated with signal-integrity noise problems, are related to one of the following four unique families of noise sources:

1. Signal quality of one net, including losses on the line

2. Cross talk between two or more nets, including ground and power balance

3. Rail collapse in the power and ground distribution

4. Electromagnetic interference and radiation from the entire system

These four families are illustrated in Figure 1-3. Once we identify the root cause of the noise associated with each of these four families of problems, and understand the essential principles behind them, the general solution for finding and fixing the problems in each family will become obvious. This is the power of being able to classify every signal-integrity noise problem into one of these four families.

These problems play a role in all interconnects, from the smallest on-chip wire to the cables connecting racks of boards and everywhere in-between. The

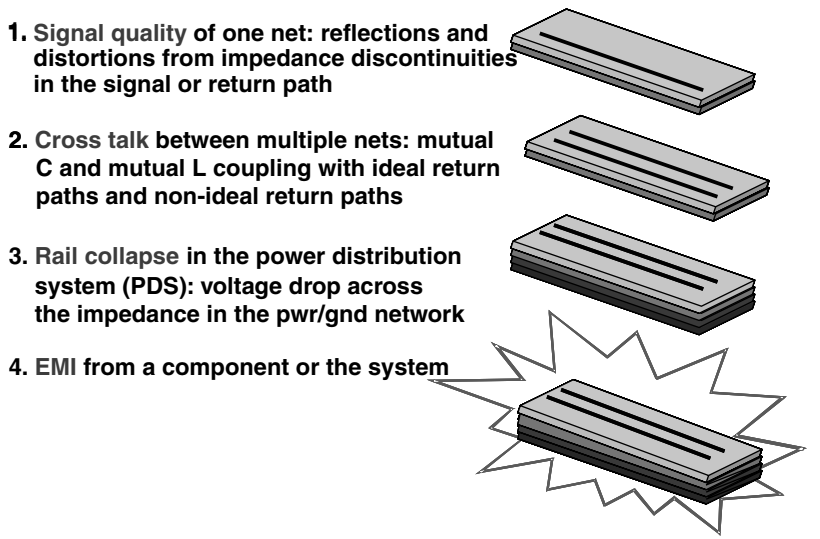


Figure 1-3 The four families of signal-integrity problems.

principles and effects are the same. The only differences in each physical structure are the specific geometrical feature sizes and the material properties.

1.2 Signal Quality on a Single Net

A net is made up of all the metal connected together in a system. For example, there may be a trace going from a clock chip's output pin to three other chips. Each piece of metal that connects all four of these pins is considered one net. In addition, the net includes not only the signal path but also the return path for the signal current. Signal quality on a single net depends as much on the physical features of the signal trace as on the return path. An example of two different nets on a circuit board is shown in Figure 1-4.

When the signal leaves the output driver, the voltage and the current, which make up the signal, see the interconnect as an electrical impedance. As the signal propagates down the net, it is constantly probing and asking, "what is the instantaneous impedance I see?" If the impedance the signal sees stays the same, the signal continues undistorted. If, however, the impedance changes, the signal will reflect from the change and continue through the rest of the interconnect distorted. If there are enough impedance changes, the distortions can cause false triggering.

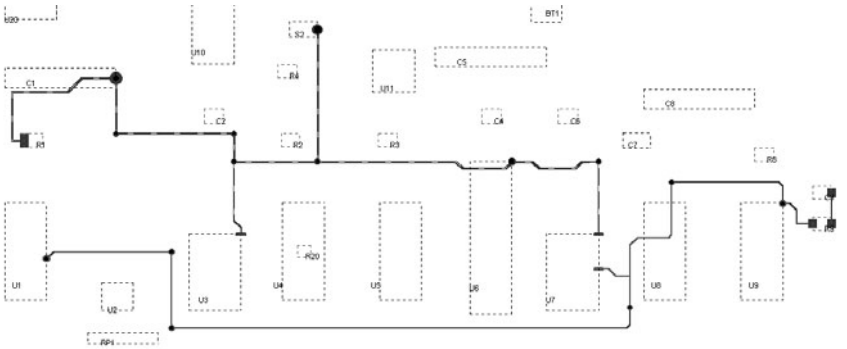


Figure 1-4 Example of two nets on a circuit board. All metal connected together is considered one net. Note: One net has a surface-mount resistor in series. Routed with Mentor Graphics Hyperlynx.

Any feature that changes the cross section or geometrical shape of the net will change the impedance the signal sees. We call any feature that changes the impedance a discontinuity. Every discontinuity will cause the signal to be distorted from its original pristine shape, to some extent. For example, some of the features that would change the impedance the signal sees include the following:

1. A line-width change

2. A layer change through a via

3. A gap in return-path plane

4. A connector

5. A branch, tee, or stub

6. The end of a net

These impedance discontinuities can arise from the cross section, the topology of the routed traces, or the added components. The most common discontinuity is what happens at the end of a trace, which is usually either a high impedance open at the receiver or a low impedance at the output driver.

TIP The way to minimize the problems associated with impedance changes is to keep the impedance the signal sees constant throughout the net.

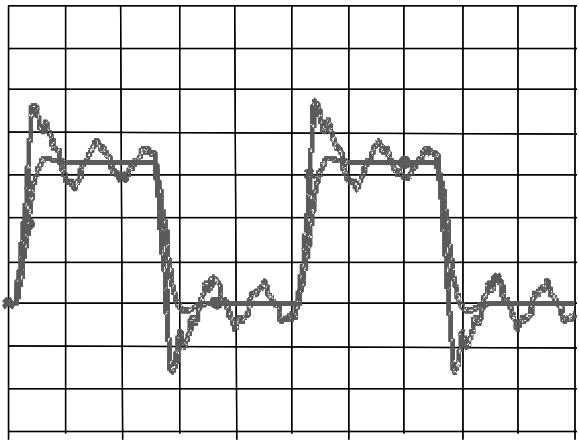


Figure 1-5 Ringing in an unterminated line and good signal quality in a source-series terminated interconnect line. The PCB trace is only two inches long in each case. Scale is 1 v/div and 2 nsec/div. Simulated with Mentor Graphics Hyperlynx.

This strategy is typically implemented by doing three things. First, use a board with constant, or "controlled," impedance traces. This usually means using uniform transmission lines. Second, use routing rules that allow the topology to maintain a constant impedance down the trace. Finally, use strategically placed resistors to manipulate the reflections and keep the received signals looking clean.

Figure 1-5 is an example of poor signal quality due to impedance changes in the same net and when a terminating resistor is used to manage the impedance changes. Often, what we think of as "ringing" is really due to reflections associated with impedance changes.

Even with perfect terminations, the precise board layout can drastically affect signal quality. When a trace branches into two paths, the impedance at the junction changes and some signal will reflect back to the source while some will continue down the branches in a reduced and distorted form. By rerouting the trace to be a daisy chain, the signal sees a constant impedance all down the path and the signal quality can be restored.

The impact on a signal from any discontinuity depends on the rise time of the signal. As the rise time gets shorter, the magnitude of the distortion will increase. This means that a discontinuity that was not a problem in a 33-MHz design may be a problem in a 100-MHz design. This is illustrated in Figure 1-6.

The higher the frequency and the shorter the rise time, the more important it is to keep the impedance the signal sees constant. One way of achieving this is by

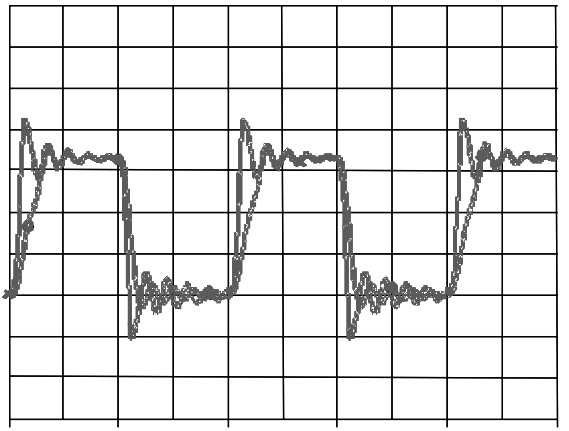


Figure 1-6 25-MHz clock waveforms with a PCB trace six inches long and unterminated. The slow rise time is a 3-nsec rise time. The ringing is from a rise time of 1 nsec. What may not have been a problem with one rise time can be a problem with a shorter rise time. Scale is 1 v/div and 5 nsec/div. Simulated with Mentor Graphics Hyperlynx.

using controlled impedance interconnects even in the packages, such as with multilayer ball grid arrays (BGAs). When the packages do not use controlled impedance, such as with lead frames, it's important to keep the leads short, such as by using chip-scale packages (CSPs).

There are two other aspects of signal quality associated with one net. Frequency-dependent losses in the line from the conductor and the dielectric cause higher frequency signal components to be attenuated more than the lower frequency components. The end result is an increase in the rise time of the signal as it propagates. When this rise time degradation approaches the period of a single bit, the digital information will be distorted. This effect is called inter-symbol interference (ISI) and is a significant source of problems in high-speed serial links, in the 1 Gigabit per second (Gbps) and higher regime.

The third aspect of signal quality problems associated with a single net is related to timing. The time delay difference between two or more signal paths is called *skew*. When a signal and clock line have a skew different than expected, false triggering and errors can result. When the skew is between the two lines that make up a differential pair, some of the differential signal will be converted into common signal and the differential signal will be distorted. This will result in ISI and false triggering.

While skew is a timing problem, it often arises due to the electrical properties of the interconnect. The first order impact on skew is from the total length of

the interconnects. This is easily controlled by careful layout to match lengths. However, the time delay is also related to the local dielectric constant each signal sees and is often a much more difficult problem to fix.

TIP Skew is the time delay between two or more nets. It can be controlled to first order by matching the length of the nets. In addition, the local variation in dielectric constant between the nets will also affect the time delay and is more difficult to control.

1.3 Cross Talk

When one net carries a signal, some of this voltage and current can pass over to an adjacent quiet net, which is just sitting there, minding its own business. Even though the signal quality on the first net (the active net) is perfect, some of the signal can couple over and appear as unwanted noise on the second, quiet net.

TIP It is the capacitive and inductive coupling between two nets that provides a path for unwanted noise from one net to the other.

Cross talk occurs in two different environments: when the interconnects are uniform transmission lines, as in most traces in a circuit board, and when they are not uniform transmission lines, as in connectors and packages. In controlled-impedance transmission lines where the traces have a wide uniform return path, the relative amount of capacitive coupling and inductive coupling is comparable. In this case, these two effects combine in different ways at the near end of the quiet line and at the far end of the quiet line. An example of the measured near- and far-end cross talk between two nets in a circuit board is shown in Figure 1-7.

Having a wide uniform plane as the return path is the configuration of lowest cross talk. Anything that changes the return path from a wide uniform plane will increase the amount of coupled noise between two transmission lines. Usually when this happens, for example, when the signal goes through a connector and the return paths for more than one signal path are now shared by one of the pins rather than by a plane, the inductively coupled noise increases much more than the capacitively coupled noise.

In this regime, where inductively coupled noise dominates, we usually refer to the cross talk as switching noise, delta I noise, dI-dt noise, ground bounce, simultaneous switching noise (SSN), or simultaneous switching output (SSO) noise. As we will see, this type of noise is generated by the coupled inductance,

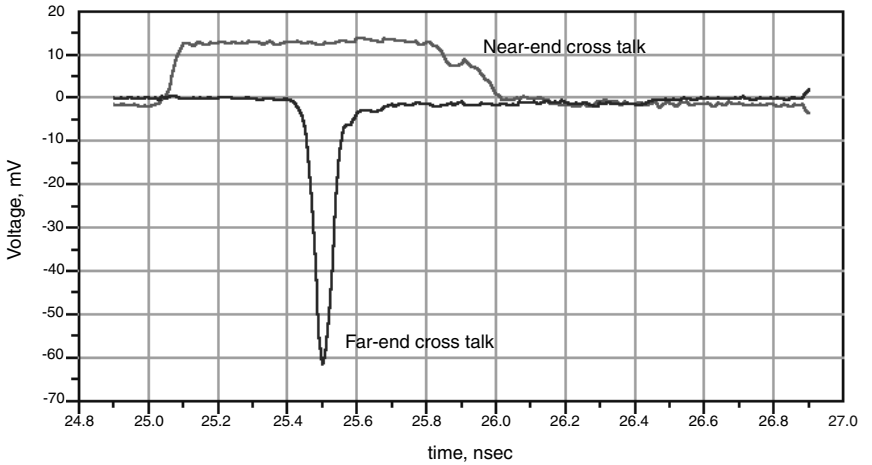


Figure 1-7 Measured voltage noise at the near end and the far end of a quiet trace when a 200-mV signal is injected in the active trace. Note the near-end noise is about 7% and the far-end noise is nearly 30% of the signal. Measurements performed with an Agilent DCA86100 with time domain reflectometer (TDR) plug-in.

which is called mutual inductance. Switching noise occurs mostly in connectors, packages, and vias, where the return path conductor is not a wide, uniform plane. Ground bounce, reviewed later in this book, is really a special case caused when return currents overlap in the same conductor and their mutual inductance is very high. An example of SSO noise from the high mutual inductance between adjacent signal and return paths in a package is shown in Figure 1-8.

TIP SSO noise, where the coupled, or mutual inductance, dominates is becoming one of the most important issues in connector and package design. It will only get worse in next generation products. The solution lies in careful design of the geometry of paths so that mutual inductance is minimized, and in the use of diferential signaling.

By understanding the nature of the capacitive and inductive coupling, it is possible to optimize the physical design of the adjacent signal traces to minimize the coupling. This usually can be as simple as spacing the traces farther apart. In addition, the use of lower dielectric constant material will decrease the cross talk for the same characteristic impedance lines. Some aspects of cross talk, especially

Figure 1-8 Top trace: measured voltage on active lines in a multiline buss. Bottom trace: measured noise on one quiet line showing the switching noise due to mutual inductance between the active and quiet nets in the package.

switching noise, increase with the length of the interconnect and with decreasing rise time. Shorter rise-time signals will create more cross talk. Keeping interconnects short, such as by using CSPs and high-density interconnects (HDI), will help to minimize cross talk.

1.4 Rail-Collapse Noise

Noise is generated, and is a problem, in more than just the signal paths. It can also be a disaster in the power- and ground-distribution network that feeds each chip. When current through the power- and ground-path changes, as when a chip switches its outputs or core gates switch, there will be a voltage drop across the impedance of the power and ground paths. This voltage drop will mean less voltage gets to the chip, which will see a decrease, or collapse, of the voltage between the power and ground rails. One example of the change in the voltage across a microprocessor's power rails is shown in Figure 1-9.

In high-performance processors, FPGAs, and some ASICs, the trend is for lower power-supply voltage, but higher power consumption. This is primarily due to more gates on a chip switching faster. In each cycle, a certain amount of energy is consumed. When the chip switches faster, the same energy is consumed in each cycle, but consumed more often, leading to higher than average power consumption.

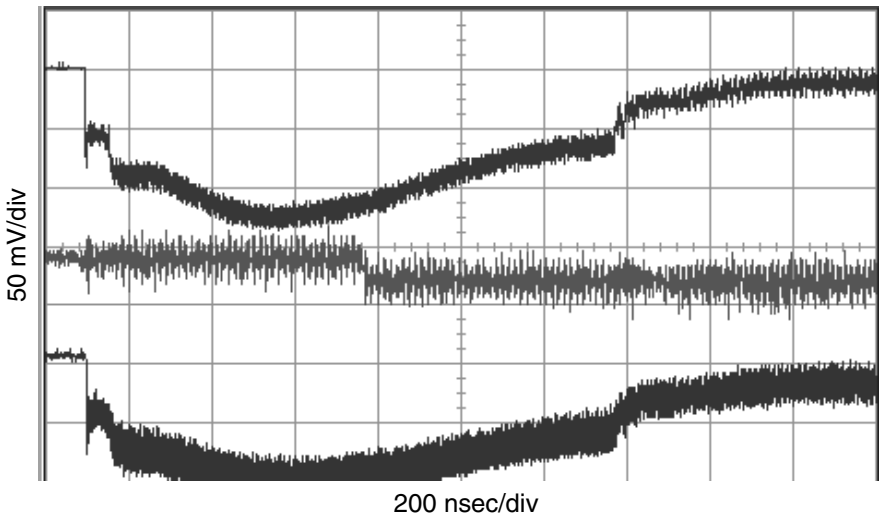


Figure 1-9 Measured Vcc voltage at three pin locations on the package of a microprocessor just coming from a "stop-clock" state. Nominal voltage is supposed to be 2.5 V, but due to voltage drops in the power distribution system, the delivered voltage collapses by almost 125 mV.

These factors combine to mean higher currents will be switching in shorter amounts of time, and the amount of noise that can be tolerated will decrease. As the drive voltage decreases and the current level increases, any voltage drops associated with rail collapse become a bigger and bigger problem.

TIP The goal in designing the power and ground distribution is to try to minimize the impedance of the power-distribution system (PDS), sometimes also called the power delivery network (PDN).

In this way, even though there is current switching in the PDN, the voltage drop across a lower impedance may be kept to an acceptable level. The requirements for the impedance of the PDN have been evaluated for high-end processors by Sun Microsystems. Their estimate of the required impedance of the PDN is shown in Figure 1-10. Lower impedance in the PDN is increasingly important and harder to achieve.

If we understand how the physical design of the interconnects affects their impedance, we can optimize the design of the PDN for low impedance. As we will see, designing a low-impedance PDN means including features such as the following:

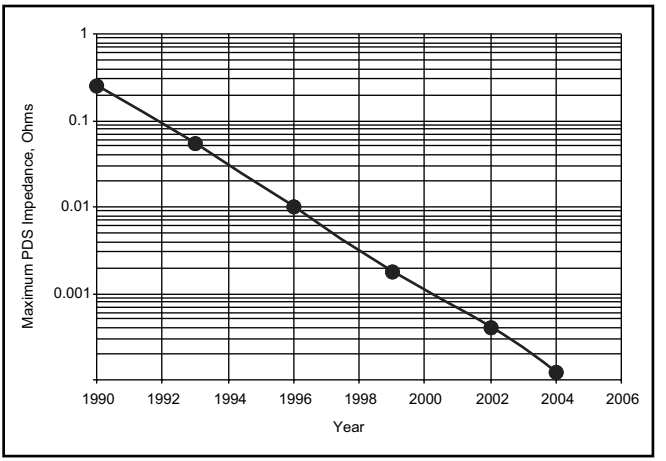


Figure 1-10 Trend in the maximum allowable impedance of the power distribution system for high-end processors. Source, Sun Microsystems.

1. Closely spaced adjacent planes for the power and ground distribution with as thin a dielectric between them as possible, near the surface of the board
2. Multiple, low-inductance decoupling capacitors
3. Multiple, very short power and ground leads in packages
4. On-chip decoupling capacitance

An innovative technology to help minimize rail collapse can be seen in the new ultrathin, high-dielectric constant laminates for use between power and ground layers. One example is C-Ply from 3M Corp. This material is 8 microns thick and has a dielectric constant of 20. Used as the power and ground layers in an otherwise conventional board, the ultralow loop inductance and high distributed capacitance dramatically reduce the impedance of the power and ground distribution. One example of the rail-collapse noise on a small test board using conventional layers and a board with this new C-Ply is shown in Figure 1-11.

1.5 Electromagnetic Interference (EMI)

With board-level clock frequencies in the 100-MHz to 500-MHz range, the first few harmonics are within the common communications bands of TV, FM radio, cell phone, and personal communications services (PCS). This means there is the very real possibility of electronic products interfering with communications unless their electromagnetic emissions are kept below acceptable levels. Unfortu-

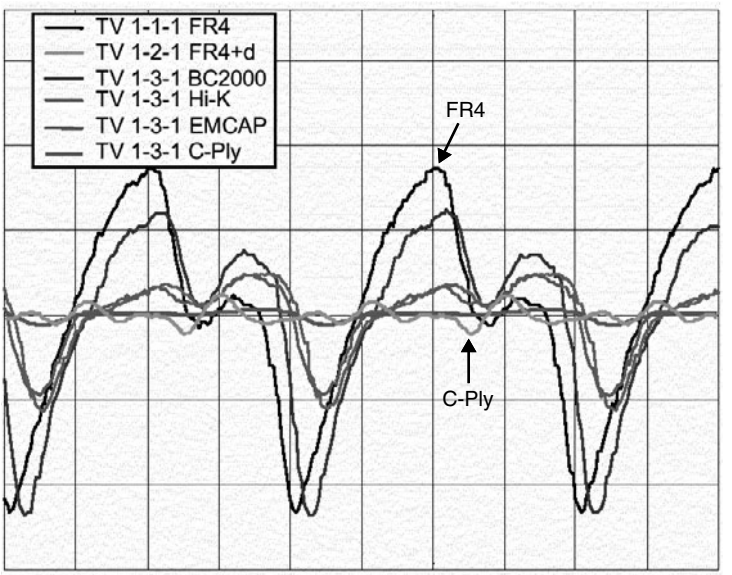


Figure 1-11 Measured rail voltage noise on a small digital board, with various methods of decoupling. The worst case is no decoupling capacitors on FR4. The best case is the 3M C-Ply material, showing virtually no voltage noise. Courtesy of National Center for Manufacturing Science. Scale is 0.5 v/div and 5 nsec/div.

nately, without special design considerations, EMI will get worse at higher frequencies. The radiated far-field strength from common currents will increase linearly with frequency and from differential currents will increase with the square of the frequency. With clock frequencies increasing, the radiated emissions level will inevitably increase as well.

It takes three things to have an EMI problem: a source of noise, a pathway to a radiator, and an antenna. Every source of signal-integrity problem mentioned above will be a source of EMI. What makes EMI so challenging is that even if the noise is low enough to meet the signal-integrity noise budget, it may still be large enough to cause serious radiated emissions.

TIP The two most common sources of EMI are (1) the conversion of some differential signal into a common signal, which eventually gets out on an external-twisted pair cable, and (2) ground bounce on a circuit board generating common currents on external single-ended shielded cables. Additional noise can come from internally generated radiation leaking out of the enclosure.

Most of the voltage sources that drive radiated emissions come from the power- and ground-distribution networks. Often, the same physical designs that contribute to low rail-collapse noise will also contribute to lower emissions.

Even with a voltage noise-source that can drive radiation, it is possible to isolate it by grouping the high-speed sections of a board away from where they might exit the product. Shielding the box will minimize the leakage of the noise from an antenna. Many of the ills of a poorly designed board can be fixed with a good shield.

A product with a great shield will still need to have cables connecting it to the outside world for communications, for peripherals, or for interfacing. Typically, the cables extending outside the shielded enclosure act as the antennas, and radiate. The correct use of ferrites on all connected cables, especially on twisted pair, will dramatically decrease the efficiency of the cables as antennas. A close-up of the ferrite around a cable is shown in Figure 1-12.

The impedance associated with the I/O connectors, especially the impedance of the return-path connections, will dramatically affect the noise voltages that can drive radiating currents. The use of shielded cables with low-impedance connections will go a long way to minimize EMI problems.

Unfortunately, for the same physical system, increasing the clock frequency generally will also increase the radiated emissions level. This means that EMI problems will be harder to solve as clock frequencies increase.

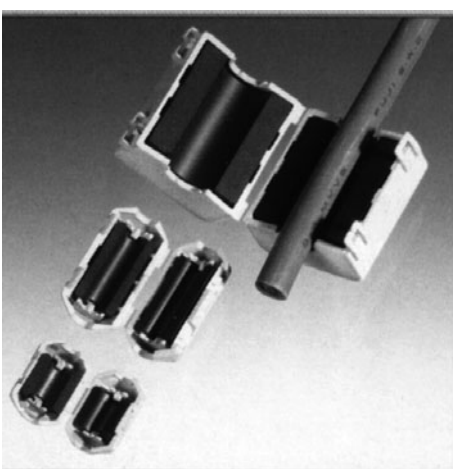


Figure 1-12 Ferrite choke around a cable, split apart. Ferrites are commonly used around cables to decrease common currents, a major source of radiated emissions. Courtesy of IM Intermark.

1.6 Two Important Signal Integrity Generalizations

Two important generalizations should be clear from looking at the four signal-integrity problems above.

First, each of the four families of problems gets worse as rise times decrease. All the signal-integrity problems above scale with how fast the current changes or with how fast the voltage changes. This is often referred to as dI/dt or dV/dt. Shorter rise times mean higher dI/dt and dV/dt.

It is unavoidable that as rise times decrease, the noise problems will increase and be more difficult to solve. And, as a consequence of the general trends in the industry, the rise times found in *all* electronic products will continually decrease. This means that what might not have caused a problem in one design may be a killer problem in the next design with the next generation chip sets operating with a shorter rise time. This is why it is often said, "There are two kinds of engineers, those that have signal integrity problems and those that will."

The second important generalization is that effective solutions to signal-integrity problems are based heavily on understanding the impedance of interconnects. If we have a clear intuitive sense of impedance and can relate the physical design of the interconnects with their impedance, many signal-integrity problems can be eliminated during the design process.

This is why an entire chapter in this book is devoted to understanding impedance from an intuitive and engineering perspective and why much of the rest of this book is about how the physical design of interconnects affects the impedance seen by signals and the PDN.

1.7 Trends in Electronic Products

A few trips to the local computer store over the last 10 years will have given anyone a good sense of the incredible treadmill of progress in computer performance. One measure of performance is the clock frequency of a processor chip. The trend for Intel processor chips, as illustrated in Figure 1-13, shows a doubling in clock frequency about every two years.

This trend toward ever-higher clock frequency is driven by the same force that drives the semiconductor revolution—photolithography. As the gate channel length of transistors is able to be manufactured at smaller size, the switching speed of the transistor increases. The electrons and holes have a shorter distance to travel and can transit the gate, effecting transitions, in a shorter time when the channel length is shorter.

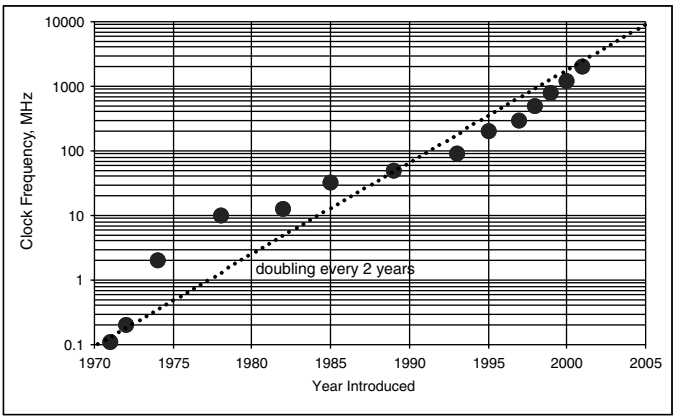


Figure 1-13 Historical trend in the clock frequency of Intel processors based on year of introduction. The trend is a doubling in clock frequency every two years. Source is Intel Corp.

When we refer to a technology generation as 0.18 microns or 0.13 microns, we are really referring to the smallest channel length that can be manufactured. The shorter switching time for smaller channel-length transistors has two important consequences for signal integrity.

The minimum time required for one clock cycle is limited by all the operations that need to be performed in one cycle. Usually, there are three main factors that contribute to this minimum time: the intrinsic time for all the gates that need to switch in series, the time for the signals to propagate through the system to all the gates that need to switch, and the setup and hold times needed for the signals at the inputs to be read by the gates.

In single-chip microprocessor-based systems, such as in personal computers, the dominant factor influencing the minimum cycle time is the switching speed of the transistors. If the switching time can be reduced, the minimum total time required for one cycle can be reduced. This is the primary reason clock frequencies have increased as feature size has been reduced.

TIP It is inevitable that as transistor feature size continues to reduce, rise times will continue to decrease and clock frequencies will continue to increase.

The projections from the 2001 Semiconductor Industry Association (SIA) International Technology Roadmap for Semiconductors (ITRS) for future on-chip clock frequencies, based on projected feature size reductions, compared with the

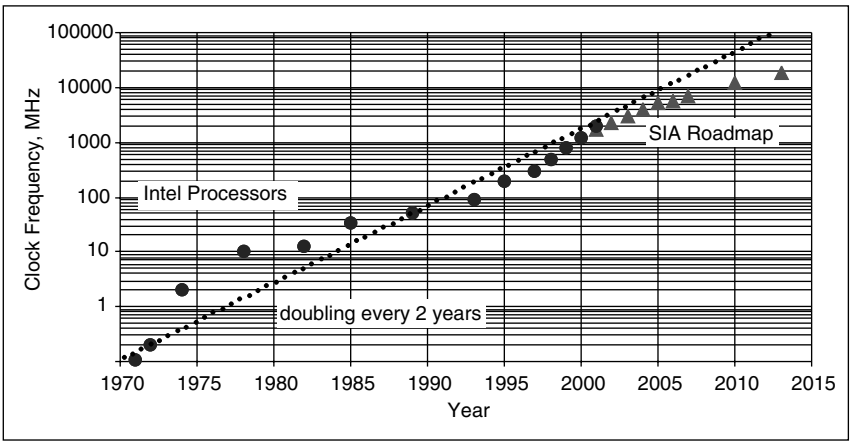


Figure 1-14 Historical trend in the clock frequency of Intel processors based on year of introduction. The trend is a doubling in clock frequency every two years. Also included is the Semiconductor Industry Association roadmap expectations. Source is Intel Corp. and SIA.

Intel processor trend are shown in Figure 1-14. This shows the projected trend for clock frequency increasing at a slightly diminishing but still growing rate for the next 15 years as well.

As the clock frequency increases, the rise time of the signals must also decrease. Each gate that reads either the data lines or the clock lines needs enough time with the signal being either in the high state or the low state, to read it correctly.

This means only a short time is left for the signal to be in transition. We usually measure the transition time, either the rise time or the fall time, as the time it takes to go from 10% of the final state to 90% of the final state. This is called the 10–90 rise time. Some definitions use the 20% to 80% transition points and this rise time is referred to as the 20–80 rise time. An example of a typical clock waveform and the time allocated for the transition are shown in Figure 1-15. In most high-speed digital systems, the time allocated to the rise time is about 10% of the clock cycle time, or the clock period. Based on this generalization, the rise time is roughly related to the clock frequency by:

$$RT = \frac{1}{10 \times F_{clock}} \tag{1-1}$$

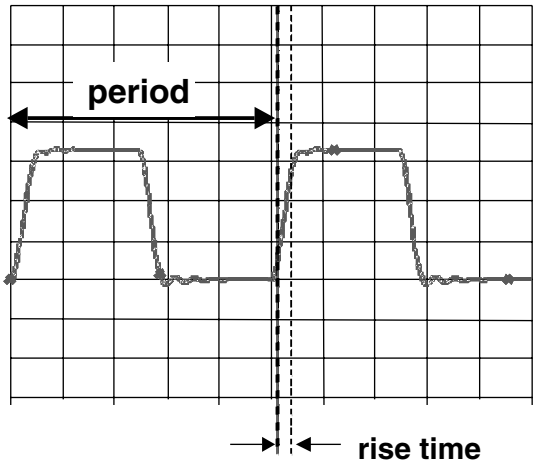


Figure 1-15 The 10–90 rise time for a typical clock waveform is roughly 10% of the period. Scale is 1 v/div and 2 nsec/div, simulated with Mentor Graphics Hyperlynx.

where:

RT = the rise time, in nsec

F_{clock} = the clock frequency, in GHz

For example, when the clock frequency is 1 GHz, the rise time of the associated signals is about 0.1 nsec, or 100 psec. When the clock frequency is 100 MHz, the rise time is roughly 1 nsec. This relationship is shown in Figure 1-16.

T I P The treadmill-like advance of ever-increasing clock frequency means an ever-decreasing rise time and signal-integrity problems that are harder to solve.

Even if the clock frequency of a product is low, there is still the danger of shorter rise times as a direct consequence of the chip technology. A chip-fabrication factory, usually called a "fab," will try to standardize all their wafers on one process, to increase the overall yield. The smaller the chip size, the more chips can be fit on a wafer and the lower the cost per chip. Even if the chip is going to be used in a slow-speed product, it may be fabricated on the same line as a leading-edge ASIC, with the same small-feature size.

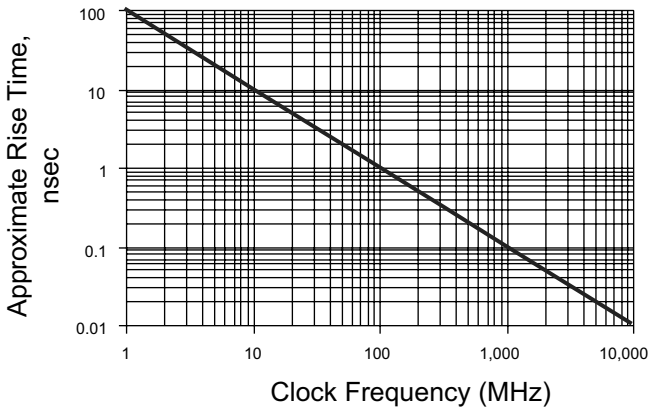


Figure 1-16 Rise time decreases as the clock frequency increases. Signal-integrity problems usually arise at rise times less than 1 nsec or at clock frequencies greater than 100 MHz.

Ironically, the lowest-cost chips will always have ever-shorter rise times, even if they don't need it for the specific application. This has an unintended, scary consequence of Moore's Law. If you have designed a chip set into your product and the rise time is 2 nsec, for example, with a 50-MHz clock, there may be no signal-integrity problems. When your chip supplier upgrades their fab line with a finer-feature process, they may provide you with lower-cost chips. You may think you are getting a good deal. However, these lower cost chips may now have a rise time of 1 nsec. This shorter rise time may cause reflection noise, excess cross talk, rail collapse, and they may fail the Federal Communications Commission (FCC) EMI certification tests. The clock frequency of your product hasn't changed, but, unknown to you, the rise time of the chips supplied to you has decreased with the newer, finer-feature manufacturing process.

TIP As all fabs migrate to lower-cost, finer-feature processes, all fabricated chips will have shorter rise times, and signal integrity problems have the potential to arise in all products, even those with clock frequencies below 50 MHz.

It's not just microprocessor-based digital products that are increasing in clock frequency and decreasing in rise times. The data rate and clock frequencies used in high-speed telecommunications products are blowing past the clock frequencies of microprocessor digital products.

One of the most common specifications for defining the speed of a high-speed serial link is the optical carrier, or OC, spec. This is actually a data rate, with OC-1 corresponding to about 50 Mbits/sec (Mbps). OC-48, with a data rate of 2.5 Gbits/sec (Gbps), is in high-volume deployment and widely implemented. OC-192, at 10 Gbps, is just ramping up. In the near future, OC-768, at 40 Gbps, will be in wide-scale deployment.

The OC designation is a specification for data rate, not for clock frequency. Depending on how a bit is encoded in the data stream, the actual clock frequency of a system can be higher or lower than the data rate.

For example, if there is one bit per clock cycle, then the actual clock frequency will be the data rate. If two data bits are encoded per clock cycle, then a 2.5-Gbps data rate can be obtained with a 1.25-GHz clock. The more bits in the data stream used for error correction and overhead, the lower the data rate, even though the clock frequency is constant.

To keep signal-integrity problems at a minimum, the lowest clock frequency and longest edge rate should be used. This has inspired a growing trend to encode four to eight data bits per clock cycle, using multilevel signaling and multiple signal lines in parallel. This technique of encoding bits in the amplitude is called pulse amplitude module (PAM).

Given these caveats, as a general rule of thumb, most high-speed serial links encode two data bits per clock cycle. The clock frequency of a high-speed serial link can be roughly approximated as half the data rate. The trend is definitely toward higher data rates. In the near future, data rates of 20 Gbps, with clock frequencies of 10 GHz will be common in high-speed serial links. The entire clock period of a 10-GHz clock signal is only 100 psec. This means that the rise time must be significantly less than 20 psec. This is a short rise time and will require extremely careful design practices.

High-speed serial links are not just limited to telecommunications applications. All high-speed digital buses will soon migrate to clock frequencies above 1 GHz. Many high-speed serial link buses have been proposed for on-board data flow for general high-speed digital products. Whatever bit rate they may start at, they also have a migration path mapped to at least three generations, each one doubling in bit rate. Among these are the following:

- PCI-express, starting at 2.5 Gbps and moving to 5 Gbps and 8 Gbps
- Infiniband, starting at 2.5 Gbps and moving to 5 Gbps and 10 Gbps
- Serial ATA, starting at 1.25 Gbps and moving to 2.5 Gbps and 5 Gbps

- XAUI, starting at 3.125 Gbps and moving to 6.25 Gbps
- Gigabit Ethernet, starting at 1 Gbps and moving to 10 Gbps

1.8 The Need for a New Design Methodology

We've painted a scary picture of the future. The situation analysis is as follows:

- Signal integrity problems can prevent the correct operation of a high-speed digital product.
- These problems are a direct consequence of shorter rise time and higher clock frequencies.
- Rise times will continue their inevitable march toward shorter values and clock frequencies will continue to increase.
- Even if we limit the clock frequency, using the lowest-cost chips means even low-speed systems will have chips with very short rise times.
- We have less time in the product-design cycle to get the product to market—it must work the first time.

What are we to do? How are we to efficiently design high-speed products in this new era? In the good old days of 10-MHz clock systems, when the interconnects were transparent, we did not have to worry about signal-integrity effects. We could get away with designing the product for functionality and ignore signal integrity. In today's products, ignoring signal integrity invites schedule slips, higher development costs, and the possibility of never being able to build a functional product.

It will always be more profitable to pay extra to design a product right the first time than to try to fix it later. In the product life cycle, often, the first six months in the market are the most profitable. If your product is late, a significant share of the life-cycle profits may be lost. Time really is money.

TIP A new product-design methodology is required that ensures signal-integrity problems are identified and eliminated from the product as early in the design cycle as possible. To meet ever-shorter design cycle times the product must meet performance specifications the first time.

1.9 A New Product Design Methodology

There are five key ingredients to this new methodology:

1. Understand the root cause of signal-integrity problems and the general guidelines to minimize these problems.
2. Translate the general guidelines into specific design rules for each specific custom product.
3. Predict performance early in the design cycle by creating equivalent electrical circuit models for each component, critical net, and the entire system and by performing local and system-level simulation.
4. Optimize the performance of the design for cost, schedule, and risk by modeling and simulating at every step of the design cycle, especially at the beginning.
5. Use characterization measurements throughout the design cycle to reduce the risk and increase confidence of the quality of the predictions.

In addition to identifying the root cause of signal integrity problems, the other key processes in this new design methodology are modeling, simulation, and characterization.

Simulation is about predicting the performance of the system before building the hardware. It used to be, only those nets in a system that were sensitive to signal-integrity effects were simulated. These nets are called "critical nets." Typically, they were clock lines and maybe a few high-speed bus lines. In 100-MHz clock frequency products, maybe only 5%–10% of the nets were critical nets. In products with clock frequencies at 200 MHz and higher, more than 50% of the nets may be critical, and the entire system needs to be included in the simulation.

TIP In all high-speed products today, system-level simulations must be performed to accurately predict signal-integrity effects.

In order to predict the electrical performance, which is typically the actual voltage and current waveforms at various nodes, we need to translate the physical design into an electrical description. This can be accomplished by one of two paths. The physical design can be converted into an equivalent circuit model and then a circuit simulator can be used to predict the voltages and currents at any node.

Alternatively, an electromagnetic simulator can be used to simulate the electric and magnetic fields everywhere in space, based on the physical design. From the electric and magnetic fields, a behavioral model of the interconnects can be generated, which can then be used in a circuit simulator, or the electric and magnetic fields can be converted into voltage and currents to show performance.

1.10 Simulations

There are three types of electrical simulation tools that predict the analog effects of the interconnects on signal behavior:

1. Electromagnetic (EM) simulators, which solve Maxwell's Equations and simulate the electric and magnetic fields at various locations in the time or frequency domains
2. Circuit simulators, which solve the differential equations corresponding to various circuit elements and include Kirchhoff's current and voltage relationships to predict the voltages and currents at various circuit nodes, in the time or frequency domains
3. Behavioral simulators, which use models based on tables and transmission lines and other passive-element models based on transfer functions, which quickly predict the voltages and currents at various nodes, typically in the time domain

Blame signal integrity on Maxwell's Equations. These four equations describe how conductors and dielectrics interact with electric and magnetic fields. After all, signals are nothing more than propagating electric and magnetic fields. When the electric and magnetic fields themselves are simulated, the interconnects and all passive components must be translated into conductors and dielectrics, with their associated geometries and material properties.

A device driver is converted into an incident electromagnetic wave, and Maxwell's Equations are used to predict how this wave interacts with the conductors and dielectrics. The material geometries and properties define the boundary conditions in which Maxwell's Equations are solved.

Though Figure 1-17 shows the actual Maxwell's Equations, it is never necessary for any practicing engineer to solve them by hand. They are shown here only for reference and to demonstrate that there really is a set of just a few simple equations that describe absolutely everything there is to know about electromagnetic fields. How the incident electromagnetic field interacts with the geometry and

Time Domain	*Frequency Domain*
$\nabla \bullet \varepsilon E = \frac{\rho}{\varepsilon_0}$	$\nabla \bullet \varepsilon E = \frac{\rho}{\varepsilon_0}$
$\nabla \bullet B = 0$	$\nabla \bullet \mu H = 0$
$\nabla x E + \frac{\partial B}{\partial t} = 0$	$\nabla x E + j\omega\mu H = 0$
$\nabla x B - \frac{\mu\varepsilon}{c^2}\frac{\partial E}{\partial t} = \mu_0$	$\nabla x H - j\omega\varepsilon E = J$

Figure 1-17 Maxwell's Equations in the time and frequency domains. These equations describe how the electric and magnetic fields interact with materials through time and space. They are provided here just for reference.

materials, as calculated from these equations, can be displayed at every point in space. These fields can be simulated either in the time domain or the frequency domain.

An example of the simulated electric-field intensity inside a 208-pin plastic quad flat pack (PQFP) for an incident voltage sine wave on one pin at 2.0 GHz and 2.3 GHz is shown in Figure 1-18. The different shadings show the field intensity. This simulation illustrates that if a signal has frequency components at 2.3

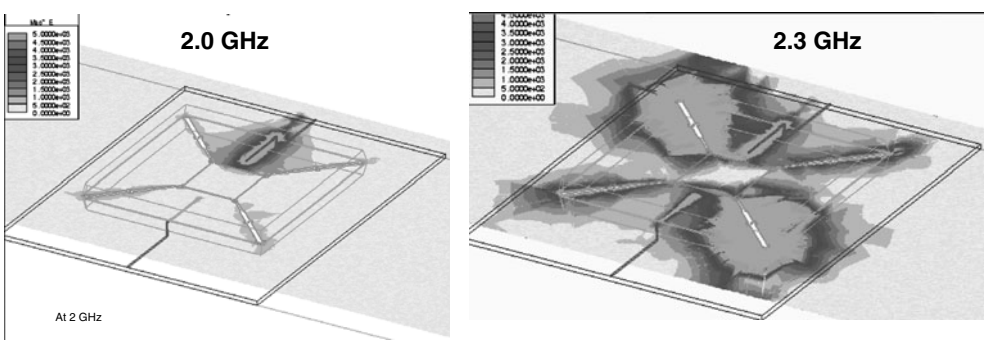


Figure 1-18 Example of an electromagnetic simulation of the electric fields in a 208-pin PQFP excited at 2.0 GHz (left) and at 2.3 GHz (right) showing a resonance. Simulation done with Ansoft's High Frequency Structure Simulator (HFSS).

GHz, it will cause large field distributions inside the package. These are called resonances and can be disastrous for a product. These resonances will cause signal-quality degradation, enhanced cross talk, and enhanced EMI. Resonances will always limit the highest bandwidth for which the part can be used.

Some effects can only be simulated with an electromagnetic simulator. Usually, an EM simulator is needed when interconnects are very nonuniform and electrically long (such as traces over gaps in the return path), when electromagnetic-coupling effects dominate (such as resonances in packages and connectors), or when it's necessary to simulate EMI effects.

Though all the physics can be taken into account with Maxwell's Equations, with the best of today's hardware and software tools, it is not practical to simulate the electromagnetic effects of other than the simplest structures. An additional limitation is that many of the current tools require a skilled user with experience in electromagnetic theory.

An alternative simulation tool, which is easier and quicker to use, is a circuit simulator. This simulation tool represents signals as voltages and currents. The various conductors and dielectrics are translated into their fundamental electrical circuit elements of resistances, capacitances, inductances, and their coupling.

Circuit theory is no less correct than the electromagnetic approach. It's just that some problems in signal integrity are more easily understood and their solutions are more easily identified by using the circuit description over the EM theory description. There are some limitations to what can be simulated by a circuit simulator. Electromagnetic effects such as EMI, resonances, and nonuniform wave propagation cannot be taken into account by a circuit simulator. However, effects such as near-field cross talk, transmission line propagation, and switching noise can be accurately accounted for. An example of a circuit and the resulting simulated waveforms are shown in Figure 1-19.

A circuit diagram that contains combinations of fundamental circuit elements is called a schematic. If you can draw the schematic, a circuit simulator will be able to calculate the voltages and currents at every node.

The most popular circuit simulator is generically called SPICE (short for simulation program with integrated circuit emphasis). The first version was created at UC Berkeley in the early 1970s as a tool to predict the performance of transistors based on their geometry and material properties. It is fundamentally a circuit simulator. If you input a schematic, in a specialized format, the tool will solve the differential equations each circuit element represents, then calculate the voltages and currents either in the time domain, called a transient simulation, or in

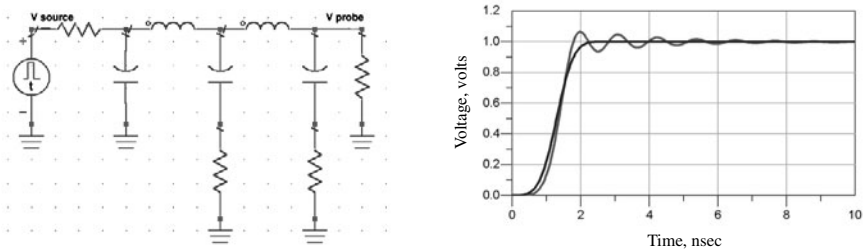


Figure 1-19 Example of a circuit model for the probe tip of a typical scope probe, approximately 5 cm long, and the resulting circuit simulation response of a clean, 1-nsec rise-time signal. The ringing is due to the excessive inductance of the probe tip.

the frequency domain, called an AC simulation. There are over 30 commercially available versions of SPICE, with some free student/demo versions available for download from the Web.

Behavioral simulators use tables and specialized transfer functions to simulate voltages and currents. Their chief advantage over circuit simulators is in their computation speed. Many of the behavioral simulators use proprietary simulation engines and are optimized for particular types of circuits, such as lossless, coupled transmission lines.

1.11 Modeling and Models

Modeling refers to creating an electrical representation of a device or component that a simulator can interpret and use to predict voltage and current waveforms. The models for active devices, such as transistors and output drivers, are radically different from the models of passive devices, such as all interconnects and discrete components. For active devices, a model is typically either a SPICE-compatible model or an input/output buffer interface spec (IBIS) compatible model.

A SPICE model of an active device will use either combinations of ideal sources and passive elements or specialized transistor models based on the geometry of the transistors. This allows easy scaling of the transistor's behavior if the process technology changes. A SPICE model contains information about the specific features and process technology of a driver. For this reason, most vendors are reluctant to give out SPICE models of their chips since they contain such valuable information.

IBIS is a format that defines the response of input or output drivers in terms of their V-I and V-t characteristics. A behavioral simulator will take the V-I and V-t curves of the active devices and simulate how these curves might change as they are affected by the transmission lines and lumped resistor (R), inductor (L), and capacitor (C) elements, which represent the interconnects. The primary advantage of an IBIS model for an active device is that an IC vendor can provide an IBIS model of its device drivers without revealing any proprietary information about the geometry of the transistors.

It is much easier to obtain an IBIS model than a SPICE model from an IC supplier. For system-level simulation, where 1,000 nets and 100 ICs may be simulated at the same time, IBIS models are typically used because they are more available and typically run faster than SPICE models.

The most important limitation to the accuracy of any simulation, SPICE or behavioral based, is the quality of the models. While it is possible to get an IBIS model of a driver that compares precisely with a SPICE model and matches the actual measurement of the device perfectly, it is difficult to routinely get good models of every device.

TIP In general, as end users, we must continually insist that our vendors supply us with some kind of verification of the quality of the models they provide for their components.

Another problem with device models is that a model that applies to one generation of chips will not match the next generation. With the next die shrink, which happens every six to nine months, the channel lengths are shorter, the rise times are shorter, and the V-I curves and transient response of the drivers change. An old model of a new part will give low estimates of the signal-integrity effects. As users, we must always insist the vendor provide current, accurate, and verified models of all drivers they supply.

TIP Though the intrinsic accuracy of all SPICE or behavioral simulators is typically very good, the quality of the simulations are only as good as the quality of the models that are simulated. The expression “garbage in, garbage out (GIGO)” was invented to describe circuit simulations.

For this reason, it is critically important to verify the accuracy of the models of all device drivers, interconnects, and passive components. Only in this way can

we have confidence in the results from the simulations. Though the models for the active devices are vitally important, this book deals with models for the passive devices and interconnects. Other references listed in the bibliography discuss active-device models.

Obtaining models for all the elements that make up the system is critically important. The only way to simulate signal-integrity effects is to include models of the interconnects and passive components, such as the board-level transmission lines, package models, connector models, decoupling capacitors, and terminating resistors.

Of course, the circuit model can only use elements that the simulator understands. For most behavioral simulators, this means interconnects are described by resistors, capacitors, inductors, and transmission lines. For SPICE simulators, interconnects and passive components can be described by resistors, capacitors, inductors, mutual inductors, and transmission lines. In some SPICE and behavioral simulators, new ideal circuit elements have been introduced that include ideal coupled transmission lines and ideal lossy transmission lines as basic ideal circuit elements.

An example of a physical component, two surface-mount terminating resistors, and their equivalent electrical circuit model is shown in Figure 1-20. This model includes their inductive coupling, which gives rise to switching noise. Every electrical quality of their behavior can be described by their circuit model. This circuit model, or schematic, can accurately predict any measurable effect.

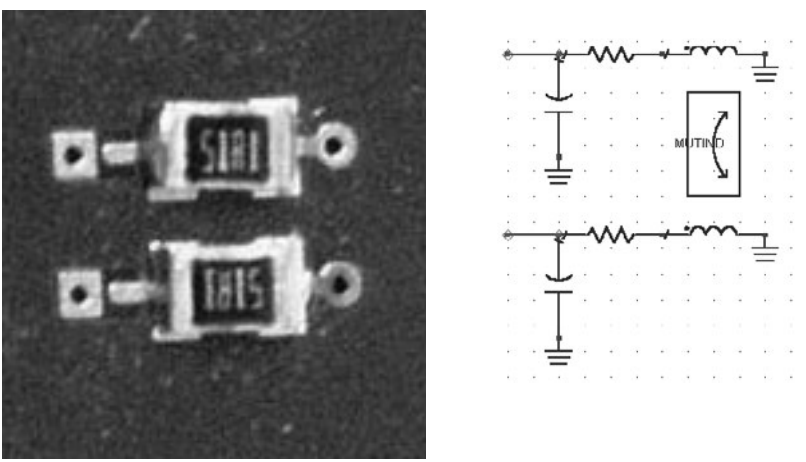


Figure 1-20 Two surface-mount 0805 resistors and their equivalent circuit model. This model has been verified up to 5 GHz.

There are two basic approaches to creating accurate circuit models of interconnects: by calculation and by measurements. We usually refer to creating a model from calculation as analysis and creating a model from a measurement as characterization.

1.12 Creating Circuit Models from Calculation

So much of life is a constant balancing act between the value received and the cost (in time and money). The analysis of all signal integrity design problems, as well as in most other fields, is no exception. We are constantly balancing the quality of an answer, as measured by its accuracy, for example, with how long it will take and how much it will cost to get the answer.

> **TIP** The goal in virtually all product-development programs in today's globally competitive market, is to get to an acceptable design that meets the performance spec, while staying within the time, cost, and risk budget.

This is a tough challenge. The engineer involved in signal integrity and interconnect design can benefit from skill and versatility in selecting the best technology and establishing the optimum design rules as early in the design cycle as possible.

The most important tool in any engineer's toolbox is the flexibility to evaluate trade-offs quickly. These are really trade-offs between how choices of geometry, material properties, and design rules will affect system performance.

> **TIP** The earlier in the design cycle the right trade-offs can be established, the shorter the development time and the lower the development costs.

To aid in the trade-off analysis, there are three levels of approximation used to predict performance or electrical properties. These three approaches are as follows:

1. Rules of thumb
2. Analytical approximations
3. Numerical simulations

Each approach represents a different balance of closeness to reality (i.e., accuracy) and the time and effort required to get the answer. This is illustrated in Figure 1-21. Of course, these approaches are not substitutes for actual measurements. However, the correct application of the right analysis technique can sometimes shorten the design-cycle time to 10% of its original value, compared to relying on the build it/test it/redesign it approach.

Rules of thumb are simple relationships that are easy to remember and help feed your intuition. An example of a rule of thumb is, the self-inductance of a short wire is about 25 nH/inch. A wire bond 0.1 inches long, therefore, would have a self-inductance of about 2.5 nH.

A rule of thumb should be the first step in any analysis problem, if only to provide a sanity-check number that every answer would be compared to. When you are brainstorming design/technology/cost trade-offs or looking for rough estimates, comparisons, or plausibility arguments, using rules of thumb can accelerate your progress better than tenfold. After all, the closer to the beginning of the product-development cycle the right design and technology decisions can be made, the more time and money will be saved in the project.

Rules of thumb are not meant to be accurate, they are meant to give an answer quickly. They should never be used when signing off on a design. They should be used to calibrate your intuition and provide guidance to help make high-level trade-offs. Appendix A contains a summary of many of the important rules of thumb used in signal integrity.

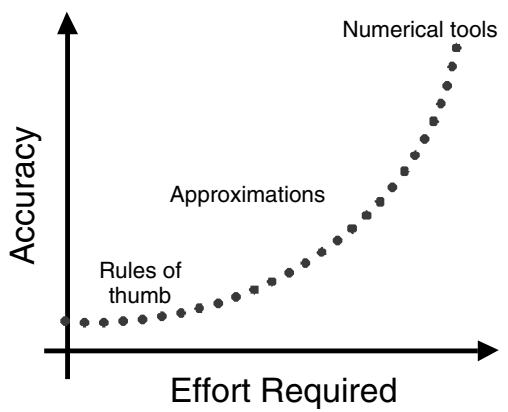


Figure 1-21 The balance between accuracy returned and effort required for the three levels of analysis. Each tool has its appropriate time and place.

Analytical approximations are equations or formulas. For example, an approximation for the loop self-inductance of a circular loop of wire is:

$$L_{self} = 32 \times R \times \ln\left(\frac{4R}{D}\right) nH \quad (1\text{-}2)$$

where:

L_{self} = the self-inductance, in nH

R = the radius of the loop, in inches

D = the diameter of the wire, in inches

For example, a round loop, 1/2 inch in radius or 1 inch in diameter, made from 10-mil-thick wire, would have a loop inductance of about 85 nH. Put your index finger and thumb together in a circle. If they were made of 20-gauge copper wire, the loop inductance would be about 85 nH.

Approximations have value in that they can be implemented in a spreadsheet, and they can answer what-if questions quickly. They identify the important first-order terms and how they are related. The approximation above illustrates that the inductance scales a little faster than the radius. The larger the radius of the loop, the larger the inductance of the loop. Also, the larger the wire thickness, the smaller the loop inductance, but only slightly, as the loop inductance varies inversely with the natural log of the thickness, a weak function.

TIP It is important to note that with very few exceptions, every equation you see being used in signal-integrity analysis is either a definition or an approximation.

A definition establishes an exact relationship between two or more terms. For example, the relationship between clock frequency and clock period, F = 1/T, is a definition. The relationship between voltage, current, and impedance, Z = V/I, is a definition.

You should always worry about the accuracy of an approximation, which may vary from 1% to 50% or more. Just because a formula allows evaluation on a calculator to five decimal places doesn't mean it is accurate to five decimal places. You can't tell the accuracy of an approximation by its complexity or its popularity.

How good are approximations? If you don't know the answer to this question for your specific case, you may not want to base a design sign-off—where a variance of more than 5% might mean the part won't work in the application—on an approximation with unknown quality. The first question you should always ask of every approximation is: How accurate is it?

One way of verifying the accuracy of an approximation is to build well-characterized test vehicles and perform measurements that can be compared to the calculation. Figure 1-22 shows the very good agreement between the approximation for loop inductance offered above and the measured values of loop inductance based on building loops and measuring them with an impedance analyzer. The agreement is seen to be better than 2%.

Approximations are extremely important when exploring design space or performing a tolerance analysis. They are wonderful when balancing trade-offs. However, when being off will cost a significant amount of time, money, or resources, you should never rely on approximations whose accuracy is uncertain.

There is a more accurate method for calculating the parameter values of the electrical circuit elements of interconnects from the geometry and material properties. It is based on the numerical calculation of Maxwell's Equations. These tools are called *field solvers*, in that they solve for the electric and magnetic fields based on Maxwell's Equations, using as boundary conditions the distribution of

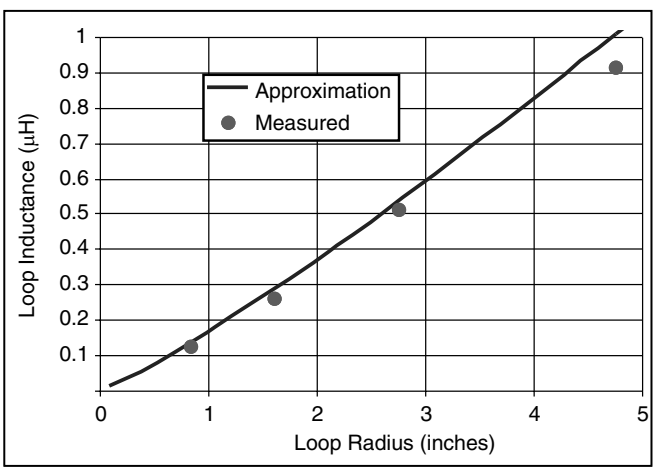


Figure 1-22 Comparison of the measured and calculated loop inductance of various circular wire loops as measured with an impedance analyzer. The accuracy of the approximation is seen to be about 2%.

conductors and dielectrics. A type of field solver that converts the calculated fields into the actual parameter values of the equivalent circuit-model elements, such as the R, L, or C values, is called a *parasitic extraction tool*.

When the geometry of the interconnect is uniform down its length, it can be described by its cross section, and a 2D field solver can be used to extract its transmission line properties. An example of a typical cross section of a microstrip transmission line and the simulated electric field lines and equipotentials is shown in Figure 1-23. For this structure, the extracted parameter values were Z_0 = 50.3 Ohms and TD = 142 psec.

When the cross section is nonuniform, such as in a connector or IC package, a 3D field solver is needed for the most accurate results.

Before relying on any numerical-simulation tool, it is always important to have its results verified for some test cases similar to the end application for which you will be using it. Every user should insist on vendor verification that the tool is suitably accurate for your typical application. In this way, you gain confidence in the quality of your results. The accuracy of some field solvers has been verified to better than 1%. Obviously, not all field solvers are always this accurate.

When accuracy is important, as, for example, in any design sign-off, a numerical-simulation tool, such as a parasitic extraction tool, should be used. It may take longer to construct a model using a numerical-simulation tool than it would using a rule of thumb or even using an analytic approximation. More effort in time and in expertise is required. But they offer greater accuracy and a higher

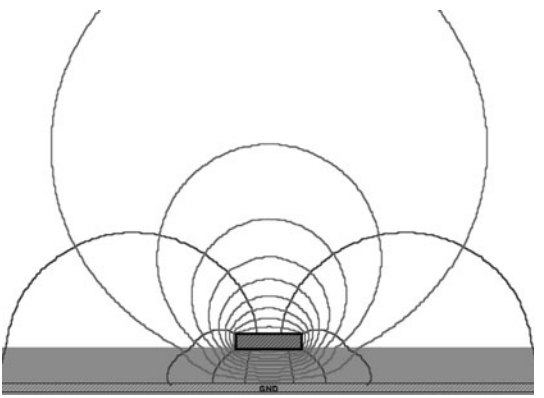


Figure 1-23 The results of a 2D field solver used to calculate the electric fields in a microstrip transmission line. The parasitic extraction tool used was Hyperlynx from Mentor Graphics. The accuracy of this tool has been independently verified to be better than 2%.

confidence the as-manufactured part will match the predicted performance. As new commercially available numerical-simulation tools advance, market pressures will always drive them to be easier to use.

Using a combination of these three analysis techniques, the trade-offs between the costs in time, money, and risk can be balanced with a very good prediction of the possible performance gain.

1.13 Three Types of Measurements

> **TIP** Though calculations play the critical role of offering a prediction of performance before the product is built, measurements play the critical role of risk reduction. The ultimate test of any calculation result is a measurement.

When it comes to measuring passive interconnects, as distinct from active devices, the measuring instrument must create a precision reference signal, apply it to the device under test, and measure the response. Ultimately, this response is related to the impedance of the device. In active devices, which create their own signals, the measurement instrument can be passive, merely measuring the created voltages or currents. There are three primary instruments used to perform measurements on passive components:

1. Impedance analyzer
2. Vector-network analyzer (VNA)
3. Time-domain reflectometer (TDR)

An impedance analyzer, typically a four-terminal instrument, operates in the frequency domain. One pair of terminals is used to generate a constant-current sine wave through the device under test (DUT). The second pair of terminals is used to measure the sine-wave voltage across the DUT.

The ratio of the measured voltage to the measured current is the impedance. The frequency is typically stepped from the low 100-Hz range to about 40 MHz. The magnitude and phase of the impedance at each frequency point is measured based on the definition of impedance.

The vector-network analyzer also operates in the frequency domain. Each terminal or port emits a sine-wave voltage at frequencies that can range in the low kHz to over 50 GHz. At each frequency, the incident-voltage amplitude and phase, as well as the reflected amplitude and phase, are measured.

The reflected signal will depend on the incident signal and the impedance change in going from the VNA to the DUT. The output impedance of a VNA is typically 50 Ohms. By measuring the reflected signal, the impedance of the device under test can be determined at each frequency point. The relationship between the reflected signal and the impedance of the DUT is:

$$\frac{V_{reflected}}{V_{incident}} = \frac{Z_{DUT} - 50\Omega}{Z_{DUT} + 50\Omega} \quad (1\text{-}3)$$

where:

$V_{reflected}$ = the amplitude and phase of the reflected sine-wave voltage

$V_{incident}$ = the amplitude and phase of the incident sine-wave voltage

Z_{DUT} = the impedance of the device under test

50 Ω = impedance of the VNA

The ratio of the reflected to the incident voltage, at each frequency, is often referred to as one of the scattering, or S, parameters, signified as S11. Measuring S11 and knowing the source impedance is 50 Ohms allow us to extract the impedance of the device under test at any frequency. An example of the measured impedance of a short-length transmission line is shown in Figure 1-24.

The time-domain reflectometer (TDR) is similar to the VNA but operates in the time domain. It emits a fast rise-time step signal, typically 35 psec to 150 psec, and measures the reflected transient amplitude. Again, using the reflected voltage, the impedance of the DUT can be extracted. In the time domain, the impedance measured represents the instantaneous impedance of the DUT. For an interconnect that is electrically long, such as a transmission line, the TDR can map the impedance profile. Figure 1-25 is an example of a TDR profile from a 4-inch-long transmission line with a small gap in the return plane showing a higher impedance where the gap occurs.

TIP Though the same impedance of a DUT can be displayed in the frequency domain or the time domain, it is really a different impedance in each case. When displayed in the frequency domain, the impedance is the total, integrated impedance of the entire DUT at each frequency. When displayed in the time domain, it is the instantaneous impedance at each spatially distinct point on the DUT that is displayed.

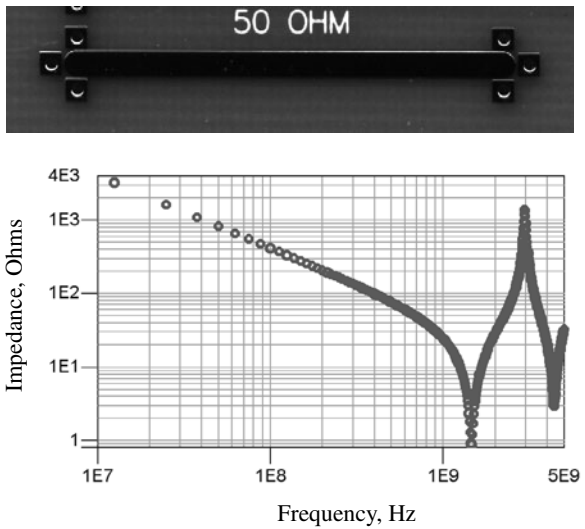


Figure 1-24 Measured impedance of a 1-inch length of transmission line. The network analyzer measured the reflected sine-wave signal between the front of the line and a via to the plane below the trace. This reflected signal was converted into the magnitude of the impedance. The phase of the impedance was measured, but is not displayed here. The frequency range is from 12 MHz to 5 GHz. Measured with a GigaTest Labs Probe Station.

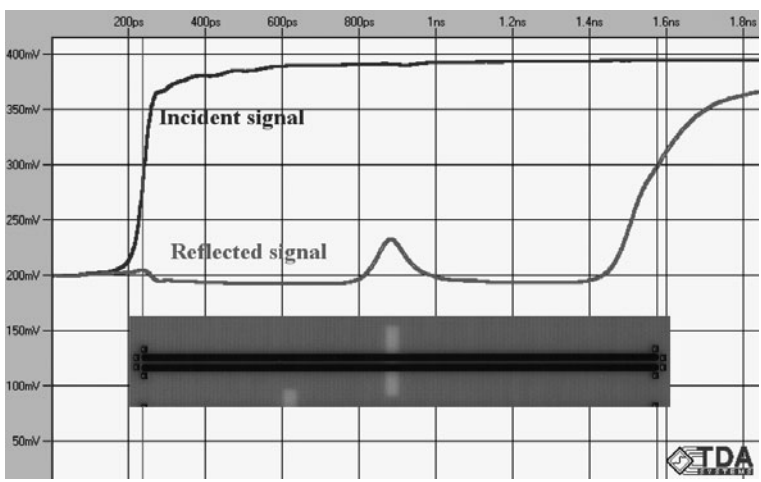


Figure 1-25 Measured TDR profile of a 4-inch-long uniform transmission line with a gap in the return path near the middle. The far end of the line is open. Measured with an Agilent 86100 DCA TDR and recorded with TDA Systems IConnect software, using a GigaTest Labs Probe Station.

1.14 The Role of Measurements

If it is possible to calculate the expected electrical performance of a component or system, why bother with a measurement? Why not just rely on the modeling and simulation tool? Measurements can only be performed on real devices. Isn't the goal to avoid the build-it/measure-it/redesign-it iteration loop that takes so much time?

Measurements such as the ones illustrated above play five critically important basic roles, at various stages in the product life cycle, all related to risk reduction and establishing higher confidence in the accuracy of the simulations. Measurements allow designers to do the following:

1. Verify the accuracy of the design/modeling/simulation process before significant resources are expended using an unverified process.
2. Verify an as-fabricated component meets the performance spec.
3. Create an equivalent electrical model for a component at any stage of the design cycle as parts are made available or acquired from a vendor.
4. Emulate system performance of a component as a quick way of determining expected performance without building an electrical model, at any stage of the design cycle as parts are made available or acquired from a vendor.
5. Debug a functional part or system, at any stage of the design cycle as parts are made available or acquired from a vendor.

An example from Delphi Electronics illustrates the incredible power of the combination of a design/modeling/simulation process that has been verified using measurements. One of the products Delphi Electronics makes is a custom flexible connector connecting two circuit boards for high-speed signals. An example is shown in Figure 1-26. They are used in servers, computers, and switching systems. The electrical performance of this connector is critical to the correct function of the system.

A customer comes to them with a set of performance specifications and Delphi is to deliver a part that meets these specs. The old way of designing the product was to make a best guess, manufacture the part, bring it back to the lab and perform measurements on it, compare to the specs, and redesign it. This is the old, build-it/test-it/redesign-it approach. In this old process, one iteration took almost nine weeks because of the long CAD and manufacturing cycle. Sometimes, the

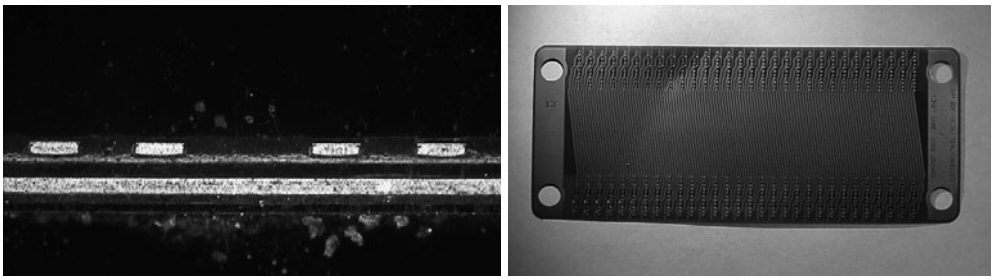


Figure 1-26 Gold Dot Connector from Delphi Automotive. Left: cross section of the two-metal-layer flex substrate. Right: top view of the hundreds of conductors connecting one board to the other, each with controlled electrical performance. Photo courtesy of Laurie Taira-Griffin, Delphi Electronics.

first design did not meet the customer specs and a second cycle was required. This meant an 18-week development cycle.

To shorten this design-cycle time, Delphi implemented a 2D modeling tool that allowed predicting the electrical properties of the connector based on the geometry and material properties. Through a few cycles of experimenting, using measurements with a TDR and VNA as the verification process, Delphi was able to fine-tune its modeling process to ensure accurate predictability of its product, before the part was sent to manufacturing. An example of the final agreement between what its modeling tool is able to predict and what was actually measured for a connector, compared with the customer's original specification, is shown in Figure 1-27.

Once the modeling/simulation process was in place and Delphi had confidence in its ability to accurately predict the final performance of the manufactured connectors, Delphi was able to reduce the design-cycle time to less than four hours. Nine weeks to four hours is a reduction of more than 100x. Measurements provided the critical verification step for this process.

1.15 The Bottom Line

1. Signal integrity problems relate to how the physical interconnects screw up pristine signals coming from the integrated circuits.
2. There are four general families of signal-integrity problems: signal quality on one net, cross talk between adjacent nets, rail collapse, and EMI.

Parameter	Simulation	Measured	Goal
Single-Ended Impedance	52.1 Ohms	53 Ohms	50 +/- 10% Ohms
Differential Impedance	95.2 Ohms	98 Ohms	100 +/- 10% Ohms
Attenuation (5GHz)	<.44 dB/inch	<.44 dB/inch	<.5 dB/inch
Propagation Delay	152 ps/inch	158 ps/inch	170 ps/inch
Single-Ended NEXT	<4.5%	<4.5%	<5%
Differential NEXT	<.3%	<.3%	<.5%
Data Rate	>5 Gbps	>5 Gbps	5 Gbps

Figure 1-27 Summary of the predicted and measured electrical specifications of the connector compared with the requirements for a particular connector. After the modeling/simulation process was optimized, the ability to predict performance was excellent.

3. Each of these problems gets worse and harder to solve as rise times decrease or clock frequencies increase.

4. It is inevitable that as transistors get smaller, their rise times will get shorter and signal integrity will be a greater problem.

5. To find, fix, and prevent signal-integrity problems, it is essential to be able to convert the physical design into its equivalent electrical circuit model and to use this model to simulate waveforms so as to predict performance before the product is built.

6. Three levels of analysis can be used for calculating electrical effects—rules of thumb, approximations, and numerical tools. These can be applied to modeling and simulation.

7. Three general instruments can be used to measure electrical properties of passives and interconnects: an impedance analyzer, a network analyzer, and a time-domain reflectometer.

8. These instruments play the important role of reducing the risk and increasing the confidence level in the accuracy of the modeling and simulation process.

9. Understanding the four signal-integrity problems leads to the most important ways to design them out of the product. Figure 1-28 summarizes the general solutions for the four families of signal integrity problems.

The rest of this book discusses the fundamental principles required to understand these problems and the specific techniques to minimize them in your product design.

Noise Category	*Design Principle*
Signal Quality	Signals should see the same impedance through all interconnects
Cross Talk	Keep spacing of traces greater than a minimum value, minimize mutual inductance with non-ideal returns
Rail Collapse	Minimize the impedance of the power/ground path and the delta I
EMI	Minimize bandwidth, minimize ground impedance, and shield

Figure 1-28 Summary of the four families of signal-integrity problems and the general design guidelines to minimize these problems. Even if these guidelines are followed, it is still essential to model and simulate the system to evaluate whether the design will meet the performance requirements.

This page intentionally left blank

CHAPTER 2

Time and Frequency Domains

In this chapter, we explore the basic properties of signals in preparation for looking at how they interact with interconnects. We will find that there are multiple ways of looking at a signal, each providing a different perspective. The quickest path to the answer may not be the most obvious path. The different perspectives we will use to look at signals are called domains. In particular we'll use the time domain and the frequency domain.

We will find that while we may generally be more familiar with the time domain, the frequency domain can provide valuable insight to understand and master many signal-integrity effects such as impedance, lossy lines, the power-distribution network, measurements, and models.

After introducing the two domains, we will look at how to translate between the two for some special cases. We will apply what we learn to relate two important quantities: rise time and bandwidth. The first is a time-domain term and the second a frequency-domain term. However, as we will see, they are intimately related.

Finally, we'll apply this concept of bandwidth to interconnects, models, and measurements.

2.1 The Time Domain

We use the term a lot—*the time domain*. But what do we really mean? What is the time domain? What are the features that are special about the time domain that make it useful? These are surprisingly difficult questions to answer because they seem so obvious and we rarely think about what we really mean by the time domain.

TIP The time domain is the real world. It is the only domain that actually exists.

We take it for granted because from the moment we are born, our experiences are developed and calibrated in the time domain. We are used to seeing events happen with a time stamp and ordered sequentially.

The time domain is the world of our experiences and is the domain in which high-speed digital products perform. When evaluating the behavior of a digital product, we typically do the analysis in the time domain because that's where performance is ultimately measured.

For example, two important properties of a clock waveform are clock period and rise time. Figure 2-1 illustrates these features.

The clock period is the time interval to repeat one clock cycle, usually measured in nanoseconds (nsec). The clock frequency, F_{clock}, or how many cycles per second the clock goes through, is the inverse of the clock period, T_{clock}.

$$F_{clock} = \frac{1}{T_{clock}} \tag{2-1}$$

where:

F_{clock} = the clock frequency, in GHz

T_{clock} = the clock period, in nsec

For example, a clock with a period of 10 nsec will have a clock frequency of 1/10 nsec = 0.1 GHz or 100 MHz.

The rise time is related to how long it takes for the signal to transition from a low value to a high value. There are two popular definitions of rise time. The 10–90 rise time is how long it takes for the signal to transition from 10% of its final value to 90% of its final value. This is usually the default meaning of rise time. It can be read directly off the time domain plot of a waveform.

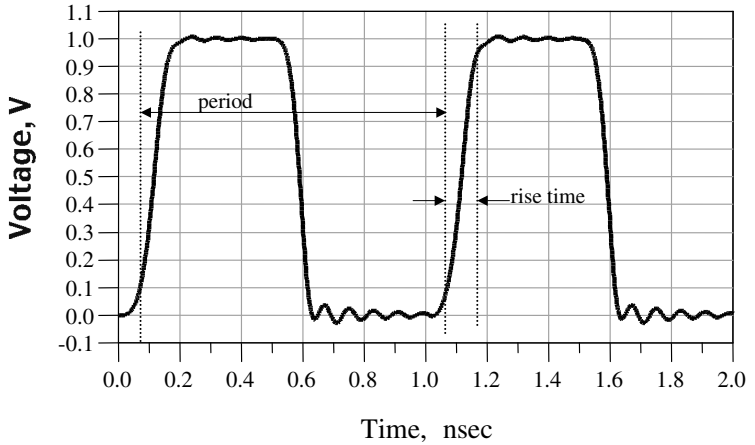


Figure 2-1 Typical clock waveform showing the clock period and the 10–90 rise time for a 1-GHz clock. The fall time is typically slightly shorter than the rise time and sometimes creates more noise.

The second definition is the 20–80 rise time. This is the time it takes for the signal to transition from 20% of its final value to 80% of its final value. Of course, for the same waveform the 20–80 rise time is shorter than the 10–90 rise time. Some IBIS models of real devices use the 20–80 definition of rise time. This makes it confusing. To remove ambiguity, it's often good practice referring explicitly to the 10–90 rise time or the 20–80 rise time.

There is a corresponding value for the fall time of a time-domain waveform. Depending on the logic family, the fall time is usually slightly shorter than the rise time. This is due to the design of typical CMOS output drivers. In a typical output driver, a p and an n transistor are in series between the V_{CC} (+) and the V_{SS} (–) power rails. The output is connected to the center, between them. Only one transistor is on at any one time, depending on whether the output is a low or a high.

When the driver switches from a low to a high (i.e., rising edge), the n transistor turns off and the p transistor turns on. The rise time is related to how fast the p transistor can turn on. When switching from the high to the low state (i.e., a falling edge), the p transistor turns off and the n transistor turns on. In general, for the same feature-size transistor, an n transistor can turn on faster than a p transistor. This means switching from high to low, the falling edge will be shorter than the rising edge. In general, signal-integrity problems are more likely to occur when switching from a high to low transition than from a low to high transition. By

making the n channel transistor larger than the p channel, the rising and falling edges can be closely matched.

Having established an awareness of the time domain as a distinct way of looking at events, we can turn our attention to one of a number of alternative ways of analyzing the world—the frequency domain.

2.2 Sine Waves in the Frequency Domain

We hear the term *frequency domain* quite a bit, especially when it involves radio frequency (rf) or communications systems. We will also encounter the frequency domain in high-speed digital applications. There are few engineers who have not heard of and used the term multiple times. Yet, what do we really mean by the frequency domain? What is the frequency domain and what makes it special and useful?

> **TIP** The most important quality of the frequency domain is that it is not real. It is a mathematical construct. The only reality is the time domain. The frequency domain is a mathematical world where very specific rules are followed.

The most important rule in the frequency domain is that the only kind of waveforms that exist are sine waves. Sine waves are the language of the frequency domain.

There are other domains that use other special functions. For example, the JPEG picture-compression algorithm takes advantage of special waveforms that are called wavelets. The wavelet transform takes the space domain, with a lot of x-y amplitude information content, and translates it into a different mathematical description that is able to use less than 10% of the memory to describe the same information. It is an approximation, but a very good one.

It's common for engineers to think that we use sine waves in the frequency domain because we can build any time-domain waveform from combinations of sine waves. This is a very important property of sine waves. However, there are many other waveforms with this property. It is not a property that is unique to sine waves.

In fact, there are four properties that make sine waves very useful for describing any other waveform. These properties are as follows:

1. Any waveform in the time domain can be completely and uniquely described by combinations of sine wave.
2. Any two sine waves with different frequencies are orthogonal to each other. If you multiply them together and integrate over all time, they integrate to zero. This means you can separate each component from every other.
3. They are well defined mathematically.
4. They have a value everywhere with no infinities and they have derivatives that have no infinities anywhere. This means they can be used to describe real world waveforms, since there are no infinities in the real world.

All of these properties are vitally important, but are not unique to sine waves. There is a whole class of functions called *orthonormal functions*, or sometimes called *eigenfunctions* or *basis functions*, which could be used to describe any time-domain waveform. Other orthonormal functions are Hermite Polynomials, Legendre Polynomials, Laguerre Polynomials, and Bessel Functions.

Why did we choose sine waves as our functions in the frequency domain? What's so special about sine waves? The real answer is that by using sine waves, some problems related to the electrical effects of interconnects will be easier to understand and solve using sine waves. If we switch to the frequency domain and use sine-wave descriptions, we can sometimes get to an answer faster than staying in the time domain.

TIP After all, if the time domain is the real world, we would never leave it unless the frequency domain provides a faster route to an acceptable answer.

Sine waves can sometimes provide a faster path to an acceptable answer because of the types of electrical problems we often encounter in signal integrity. If we look at the circuits that describe interconnects, we find that they will often include combinations of resistors (R), inductors (L), and capacitors (C). These elements in a circuit can be described by a second-order linear differential equation. The solution to this type of differential equation is a sine wave. In these circuits, the naturally occurring waveforms will be combinations of the waveforms that are solutions to the differential equation.

We find that in the real world, if we build circuits that contain Rs, Ls, and Cs and send any arbitrary waveform in, more often than not, we get waveforms out that look like sine waves and can more simply be described by a combination of a few sine waves. An example of this is shown in Figure 2-2.

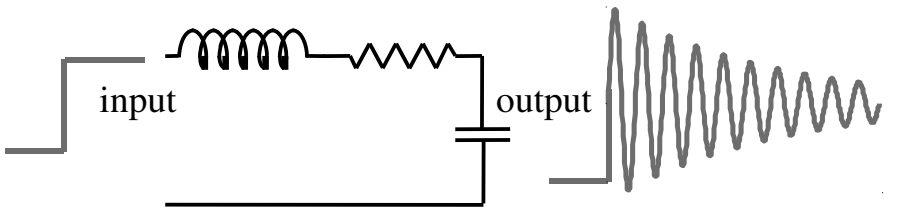


Figure 2-2 Time-domain behavior of a fast edge interacting with an ideal RLC circuit. Sine waves are naturally occurring when digital signals interact with interconnects, which can often be described as combinations of ideal RLC circuit elements.

2.3 Shorter Time to a Solution in the Frequency Domain

TIP The only reason we would ever want to move to another domain is to get to an acceptable answer faster.

In some situations, if we use the naturally occurring sine waves in the frequency domain rather than in the time domain, we may arrive at a simpler description to a problem and get to a solution faster.

It is important to keep in mind that there is fundamentally no new information in the frequency domain. The time- and the frequency-domain descriptions of the same waveforms will each have exactly the same information content.

However, some problems are easier to understand and describe in the frequency domain than in the time domain. For example, the concept of bandwidth is intrinsically a frequency-domain idea. We use this term to describe the most significant sine-wave frequency components associated with a signal, a measurement, a model, or an interconnect.

Impedance is defined in both the time and the frequency domain. However, it is far easier to understand, to use, and to apply the concepts of impedance in the frequency domain. We need to understand impedance in both domains, but we will often get to an answer faster by solving an impedance problem in the frequency domain first.

Looking at the impedance of the power and ground distribution in the frequency domain will allow a simpler explanation and solution to rail-collapse problems. As we shall see, the design goal for the power-distribution system is to keep its impedance below a target value from direct current (DC) up to the bandwidth of the typical signals.

When dealing with EMI issues, both the FCC specifications and the methods of measuring the electromagnetic compliance of a product are more easily performed in the frequency domain.

With today's current capabilities of hardware and software tools, the quality of the measurements and the computation speed of the numerical-simulation tools can sometimes be better in the frequency domain.

A high signal-to-noise ratio (SNR) means higher quality measurements. The SNR of a vector-network analyzer (VNA), which operates in the frequency domain, is constant over its entire frequency range, which can be –130 dB from 10 MHz up to 50 GHz and more. For a time-domain reflectometer (TDR), the effective bandwidth may be as high as 20 GHz, but the SNR starts at –70 dB at low frequency and drops to as low as –30 dB at 20 GHz.

Many of the effects related to lossy transmission lines are more easily analyzed, measured, and simulated by using the frequency domain. The series resistance of a transmission line increases with the square root of frequency, and the shunt AC leakage current in the dielectric increases linearly with frequency. The transient (time-domain) performance of lossy transmission lines is often more easily obtained by first transforming the signal into the frequency domain, looking at how the transmission line affects each frequency component separately, and then transforming the sine-wave components back to the time domain.

2.4 Sine Wave Features

As we now know, by definition, the only waveforms that exist in the frequency domain are sine waves. We should also be familiar with the description of a sine wave in the time domain. It is a well-defined mathematical curve that has three terms that fully characterize absolutely everything you could ever ask about it. An example is shown in Figure 2-3.

The following three terms fully describe a sine wave:

- Frequency
- Amplitude
- Phase

The frequency, usually identified using a small f, is the number of complete cycles per second made by the sine wave, in Hertz. Angular frequency is measured in radians per second. A radian is like degrees, describing a fraction of a cycle. There are $2 \times \pi$ radians in one complete cycle. The Greek letter ω is often

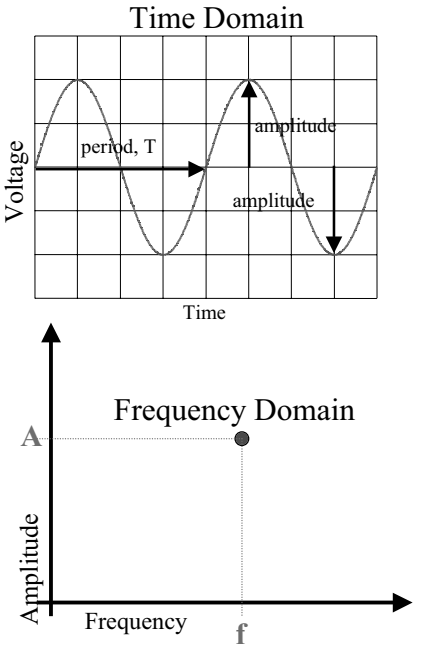


Figure 2-3 Top: Description of a sine wave in the time domain. It is composed of over one thousand voltage-versus-time data points. Bottom: Description of a sine wave in the frequency domain. Only three terms define a sine wave, which is a single point in the frequency domain.

used to refer to the angular frequency, measured in radians per second. The sine-wave frequency and the angular frequency are related by:

$$\omega = 2\pi \times f \tag{2-2}$$

where:

ω = angular frequency, in radians/sec

π = constant, 3.14159...

f = sine-wave frequency, in Hz

For example, if the frequency of a sine wave is 100 MHz, the angular frequency is $2 \times 3.14159 \times 100$ MHz ~ 6.3×10^8 radians/sec.

The amplitude is the maximum value of the peak height above the center value. The wave peak goes below the horizontal just as much as it goes above.

The phase is more complicated and identifies where the wave is in its cycle at the beginning of the time axis. The units of phase are in cycles, radians, or degrees, with 360 degrees in one cycle. While phase is important in mathematical analysis, we will minimize the use of phase in most of our discussion to concentrate on the more important aspects of sine waves.

In the time domain, describing a sine wave requires plotting a lot of voltage-versus-time data points to draw the complete sine-wave curve. However, in the frequency domain, describing a sine wave is much simpler.

In the frequency domain, we already know that the only waveforms we can talk about are sine waves, so all we have to identify are the amplitude, frequency, and phase. If there is only one sine wave we are describing, all we need are these three values and we have identified a complete description of the sine wave.

Since we are going to ignore phase for right now, we really only need two terms to completely describe a sine wave: its amplitude and its frequency. These two values are plotted with the frequency as one axis and the amplitude as the other axis, as shown in Figure 2-3. Of course, if we were including phase, we'd have a third axis. A sine wave, plotted in the frequency domain, is just one single data point. This is the key reason why we will go into the frequency domain. What might have been a thousand voltage-versus-time data points in the time domain is converted to a single amplitude-versus-frequency data point in the frequency domain.

When we have multiple frequency values, the collection of amplitudes is called the spectrum. As we will see, every time-domain waveform has a particular pattern to its spectrum. The only way to calculate the spectrum of a waveform in the time domain is with the Fourier Transform.

2.5 The Fourier Transform

The starting place for using the frequency domain is being able to convert a waveform from the time domain into a waveform in the frequency domain. We do this with the Fourier Transform. There are three types of Fourier Transforms:

- Fourier Integral (FI)
- Discrete Fourier Transform (DFT)
- Fast Fourier Transform (FFT)

The Fourier Integral (FI) is a mathematical technique of transforming an ideal mathematical expression in the time domain into a description in the frequency domain. For example, if the entire waveform in the time domain were just a short pulse, and nothing else, the Fourier Integral would be used to transform to the frequency domain.

This is done with an integral over all time from – infinity to + infinity. The result is a frequency-domain function that is also continuous from 0 to + infinity frequencies. There is a value for the amplitude at every continuous frequency value in this range.

For real-world waveforms, the time-domain waveform is actually composed of a series of discrete points, measured over a finite time, T. For example, a clock waveform may be a signal from 0 v to 1 v and have a period of 1 nsec and a repeat frequency of 1 GHz. To represent one cycle of the clock, there might be as many as 1000 discrete data points, taken at 1-psec intervals. An example of a 1-GHz clock wave in the time domain is shown in Figure 2-4.

To transform this waveform into the frequency domain, the Discrete Fourier Transform (DFT) would be used. The basic assumption is that the original time-domain waveform is periodic and repeats every T seconds. Rather than integrals, just summations are used so any arbitrary set of data can be converted to the frequency domain using simple numerical techniques.

Finally, there is the Fast Fourier Transform (FFT). It is exactly the same as a DFT, except that the actual algorithm used to calculate the amplitude values at each frequency point uses a trick of very fast matrix algebra. This trick works only if the number of time-domain data points is a power of two, for example 256 points, or 512 points, or 1024 points. The result is a DFT, only calculated 100–10,000 times faster than the general DFT algorithm, depending on the number of voltage points.

In general, it is common in the industry to use all three terms, FI, DFT, and FFT, synonymously. We now know there is a difference between them, but they have the same purpose—to translate a time-domain waveform into its frequency-domain spectrum.

TIP Once in the frequency domain, the description of a waveform is a collection of sine-wave frequency values. Each frequency component has an amplitude and phase associated with it. We call the entire set of frequency values and their amplitudes the spectrum of the waveform.

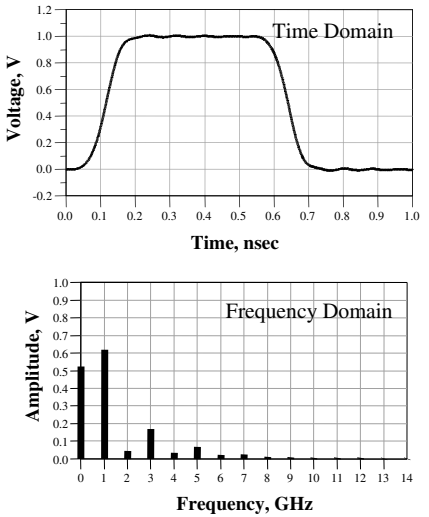


Figure 2-4 One cycle of a 1-GHz clock signal in the time domain (top) and frequency domain (bottom).

An example of a simple time-domain waveform and its associated spectrum, calculated by using a DFT, is shown in Figure 2-4.

At least once in his or her life, every serious engineer should calculate a Fourier Integral by hand, just to see the details. After this, we never again need to do the calculation manually. We can always get to an answer faster by using one of the many commercially available software tools that calculate Fourier Transforms for us.

There are a number of relatively easy-to-use, commercially available software tools that calculate the DFT or FFT of any waveform entered. Every version of SPICE has a function called the .FOUR command that will generate the amplitude of the first nine frequency components for any waveform. Most versions of the more advanced SPICE tools will also compute the complete set of amplitude and frequency values using a DFT. Microsoft Excel has an FFT function, usually found in the "engineering add-ins."

2.6 The Spectrum of a Repetitive Signal

In practice, the DFT or FFT is used to translate a real waveform from the time domain to the frequency domain. It is possible to take a DFT of any arbitrary, measured waveform. A key requirement of the waveform is that it be repetitive. We usually designate the repeat frequency of the time-domain waveform with the capital letter F.

For example, an ideal square wave might go from 0 v to 1 v, with a repeat time of 1 nsec and a 50% duty cycle. As an ideal square wave, the rise time to transition from 0 v to 1 v is precisely 0 sec. The repeat frequency would be 1/1 nsec = 1 GHz.

If a signal in the time domain is some arbitrary waveform over a time interval from t = 0 to t = T, it may not look repetitive. However, it can be turned into a repetitive signal by just repeating the interval every T seconds. In this case, the repeat frequency would be F = 1/T. Any arbitrary waveform can be made repetitive and the DFT used to convert it to the frequency domain. This is illustrated in Figure 2-5.

For a DFT, only certain frequency values exist in the spectrum. These values are determined by the choice of the time interval or the repeat frequency. When using an automated DFT tool, such as in SPICE, it is recommended to choose a value for the period equal to the clock period. This will simplify the interpretation of the results.

The only sine-wave frequency values that will exist in the spectrum will be multiples of the repeat frequency. If the clock frequency is 1 GHz, for example, the DFT will only have sine wave components at 1 GHz, 2 GHz, 3 GHz, etc.

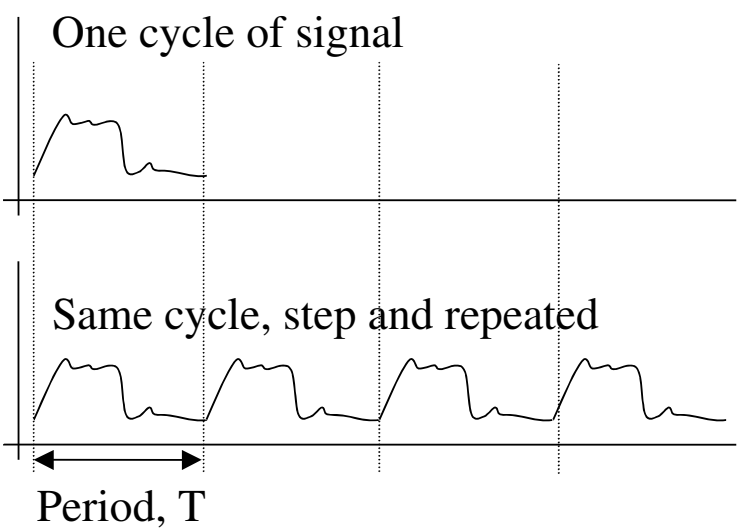


Figure 2-5 Any arbitrary waveform can be made to look repetitive. A DFT can be performed only on a repetitive waveform.

The first sine-wave frequency is called the first harmonic. The second sine-wave frequency is called the second harmonic, and so on. Each harmonic will have a different amplitude and phase associated with it. The collection of all the harmonics and their amplitudes is called the spectrum.

The actual amplitudes of each harmonic will be determined by the values calculated by the DFT. Every specific waveform will have its own spectrum.

2.7 The Spectrum of an Ideal Square Wave

An ideal square wave has a zero rise time, by definition. It is not a real waveform; it is an approximation to the real world. However, useful insight can be gained by looking at the spectrum of an ideal square wave and using this to evaluate real waveforms later. An ideal square wave has a 50% duty cycle, is symmetrical, and has a peak voltage of 1 v. This is illustrated in Figure 2-6.

If the ideal square-wave repeat frequency is 1 GHz, the sine-wave frequency values in its spectrum will be multiples of 1 GHz. We expect to see components at f = 1 GHz, 2 GHz, 3 GHz, and so on. But what are the amplitudes of each sine wave? The only way to determine this is to perform a DFT on the ideal square wave. Luckily, it is possible to calculate the DFT exactly for this special case of an ideal square wave. The result is relatively simple.

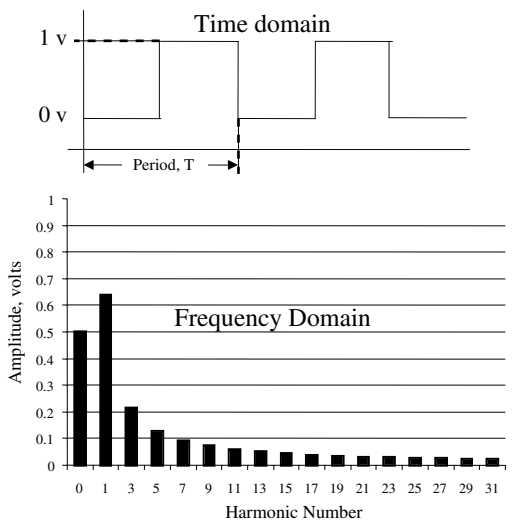


Figure 2-6 Time and frequency domain views of an ideal square wave.

The amplitudes of all the even harmonics (e.g., 2 GHz, 4 GHz, 6 GHz) are all zero. It is only odd harmonics that have values. The amplitudes, A_n, of the odd harmonics are given by:

$$A_n = \frac{2}{\pi \times n} \tag{2-3}$$

where:

A_n = the amplitude of the n^{th} harmonic

π = the constant, 3.14159...

n = the harmonic number, only odd allowed

For example, an ideal square wave with 50% duty-cycle and 0 v to 1 v transition has a first harmonic amplitude of 0.63 v. The amplitude of the third harmonic is 0.21 v. We can even calculate the amplitude of the 1001^{st} harmonic. It is 0.00063 v. It is important to note that the amplitudes of higher sine-wave-frequency components decrease with 1/f.

If the transition-voltage range of the ideal square wave were to double to 0 v to 2 v, the amplitudes of each harmonic would double as well.

There is one other special frequency value, 0 Hz. Since sine waves are all centered about zero, any combination of sine waves can only describe waveforms in the time domain that are centered about zero. To allow a DC offset, or a nonzero average value, the DC component is stored in the zero-frequency value. This is sometimes called the zeroth harmonic. Its amplitude is equal to the average value of the signal. In the case of the 50% duty-cycle square wave, the zeroth harmonic is 0.5 v.

To summarize:

- The collection of sine-wave-frequency components and their amplitudes is called the spectrum. Each component is called a harmonic.
- The zeroth harmonic is the DC value.
- For the special case of a 50% duty-cycle ideal square wave, the even harmonics have an amplitude of zero.
- The amplitude of any harmonic can be calculated as $2/(\pi \times n)$.

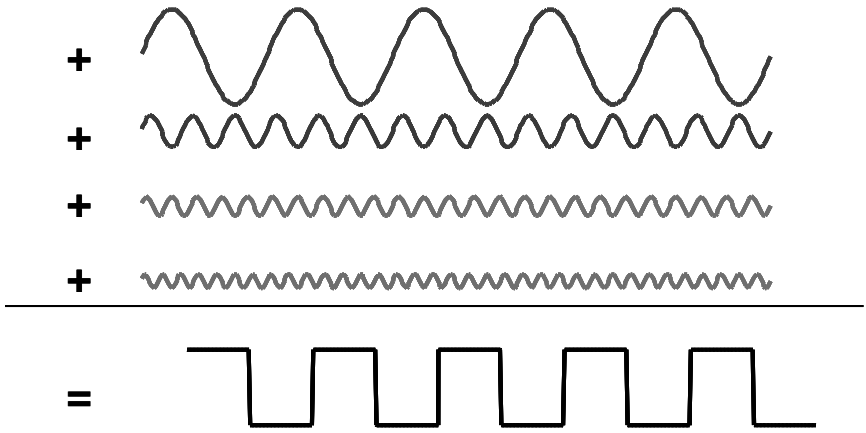


Figure 2-7 Convert the frequency-domain spectrum into the time-domain waveform by adding up each sine-wave component.

2.8 From the Frequency Domain to the Time Domain

The spectrum, in the frequency domain, represents all the sine-wave-frequency amplitudes of the time-domain waveform. If we have a spectrum and want to look at the time-domain waveform, we simply take each frequency component, convert it into its time-domain sine wave, then add it to all the rest. This process is called the Inverse Fourier Transform. It is illustrated in Figure 2-7.

Each component in the frequency domain is a sine wave in the time domain, defined from t = – infinity to t = + infinity. To re-create the time-domain waveform, we take each of the sine waves described in the spectrum and add them up in the time domain at each time-interval point. We start at the low-frequency end and add each harmonic based on the spectrum.

For a 1-GHz ideal-square-wave spectrum, the first term in the frequency domain is the zeroth harmonic, with amplitude of 0.5 v. This component describes a constant DC value in the time domain.

The next component is the first harmonic, which is a sine wave in the time domain with a frequency of 1 GHz and an amplitude of 0.63 v. When this is added to the previous term, the result in the time domain is a sine wave, offset to 0.5 v. It is not a very good approximation to the ideal square wave. This is shown in Figure 2-8.

The next term is the third harmonic. The amplitude of the 3-GHz sine-wave-frequency component is 0.21 v. When we add this to the existing time-domain

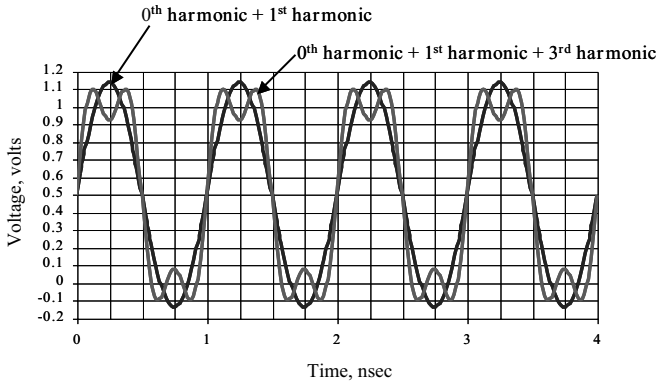


Figure 2-8 The time-domain waveform is created by adding together the zeroth harmonic and first harmonic and then the third harmonic, for a 1-GHz ideal square wave.

waveform, we see that it changes the shape of the new waveform slightly. The top is a bit more flat, better approximating a square wave, and the rise time is a little sharper. As we go through this process, adding each successive higher harmonic to re-create the ideal square wave, the resulting waveform begins to look more and more like a square wave. In particular, the rise time of the resulting time-domain waveform changes as we add higher harmonics.

To illustrate this in more detail, we can zoom in on the rise time of the waveform, centered about the beginning of a cycle. As we add all the harmonics up to the seventh harmonic, and then all the way up to the nineteenth, and finally, all the way up to the thirty-first harmonic, we see that the rise time of the resulting waveform in the time domain continually gets shorter. This is shown in Figure 2-9.

Depending on how the DFT was set up, there could be over 100 different harmonics listed in the spectrum. The logical question to ask is, do we have to include all of them, or can we still re-create a "good enough" representation of the original time-domain waveform with just a limited number of harmonics? What really is the impact of limiting the highest harmonic included in the re-created time-domain waveform? Is there a highest sine-wave-frequency component at which we can stop?

2.9 Effect of Bandwidth on Rise Time

The term *bandwidth* is used for the highest sine-wave-frequency component that is significant in the spectrum. This is the highest sine-wave frequency we need to include to adequately approximate the important features of the time-domain

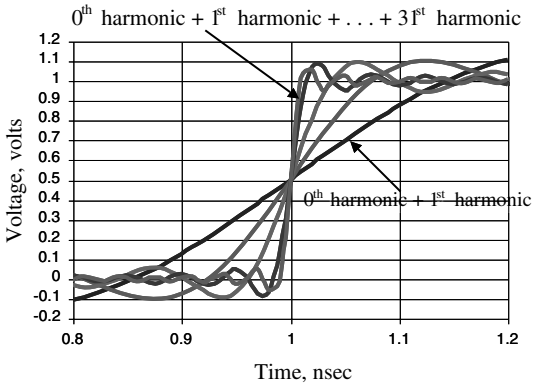


Figure 2-9 The time domain waveform created by adding together the zeroth harmonic and first harmonic, then the third harmonic and then up to the seventh harmonic, then up to the nineteenth harmonic, and then all harmonics up to the thirty-first harmonic, for a 1-GHz ideal square wave.

waveform. All frequency components of higher frequency than the bandwidth can be ignored. In particular, as we will see, the bandwidth we choose will have a direct effect on the shortest rise time of the signal we are able to describe in the time domain.

The term *bandwidth* historically is used in the rf world to refer to the range of frequencies in a signal. In rf applications, a carrier frequency is typically modulated with some amplitude or phase pattern. The spectrum of frequency components in the signal falls within a band. The range of frequencies in the rf signal is called the bandwidth. Typical rf signals might have a carrier frequency of 1.8 GHz with a bandwidth about this frequency of 100 MHz. The bandwidth of an rf signal defines how dense different communications channels can fit.

With digital signals, bandwidth also refers to the range of frequencies in the signal's spectrum. It's just that for digital signals, the low frequency range starts at DC and extends to the highest frequency component. In the world of digital signals, since the lowest frequency will always be DC, bandwidth will always be a measure of the highest sine wave frequency component that is significant.

When we created a time-domain waveform from just the zeroth, the first, and the third harmonics included, as in Figure 2-8, the bandwidth of the resulting waveform was just up to the third harmonic, or 3 GHz in this case. By design, the highest sine-wave-frequency component in this waveform is 3 GHz. The amplitude of all other sine-wave components in this time-domain waveform is exactly 0.

When we added higher harmonics to create the waveforms in Figure 2-9, we designed their bandwidths to be 7 GHz, 19 GHz, and 31 GHz. If we were to take the shortest rise-time waveform in Figure 2-9 and transform it back into the frequency domain, its spectrum would look exactly like that shown in Figure 2-6. It would have components from the zeroth to the thirty-first harmonics. Beyond the thirty-first harmonic, all the components would be zero. The highest sine-wave-frequency component that is significant in this waveform is the thirty-first harmonic, or the waveform has a bandwidth of 31 GHz.

In each case, we created a waveform with a higher bandwidth, using the ideal-square-wave's spectrum as the starting place. And, in each case, the higher-bandwidth waveform also had a shorter 10–90 rise time. The higher the bandwidth, the shorter the rise time and the more closely the waveform approximates an ideal square wave. Likewise, if we do something to a short rise-time signal to decrease its bandwidth (i.e., eliminate high-frequency components), its rise time will increase.

For example, it is initially difficult to evaluate the time-domain response of a signal propagating down a lossy transmission line in FR4. As we will see, there are two loss mechanisms: conductor loss and dielectric loss. If each of these processes were to attenuate low-frequency components the same as they do high-frequency components, there would simply be less signal at the far end, but the pattern of the spectrum would look the same coming out as it does going in. There would be no impact on the rise time of the waveform.

However, both conductor loss and dielectric loss will attenuate the higher-frequency components more than the low-frequency components. By the time the signal has traveled through even four inches of trace, the high-frequency components, above about 8 GHz, can have lost more than 50% of their power, leaving the low-frequency terms less affected. In Figure 2-10 (top), we show the measured attenuation of sine-wave-frequency components through a four-inch length of transmission line in FR4. This transmission line happens to have a 50-Ohm characteristic impedance and was measured with a network analyzer. Frequency components below 2 GHz are not attenuated more than –1 dB, while components at 10 GHz are attenuated by –4 dB.

This preferential attenuation of higher frequencies has the impact of decreasing the bandwidth of a signal that would propagate through the interconnect. Figure 2-10 (bottom) is an example of the measured rise time of a 50-psec signal entering a 36-inch-long trace in FR4 and this same waveform when it exits the trace. The rise time has been increased from 50 psec to nearly 1.5 nsec, due to the higher attenuation of the high-frequency components. Thirty-six inches is a typi-

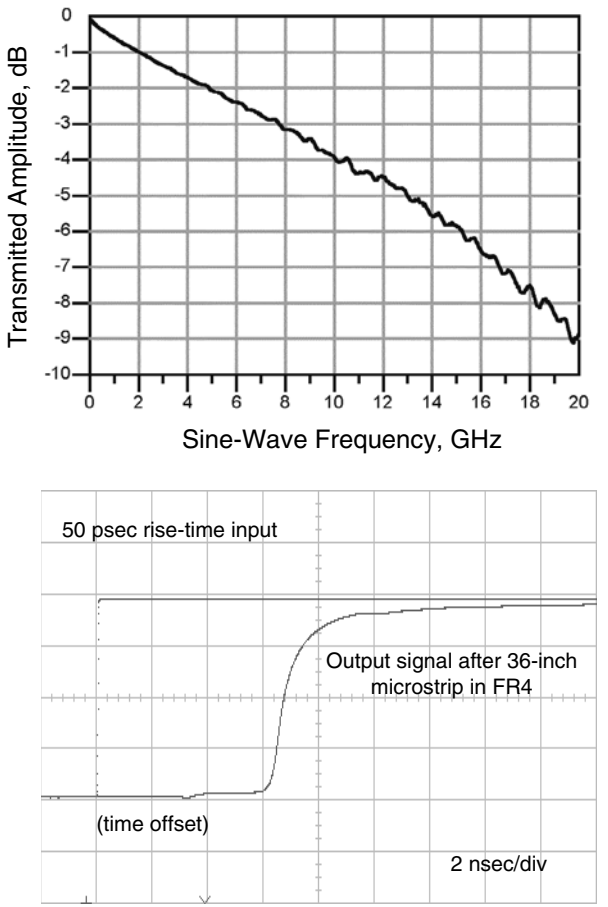


Figure 2-10 Top: The measured attenuation through a 4-inch length of 50-Ohm transmission line in FR4 showing the higher attenuation at higher frequencies. Bottom: The measured input and transmitted signal through a 36-inch 50-Ohm transmission line in FR4, showing the rise time to have degraded from 50 psec to more than 1.5 nsec.

cal length for a trace that travels over two 6-inch-long daughter cards and 24 inches of backplane. This rise-time degradation is the chief limitation to the use of FR4 laminate in high-speed serial links above 1 GHz.

TIP In general, a shorter rise-time waveform in the time domain will have a higher bandwidth in the frequency domain. If something is done to the spectrum to decrease the bandwidth of a waveform, the rise time of the waveform will be increased.

The connection between the highest sine-wave-frequency component that is significant in a spectrum and the corresponding rise time of the waveform in the time domain is a very important property.

2.10 Bandwidth and Rise Time

The relationship between rise time and bandwidth for a re-created ideal square wave can be quantified. In each synthesized waveform in the previous example re-creating an ideal square wave, the bandwidth is explicitly known because each waveform was artificially created by including sine-wave-frequency components only up to a specified value. The rise time, defined as the time from the 10% point to the 90% point, can be measured from time-domain plots.

When we plot the measured 10–90 rise time and the known bandwidth for each waveform, we see that empirically there is a simple relationship. This is a fundamental relationship for all signals and is shown in Figure 2-11.

For the special case of a re-created square wave with only some of the higher harmonics included, the bandwidth is inversely related to the rise time. We can fit a straight-line approximation through the points and find the relationship between bandwidth and rise time as:

$$BW = \frac{0.35}{RT} \tag{2-4}$$

where:

BW = the bandwidth, in GHz

RT = the 10–90 rise time, in nsec

For example, if the rise time of a signal is 1 nsec, the bandwidth is about 0.35 GHz or 350 MHz. Likewise, if the bandwidth of a signal is 3 GHz, the rise time of the signal will be about 0.1 nsec. A signal with a rise time of 0.25 nsec, such as might be seen in a DDR3-based system, has a bandwidth of 0.35/0.25 nsec = 1.4 GHz.

There are other ways of deriving this relationship for other waveforms, such as with Gaussian or exponential edges. The approach we took here for square waves is purely empirical and makes no assumptions. It is one of the most useful rules of thumb in our toolbox.

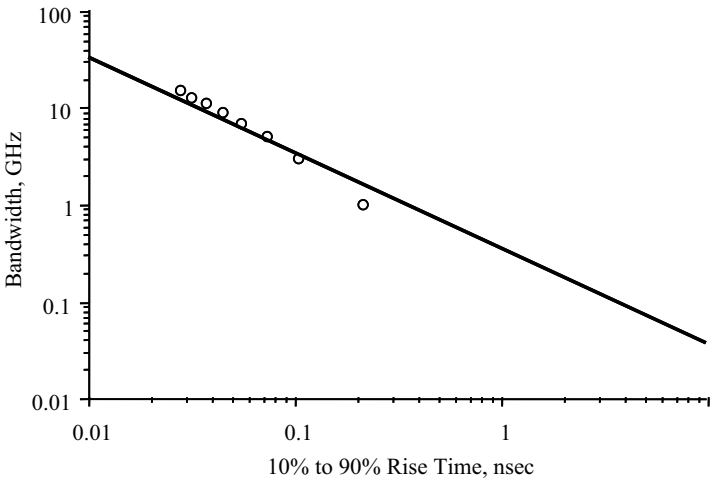


Figure 2-11 Empirical relationship between the bandwidth of a signal and its 10–90 rise time, as measured from a re-created ideal square wave with each harmonic added one at a time. Circles are the values extracted from the data; line is the approximation of BW = 0.35/rise time.

It is important to keep the units straight. When rise time is in microseconds, the bandwidth is in MHz. For example, a very long rise time of 10 microseconds has a bandwidth of about 0.35/10 microsec = 0.035 MHz. This is equivalent to 35 kHz.

When the rise time is in nanoseconds, the bandwidth is in GHz. A 10-nsec rise time, typical of a 10-MHz clock frequency, has a bandwidth of about 0.35/10 nsec = 0.035 GHz or 35 MHz.

2.11 What Does *Significant* Mean?

We defined the bandwidth of a signal as the highest sine-wave-frequency component that is significant. In the example, where we started with an ideal square wave and limited the high-frequency components, there was absolutely no ambiguity about what *significant* meant. We explicitly cut off all higher frequency sine-wave components in the frequency domain so that the highest significant component was the last harmonic in the spectrum.

We simply showed that if we include 100% of all the frequency components of an ideal square wave, up to the bandwidth, we would be able to re-create a square wave with a limited rise time, where the relationship of rise time = 0.35/BW. But what is the impact from adding only a fraction of the next component?

For example, if we take an ideal-square-wave clock signal with clock frequency of 1 GHz, its first harmonic will be a 1-GHz sine-wave frequency. If we were to include 100% of every component up to the twenty-first harmonic, the bandwidth would be 21 GHz and the resulting rise time of the re-created signal would be 0.35/21 GHz = 0.0167 nsec or 16.7 psec.

How would the rise time change if we added the twenty-third harmonic? The rise time would be 0.35/23 GHz = 0.0152 nsec or 15.2 psec. The rise time dropped by 1.5 psec. This is about 10% of the rise time, which is consistent, because we increased the bandwidth by 10%. The magnitude of the component we added was just 0.028 v, compared with the first harmonic of 0.63 v. Even though this amplitude is a small amount, less than 5% of the first harmonic amplitude and less than 3% of the peak value of the original square wave, it had the impact of dropping the rise time by 10%.

The spectrum of an ideal square wave has components that extend to infinite frequency. In order to achieve the zero rise time of an ideal square wave, each of these components is needed and is significant.

For a real time-domain waveform, the spectral components will almost always drop off in frequency faster than those of an ideal square wave of the same repeat frequency. The question of significance is really about the frequency at which amplitudes of the higher harmonics become small compared to the corresponding amplitudes of an ideal square wave.

By "small," we usually mean when the power in the component is less than 50% of the power in an ideal square wave's amplitude. A drop of 50% in power is the same as a drop to 70% in amplitude. This is really the definition of significant. Significant is when the amplitude is still above 70% of an ideal square wave's amplitude of the same harmonic.

TIP For any real waveform that has a finite rise time, *significant* refers to the point at which its harmonics are still more than 70% of the amplitude of an equivalent repeat-frequency ideal square wave's.

In a slightly different view, we can define *significant* as the frequency at which the harmonic components of the real waveform begin to drop off faster than 1/f. The frequency at which this happens is sometimes referred to as the *knee frequency*. The harmonic amplitudes of an ideal square wave will initially drop off similarly as 1/f. The frequency at which the harmonic amplitudes of a real waveform begin to significantly deviate from an ideal square wave is the knee frequency.

To evaluate the bandwidth of a time-domain waveform, we are really asking what is the highest frequency component that is just barely above 70% of the same harmonic of an equivalent ideal square wave. When the harmonic amplitudes of the real waveform are significantly less than an ideal square wave's, these lower amplitude harmonics will not contribute significantly to decreasing the rise time and we can ignore them.

For example, we can compare, in the time-domain waveform, two clock waves with a repeat frequency of 1 GHz: an ideal square wave and an ideal trapezoidal waveform, which is a non-ideal square wave with a long rise time. In this example, the 10–90 rise time is about 0.08 nsec, which is a rise time of about 8% of the period, typical of many clock waveforms. These two waveforms are shown in Figure 2-12.

If we compare the frequency components of these waveforms, at what frequency will the trapezoid's spectrum start to differ significantly from the ideal square wave's? We would expect the trapezoid's higher frequency components to begin to become insignificant at about 0.35/0.08 nsec = about 5 GHz. This is the fifth harmonic. After all, we could create a non-ideal square wave with this rise time if we were to take the ideal-square-wave spectrum and drop all components above the fifth harmonic, as we saw earlier.

When we look at the actual spectrum of the trapezoid compared to the square wave, we see that the first and third harmonics are about the same for each. The trapezoid's fifth harmonic is about 70% of the square wave's, which is still a large fraction. However, the trapezoid's seventh harmonic is only about 30% of the ideal square wave's. This is illustrated in Figure 2-12.

We would conclude, by simply looking at the spectra of the trapezoid, that harmonics above the fifth harmonic (i.e., the seventh and beyond) are contributing only a very small fraction of the amount of voltage as in the ideal square wave. Thus, their ability to further affect the rise time is going to be minimal. From the spectrum, we would say that the highest sine-wave-frequency component that is significant in the trapezoid, *compared to that in the ideal square wave*, is the fifth harmonic, which is what our approximation gave us.

There are higher harmonics in the trapezoid's spectrum than the fifth harmonic. However, the largest amplitude is 30% of the square wave's and then only a few percent after this. Their magnitude is such a small fraction of the amplitude of the ideal square wave's that they will contribute very little to the decrease of the rise time and can be ignored.

The bandwidth of any waveform is always the highest sine-wave-frequency component in its spectrum that is comparable in magnitude to a corresponding ideal

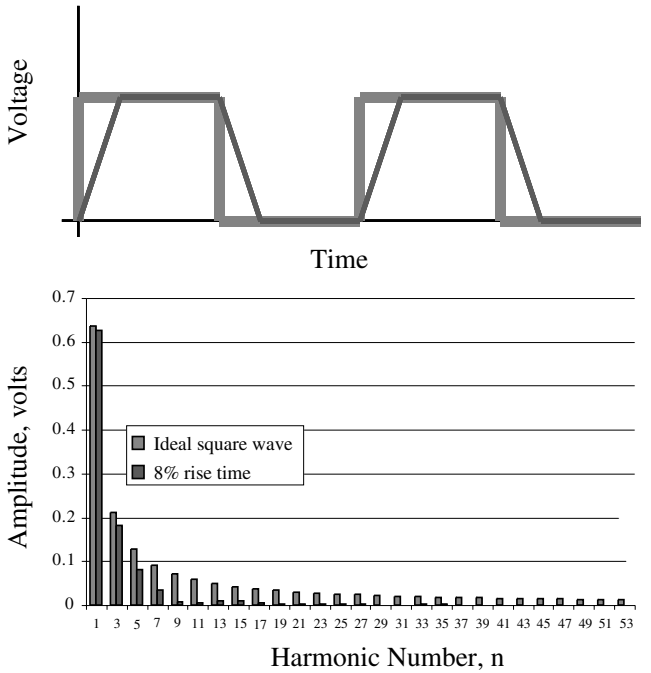


Figure 2-12 Top: Time domain waveforms of 1-GHz repeat frequency: an ideal square wave and an ideal trapezoidal wave with 0.08-nsec rise time. Bottom: Frequency-domain spectra of these waveforms showing the drop-off of the trapezoidal wave's higher harmonics, compared to the square wave's.

square wave. We can find out the bandwidth of any waveform by using a DFT to calculate its spectrum and compare it to an ideal square wave. We identify the frequency component of the waveform that is less than 70% of the ideal square wave, or we can use the rule of thumb developed earlier, that the BW is 0.35/rise time.

TIP It is important to note that this concept of bandwidth is inherently an approximation. It is really a rule of thumb, identifying roughly where the amplitude of frequency components in a real waveform begin to drop off faster than in an ideal square wave.

If you have a problem where it is important to know whether the bandwidth of a waveform is 900 MHz or 950 MHz, you should not use this term *bandwidth*. Rather, you should use the whole spectrum. The entire spectrum is always an accurate representation of the time-domain waveform.

2.12 Bandwidth of Real Signals

Other than the approximation for the bandwidth of a waveform based on its rise time, there is little calculation we can do by hand. Fourier Transforms of arbitrary waveforms can only be done using numerical simulation.

For example, the spectrum of a good-quality, nearly square wave signal has a simple behavior. If a transmission line circuit is poorly terminated, the signal may develop ringing. The resulting spectrum will have peaks at the ringing frequency. The amplitudes of the ringing frequency can be more than a factor of 10 greater than the amplitudes of the signal without ringing. This is shown in Figure 2-13.

The bandwidth of a waveform with ringing is clearly higher than one without. When ringing is present in a waveform, the bandwidth is better approximated

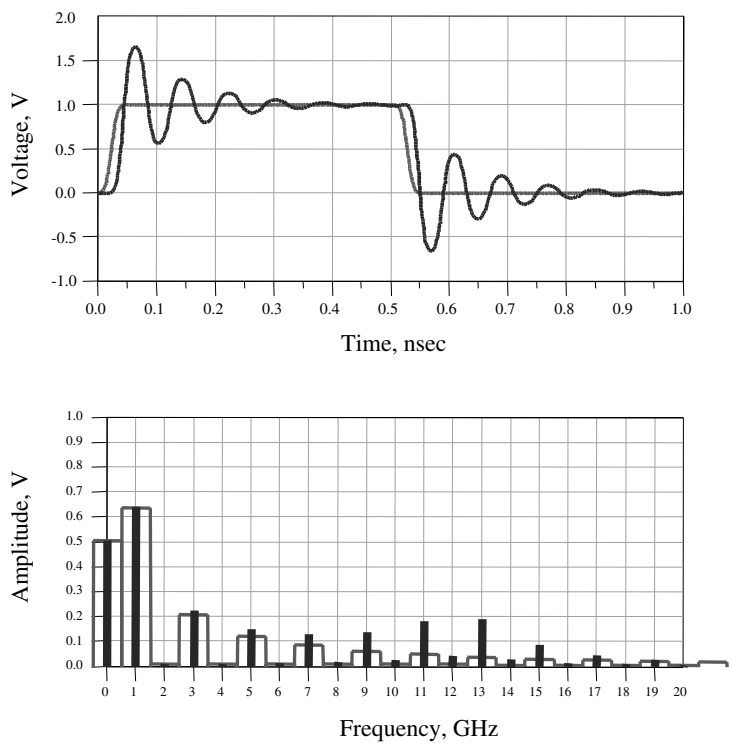


Figure 2-13 Top: The time-domain waveform of a near-square wave and one that has significant ringing due to poor termination. Bottom: the resulting DFT spectrum of these two waves, showing the effect of the ringing on the spectrum. The wide bars are for the ideal waveform while the narrow bars are for the ringing waveform.

by the ringing frequency. Just using the bandwidth to characterize a ringing signal, though, may be misleading. Rather, the whole spectrum needs to be considered.

EMI arises from each frequency component of the currents radiating. For the worst offender, the common currents, the amount of radiated emissions will increase linearly with the frequency. This means that if the current had an ideal-square-wave behavior, though the amplitude of each harmonic drops off at a rate of 1/f, the ability to radiate would increase at the rate of f, so all harmonics contribute equally to EMI. To minimize EMI, the design goal is to use the absolute lowest bandwidth possible in all signals. Above the bandwidth, the harmonic amplitudes drop off faster than 1/f, and would contribute to less radiated emissions. By keeping the bandwidth low, the radiated emissions will be kept to a minimum.

Any ringing in the circuits may increase the amplitudes of higher-frequency components and increase the magnitude of radiated emissions by a factor of 10. This is one reason why solving all signal-integrity problems is usually a starting place to minimize EMI problems.

2.13 Bandwidth and Clock Frequency

As we have seen, bandwidth relates to the rise time of a signal. It is possible to have two different waveforms, with exactly the same clock frequency but different rise times and different bandwidths. Just knowing the clock frequency cannot tell us what the bandwidth is. Figure 2-14 shows four different waveforms, each with exactly the same clock frequency of 1 GHz. However, they have different rise times and hence different bandwidths.

Sometimes, we don't always know the rise time of a signal but need an idea of its bandwidth anyway. Using a simplifying assumption, we can estimate the bandwidth of a clock wave from just its clock frequency. Still, it is important to keep in mind that it is not the clock frequency that determines the bandwidth, it is the rise time. If all we know about the waveform is the clock frequency, we can't know the bandwidth for sure; we can only guess.

To evaluate the bandwidth of a signal from just its clock frequency, we have to make a very important assumption. We need to estimate what a typical rise time might be for a clock wave.

How is the rise time related to the clock period in a real clock waveform? In principle, the only relationship is that the rise time must be less than 50% of the period. Other than this, there is no restriction, and the rise time can be any arbitrary fraction of the period. It could be 25% of the period, as in cases where the

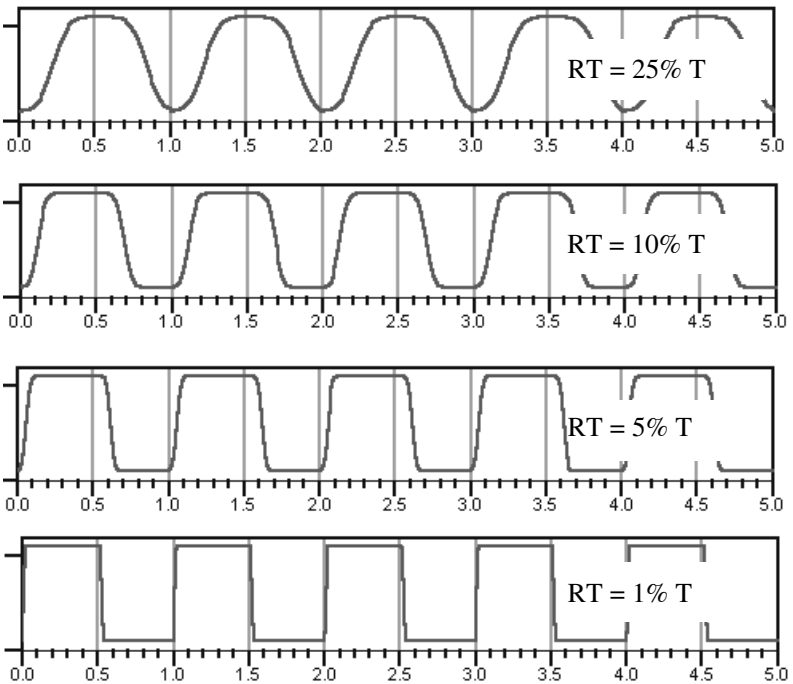


Figure 2-14 Four different waveforms, each with exactly the same 1-GHz clock frequency. Each of them has a different rise time, as a fraction of the period, and hence different bandwidths.

clock frequency is pushing the limits of the device technology, such as in 1-GHz clocks. It could be 10% of the period, which is typical of many microprocessor-based products. It could be 5% of the period, which is found in high-end FPGAs driving external low-clock-frequency memory buses. It could even be 1% if the board-level bus is a legacy system.

If we don't know what fraction of the period the rise time is, a reasonable generalization is that the rise time is 7% of the clock period. This approximates many typical microprocessor-based boards and ASICs driving board-level buses. From this, we can estimate the bandwidth of the clock waveform.

It should be kept in mind that this assumption of the rise time being 7% of the period is a bit aggressive. Most systems are probably closer to 10%, so we are assuming a rise time slightly shorter than might typically be found. Likewise, if we are underestimating the rise time, we will be overestimating the bandwidth, which is safer than underestimating it.

If the rise time is 7% of the period, then the period is 1/0.07 or 15 times the rise time. We have an approximation for the bandwidth as 0.35/rise time. We can relate the clock frequency to the clock period, because they are each the inverse of the other. Replacing the clock period for the clock frequency results in the final relationship; the bandwidth is five times the clock frequency:

$$BW_{clock} = 5 \times F_{clock} \tag{2-5}$$

where:

BW_{clock} = the approximate bandwidth of the clock, in GHz

F_{clock} = the clock repeat frequency, in GHz

For example, if the clock frequency is 100 MHz, the bandwidth of the signal is about 500 MHz. If the clock frequency is 1 GHz, the bandwidth of the signal is about 5 GHz.

This is a generalization and an approximation, based on the assumption that the rise time is 7% of the clock period. Given this assumption, it is a very powerful rule of thumb, which can give an estimate of bandwidth with very little effort. It says that the highest sine-wave-frequency component in a clock wave is typically the fifth harmonic!

It's obvious, but bears repeating, that we always want to use the rise time to evaluate the bandwidth. Unfortunately, we do not always have the luxury of knowing the rise time for a waveform. And yet, we need an answer *now*!

TIP Sometimes getting an OK answer is often more important than getting a BETTER answer LATE.

2.14 Bandwidth of a Measurement

So far, we have been using the term *bandwidth* to refer to signals, or clock waveforms. We have said that the bandwidth is the highest significant sine-wave-frequency component in the waveform's spectrum. And, for signals, we said *significant* was based on comparing the amplitude of the signal's harmonic to the amplitude of an equivalent repeat frequency ideal square wave's.

We also use this term *bandwidth* to refer to other quantities. In particular, it can relate to the bandwidth of a measurement, the bandwidth of a model, and the bandwidth of an interconnect. In each case, it refers to the highest sine-wave-

frequency component that is significant, but the definition of significant varies per application.

The bandwidth of a measurement is the highest sine-wave-frequency component that has significant accuracy. When the measurement is done in the frequency domain, using an impedance analyzer or a network analyzer, the bandwidth of the measurement is very easy to determine. It is simply the highest sine-wave frequency in the measurement.

The measured impedance of a decoupling capacitor, from 1 MHz up to 1 GHz, shows that below about 10 MHz, the impedance behaves like an ideal capacitor, but above 10 MHz, it looks like an ideal inductor. Such a measurement is shown in Figure 2-15. There is good, accurate data up to the full range of the network analyzer, in this case, up to 1 GHz. The bandwidth of the measurement is 1 GHz in this example. The measurement bandwidth is not the same as the useful application bandwidth of the device.

When the measuring instrument works in the time domain, such as a time-domain reflectometer (TDR), the bandwidth of the measurement can be found by the rise time of the fastest signal that can be launched into the DUT. After all, this is a rough measure of when the higher-frequency components are small.

In a typical TDR, a fast step edge is created and its change due to interaction with the DUT is measured. A typical rise time entering the DUT is 35 psec to 70

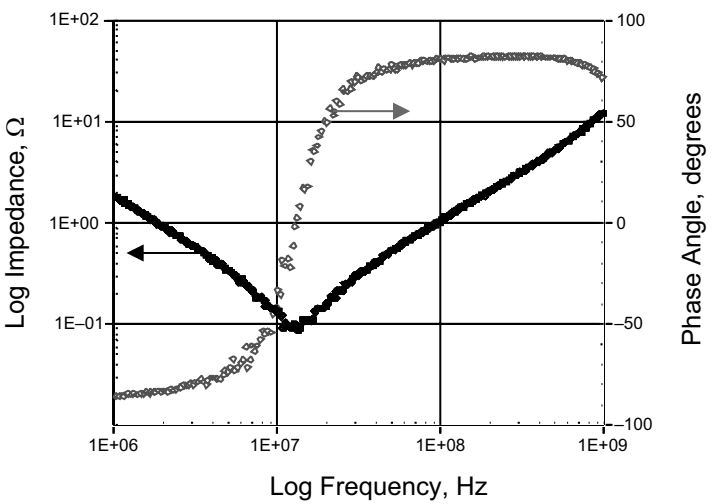


Figure 2-15 Measured impedance of a small 1206 ceramic decoupling capacitor. The measurement bandwidth for this data is 1 GHz.

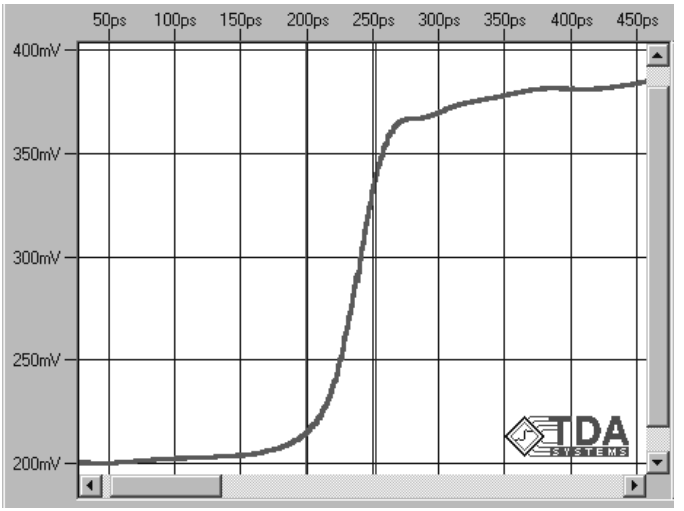


Figure 2-16 Measured TDR profile from the output of a 1-meter cable and microprobe tip, open at the end. The TDR rise time after the cable and probe is about 52 psec. The bandwidth of the measurement is about 0.35/52 psec = 7 GHz. The measurement was recorded with TDA Systems IConnect software, using a GigaTest Labs Probe Station.

psec, depending on the probes and cables used. Figure 2-16 shows the measured rise time of a TDR as about 52 psec. The bandwidth of the edge is 0.35/52 psec = 0.007 THz or 7 GHz. This is the bandwidth of the signal coming out of the TDR and is a good first order measure of the bandwidth of the measurement.

In state of the art TDRs, calibration techniques allow the bandwidth of the measurement to exceed the bandwidth of the signal. The bandwidth of the measurement is set by when the signal-to-noise ratio of a frequency component is below a reasonable value, like 10. The bandwidth of the measurement of some TDRs can exceed the signal's bandwidth by a factor of 3-5, making the bandwidth of a TDR's measurement as high as 30 GHz.

2.15 Bandwidth of a Model

TIP When we refer to the bandwidth of a model, we are referring to the highest sine-wave-frequency component where the model will accurately predict the actual behavior of the structure it is representing. There are a few tricks that can be used to determine this, but in general, only a comparison to a measurement will give a confident measure of a model's bandwidth.

The simplest starting equivalent circuit model to represent a wire bond is an inductor. Up to what bandwidth might this be a good model? The only way to really tell is to compare a measurement with the prediction of this model. Of course, it will be different for different wire bonds.

As an example, we take the case of a very long wire bond, 300 mils long, connecting two pads over a return-path plane 10 mils below. This is diagrammed in Figure 2-17. A simple starting circuit model is a single ideal inductor and ideal resistor in series, such as shown in Figure 2-18. The best values for the L and R give a prediction for the impedance that closely matches the measured impedance up to 2 GHz. The bandwidth of this simple model is 2 GHz. This is shown in Figure 2-18.

We could confidently use this simple model to predict performance of this physical structure in applications that had signal bandwidths of 2 GHz. It is surprising that for a wire bond this long, the simplest model, that of a constant ideal inductor and resistor, works so well up to 2 GHz. This is probably higher than the useful bandwidth of the wire bond, but the model is still accurate up to this high a frequency.

Suppose we wanted a model with an even higher bandwidth that would predict the actual impedance of this real wire bond to higher frequency. We might add the effect of the pad capacitance. Building a new model, a second-order model, and finding the best values for the ideal R, L, and C elements result in a simulated impedance that matches the actual impedance to almost 4 GHz. This is shown in Figure 2-18.

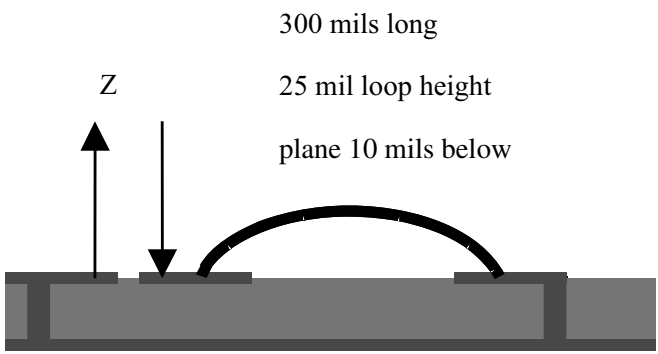


Figure 2-17 Diagram of a wire-bond loop between two pads, with a return path about 10 mils beneath the wire bond.

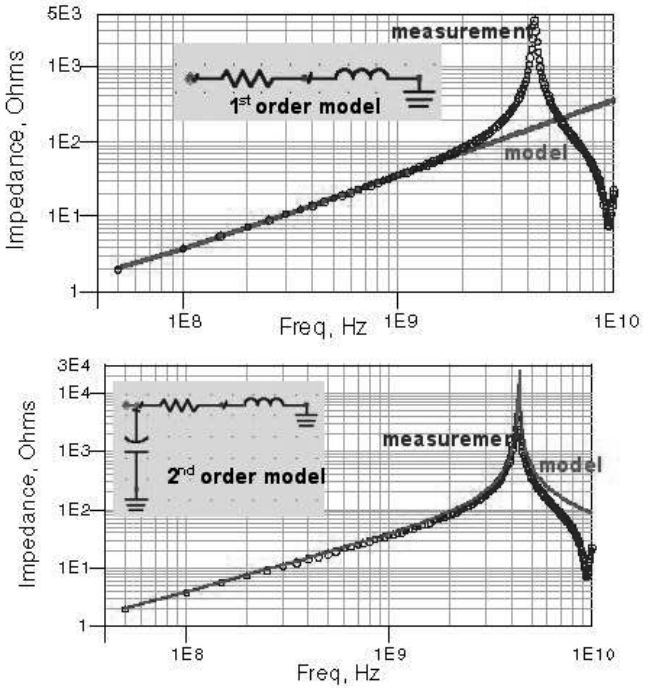


Figure 2-18 Top: Comparison of the measured impedance and the simulation based on the first-order model. The agreement is good up to a bandwidth of about 2 GHz. Bottom: Comparison of the measured impedance and the simulation based on the second-order model. The agreement is good up to a bandwidth of about 4 GHz. The bandwidth of the measurement is 10 GHz, measured with a GigaTest Labs Probe Station.

2.16 Bandwidth of an Interconnect

The bandwidth of an interconnect refers to the highest sine-wave-frequency component that can be transmitted by the interconnect without significant loss. What does *significant* mean? In some applications, a transmitted signal that is within 95% of the incident signal is considered too small to be useful. In other cases, a transmitted signal that is less than 10% of the incident signal is considered usable. In long-distance cable-TV systems, the receivers can use signals that have only 1% of the original power. Obviously, the notion of how much transmitted signal is significant is very dependent on the application and the particular specification. In reality, the bandwidth of an interconnect is the highest sine-wave frequency at which the interconnect still meets the performance specification for the application.

TIP In practice, *significant* means when the transmitted frequency-component amplitude is reduced by –3 dB, which means that its amplitude is reduced to 70% of the incident value. This is often referred to as the 3-dB bandwidth of an interconnect.

The bandwidth of an interconnect can be measured in either the time domain or the frequency domain. In general, we have to be careful interpreting the results if the source impedance is different than the characteristic impedance of the line, due to the complication of multiple reflections.

Measuring the bandwidth of an interconnect in the frequency domain is very straightforward. A network analyzer is used to generate sine waves of various frequencies. It injects the sine waves in the front of the interconnect and measures how much of each sine wave comes out at the far end. It is basically measuring the transfer function of the interconnect, and the interconnect is acting like a filter. This is also sometimes referred to as the *insertion loss* of the interconnect. The interpretation is simple when the interconnect is 50 Ohms, matched to the network analyzer's impedance.

For example, Figure 2-19 shows the measured transmitted amplitude of sine waves through a 4-inch length of a 50-Ohm transmission line in FR4. The

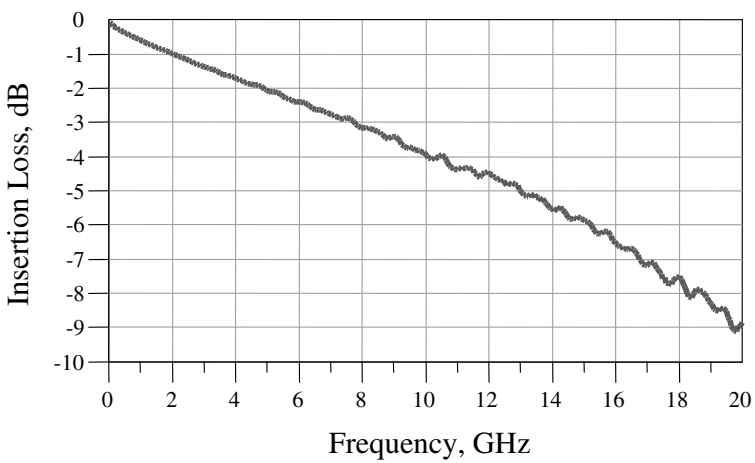


Figure 2-19 Measured transmitted amplitude of different sine-wave signals through a 4-inch-long transmission line made in FR4. The 3 dB bandwidth is seen to be about 8 GHz for this cross section and material properties. Measured with a GigaTest Labs Probe Station.

measurement bandwidth is 20 GHz in this case. The 3-dB bandwidth of the interconnect is seen to be about 8 GHz. This means that if we send in a sine wave at 8 GHz, at least 70% of the amplitude of the 8-GHz sine wave would appear at the far end. More than likely, if the interconnect bandwidth were 8 GHz, nearly 100% of a 1-GHz sine wave would be transmitted to the far end of the same interconnect.

The interpretation of the bandwidth of an interconnect is the approximation that if an ideal square wave were transmitted through this interconnect, each sine-wave component would be transmitted, with those components lower than 8 GHz having roughly the same amplitude coming out as they did going in. But the amplitude of those components above 8 GHz would be reduced to insignificance.

A signal that might have a rise time of 1 psec going into the interconnect would have a rise time of 0.35/8 GHz = 0.043 nsec or 43 psec when it came out. The interconnect will degrade the rise time.

TIP The bandwidth of the interconnect is a direct measure of the minimum rise-time signal an interconnect can transmit.

If the bandwidth of an interconnect is 1 GHz, the fastest edge it can transmit is 350 psec. This is sometimes referred to as its intrinsic rise time. If a signal with a 350-psec edge enters the interconnect, what will be the rise time coming out? This is a subtle question. The rise time exiting the interconnect can be approximated by:

$$RT^2_{out} = RT^2_{in} + RT^2_{interconnect} \tag{2-6}$$

where:

RT_{out} = the 10–90 rise time of the output signal

RT_{in} = the 10–90 rise time of the input signal

$RT_{interconnect}$ = the intrinsic 10–90 rise time of the interconnect

This assumes that both the incident spectra and the response of the interconnect correspond to a Gaussian-shaped rise time.

For example, in the case of this 4-inch-long interconnect, if a signal with a rise time of 50 psec were input, the rise time of the transmitted signal would be:

$$\text{sqrt}(50\ \text{psec}^2 + 43\ \text{psec}^2) = 67\ \text{psec}. \quad (2\text{-}7)$$

This is an increase of about 17 psec in the rise time of the transmitted waveform compared to the incident rise time.

In Figure 2-20, we show the measured time-domain response of the same 4-inch-long, 50-Ohm interconnect that was measured in the frequency domain above. The input waveform has been time shifted to lie directly at the start of the measured output waveform.

The rise time of the waveform going into the PCB trace is 50 psec. The measured 10–90 rise time of the output waveform is about 80 psec. However, this is somewhat distorted by the long roll to stabilize at the top, characteristic of the behavior of lossy lines. The extra delay at the 70% point is about 15 psec, which is very close to what our approximation above predicted.

If a 1-nsec rise-time signal enters an interconnect with an intrinsic rise time of 0.1 nsec, the rise time of the signal transmitted would be about sqrt(1 nsec^2 + 0.1 nsec^2), or 1.005 nsec, which is still basically 1 nsec. The interconnect would not affect the rise time. However, if the interconnect intrinsic rise time were 0.5 nsec, the output rise time would be 1.1 nsec, and would start to have a significant impact.

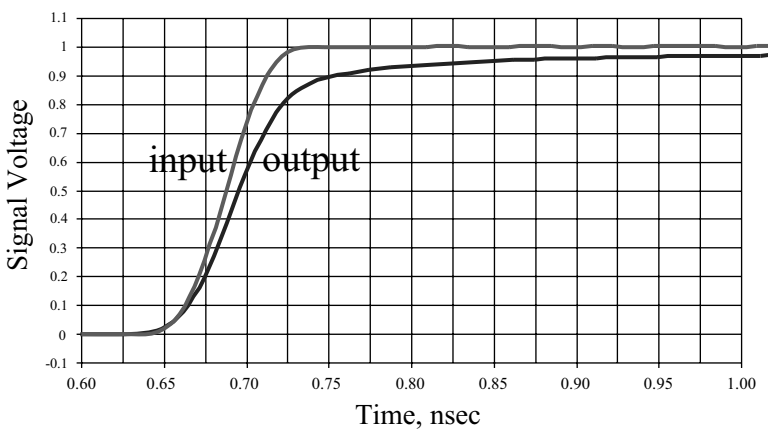


Figure 2-20 Measured input and transmitted signal through a 4-inch long, 50-Ohm transmission line in FR4 showing the rise-time degradation. The input rise time is 50 psec. The predicted output rise time is 67 psec based on the measured bandwidth of the interconnect. Measured with a GigaTest Labs Probe Station.

TIP As a simple rule of thumb, in order for the rise time of the signal to be increased by the interconnect less than 10%, the intrinsic rise time of the interconnect should be shorter than 50% of the rise time of the signal.

TIP In the frequency-domain perspective, to support the transmission of a 1-GHz bandwidth signal, we want the bandwidth of the interconnect to be at least twice as high, or 2 GHz.

It is important to keep in mind that this is a rule of thumb and it should not be used for design sign-off. It should be used only for a rough estimate or to identify a goal. If the bandwidth of an interconnect is within a factor of two of the bandwidth of the signal, it would probably be important to perform an analysis of how the interconnect affected the entire signal's spectrum.

2.17 The Bottom Line

1. The time domain is the real world and is typically where high-speed digital performance is measured.
2. The frequency domain is a mathematical construct where very specific, specialized rules apply.
3. The only reason to ever leave the time domain and use the frequency domain is to get to an answer faster.
4. The rise time of a digital signal is commonly measured from 10% of the final value to 90% of the final value.
5. Sine waves are the only waveform that can exist in the frequency domain.
6. The Fourier Transform converts a time-domain waveform into its spectrum of sine-wave-frequency components.
7. The spectrum of an ideal square wave has amplitudes that drop off at a rate of 1/f.
8. If the higher-frequency components are removed in the square wave, the rise time will increase.
9. The bandwidth of a signal is the highest sine-wave-frequency component that is significant, compared to the same harmonics in an ideal square wave with the same repeat frequency.

10. A good rule of thumb is that the bandwidth of a signal is 0.35/rise time of the signal.

11. Anything that decreases the bandwidth of a signal will increase its rise time.

12. The bandwidth of a measurement is the highest sine-wave frequency where the measurement has good accuracy.

13. The bandwidth of a model is the highest sine-wave frequency where the predictions of the model give good agreement with the actual performance of the interconnect.

14. The bandwidth of an interconnect is the highest sine-wave frequency where the performance of the interconnect still meets specifications.

15. The 3-dB bandwidth of an interconnect is the highest sine-wave frequency where the attenuation of a signal is less than –3 dB.

This page intentionally left blank

CHAPTER 3

Impedance and Electrical Models

In high-speed digital systems, where signal integrity plays a significant role, we often refer to signals as either changing voltages or changing currents. All the effects that we lump in the general category of signal integrity are due to how analog signals (those changing voltages and currents) interact with the electrical properties of the interconnects. The key electrical property with which signals interact is the impedance of the interconnects.

Impedance is defined as the ratio of the voltage to the current. We usually use the letter Z to represent impedance. The definition, which is *always* true, is Z = V/I. The manner in which these fundamental quantities, voltage and current, interact with the impedance of the interconnects determines all signal-integrity effects. As a signal propagates down an interconnect, it is constantly probing the impedance of the interconnect and reacting based on the answer.

TIP If we know the impedance of the interconnect, we can accurately predict how the signals will be distorted and whether a design will meet the performance specification, before we build it.

Likewise, if we have a target spec for performance and know what the signals will be, we can sometimes specify an impedance specification for the interconnects. If we understand how the geometry and material properties affect the

impedance of the interconnects, then we will be able to design the cross section, the topology, and the materials and select the other components so they will meet the impedance spec and result in a product that works the first time.

> **TIP** Impedance is the key term that describes every important electrical property of an interconnect. Knowing the impedance and propagation delay of an interconnect is to know almost everything about it electrically.

3.1 Describing Signal-Integrity Solutions in Terms of Impedance

Each of the four basic families of signal-integrity problems can be described based on impedance.

1. Signal-quality problems arise because voltage signals reflect and are distorted whenever the impedance the signal sees changes. If the impedance the signal sees is always constant, there will be no reflection and the signal will continue undistorted. Attenuation effects are due to series and shunt-resistive impedances.
2. Cross talk arises from the electric and magnetic fields coupling between two adjacent signal traces (and, of course, their return paths). The mutual capacitance and mutual inductance between the traces establishes an impedance, which determines the amount of coupled current and voltage.
3. Rail collapse of the voltage supply is really about the impedance in the power-distribution network (PDN). A certain amount of current must flow to feed all the ICs in the system. Because of the impedance of the power and ground distribution, a voltage drop will occur as the IC current switches. This voltage drop means the power and ground rails have collapsed from their nominal values.
4. The greatest source of EMI is from common currents, driven by voltages in the ground planes, through external cables. The higher the impedance of the return current paths in the ground planes, the greater the voltage drop, or ground bounce, which will drive the radiating currents. The most common fix for EMI from cables is the use of a ferrite choke around the cable. This works by increasing the impedance the common currents see, thereby reducing the amount of common current.

There are a number of design rules, or guidelines, that establish constraints on the physical features of the interconnects. For example, “keep the spacing between adjacent signal traces greater than 10 mils” is a design rule to minimize

cross talk. “Use power and ground planes on adjacent layers separated by less than 5 mils” is a design rule for the power and ground distribution.

> **TIP** Not only are the problems associated with signal integrity best described by the use of impedance, but the solutions and the design methodology for good signal integrity are also based on the use of impedance.

These rules establish a specific impedance for the physical interconnects. This impedance provides a specific environment for the signals, resulting in a desired performance. For example, keeping the power and ground planes closely spaced will result in a low impedance for the power distribution system and hence a lower voltage drop for a given power and ground current. This helps minimize rail collapse and EMI.

If we understand how the physical design of the interconnects affects their impedance, we will be able to interpret how they will interact with signals and what performance they might have.

> **TIP** Impedance is the Rosetta stone that links physical design and electrical performance. Our strategy is to translate system-performance needs into an impedance requirement and physical design into an impedance property.

Impedance is at the heart of the methodology we will use to solve signal-integrity problems. Once we have designed the physical system as we think it should be for optimal performance, we will translate the physical structure into its equivalent electrical circuit model. This process is called *modeling*.

It is the impedance of the resulting circuit model that will determine how the interconnects will affect the voltage and current signals. Once we have the circuit model, we will use a circuit simulator, such as SPICE, to predict the new waveforms as the voltage sources are affected by the impedances of the interconnects. Alternatively, behavioral models of the drivers or interconnects can be used where the interaction of the signals with the impedance, described by the behavioral model, will predict performance. This process is called *simulation*.

Finally, the predicted waveforms will be analyzed to determine if they meet the timing and distortion or noise specs, and are acceptable, or if the physical design has to be modified. This process flow for a new design is illustrated in Figure 3-1.

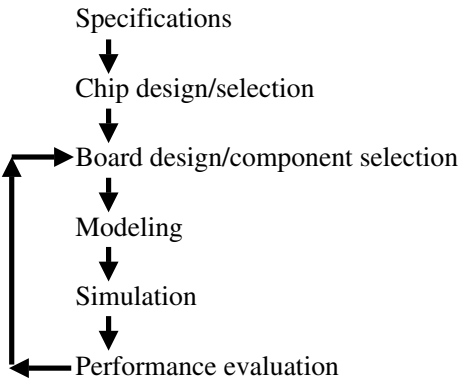


Figure 3-1 Process flow for hardware design. The modeling, simulation, and evaluation steps should be implemented as early and often in the design cycle as possible.

The two key processes, modeling and simulation, are based on converting electrical properties into an impedance, and analyzing the impact of the impedance on the signals.

If we understand the impedance of each of the circuit elements used in a schematic and how the impedance is calculated for a combination of circuit elements, the electrical behavior of *any* model and *any* interconnect can be evaluated. This concept of impedance is absolutely critical in all aspects of signal-integrity analysis.

3.2 What Is Impedance?

We use the term *impedance* in everyday language and often confuse the electrical definition with the common usage definition. As we saw earlier, the electrical term *impedance* has a very precise definition based on the relationship between the current through a device and the voltage across it: Z = V/I. This basic definition applies to any two-terminal device, such as a surface mount resistor, a decoupling capacitor, a lead in a package, or the front connections to a printed circuit-board trace and its return path. When there are more than two terminals, such as in coupled conductors, or between the front and back ends of a transmission line, the definition of impedance is the same, it's just more complex to take into account the additional terminals.

For two-terminal devices, the definition of impedance, as illustrated in Figure 3-2, is simply:

$$Z = \frac{V}{I} \tag{3-1}$$

where:

Z = the impedance, measured in Ohms

V = the voltage across the device, in units of volts

I = the current through the device, in units of Amps

For example, if the voltage across a terminating resistor is 5 v and the current through it is 0.1 A, then the impedance of the device must be 5 v/0.1 A = 50 Ohms. No matter what type of device the impedance is referring to, in both the time and the frequency domain, the units of impedance are always in Ohms.

TIP This definition of impedance applies to absolutely all situations, whether in the time domain or the frequency domain, whether for real devices that are measured or for ideal devices that are calculated.

If we always go back to this basic definition, we will never go wrong and often eliminate many sources of confusion. One aspect of impedance that is often confusing is to think of it only in terms of resistance. As we will see, the impedance of an ideal resistor-circuit element, with resistance, R, is, in fact, Z = R.

Our intuition of the impedance of a resistor is that a high impedance means less current flow for a fixed voltage. Likewise, a low impedance means a lot of current can flow for the same voltage. This is consistent with the definition that I = V/Z, and applies just as well when the voltage or current is not DC.

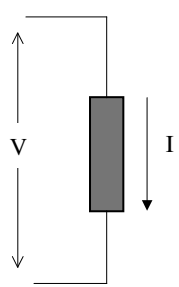


Figure 3-2 The definition of impedance for any two-terminal device showing the current through the component and the voltage across the leads.

In addition to the notion of the impedance of a resistor, the concept of impedance can apply to an ideal capacitor, an ideal inductor, a real-wire bond, a printed circuit trace, or even a pair of connector pins.

There are two special extreme cases of impedance. For a device that is an open, there will be no current flow. If the current through the device for any voltage applied is zero, the impedance is Z = 1 v / 0 A = infinite Ohms. The impedance of an open device is very, very large. When the device is a short, there will be no voltage across it, no matter what the current through it. The impedance of a short is Z = 0 v / 1 A = 0 Ohms. The impedance of a short is always 0 Ohms.

3.3 Real Versus Ideal Circuit Elements

There are two types of electrical devices, real and ideal. Real devices can be measured. They are the only things that physically exist. They are the actual interconnects or components that make up the hardware of a real system. Real devices are traces on a board, leads in a package, or discrete decoupling capacitors mounted to a board.

Ideal devices are mathematical descriptions of specialized circuit elements that have precise, specific definitions. Simulators can only simulate the performance of ideal devices. The formalism and power of circuit theory apply only to ideal devices. Models are composed of combinations of ideal devices.

It is very important to keep separate real versus ideal circuit elements. The impedance of any real, physical interconnect or passive component can be measured. However, when impedance is calculated, it is only the impedance of four very-well-defined, ideal circuit elements that can be considered. We cannot measure ideal circuit elements, nor can we calculate the impedance of any circuit elements other than ideal ones. This is why it is important to make the distinction between real components and ideal circuit elements. This distinction is illustrated in Figure 3-3.

TIP Ultimately, our goal is to create an equivalent circuit model composed of combinations of ideal circuit elements whose impedance closely approximates the actual, measured impedance of a real component.

A circuit model will always be an approximation of the real-world structure. However, it is possible to construct an ideal model with a simulated impedance that accurately matches the measured impedance of a real device. For example, Figure 3-4 shows the measured impedance of a real decoupling capacitor and the

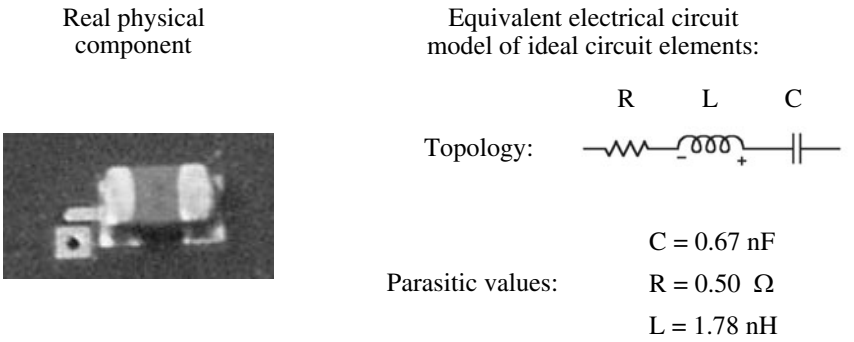


Figure 3-3 The two worldviews of a component, in this case a 1206 decoupling capacitor mounted to a circuit board and an equivalent circuit model composed of combinations of ideal circuit elements.

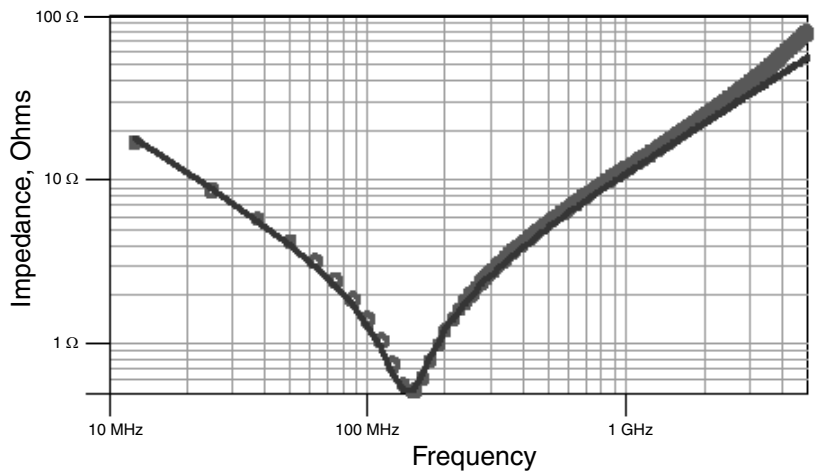


Figure 3-4 The measured impedance, as circles, and the simulated impedance, as the line, for a nominal 1-nF decoupling capacitor. The measurement was performed with a network analyzer and a GigaTest Labs Probe Station.

simulated impedance based on an RLC circuit model. These are the component and model in Figure 3-3. The agreement is excellent even up to 5 GHz, the bandwidth of the measurement.

There are four ideal, two-terminal, circuit elements that we will use in combination as building blocks to describe any real interconnect:

1. An ideal resistor
2. An ideal capacitor
3. An ideal inductor
4. An ideal transmission line

We usually group the first three elements in a category called *lumped circuit elements*, in the sense that their properties can be lumped into a single point. This is different from the properties of an ideal transmission line, which are "distributed" along its length.

As ideal circuit elements, these elements have precise definitions that describe how they interact with currents and voltages. It is very important to keep in mind that ideal elements are different from real components, such as real resistors, real capacitors, or real inductors. One is a physical component, the other an ideal element.

The properties of a transmission line are initially seen as so confusing and nonintuitive, yet so important, that we devote an entire chapter to transmission lines and their impedance. In this chapter, we will concentrate on the impedance of just R, L, and C elements.

TIP Only real devices can be measured, and only ideal elements can be calculated or simulated.

An equivalent electrical circuit model is an idealized electrical description of a real structure. It is an approximation, based on using combinations of ideal circuit elements. A good model will have a calculated impedance that closely matches the measured impedance of the real device. The better we can model the impedance of an interconnect, the better we can predict how a signal will interact with it.

When dealing with some high-frequency effects, such as lossy lines, we will need to invent new ideal circuit elements to create better models.

3.4 Impedance of an Ideal Resistor in the Time Domain

Each of the four basic circuit elements above has a definition of how voltage and current interact with it. This is different from the impedance of the ideal circuit element.

The relationship between the voltage across and the current through an ideal resistor is:

$$V = I \times R \tag{3-2}$$

where:

V = the voltage across the ends of the resistor

I = the current through the resistor

R = the resistance of the resistor, in Ohms

An ideal resistor has a voltage across it that increases with the current through it. This definition of the I-V properties of an ideal resistor applies in both the time domain and the frequency domain.

In the time domain, we can apply the definition of the impedance and, using the definition of the ideal element, calculate the impedance of an ideal resistor:

$$Z = \frac{V}{I} = \frac{I \times R}{I} = R \tag{3-3}$$

This basically says, the impedance is constant and independent of the current or voltage across a resistor. The impedance of a resistor is pretty boring.

3.5 Impedance of an Ideal Capacitor in the Time Domain

In an ideal capacitor, there is a relationship between the charge stored between the two leads and the voltage across the leads. The capacitance of an ideal capacitor is defined as:

$$C = \frac{Q}{V} \tag{3-4}$$

where:

C = the capacitance, in Farads

V = the voltage across the leads, in volts

Q = the charge stored between the leads, in Coulombs

The value of the capacitance of a capacitor describes its capacity to store charge at the expense of voltage. A large capacitance means the ability to store a lot of charge at a low voltage across the terminals.

The impedance of a capacitor can only be calculated based on the current through it and the voltage across its terminals. In order to relate the voltage across the terminals with the current through it, we need to know how the current flows through a capacitor. A real capacitor is made from two conductors separated by a dielectric. How does current get from one conductor to the other, when it has an insulating dielectric between them? This is a fundamental question and will pop up over and over in signal integrity applications.

The answer is that real current probably doesn't really flow through a capacitor, it just acts as though it does when the voltage across the capacitor changes. Suppose the voltage across a capacitor were to increase. This means that some positive charge had to be added to the top conductor and some negative charge had to be added to the bottom conductor. Adding negative charge to the bottom conductor is the same as pushing positive charge out; it is as though positive charges were added to the top terminal and positive charges were pushed out of the bottom terminal. This is illustrated in Figure 3-5. The capacitor behaves as though current flows through it, but only when the voltage across it changes.

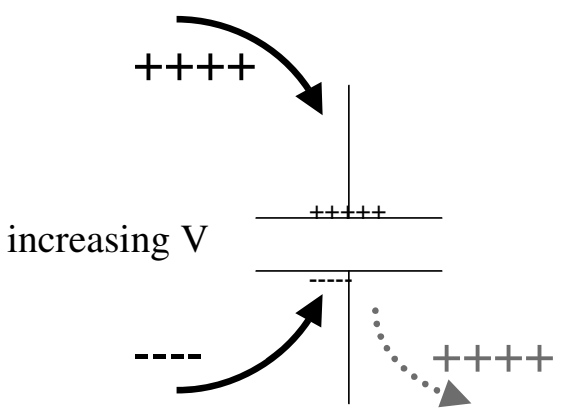


Figure 3-5 Increasing the voltage across a capacitor adds positive charge to one conductor and negative charge to the other. Adding negative charge to one conductor is the same as taking positive charges from it. It looks like positive charge enters one terminal and comes out of the other.

By taking derivatives of both sides of the previous equation, a new definition of the I-V behavior of a capacitor can be developed:

$$I = \frac{dQ}{dt} = C\frac{dV}{dt} \tag{3-5}$$

where:

I = the current through the capacitor

Q = the charge on one conductor of the capacitor

C = the capacitance of the capacitor

V = the voltage across the capacitor

This relationship points out, as we saw previously, that the only way current flows through a capacitor is when the voltage across it changes. If the voltage is constant, the current through a capacitor is zero. We also saw that for a resistor, the current through it doubled if the current doubled. However, in the case of a capacitor, the current through it doubles if the rate of change of the voltage across it doubles.

This definition is consistent with our intuition. If the voltage changes rapidly, the current through a capacitor is large. If the voltage is nearly constant, the current through a capacitor is near zero. Using this relationship, we can calculate the impedance of an ideal capacitor in the time domain:

$$Z = \frac{V}{I} = \frac{V}{C\frac{dV}{dt}} \tag{3-6}$$

where:

V = the voltage across the capacitor

C = the capacitance of the capacitor

I = the current through the capacitor

This is a complicated expression. It says that the impedance of a capacitor depends on the precise shape of the voltage waveform across it. If the slope of the waveform is large (i.e., if the voltage changes very fast), the current through it is high and the impedance is small. It also says that a large capacitor will have

a lower impedance than a small capacitor for the same rate of change of the voltage signal.

However, the precise value of the impedance of a capacitor is more complicated. It is hard to generalize what the impedance of a capacitor is other than it depends on the shape of the voltage waveform. The impedance of a capacitor is not an easy term to use in the time domain.

3.6 Impedance of an Ideal Inductor in the Time Domain

The behavior of an ideal inductor is defined by:

$$V = L\frac{dI}{dt} \tag{3-7}$$

where:

V = the voltage across the inductor

L = the inductance of the inductor

I = the current through the inductor

This says that the voltage across an inductor depends on how fast the current through it changes. If the current is constant, the voltage across the inductor will be zero. Likewise, if the current changes rapidly through an inductor, there will be a large voltage drop across it. The inductance is the proportionality constant that says how sensitive the voltage generated is to a changing current. A large inductance means that a small changing current produces a large voltage.

There is often confusion about the direction of the voltage drop that is generated across an inductor. If the direction of the changing current reverses, the polarity of the induced voltage will reverse. An easy way of remembering the polarity of the voltage is to base it on the voltage drop of a resistor.

If a DC current goes through a resistor, the terminal the current goes into is the positive side and the other terminal is the negative side. Likewise, with an inductor, the terminal the current is increasing into is the positive side and the other is the negative side for the induced voltage. This is illustrated in Figure 3-6.

Using this basic definition of inductance, we can calculate the impedance of an inductor. This is, by definition, the ratio of the voltage to the current through an inductor:

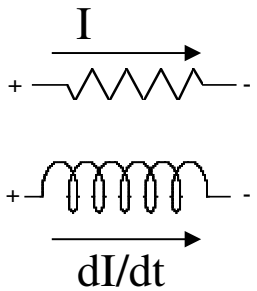


Figure 3-6 The direction of voltage drop across an inductor for a changing current is in the same direction as the voltage drop across a resistor for a DC current.

$$Z = \frac{V}{I} = L\frac{\frac{dI}{dt}}{I} \tag{3-8}$$

where:

V = the voltage across the inductor

L = the inductance of the inductor

I = the current through the inductor

Again, we see the impedance of an inductor, though well defined, is awkward to use in the time domain. The general features are easy to discern. If the current through an inductor increases rapidly, the impedance of the inductor is large. An inductor will have a high impedance when current through it tries to change. If the current through an inductor changes only slightly, its impedance will be very small. For DC current the impedance of an inductor is nearly zero. But, other than these simple generalities, the actual impedance of an inductor depends very strongly on the precise waveform of the current through it.

TIP For both the capacitor and the inductor, the impedance, in the time domain, is not a simple function at all. Impedance in the time domain is a very complicated way of describing these basic building-block ideal circuit elements. It is not wrong; it is just complicated.

This is one of the important occasions where moving to the frequency domain will make the analysis of a problem much simpler.

3.7 Impedance in the Frequency Domain

The important feature of the frequency domain is that the only waveforms that can exist are sine waves. We can only describe the behavior of ideal circuit elements in the frequency domain by how they interact with sine waves: sine waves of current and sine waves of voltage. These sine waves have three and only three features: the frequency, the amplitude, and the phase associated with each wave.

Rather than describe the phase in cycles or degrees, it is more common to use radians. There are $2 \times \pi$ radians in one cycle, so a radian is about 57 degrees. The frequency in radians per second is referred to as the *angular frequency*. The Greek letter omega (ω) is used to denote the angular frequency. ω is related to the frequency, by:

$$\omega = 2\pi \times f \tag{3-9}$$

where:

ω = the angular frequency, in radians/sec

f = the sine-wave frequency, in Hertz

We can apply sine-wave voltages across a circuit element and look at the sine waves of current through it. When we do this, we will still use the same basic definition of impedance (that is, the ratio of the voltage to the current) except that we will be taking the ratio of two sine waves, a voltage sine wave and a current sine wave.

It is important to keep in mind that all the basic building-block circuit elements and all the interconnects are linear devices. If a voltage sine wave of 1 MHz, for example, is applied across any of the four ideal circuit elements, the only sine-wave-frequency components that will be present in the current waveform will be a sine wave at 1 MHz. The amplitude of the current sine wave will be some number of Amps and it will have some phase shift with respect to the voltage wave, but it will have exactly the same frequency. This is illustrated in Figure 3-7.

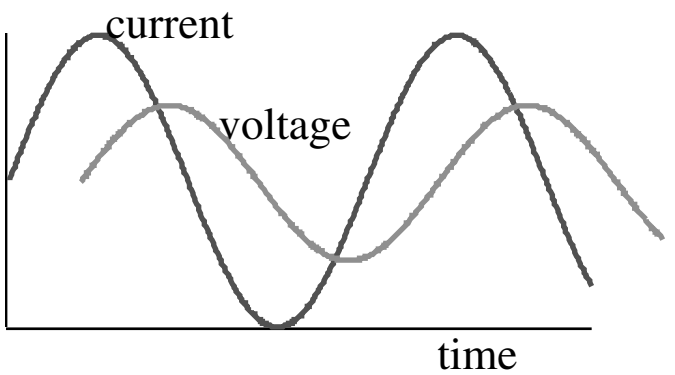


Figure 3-7 The sine-wave current through and voltage across an ideal circuit element will have exactly the same frequency but different amplitudes and some phase shift.

TIP When we take the ratio of two sine waves, we need to account for the ratio of the amplitudes and the phase shift between the two waves.

What does it mean to take the ratio of two sine waves, the voltage and the current? The ratio of two sine waves is not a sine wave. It is a pair of numbers that contains information about the ratio of the amplitudes and the phase shift, at each frequency value. The magnitude of the ratio is just the ratio of the amplitudes of the two sine waves:

$$|Z| = \frac{|V|}{|I|} \tag{3-10}$$

The ratio of the voltage amplitude to the current amplitude will have units of Ohms. We refer to this ratio as the magnitude of the impedance. The phase of the ratio is the phase shift between the two waves. The phase shift has units of degrees or radians. In the frequency domain, the impedance of a circuit element or combination of circuit elements would be of the form: at 20 MHz, the magnitude of the impedance is 15 Ohms and the phase of the impedance is 25 degrees. This means the impedance is 15 Ohms and the voltage wave is leading the current wave by 25 degrees.

The impedance of any circuit element is two numbers, a magnitude and a phase, at every frequency value. Both the magnitude of the impedance and the phase of the impedance may be frequency dependent. The ratio of the amplitudes may vary with frequency or the phase may vary with frequency. When we describe the impedance, we need to specify at what frequency we are describing the impedance.

In the frequency domain, impedance can also be described with complex numbers. For example, the impedance of a circuit can be described as having a real component and an imaginary component. The use of real and imaginary components allows the powerful formalism of complex numbers to be applied, which dramatically simplifies the calculations of impedance in large circuits. Exactly the same information is contained in the magnitude and the phase information. These are two different and equivalent ways of describing impedance.

With this new idea of working in the frequency domain, and dealing only with sine waves of current and voltage, we can take another look at impedance.

We apply a sine wave of current through a resistor and we get a sine wave of voltage across it that is simply R times the current wave:

$$V = I_0 \sin(\omega t) \times R \tag{3-11}$$

We can describe the sine wave of current in terms of sine and cosine waves or in terms of complex exponential notation.

When we take the ratio of the voltage to the current for a resistor, we find that it is simply the value of the resistance:

$$Z = \frac{V}{I} = \frac{I_0 \sin(\omega t) \times R}{I_0 \sin(\omega t)} = R \tag{3-12}$$

The impedance is independent of frequency and the phase shift is zero. The impedance of an ideal resistor is flat with frequency. This is basically the same result we saw in the time domain, still pretty boring.

When we look at an ideal capacitor in the frequency domain, we will apply a sine-wave voltage across the ends. The current through the capacitor is the derivative of the voltage, which is a cosine wave:

$$I = C \times \frac{d}{dt} V_0 \sin(\omega t) = C \times \omega V_0 \cos(\omega t) \tag{3-13}$$

This says the current amplitude will increase with frequency, even if the voltage amplitude stays constant. The higher the frequency, the larger the amplitude of the current through the capacitor. This suggests the impedance of a capacitor will decrease with increasing frequency. The impedance of a capacitor is calculated from:

$$Z = \frac{V}{I} = \frac{V_0 \sin(\omega t)}{C \times \omega V_0 \cos(\omega t)} = \frac{1}{\omega C} \times \frac{\sin(\omega t)}{\cos(\omega t)} \tag{3-14}$$

Here is where it gets confusing. This ratio is easily described using complex math, but most of the insight can also be gained from sine and cosine waves. The magnitude of the impedance of a capacitor is just $1/\omega C$. All the important information is here. As the angular frequency increases, the impedance of a capacitor decreases. This says that even though the value of the capacitance is constant with frequency, the impedance gets smaller with higher frequency. We see this is reasonable because the current through the capacitor will increase with higher frequency and hence its impedance will be less.

The phase of the impedance is the phase shift between a sine and cosine wave, which is –90 degrees. When described in complex notation, the –90 degree phase shift is represented by the complex number, –i. In complex notation, the impedance of a capacitor is $-i/\omega C$. For most of the following discussion, the phase adds more confusion than value and will generally be ignored.

A real decoupling capacitor has a capacitance of 10 nF. What is its impedance at 1 GHz? First, we assume this capacitor is an ideal capacitor. A 10-nF ideal capacitor will have an impedance of $1/(2\pi \times 1\ \text{GHz} \times 10\ \text{nF}) = 1/(6 \times 10^9 \times 10 \times 10^{-9}) = 1/60 \sim 0.016$ Ohm. This is a very small impedance. If the real decoupling capacitor behaved like an ideal capacitor, its impedance would be about 10 milliOhms at 1 GHz. Of course, at lower frequency, its impedance would be higher. At 1 Hz, its impedance would be about 16 MegaOhms.

Let's use this same frequency-domain analysis with an inductor. When we apply a sine wave current through an inductor, the voltage generated is:

$$V = L \times \frac{d}{dt} I_0 \sin(\omega t) = L \times \omega I_0 \cos(\omega t) \tag{3-15}$$

This says that for a fixed current amplitude, the voltage across an inductor gets larger at higher frequency. It takes a higher voltage to push the same current amplitude through an inductor. This would hint that the impedance of an inductor increases with frequency.

Using the basic definition of impedance, the impedance of an inductor in the frequency domain can be derived as:

$$Z = \frac{V}{I} = \frac{L \times \omega I_0 \cos(\omega t)}{I_0 \sin(\omega t)} = \omega L \times \frac{\cos(\omega t)}{\sin(\omega t)} \tag{3-16}$$

The magnitude of the impedance increases with frequency, even though the value of the inductance is constant with frequency. It is a natural consequence of the behavior of an inductor that it is harder to shove AC current through it with increasing frequency.

The phase of the impedance of an inductor is the phase shift between the voltage and the current, which is +90 degrees. In complex notation, a +90 degree phase shift is i. The complex impedance of an inductor is $Z = i\omega L$.

In a real decoupling capacitor, there is inductance associated with the intrinsic shape of the capacitor and its board-attach footprint. A rough estimate for this intrinsic inductance is 2 nH. We really have to work hard to get it any lower than this. What is the impedance of just the series inductance of the real capacitor that we will model as an ideal inductor of 2 nH, at a frequency of 1 GHz?

The impedance is $Z = 2 \times \pi \times 1\ \text{GHz} \times 2\ \text{nH} = 12$ Ohms. When it is in series with the power and ground distribution and we want a low impedance, for example less than 0.1 Ohm, 12 Ohms is a lot. How does this compare with the impedance of the ideal-capacitor component of the real decoupling capacitor? In the last problem, the impedance of the ideal capacitor element at 1 GHz was 0.01 Ohm. The impedance of the ideal inductor component is more than 1000 times higher than this and will clearly dominate the high-frequency behavior of a real capacitor.

We see that for both the ideal capacitor and inductor, the impedance in the frequency domain has a very simple form and is easily described. This is one of the powers of the frequency domain and why we will often turn to it to help solve problems.

The value of the resistance, capacitance, and inductance of ideal resistors, capacitors, and inductors are all constant with frequency. For the case of an ideal resistor, the impedance is also constant with frequency. However, for a capacitor, its impedance will decrease with frequency, and for an inductor, its impedance will increase with frequency.

TIP It is important to keep straight that for an ideal capacitor or inductor, even though its value of capacitance and inductance is absolutely constant with frequency, its impedance will vary with frequency.

3.8 Equivalent Electrical Circuit Models

The impedance behavior of real interconnects can be closely approximated by combinations of these ideal elements. A combination of ideal circuit elements is called an *equivalent electrical circuit model*, or typically, just a *model*. The drawing of the circuit model is often referred to as a *schematic*.

An equivalent circuit model has two features: It identifies how the circuit elements are connected together (called the *circuit topology*), and it identifies the value of each circuit element (referred to as the *parameter values* or *parasitic values*).

Chip designers, who like to think they produce drivers with perfect, pristine waveforms, view all interconnects as parasitics in that they can only screw up their wonderful waveforms. To the chip designer, the process of determining the parameter values of the interconnects is really parasitic extraction, and the term has stuck in general use.

TIP It is important to keep in mind that whenever we draw circuit elements, they are always ideal circuit elements. We will have to use combinations of ideal elements to approximate the actual performance of real interconnects.

There will always be a limit to how well we can predict the actual impedance behavior of real interconnects, using an ideal equivalent circuit model. This limit can often be found only by measuring the actual impedance of an interconnect and comparing it to the predictions based on the simulations of circuits containing these ideal circuit elements.

There are always two important questions to ask of every model: How good is it and what is its bandwidth? Remember, its bandwidth is the highest sine-wave frequency at which we get good agreement between the measured impedance and

the predicted impedance. As a general rule, the closer we would like the predictions of a circuit model to be to the actual measured performance, the more complex the model may have to be.

TIP It is good practice to always start the process of modeling with the simplest model possible and grow in complexity from there.

Take, for example, a real decoupling capacitor and its impedance as measured from one of the capacitor pads, through a via and a plane below it, coming back up to the start of the capacitor. This is the example shown previously in Figure 3-3. We might expect that this real device could be modeled as a simple ideal capacitor. But, at how high a frequency will the real capacitor still behave like an ideal capacitor? The measured impedance of this real device, from 10 MHz to 5 GHz, is shown in Figure 3-8, with the impedance predicted for an ideal capacitor superimposed.

It is clear that this simple model works really well at low frequency. This simple model of an ideal capacitor with a value of 0.67 nF is a very good model. It's just that it gives good agreement only up to about 70 MHz. Its bandwidth is 70 MHz.

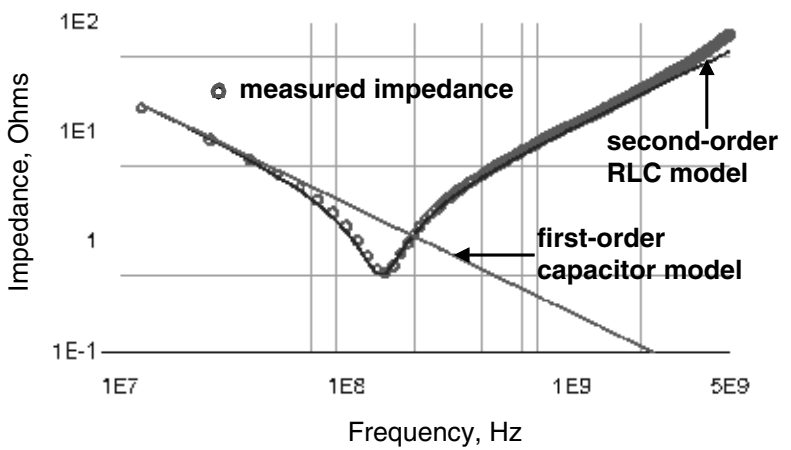


Figure 3-8 Comparison of the measured impedance of a real decoupling capacitor and the predicted impedance of a simple first-order model using a single C element and a second-order model using an RLC circuit model. Measured with a GigaTest Labs Probe Station.

If we expend a little more effort, we can create a more accurate circuit model with a higher bandwidth. A more accurate model for a real decoupling capacitor is an ideal capacitor, inductor, and resistor in series. Choosing the best parameter values, we see in Figure 3-8 that the agreement between the predicted impedance of this model and the measured impedance of the real device is excellent, all the way up to the bandwidth of the measurement, 5 GHz in this case.

We often refer to the simplest model we create as a first-order model, as it is the first starting place. As we increase the complexity, and hopefully, better agreement with the real device, we refer to each successive model as the second-order model, third-order model, and so on.

Using the second-order model for a real capacitor would let us accurately predict every important electrical feature of this real capacitor as it would behave in a system with application bandwidths at least up to 5 GHz.

TIP It is remarkable that the relatively complex behavior of real components can be very accurately approximated, to very high bandwidths, by combinations of ideal circuit elements.

3.9 Circuit Theory and SPICE

There is a well-defined and relatively straightforward formalism to describe the impedance of combinations of ideal circuit elements. This is usually referred to as *circuit theory.* The important rule in circuit theory is that when two or more elements are in series, that is, connected end-to-end, the impedance of the combination, from one end terminal to the other end terminal, is the sum of the impedances of each element. What makes it a little complicated is that when in the frequency domain, the impedances that are summed are complex and must obey complex algebra.

In the previous section, we saw that it is possible to calculate the impedance of each individual circuit element by hand. When there are combinations of circuit elements it gets more complicated. For example, the impedance of an RLC model approximating a real capacitor is given by:

$$Z(\omega) = R + i\left(\omega L - \frac{1}{\omega C}\right) \tag{3-17}$$

We could use this analytic expression for the impedance of the RLC circuit to plot the impedance versus frequency for any chosen values of R, L, and C. It

can conveniently be used in a spreadsheet and each element changed. When there are five or ten elements in the circuit model, the resulting impedance can be calculated by hand, but it can be very complicated and tedious.

However, there is a commonly available tool that is much more versatile in calculating and plotting the impedance of any arbitrary circuit. It is so common and so easy to use, every engineer who cares about impedance or circuits in general, should have access to it on their desktop. It is SPICE.

SPICE stands for *Simulation Program with Integrated Circuit Emphasis*. It was developed in the early 1970s at UC Berkeley as a tool to predict the behavior of transistors based on the as-fabricated dimensions. It is basically a circuit simulator. Any circuit we can draw with R, L, C, and T elements can be simulated for a variety of voltage or current-exciting waveforms. It has evolved and diversified over the past 30 years, with over 30 vendors each adding their own special features and capabilities. There are a few either free versions or student versions for less than $100 that can be downloaded from the Web. Some of the free versions have limited capability but are excellent tools to learn about circuits.

In SPICE, only ideal circuit elements are used and every circuit element has a well-defined, precise behavior. There are two basic types of elements: active and passive. The active elements are the signal sources, current or voltage waveforms, or actual transistor or gate models. The passive elements are all the ideal circuit elements described above. One of the distinctions between the various forms of SPICE is the variety of ideal circuit elements they provide. Every version of SPICE includes at least the R, L, C, and T (transmission-line) elements.

SPICE simulators allow the prediction of the voltage or current at every point in a circuit, simulated either in the time-domain or the frequency domain. A time-domain simulation is called a *transient simulation* and a frequency-domain simulation is called an *AC simulation*. SPICE is an incredibly powerful tool.

For example, a driver connected to two receivers located very close together can be modeled with a simple voltage source and an RLC circuit. The R is the impedance of the driver, typically about 10 Ohms. The C is the capacitance of the interconnect traces and the input capacitance of the two receivers, typically about 5 pF total. The L is the total loop inductance of the package leads and the interconnect traces, typically about 7 nH. The setup of this circuit in SPICE and the resulting time-domain waveform, showing the ringing that might be found in the actual circuit, is shown in Figure 3-9.

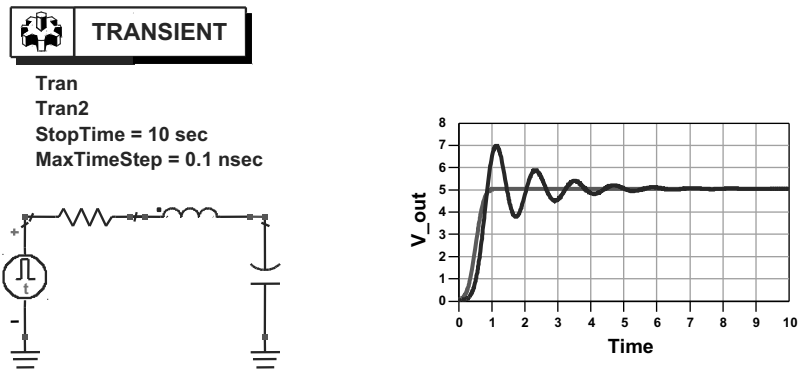


Figure 3-9 Simple equivalent circuit model to represent a driver and receiver fanout of two, including the packaging and interconnects, as set up in Agilent's Advanced Design System (ADS), a version of SPICE, and the resulting simulation of the internal-voltage waveform and the voltage at the input of the receivers. The rise time simulated is 0.5 nsec. The lead and interconnect inductance plus the input-gate capacitance dominate the source of the ringing.

TIP If the circuit schematic can be drawn, SPICE can simulate the voltage and current waveforms. This is the real power of SPICE for general electrical engineering analysis.

SPICE can be used to calculate and plot the impedance of any circuit in the frequency domain. Normally, it plots only the voltage or current waveforms at every connection point, but a trick can be used to convert this into impedance.

One of the circuit elements SPICE has in its toolbox for AC simulation is a constant-current sine-wave-current source. This current source will output a sine wave of current, with a constant amplitude, at a predetermined frequency. When running an AC analysis, the SPICE engine will step the frequency of the sine-wave-current source from the start frequency value to the stop frequency value with a number of intermediate frequency points.

It generates the constant-current amplitude by outputting a sine wave with some voltage amplitude. The amplitude of the voltage wave is automatically adjusted to result in the specified constant amplitude of current.

To build an impedance analyzer in SPICE, we set the current source to have a constant amplitude of 1 Amp. No matter what circuit elements are connected to the current source, SPICE will adjust the voltage amplitude to result in 1-Amp

current amplitude through the circuit. If the constant-current source is connected to a circuit that has some impedance associated with it, $Z(\omega)$, then to keep the amplitude of the current constant, the voltage it applies will have to adjust. The voltage applied to the circuit, from the constant-current source, with a 1-Amp current amplitude, is $V(\omega) = Z(\omega) \times 1$ Amp. The voltage across the current source, in volts, is numerically equal to the impedance of the circuit attached, in Ohms.

For example, if we attach a 1-Ohm resistor across the terminals, in order to maintain the constant current of 1 Amp, the voltage amplitude generated must be $V = 1$ Ohm $\times$ 1A = 1 v. If we attach a capacitor with capacitance C, the voltage amplitude at any frequency will be $V = 1/\omega C$. Effectively, this circuit will emulate an impedance analyzer. Plotting the voltage versus the frequency is a measure of the magnitude of the impedance versus frequency for any circuit. The phase of the voltage is also a measure of the phase of the impedance.

To use SPICE to plot an impedance profile, we construct an AC constant-current source with amplitude of 1 A and connect the circuit under test across the terminals. The voltage measured across the current source is a direct measure of the impedance of the circuit. An example of a simple circuit is shown in Figure 3-10. As a trivial example, we connect a few different circuit elements to the impedance analyzer and plot their impedance profiles.

We can use this impedance analyzer to plot the impedance of any circuit model. Impedance is complex. It has not only magnitude information but also phase information. We can plot each of these separately in SPICE. The phase is

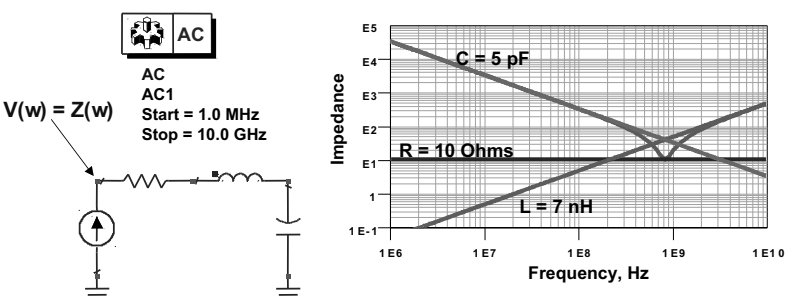


Figure 3-10 Left: An impedance analyzer in SPICE. The voltage across the constant-current source is a direct measure of the impedance of the circuit connected to it. Right: An example of the magnitude of the impedance of various circuit elements, calculated with the impedance analyzer in SPICE.

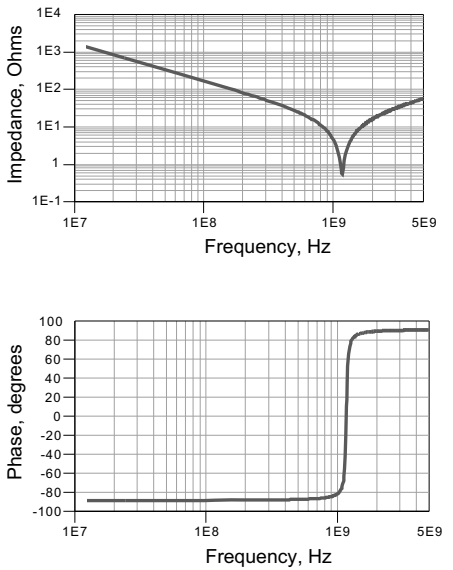


Figure 3-11 Simulated magnitude and phase of an ideal RLC circuit. The phase shows the capacitive behavior at low frequency and the inductive behavior at high frequency.

also available in an AC simulation in SPICE. In Figure 3-11, we illustrate using the impedance analyzer to simulate the impedance of an RLC circuit model, approximating a real capacitor, plotting the magnitude and phase of the impedance across a wide frequency range.

It is exactly as expected. At low frequency, the phase of the impedance is –90 degrees, suggesting capacitive behavior. At high frequency the phase of the impedance is +90 degrees, suggesting inductive behavior.

3.10 Introduction to Modeling

As pointed out in Chapter 1, equivalent circuit models for interconnects and passive devices can be created based either on measurements or on calculations. In either case, the starting place is always some assumed topology for the circuit model. How do we pick the right topology? How do we know what is the best circuit schematic with which to start?

The strategy for building models of interconnects or other structures is to follow the principle that Albert Einstein articulated when he said, “Everything should be made as simple as possible, but not simpler.” Always start with the simplest model first, and build in complexity from there.

Building models is a constant balancing act between the accuracy and bandwidth of the model required and the amount of time and effort we are willing to expend in getting the result. In general, the more accuracy required, the more expensive the cost in time, effort, and dollars. This is illustrated in Figure 3-12.

TIP When constructing models for interconnects, it is always important to keep in mind that sometimes an OK answer, NOW, is better than a more accurate answer late. This is why Einstein's advice should be followed: Start with the simplest model first and build in complexity from there.

It is always a good idea to start with an ideal transmission line as the first order model for any interconnect. An ideal transmission line is a good low frequency and high frequency model for an interconnect, which is why we devote an entire chapter to this ideal circuit element.

However, sometimes the question you want answered is phrased in terms of lumped circuit elements, such as what is the inductance of a package lead or what is the capacitance of an interconnect trace. When the interconnect structure is electrically short, or at low frequency, interconnects can be approximated by ideal R, L, or C elements. This concept of electrical length is described in a later chapter.

The simplest lumped circuit model is just a single R, L, or C circuit element. The next simplest are combinations of two of them, and then three of them, and so on. The key factor that determines when we need to increase the complexity of a

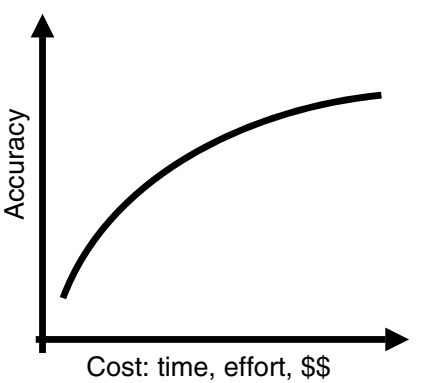


Figure 3-12 Fundamental trade-off between the accuracy of a model and how much effort is required to achieve it. This is a fundamental relationship for most issues in general.

model is the bandwidth of the model required. As a general trend, the higher the bandwidth, the more complex the model. However, every high-bandwidth model must still give good agreement at low frequency; otherwise, it will not be accurate for transient simulations that can have low-frequency components in the signals.

For discrete passive devices, such as surface-mount technology (SMT) terminating resistors, decoupling capacitors, and filter inductors, the low-bandwidth and high-bandwidth ideal circuit model topologies are illustrated in Figure 3-13. As we saw earlier for the case of decoupling capacitors, the single-element circuit model worked very well at low frequency. The higher-bandwidth model for a real decoupling capacitor worked even up to 5 GHz for the specific component measured. The bandwidth of a circuit model for a real component is not easy to estimate except from a measurement.

For many interconnects that are electrically short, simple circuit models can also be used. The simplest starting place for a printed-circuit trace over a return plane in the board, which might be used to connect one driver to another, is a single capacitor. Figure 3-14 is an example of the measured impedance of a 1-inch interconnect and the simulated impedance of a first-order model consisting of a

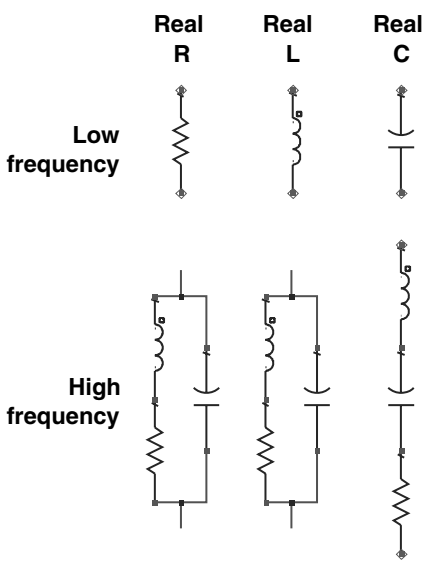


Figure 3-13 Simplest starting models for real components or interconnect elements, at low frequency and for higher bandwidth.

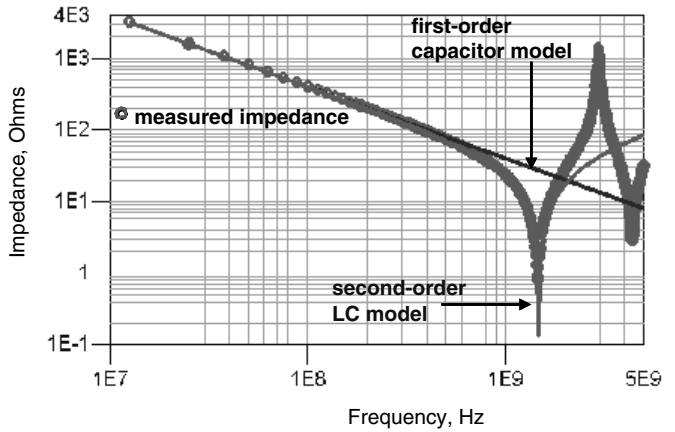


Figure 3-14 Measured impedance of a 1-inch-long microstrip trace and the simulated impedance of a first- and second-order model. The first-order model is a single C element and has a bandwidth of about 1 GHz. The second-order model uses a series LC circuit and has a bandwidth of about 2 GHz.

single C element model. In this case, the agreement is excellent up to about 1 GHz. If the application bandwidth was less than 1 GHz, a simple ideal capacitor could be used to accurately model this 1-inch-long interconnect.

For a higher bandwidth model, a second-order model consisting of an inductor in series with the capacitor can be used. The agreement of this higher-bandwidth model is up to about 2 GHz.

As we show in a later chapter, the best model for any uniform interconnect is an ideal transmission line model. This T element works at low frequency and at high frequency. Figure 3-15 illustrates the excellent agreement between the measured impedance and the simulated impedance of an ideal T element across the entire bandwidth of the measurement.

An ideal resistor-circuit element can model the actual behavior of real resistor devices up to surprisingly high bandwidth. There are three general technologies for resistor components, such as those used for terminating resistors: axial lead, SMT, and integrated passive devices (IPDs). The measured impedance of a representative of each technology is shown in Figure 3-16.

An ideal resistor will have an impedance that is constant with frequency. As can be seen, the IPD resistors match the ideal resistor-element behavior up to the full-measurement bandwidth of 5 GHz. SMT resistors are well approximated by an ideal resistor up to about 2 GHz, depending on the mounting geometry and board

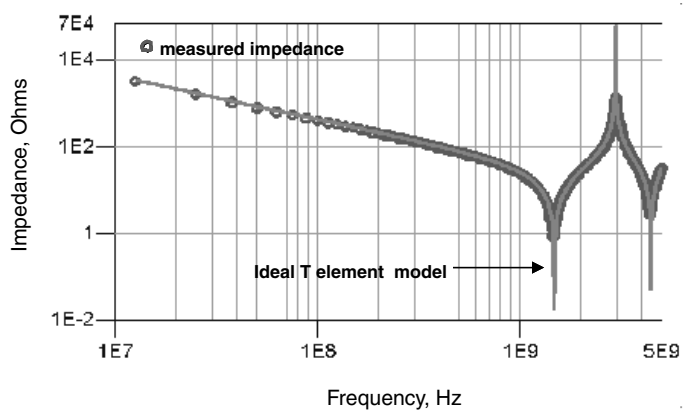


Figure 3-15 Measured impedance of a 1-inch-long microstrip trace and the simulated impedance of an ideal T element model. The agreement is excellent up to the full bandwidth of the measurement. Agreement is also excellent at low frequency.

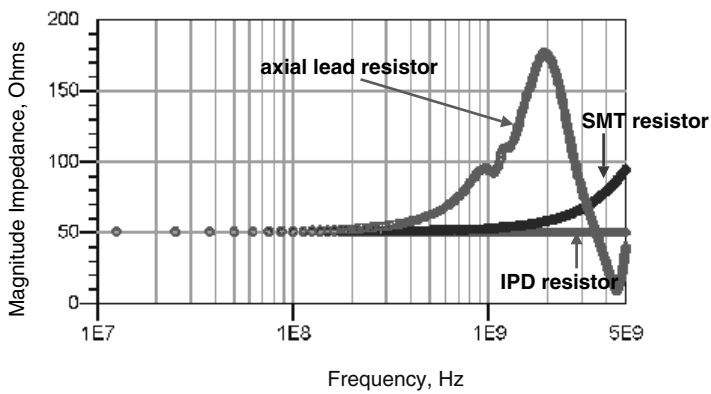


Figure 3-16 Measured impedance of three different resistor components, axial lead, surface-mount (SMT), and integrated passive device (IPD). An ideal resistor element has an impedance constant with frequency. This simple model matches each real resistor at low frequency but has limited bandwidth depending on the resistor technology.

stack-up, and axial-lead resistors can be approximated to about 500 MHz by an ideal resistor. In general, the primary effect that arises at higher frequency is the impact from the inductive properties of the real resistors. A higher-bandwidth model would have to include inductor elements and maybe also capacitor elements.

Having the circuit-model topology is only half of the solution. The other half is to extract the parameter values, either from a measurement or with a calculation. Starting with the circuit topology, we can use rules of thumb, analytic approximations, and numerical-simulation tools to calculate the parameter values from the geometry and material properties for each of the circuit elements. This is detailed in the next chapters.

3.11 The Bottom Line

1. Impedance is a powerful concept to describe all signal-integrity problems and solutions.
2. Impedance describes how voltages and currents are related in an interconnect or component. It is fundamentally the ratio of the voltage across a device to the current through it.
3. Real components that make up the actual hardware are not to be confused with ideal circuit elements that are the mathematical description of an approximation to the real world.
4. Our goal is to create an ideal circuit model that adequately approximates the impedance of the real physical interconnect or component. There will always be a bandwidth beyond which the model is no longer an accurate description, but simple models can work to surprisingly high bandwidth.
5. The resistance of an ideal resistor, the capacitance of an ideal capacitor, and the inductance of an ideal inductor are all constant with frequency.
6. Though impedance has the same definition in the time and frequency domains, the description is simpler and easier to generalize for C and L components in the frequency domain.
7. The impedance of an ideal R is constant with frequency. The impedance of an ideal capacitor varies as $1/\omega C$, and the impedance of an ideal inductor varies as ωL.
8. SPICE is a very powerful tool to simulate the impedance of any circuit or the voltage and current waveforms expected in both the time and frequency domains. All engineers who deal with impedance should have a version of SPICE available to them on their desktop.

9. When building equivalent circuit models for real interconnects, it is always important to start with the simplest model possible and build in complexity from there. The simplest starting models are single R, L, C, or T elements. Higher-bandwidth models use combinations of these ideal circuit elements.
10. Real components can have very simple equivalent circuit models with bandwidths in the GHz range. The only way to know what the bandwidth of a model is, however, is to compare a measurement of the real device to the simulation of the impedance using the ideal circuit model.

This page intentionally left blank

CHAPTER 4

The Physical Basis of Resistance

The electrical description of every interconnect and passive device is based on using just three ideal lumped circuit elements (resistors, capacitors, and inductors) and one distributed element (a transmission line). The electrical properties of the interconnects are all due to the precise layout of the conductors and dielectrics and how they interact with the electric and magnetic fields of the signals.

Understanding the connection between geometry and electrical properties will give us insight into how signals are affected by the physical design of interconnects and feed our intuition about manipulating signal integrity performance by design.

TIP The key to optimizing the physical design of a system for good signal integrity is to be able to accurately predict the electrical performance from the physical design and to efficiently optimize the physical design for a target electrical performance.

All the electrical properties of interconnects can be completely described by the application of Maxwell's Equations. These four equations describe how electric and magnetic fields interact with the boundary conditions: conductors and dielectrics in some geometry. In principle, with optimized software and a powerful

enough computing platform, we should be able to input the precise layout of a circuit board and all the various initial voltages coming out of the devices, push a button, and see the evolution of all the electric and magnetic fields. In principle. After all, there are no new physics or unknown, mysterious effects involved in signal propagation. It's all described by Maxwell's Equations.

There are some software tools available that, running on even a PC, will allow small problems to be completely simulated by Maxwell's Equations. However, there is no way of simulating an entire board directly with Maxwell's Equations—yet. Even if there were, we would be able to perform only a final system verification telling us that the board met or failed the performance spec. Just being able to solve for all the time-varying electric and magnetic fields doesn't give us insight into what should be changed or done differently in the next design.

TIP The design process is a very intuitive process. New ideas come from imagination and creativity. These are fed, not by numerically solving a set of equations but by understanding, at an intuitive level, the meaning of the equations and what they tell us.

4.1 Translating Physical Design into Electrical Performance

As we saw in the last chapter, the simplest starting place in thinking about the electrical performance of an interconnect is with its equivalent electrical circuit model. Every model has two parts, the circuit topology and the parameter values of each circuit element. And, the simplest starting place in modeling any interconnect is using some combination of the three ideal lumped circuit elements (resistors, capacitors, and inductors) or the distributed element (an ideal transmission line circuit element).

Modeling is the process of translating the physical design of line widths, lengths, thickness, and material properties, into the electrical view of R, L, and C elements. Figure 4-1 shows this relationship between the physical view and the electrical view for the special case of a generic RLC model.

Once we have established the topology of the circuit model for an interconnect, the next step is to extract the parameter values. This is sometimes called parasitic extraction. The task then is to take the geometry and material properties and ask how they translate into the equivalent parameter values of the ideal R, C, L, or T elements. We will use rules of thumb, analytic approximations, and numerical simulation tools to do this.

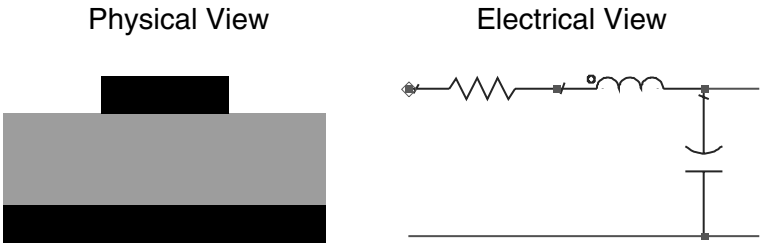


Figure 4-1 The physical worldview and the electrical worldview of the special case of a microstrip interconnect, illustrating the two different perspectives.

In this chapter, we look at how resistance is determined by the geometry and material properties. In the next three chapters, we look at the physical basis of capacitance, inductance, and transmission lines.

4.2 The Only Good Approximation for the Resistance of Interconnects

When we take the two ends of any conductor, such as a copper trace on a board, and apply a voltage across them, we get a current through the conductor. Double the voltage and the current doubles. The impedance across the ends of the real copper trace behaves very much like an ideal resistor. It has an impedance that is constant in time and frequency.

> **TIP** When we extract the resistance of an interconnect, what we are really doing is first implicitly assuming that we will model the interconnect as an ideal resistor.

Once we've established the circuit topology to be an ideal resistor element, we apply one of the three analysis techniques to extract the parameter value based on the specific geometry of the interconnect. The initial accuracy of the model will depend on how well we can translate the actual geometry into one of the standard patterns for which we have good approximations, or how well we can apply a numerical simulation tool. When we want just a rough, approximate, ball-park number, we can apply a rule of thumb.

There is only one good analytical approximation for the resistance of an interconnect. This approximation is for a conductor that has a uniform cross section down its length. For example, a wire bond, a hook-up wire, and a trace on a circuit board have the same diameter or line width all down their length. This

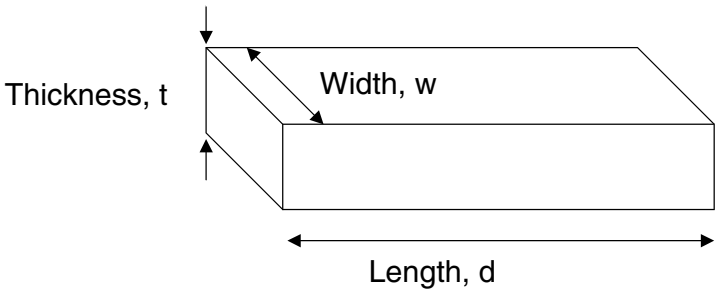


Figure 4-2 Description of the geometrical features for an interconnect that will be modeled by an ideal R element. The resistance is between the two end faces, spaced a distance, d, apart.

approximation will be a good match in these cases. Figure 4-2 illustrates the geometrical features for this approximation.

For the special case of a conductor with the same cross section down its length, the resistance can be approximated by:

$$R = \rho\frac{d}{A} \tag{4-1}$$

where:

R = the resistance, in Ohms

ρ = the bulk resistivity of the conductor, in Ohm-cm

d = the distance between the ends of the interconnect, in cm

A = the cross-section area, in cm^2

For example, if a wire bond has a length of 0.2 cm, or about 80 mils, and a diameter of 0.0025 cm, or 1 mil, and is composed of gold with a resistivity of 2.5 microOhm-cm, then the resistance from one end to the other will be:

$$R = \rho\frac{d}{A} = 2.5 \times 10^{-6} \times \frac{0.2}{(\pi/4)0.0025^2} = 0.1\Omega \tag{4-2}$$

TIP This is a good rule of thumb to remember: The resistance of a 1-mil-diameter wire bond, 80 mils long, is about 0.1 Ohm.

This approximation says that the value of the resistance will increase linearly with length. Double the length of the interconnect and the resistance will double. It also varies inversely with the cross-sectional area. If we make the cross section larger, the resistance decreases. This matches what we know about water flowing through a pipe. A wider pipe means less resistance to water flow. A longer pipe means more resistance.

The parameter value of the resistance of the equivalent ideal resistor depends on the geometry of the structure and its material property (that is, its bulk resistivity). If we change the shape of the wire, the equivalent resistance will change. If the cross section of the conductor changes down its length, like it does in a lead frame of a plastic quad flat pack (PQFP), we must find a way to approximate the actual cross section in terms of a constant cross section, or we cannot use this approximation.

Consider a lead in a 25-mil pitch, 208-pin PQFP. The lead has a total length of 0.5 inch but changes its shape and does not have a constant cross section. It always has a thickness of 3 mils, but its width starts out at 10 mils and widens to 20 mils on the outside edge. How are we to estimate the resistance from one end to the other? The key term in this question is *estimate*. If we need the most accurate result, we would probably want to take a precise profile of the shape of the conductor and use a 3D-modeling tool that would calculate the resistance of each section, taking into account the changes in width.

The only way we can use the approximation above is if we have a structure with a constant cross section. We must approximate the real variable-width PQFP lead with a geometry that has a constant cross section. One way of doing this is to assume that the lead tapers uniformly down its length. If it is 10 mils wide on one end and 20 mils wide at the other, the average width is 15 mils. As a first pass approximation, we can assume a constant cross section of 3 mils thick and 15 mils wide. Then, using the resistivity of copper, the resistance of the lead is:

$$R = \rho\frac{d}{A} = 1.8 \times 10^{-6}\,\mathrm{Ohm-cm} \times \frac{0.5\ \mathrm{inch}}{0.003\ \mathrm{inch} \times 0.015\ \mathrm{inch}} \times \frac{1\ \mathrm{inch}}{2.54\ \mathrm{cm}} = 8\ \mathrm{m\Omega} \qquad (4\text{-}3)$$

It is important to be careful with the units and always be consistent. Resistance will always be measured in Ohms.

4.3 Bulk Resistivity

Bulk resistivity is a fundamental material property that all conductors have. It has units of Ohms-length, such as Ohms-inches or Ohms-cm. This is very confusing. We might expect resistivity to have units of Ohms/cm, and in fact, we often see the bulk resistivity incorrectly reported with these units. However, do not confuse the intrinsic material property with the extrinsic resistance of a piece of interconnect.

Bulk resistivity must have the units it does so that the resistance of an interconnect has units of Ohms. For resistivity × length / (length × length) to equal Ohms, resistivity must have units of Ohms-length.

TIP Bulk resistivity is not a property of the structure or object made from a material, it is a property of the material.

Bulk resistivity is an intrinsic material property, independent of the size of the chunk of material we look at. It is a measure of the intrinsic resistance to current flow of a material. The copper in a chunk 1 mil on a side will have the same bulk resistivity as the copper in a chunk 10 inches on a side.

The worse the conductor, the higher the resistivity. Usually, the Greek letter ρ is used to represent the bulk resistivity of a material. There is another term, *conductivity*, that is sometimes used to describe the electrical resistivity of a material. Usually, the Greek letter σ is used for the conductivity of a material. It is not surprising that a more conductive material will have a higher conductivity. Numerically, resistivity and conductivity are inversely related to each other:

$$\rho = \frac{1}{\sigma} \tag{4-4}$$

While the units of resistivity are Ohms-m, for example, the units for conductivity are 1/(Ohms-m). The unit of 1/Ohms is given the special name *Siemens*. The units of conductivity are Siemens/meter. Figure 4-3 lists the values of many common conductors used in interconnects and their resistivity. It is important to note that the bulk resistivity of most interconnect metals will vary as much as 50% due to different processing conditions. For example, the bulk resistivity of copper is reported as between 1.8 and 4.5 μOhms-cm, depending on whether it is electroplated, electrolessly deposited, sputtered, rolled, extruded, or annealed. The more porous it is, the higher the resistivity. If it is important to know the bulk

Material	Resistivity μ Ohms-cm
Silver	1.47
Copper	1.58
Gold	2.01
Aluminum	2.61
Molybdenum	5.3
Tungsten	5.3
Nickel	6.2
Silver-filled glass	~10
Tin	10.1
Eutectic Pb/Sn solder	15
Lead	19.3
Kovar	49
Alloy42	57
Silver-filled epoxy	~300

Figure 4-3 Typical bulk resistivities of common interconnect materials.

resistivity of the conductor to better than 10%, it should be measured for the specific sample.

We sometimes use the terms *bulk resistivity* or *volume resistivity* to refer to this intrinsic material property. This is to distinguish it from two other resistance-related terms, the *resistance per length* and the *sheet resistance*.

4.4 Resistance per Length

When the cross section of the conductor is uniform down the length, such as in any wire or even in a trace on a circuit board, the resistance of the interconnect will be directly proportional to the length. Using the approximation above, we can see that for a uniform cross-section conductor, the resistance per length is constant and given by:

$$R_L = \frac{R}{d} = \frac{\rho}{A} \tag{4-5}$$

where:

R_L = resistance per length

d = interconnect length

ρ = bulk resistivity

A = cross-section area current travels through

For example, for a wire bond with a diameter of 1 mil, the cross section is constant down the length and the cross sectional area is $A = \pi/4 \times 1 \text{ mil}^2 = 0.8 \times 10^{-6} \text{ inches}^2$. With a bulk resistivity of gold of roughly 1 μOhm-inch, the resistance per length is calculated as $R_L = 1 \text{ μOhm-inch}/0.8 \times 10^{-6} \text{ inches}^2 = 0.8 \text{ ohms/inch} \sim 1 \text{ Ohm/inch}$.

This is an important number to keep in mind and is often called a rule of thumb: The resistance per length of a wire bond is about 1 Ohm/inch. The typical length of a wire bond is about 0.1 inch, so the typical resistance is about 1 Ohm/inch × 0.1 inch = 0.1 Ohm. A wire bond 0.05 inches long would have a resistance of 1 Ohm/inch × 0.05 inch = 0.05 Ohm or 50 mOhms.

The diameter of wires is measured by a standard reference number referred to as the American Wire Gauge (AWG). Figure 4-4 lists some gauge values and the equivalent diameter. From the diameter and assuming copper wire, we can estimate the resistance per length. For example, 22-gauge wire, typical in many personal computer boxes, has a diameter of 25 mils. Its resistance per length is $R_L = 1.58 \text{ μOhm-cm}/(2.54 \text{ cm/inch})/(\pi/4 \times 25 \text{ mil}^2) = 1.2 \times 10^{-3}$ Ohms/inch, or about 15×10^{-3} Ohms/foot or 15 Ohms per 1000 ft.

AWG wire size	Diameter (inches)	Resistance per 1000 ft (Ohms) (assumes ρ = 1.74 μOhms-cm)
24	0.0201	25.67
22	0.0254	16.14
20	0.0320	10.15
18	0.0403	6.385
16	0.0508	4.016
14	0.0640	2.525
12	0.0808	1.588
10	0.1019	0.999

Figure 4-4 AWG and their diameter and resistance per length.

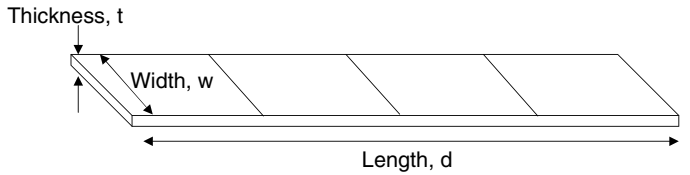


Figure 4-5 A uniform trace cut from a sheet can be divided into a number of squares, n = d/w.

4.5 Sheet Resistance

Many interconnect substrates, such as printed circuit boards, cofired ceramic substrates, and thin film substrates, are fabricated with uniform sheets of conductor that are patterned into traces. All the conductors on each layer have exactly the same thickness. For the special case, where the width of a trace is uniform, as illustrated in Figure 4-5, the resistance of the trace is given by:

$$R = \rho\frac{d}{t \times w} = \left(\frac{\rho}{t}\right) \times \left(\frac{d}{w}\right) \tag{4-6}$$

The first term, (ρ/t), is constant for every trace built on the layer with thickness, t. After all, every trace on the same layer will have the same bulk resistivity and the same thickness. This term is given the special name *sheet resistance* and is designated by R_{sq}.

The second term, (d/w), is the ratio of the length to the width for a specific trace. This is the number of squares that can be drawn down the trace. It is referred to as n and is a dimensionless number. The resistance of a rectangular trace can be rewritten as:

$$R = R_{sq} \times n \tag{4-7}$$

where:

R_{sq} = the sheet resistance

n = the number of squares down the trace

Interestingly, the units of sheet resistance are just Ohms. It has units of resistance, but what resistance does sheet resistance refer to? The simplest way of thinking about sheet resistance is to consider the resistance between the two ends of a section of sheet that is one square in shape (i.e., the length equals the width). In this case, n = 1, and the resistance between the ends of the square trace is just the sheet resistance. Sheet resistance refers to the resistance of one square of conductor.

Surprisingly, whether the square is 10 mils on a side or 10 inches on a side, the resistance across opposite ends of the square is constant. If the length of the square is doubled, we might expect the resistance to double. However, the width would also double and the resistance would be cut in half. These two effects cancel and the net resistance is constant as we change the size of a square.

TIP All square pieces cut from the same sheet of conductor have the same resistance between opposite ends and we call this resistance the sheet resistance, measured in Ohms, and often referred to as Ohms per square.

The sheet resistance will depend on the bulk resistivity of the conductor and the thickness of the sheet. In typical printed circuit boards, fabricated with layers of copper conductor, the thickness of copper is described by the weight of copper per square foot. This is a holdover from the days when the plating thickness was measured by taking a 1-square-foot panel and weighing it. A 1-ounce copper sheet gives 1 ounce of weight of copper per square foot of board. The thickness of 1-ounce copper is about 1.4 mils, or 35 microns. A 1/2-ounce copper sheet has a thickness of 0.7 mil or 17.5 microns. Based on the thickness and the bulk resistivity of copper, the sheet resistance of 1-ounce copper is $R_{sq} = 1.6 \times 10^{-6}$ Ohm-cm $/ 35 \times 10^{-4}$ cm = 0.5 mOhm per sq.

If the copper thickness is cut in half to 1/2-ounce copper, the sheet resistance doubles. Half ounce copper has a sheet resistance of 1 mOhm per square.

TIP A simple rule of thumb to remember is the sheet resistance of 1/2-ounce copper is 1 mOhm/sq. A trace 5 mil-wide and 5 inches long has 1000 squares in series and a resistance of 1 Ohm.

Sheet resistance is an important characteristic of the metallization of a layer. If the thickness and the sheet resistance are measured, then the bulk resistivity of the deposited metal can be found. Sheet resistance is measured by using a spe-

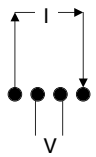


Figure 4-6 Four probes all in a row can be used to measure sheet resistance.

cially designed four-point probe. The four tips are usually mounted to a rigid fixture so they are held in a line with equal spacing. These four probes are placed in contact with the sheet being measured and connected to a four-point impedance analyzer or Ohmmeter. When a constant current is applied to the outer two points and the voltage is measured between the two inner points, the resistance is measured as $R_{meas} = V/I$. Figure 4-6 illustrates the alignment of the probe points and their connection.

As long as the probes are far from the edge (i.e., at least four probe spacings from any edge), the measured resistance is completely independent of the actual spacing of the probe points. The sheet resistance, R_{sq}, can be calculated from the measured resistance using:

$$R_{sq} = 4.53 \times R_{meas} \tag{4-8}$$

If we know the sheet resistance of the sheet, we can use this to calculate the resistance per length and the total resistance of any conductor made in the sheet. Typically, a trace will be defined by a line width, w, and a length, d. The resistance per length of a trace is given by:

$$R_L = \frac{R}{d} = R_{sq} \times \frac{1}{w} \tag{4-9}$$

where:

R_L = resistance per length

R = trace resistance

R_{sq} = sheet resistance

w = line width

d = length of the trace

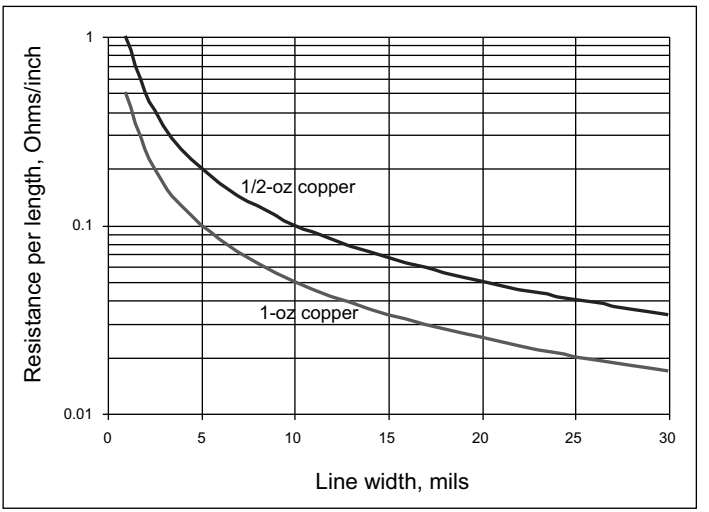


Figure 4-7 The resistance per length for traces with different line widths in 1-ounce copper and 1/2-ounce copper.

Figure 4-7 illustrates the resistance per length of different line widths for 1-ounce and 1/2-ounce copper traces. The wider the line, the lower the resistance per length, as expected. For a 5-mil-wide trace, typical of many backplane applications, a 1/2-ounce copper trace would have a resistance per length of 0.2 Ohm/inch. A 10-inch trace would have a resistance of 0.2 Ohm/inch × 10 inches = 2 Ohms.

It is important to keep in mind that these resistances calculated so far are all resistances at DC, or at least at low frequency. As we show in a later chapter, the resistance of a trace will increase with frequency due to skin-depth-related effects. The bulk resistivity of the conductor does not change; the current distribution through the conductor changes. Higher-frequency signal components will travel through a thin layer near the surface, decreasing the cross-sectional area. For 1-ounce copper traces, the resistance begins to increase at about 20 MHz and will increase roughly with the square root of frequency. It's all related to inductance.

4.6 The Bottom Line

1. Translating physical features into an electrical model is a key step in optimizing system electrical performance.
2. The first step in calculating the resistance of an interconnect is assuming the equivalent circuit model is a simple ideal resistor.

3. The most useful approximation for the end-to-end resistance of an interconnect is R = ρ × length/cross-sectional area.
4. Bulk resistivity is an intrinsic material property, independent of the amount of material.
5. If the structure is not uniform in cross section, it must either be approximated as uniform or a field solver should be used to calculate its resistance.
6. Resistance per length of a uniform trace is constant. A 10-mil-wide trace in 1/2-ounce copper has a resistance per length of 0.1 Ohm/inch.
7. Every square cut from the same sheet will have the same edge-to-edge resistance.
8. Sheet resistance is a measure of the edge-to-edge resistance of one square of conductor cut from the sheet.
9. For 1/2-ounce copper, the sheet resistance is 1 milliOhm/square.
10. The resistance of a conductor will increase at higher frequency due to skin-depth effects. For 1-ounce copper, this begins above 20 MHz.

This page intentionally left blank

CHAPTER 5

The Physical Basis of Capacitance

A capacitor is physically made up of two conductors and between every two conductors there is some capacitance.

> **TIP** The capacitance between any two conductors is basically a measure of their capacity to store charge at the cost of a voltage between them.

If we take two conductors and add positive charge to one of them and negative charge to the other, as illustrated in Figure 5-1, there will be a voltage between them. The capacitance of the pair of conductors is the ratio of the amount of charge stored on each conductor per voltage between them:

$$C = \frac{Q}{V} \tag{5-1}$$

where:

C = the capacitance, in Farads

Q = total amount of charge, in Coulombs

V = voltage between the conductors, in volts

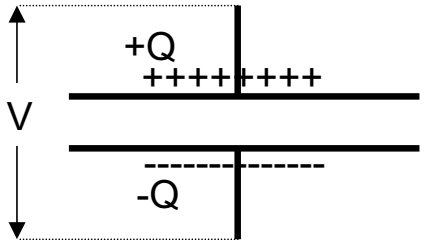


Figure 5-1 Capacitance is a measure of the capacity to store charge for a given voltage between the conductors.

Voltage is the price paid to store charge. The more charge that can be stored for a fixed voltage, the higher the capacitance of the pair of conductors.

The actual capacitance between two conductors is determined by their geometry and the material properties of any dielectrics nearby. It is completely independent of the voltage applied. If the voltage doubles, the amount of charge stored doubles and the ratio stays the same. However, the closer the conductors are brought together, or the more their areas overlap, the greater their capacitance.

Capacitance plays a key role in describing how signals interact with interconnects and is one of the four fundamental ideal circuit elements used to model interconnects.

TIP What makes capacitance so subtle is that even though there may be no direct wire connection between two conductors, which might be two different signal traces, there will always be some capacitance between them. This capacitance will allow a current flow in some cases, which can contribute to cross talk and other signal-integrity problems.

By understanding the physical nature of capacitance, we will be able to see with our mind's eye this sneak path for current flow.

5.1 Current Flow in Capacitors

There is no DC path between the two conductors that are separated by a dielectric material in an ideal capacitor. Normally, we would think there could not be any current flow through a real capacitor. After all, there is insulating dielectric between the conductors. How could we get current flow through the insulating dielectric? As we showed in an earlier chapter, it is possible to get current through

a capacitor, but only in the special case when the voltage between the conductors changes.

The current through a capacitor is related by:

$$I = \frac{\Delta Q}{\Delta t} = C\frac{dV}{dt} \tag{5-2}$$

where:

I = the current through the capacitor

ΔQ = the change in charge on the capacitor

Δt = the time it takes the charge to change

C = the capacitance

dV = the voltage change between the conductors

dt = the time period for the changing voltage

TIP Capacitance is also a measure of how much current we can get through a pair of conductors when the voltage between them changes.

If the capacitance is large, we get a lot of current through it for a fixed dV/dt. The impedance, in the time domain, would be low if the capacitance were high.

How does the current flow through the empty space between the conductors? There is an insulating dielectric. Of course, we can't have any real current through the insulating dielectric, but it sure looks that way.

Rather, there is apparent current flow. If we increase the voltage on the conductors, for example, we must add + charges to one conductor and push out + charges from the other. It looks like we add the charges to one conductor and they come out of the other. Current effectively flows through a capacitor when the voltage across the conductors changes.

We often refer to the current that effectively flows through the empty space of the capacitor as *displacement current*. This is a term we inherited from James Clerk Maxwell, the Father of Electromagnetism. In his mind, the current through the empty space between the conductors, when the voltage changed, was due to the increasing separation of charges in the *ether*. To his 1870s mind, the vacuum was not empty. Filling the vacuum, and the medium for light to propagate, was a tenuous medium called ether. When the voltage between conductors changed,

charges in the ether were pulled apart slightly, or displaced. It was their motion, while they were being displaced, that he imagined to be the current and he termed, the displacement current.

5.2 The Capacitance of a Sphere

The actual amount of capacitance between two conductors is related to how many electric field lines would connect the two conductors. The closer the spacing, the greater the area of overlap, the more field lines would connect the conductors, and the larger the capacity to store charge.

Each specific geometrical configuration of conductors will have a different relationship between the dimensions and the resulting capacitance. With few exceptions, most of the formulae that relate geometry to capacitance are approximations. In general, we can always use a field solver to accurately calculate the capacitance between every pair of conductors in any arbitrary collection of conductors, such as with multiple pins in a connector. There are a few special geometries where good approximations exist. One of these is for the capacitance between two concentric spheres, one inside the other.

The capacitance between the two spheres is:

$$C = 4\pi\varepsilon_0 \frac{r r_b}{r_b - r} \tag{5-3}$$

where:

C= the capacitance, in pF

ε_0 = permittivity of free space = 0.089 pF/cm or 0.225 pF/inch

r = the radius of the inner sphere, in inches or cm

r_b = the radius of the bigger, outer sphere, in inches or cm

When the outer sphere's radius is more than 10 × the inner sphere's, the capacitance of the sphere is approximated by:

$$C \approx 4\pi \times \varepsilon_0 \times r \tag{5-4}$$

where:

C = the capacitance, in pF

ε_0 = permittivity of free space = 0.089 pF/cm or 0.225 pF/inch

r = the radius of the sphere, in inches or cm

For example, a sphere, 0.5 inch in radius, or 1 inch in diameter, has a capacitance of C = $4\pi \times 0.225$ pF/in $\times$ 0.5 in = 1.8 pF. As a rough rule of thumb, a sphere with a 1-inch diameter has a capacitance of about 2 pF.

TIP This subtle relationship says that any conductor, just sitting, isolated in space, has some capacitance with respect to even the earth's surface. It does not get smaller and smaller. It has a minimum amount of capacitance, related to its diameter. The closer to a nearby surface, the higher the capacitance might be, above this minimum.

This means that if a small wire pigtail, even only a few inches in length, were sticking outside a box, it could have a stray capacitance of at least 2 pF. At 1 GHz, the impedance to the earth, or the chassis, would be about 100 Ohms. This is an example of how subtle capacitance can be and how it can create significant sneak current paths, especially at high frequency.

5.3 Parallel Plate Approximation

A very common approximation is the parallel plate approximation. For the case of two flat plates, as shown in Figure 5-2, separated by a distance, h, with total area, A, with just air between the plates, the capacitance is given by:

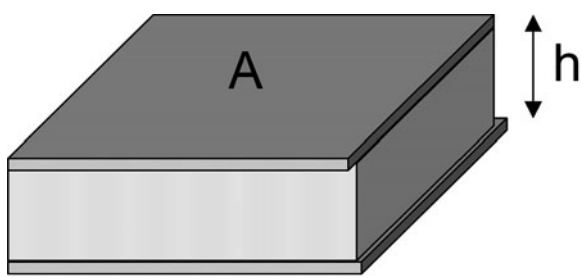


Figure 5-2 Most common approximation for the capacitance and geometry is for a pair of parallel plates with plate area, A, and separation, h.

$$C = \varepsilon_0 \frac{A}{h} \tag{5-5}$$

where:

C = capacitance, in pF

ε_0 = permittivity of free space = 0.089 pF/cm or 0.225 pF/inch

A = area of the plates

h = separation between the plates

For example, a pair of plates that look like the faces of a penny, about 1 cm^2 in area, separated by 1 mm of air, has a capacitance of C = 0.089 pF/cm × 1 cm^2/ 0.1 cm = 0.9 pF. This is a good way of getting a feel for 1 pF of capacitance. It is roughly the capacitance associated with plates of comparable size to the faces of a penny.

TIP This relationship points out the important geometrical features for all capacitors. The farther apart the conductors, the lower the capacitance, and the larger the area of overlap, the greater the capacitance.

It is always important to keep in mind that with the exception of only a few equations, every equation used in signal integrity is either a definition or an approximation. The parallel plate approximation is an approximation. It assumes that the fringe fields around the perimeter of the plates are negligible. The thinner the spacing or the wider the plates, the better this approximation is. For the case of plates that are square and of dimension w on a side, the approximation gets better as w/h gets larger.

In general, the parallel plate approximation is going to underestimate the capacitance. The actual capacitance is going to be larger than the approximation because of the contribution of the fringe fields from the edge. As a rough rule of thumb, when the distance between the plates is equal to a lateral dimension, so the plates look like a cube, the actual capacitance between the two plates is roughly twice what the parallel plate approximation predicts. In other words, the fringe fields from the edge contribute an equal amount of capacitance as the parallel plate approximation predicts, when the spacing is comparable to the width of a plate.

5.4 Dielectric Constant

The presence of an insulating material between the conductors will increase the capacitance between them. The special material property that causes the capacitance to increase is called the *relative dielectric constant*. We usually use the Greek letter epsilon plus a subscript r (ε_r) to describe the relative dielectric constant of a material. Alternatively, we use the abbreviation, Dk, to designate the dielectric constant of a material. It is the dielectric constant, relative to air, which has a dielectric constant of 1. As a ratio, there are no units for relative dielectric constant. We often leave off the term *relative*.

Dielectric constant is an intrinsic bulk property of an insulator. A small piece of epoxy will have the same dielectric constant as a large chunk of the same material. The way to measure the dielectric constant of an insulator is to compare the capacitance of a pair of conductors when they are surrounded by air, C_0, and when they are completely surrounded by the material, C. The dielectric constant of the material is defined as:

$$\varepsilon_r = \frac{C}{C_0} \tag{5-6}$$

where:

ε_r = relative dielectric constant of the material

C = capacitance when conductors are completely surrounded by the material

C_0 = capacitance when air completely surrounds the conductors

The higher the dielectric constant, the greater the capacitance between the fixed conductors is increased by the material. The dielectric constant will increase the capacitance of any two conductors, completely independent of their shape (whether they are shaped like a parallel plate, two rods, or one wire near a wide plane), provided all space in the vicinity of the conductors is uniformly filled with the material.

Figure 5-3 lists the dielectric constant of many common insulators used in interconnects. For most polymers, the dielectric constant is about 3.5–4.5. This says that the capacitance between two electrodes is increased by a factor of roughly four by the addition of a polymer material. It is important to note that the dielectric constant of most polymers will vary due to processing conditions, degree of cure, and any filler materials used. There may be some frequency dependence as

Material	Dielectric Constant
Air	1
Teflon	2.1
Polyethylene	2.3
BCB	2.6
PTFE	2.8
Polyimide	3.4
GETEK	3.6–4.2
BT/Glass	3.7–3.9
Quartz	3.8
Kapton	4
FR4	4–4.5
Glass-Ceramic	5
Diamond	5.7
Alumina	9–10
Barium Titanate	5,000

Figure 5-3 Dielectric constant of commonly used interconnect dielectric materials.

well. If it is important to know the dielectric constant to better than 10%, the dielectric constant of the sample should be measured.

The dielectric constant of a material is roughly related to the number of dipoles and their size. A material having molecules with a lot of dipoles, such as water, will have a high dielectric constant (over 80). A material with very few dipoles, such as air, will have a low dielectric constant (1). The lowest dielectric constant of any solid, homogeneous material is about 2, which is for Teflon. The dielectric constant can be decreased by adding air to the material. Foams have a dielectric constant that can approach 1. At the other extreme, some ceramics, such as barium titanate, have dielectric constants as large as 5,000.

The dielectric constant will sometimes vary with frequency. For example, from 1 kHz to 10 MHz, the dielectric constant of FR4 can vary from 4.8 to 4.4. However, from 1 GHz to 10 GHz, the dielectric constant of FR4 can be very constant. The exact value of the dielectric constant for FR4 varies depending on the relative amount of epoxy resin and glass weave. To remove the ambiguity, it is important to specify the frequency at which the dielectric constant is measured.

5.5 Power and Ground Planes and Decoupling Capacitance

One of the most important applications of the parallel plate approximation is to analyze the capacitance between the power and ground planes in an IC package or in a multilayer printed circuit board.

As we will show later, in order to reduce the voltage rail collapse in the power-distribution system, it is important to have a lot of decoupling capacitance between the power and ground return. A capacitance, C, will prevent the droop in the power voltage for a certain amount of time, δt. If the power dissipation of the chip is P, the time until the voltage droop increases to 5% of the supply voltage, because of the decoupling capacitance, is approximately:

$$\delta t = C \times 0.05 \times \frac{V^2}{P} \tag{5-7}$$

where:

δt = time in seconds before droop exceeds 5%

C = decoupling capacitance, in Farads

0.05 = the 5% voltage droop allowed

P = the average power dissipation of the chip, in watts

V = the supply voltage, in volts

For example, if the chip power is 1 watt, and the capacitance available for decoupling is 1 nF, with a supply voltage of 3.3 v, the time the capacitance will provide decoupling is $\delta t = 1\ nF \times 0.05 \times 3.3^2/1 = 0.5$ nsec. This is not a very long period of time, compared to what is required.

More typically, enough decoupling capacitance is required to provide decoupling for at least 5 μsec, until the power supply regulator can provide adequate current. In this example, we would actually need more than 10,000 times this decoupling capacitance, or 10 μF, to provide adequate decoupling.

It is often erroneously assumed that the capacitance found in the power and ground planes of the circuit board will provide a significant amount of decoupling. By using the parallel plate approximation, we can put in the numbers and evaluate just how much decoupling capacitance they provide and how long the planes can decouple a chip.

In a multilayer circuit board, with the power plane on an adjacent layer to the ground plane, we can estimate the capacitance between the layers, per square inch of area. The capacitance is given by:

$$C = \varepsilon_0 \varepsilon_r \frac{A}{h} \tag{5-8}$$

where:

C = capacitance, in pF

ε_0 = permittivity of free space = 0.089 pF/cm or 0.225 pF/inch

ε_r = relative dielectric constant of the FR4, typically ~ 4

A = area of the planes

h = separation between the planes

For the case of FR4, having a dielectric constant of 4, the capacitance for a 1 in^2 of planes is C = 0.225 pF/inch × 4 × 1 in^2/h ~ 1000 pF/h, with h in mils. If the dielectric spacing is 10 mils, a very common thickness, the capacitance between the power and ground planes is only 100 pF for 1 in^2 of planes.

If there are 4 in^2 of board area allocated for the ASIC, then the total plane-to-plane decoupling capacitance available in the power and ground planes of the board is only 0.4 nF. This is more than four orders of magnitude below the required 10-μF capacitance.

How long will this amount of capacitance provide decoupling? Using the relationship above, the amount of time the 0.4-nF capacitance in the planes would provide decoupling is 0.2 nsec. This is not a very significant amount of time. Furthermore, the chip would have to see this small capacitance through its package leads. The impedance of the package leads would make this 0.4-nF plane capacitance almost invisible. In addition, the capacitance already integrated on a chip is typically more than 100 times this plane-to-plane capacitance.

TIP Though there is plane-to-plane capacitance in a multilayer circuit board, in general, it is too small to play a significant role in power management. As we show in the next chapter, the real role of the power and ground planes is to provide a low inductance path between the chip and the bulk decoupling capacitors. It is not to provide decoupling capacitance.

What could be done to dramatically increase the capacitance in the power and ground planes? The parallel plate approximation points out only two knobs that affect the capacitance: the dielectric thickness and the dielectric constant. The thinnest layer of FR4 in commercial production is 2 mils. This results in a capacitance per area of about 1000 pF/in^2/2 = 500 pF/in^2. For this example, if the area allocated for decoupling a chip is two inches on a side, the total capacitance would be 500 pF/in^2 × 4 in^2, or 2 nF. The time this would keep the voltage from collapsing is about 1 nsec, still not very significant.

However, if the dielectric thickness can be made thin enough and the dielectric constant high enough, a significant amount of capacitance can be designed into the power and ground planes. Figure 5-4 shows the capacitance per square inch as the dielectric thickness changes, for four different dielectric constants of 1, 4, 10, and 20. Obviously, if the goal is to increase the capacitance per area, the way to do this is use thin dielectric layers and high dielectric constant.

An example of such a material is in development by 3M, under the trade name C-Ply. It is composed of ground-up barium titanate in a polymer matrix. The dielectric constant is 20 and the layer thickness is 8 microns, or 0.33 mil. With 1/2-ounce copper layers laminated to each side, the capacitance per area of

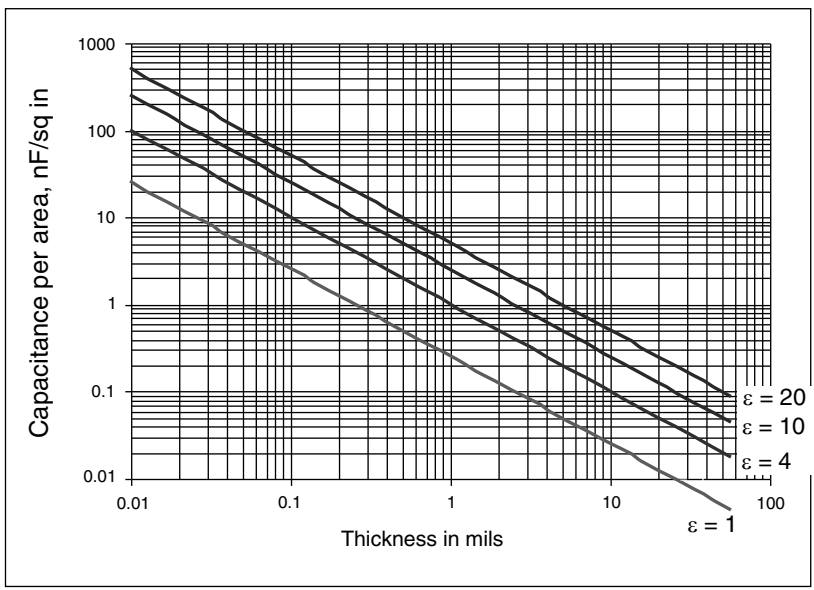


Figure 5-4 Capacitance per area for power and ground planes with four different dielectric constants and different plane-to-plane thicknesses.

one layer is C/A = 0.225 pF/inch × 20 /0.33 mil ~ 14 nF/in^2. This is about 30 times greater than the best alternative.

A 4-in^2 region of a board with a layer of C-Ply would have about 56 nF of decoupling capacitance. For the 1-watt chip, this would provide decoupling for about 28 nsec, a significant amount of time.

5.6 Capacitance per Length

Most uniform interconnects have a signal path and a return path with a fixed cross section. In this case, the capacitance between the signal trace and its return path scales with the length of the interconnect. If the interconnect length doubles, the total capacitance of the trace will double. It is convenient to describe the capacitance of the line by the capacitance per length. As long as the cross section remains uniform, the capacitance per length will be constant.

In the special case of uniform cross-section interconnects, the total capacitance between the signal and return path is related to:

$$C = C_L \times Len \tag{5-9}$$

where:

C = total capacitance of the interconnect

C_L = capacitance per length

Len = length of the interconnect

There are three cross sections that have a very good approximation for the capacitance per length based on the cross section. These are shown in Figure 5-5. In addition, there are many other approximations for other geometries but they are not as accurate.

TIP In general, when the cross section is uniform, a 2D field solver can be used to very accurately calculate the capacitance per length of any arbitrary shape.

A coax cable is an interconnect with a central conductor surrounded by dielectric material and then enclosed by an outer round conductor. The center conductor is usually termed the signal path, and the outer conductor is termed the

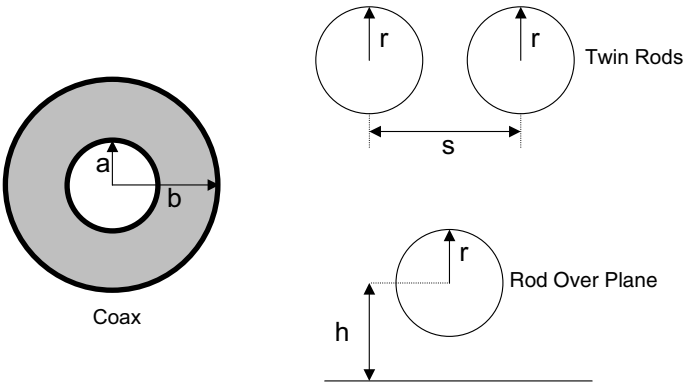


Figure 5-5 The three cross-section geometries for which there are very good approximations for the capacitance per length: coax, twin rods, and rod over plane.

return path. The capacitance per length between the inner conductor and the outer conductor is given exactly by:

$$C_L = \frac{2\pi\varepsilon_0\varepsilon_r}{\ln\left(\frac{b}{a}\right)} \tag{5-10}$$

where:

C_L = capacitance per length

ε_0 = permittivity of free space = 0.089 pF/cm or 0.225 pF/inch

ε_r = relative dielectric constant of the insulation

a = inner radius of the signal conductor

b = outer radius of the return conductor

For example, in an RG58 coax cable (the most common type of coax cable, typically with Berkeley Nuclear Corp. [BNC] connectors on the ends), the ratio of the outer to the inner diameters is (1.62 mm/0.54 mm) = 3, with a dielectric constant of 2.3 for polyethylene, the capacitance per length is

$$C_L = \frac{2\pi \times 0.225 \times 2.3}{\ln(3)} = 2.9\frac{\text{pF}}{\text{inch}} \tag{5-11}$$

A second exact relationship is for the capacitance between two parallel rods. It is given by:

$$C_L = \frac{\pi \varepsilon_0 \varepsilon_r}{\ln\left\{\frac{s}{2r}\left[1 + \sqrt{1 - \left(\frac{2r}{s}\right)^2}\right]\right\}} \tag{5-12}$$

where:

C_L = capacitance per length

ε_0 = permittivity of free space = 0.089 pF/cm or 0.225 pF/inch

ε_r = relative dielectric constant of the insulation

s = center to center separation of the rods

r = radius of the two rods

If the spacing between the rods is large compared with their radius (i.e., s >> r), this relatively complex relationship can be approximated by:

$$C_L = \frac{\pi \varepsilon_0 \varepsilon_r}{\ln\left\{\frac{s}{r}\right\}} \tag{5-13}$$

Both cases assume that the dielectric material surrounding the two rods is uniform everywhere. Unfortunately, this is not often the case, and this approximation is not very useful except in special cases such as wire bonds in air. For example, two parallel wire bonds, both of radius 0.5 mil and 5-mil center-to-center separation, have a capacitance per length of about:

$$C_L = \frac{\pi \varepsilon_0 \varepsilon_r}{\ln\left\{\frac{s}{r}\right\}} = \frac{3.14 \times 0.225 \times 1}{\ln\left\{\frac{5}{0.5}\right\}} = 0.3 \frac{pF}{inch} \tag{5-14}$$

If they are 40 mils long, the total capacitance is $0.3 \times 0.04 = 0.012$ pF.

The third good approximation is for a rod over a plane. The capacitance, when the rod is far from the plane (i.e., h >> r), is approximately:

$$C_L = \frac{2\pi\varepsilon_0\varepsilon_r}{\ln\left\{\frac{2h}{r}\right\}} \tag{5-15}$$

where:

C_L = capacitance per length

ε_0 = permittivity of free space = 0.089 pF/cm or 0.225 pF/inch

ε_r = relative dielectric constant of the insulation

h = center of the rod to surface of the plane

r = radius of the rod

There are two other useful approximations for cross sections that are commonly found in circuit-board interconnects. These are for microstrip and stripline interconnects, as illustrated in Figure 5-6. In microstrip interconnects, a signal trace rests on top of a dielectric layer that has a plane below. This is the common geometry for surface traces in a multilayer board. In a stripline, two planes provide the return path. Whether or not the two planes actually have a DC connection between them, for high-frequency signals, they are effectively shorted together and can be considered as connected. A signal trace is symmetrically spaced between them. A dielectric material, the board laminate, completely surrounds the signal conductors. In both cases, the capacitance per length between the signal trace and return path is calculated.

Though many approximations exist in the literature, the two offered here are recommended by the IPC, the industry association for the printed circuit board industry. The capacitance per length of a microstrip is given by:

$$C_L = \frac{0.67(1.41 + \varepsilon_r)}{\ln\left\{\frac{5.98 \times h}{0.8 \times w + t}\right\}} \cong \frac{0.67(1.41 + \varepsilon_r)}{\ln\left\{7.5\left(\frac{h}{w}\right)\right\}} \tag{5-16}$$

where:

C_L = capacitance per length, in pF/inch

ε_r = relative dielectric constant of the insulation

h = dielectric thickness, in mils

w = line width, in mils

t = thickness of the conductor, in mils

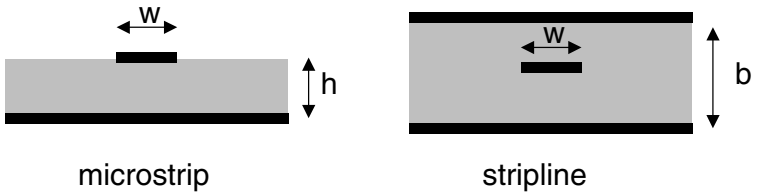


Figure 5-6 The cross-section geometries for microstrip and stripline interconnects illustrating the important geometrical features.

It is important to keep in mind that though there is a parameter to include the thickness of the trace, if a problem requires the level of accuracy where the impact from the trace thickness is important, this approximation should not be used. Rather, a 2D field solver should be used. For all purposes, the accuracy of this tool will not be affected if the trace thickness is assumed to be 0.

If the line width is twice the dielectric thickness (i.e., $w = 2 \times h$), and the dielectric constant is four, the capacitance per length is about $C_L = 2.7$ pF/inch. These are the dimensions of a microstrip that is approximately a 50-Ohm transmission line.

The capacitance per length of a stripline, as illustrated in Figure 5-6, is approximated by:

$$C_L = \frac{1.4\varepsilon_r}{\ln\left\{\frac{1.9 \times b}{0.8 \times w + t}\right\}} \cong \frac{1.4\varepsilon_r}{\ln\left\{2.4\left(\frac{b}{w}\right)\right\}} \tag{5-17}$$

where:

C_L = capacitance per length in pF/inch

ε_r = relative dielectric constant of the insulation

b = total dielectric thickness, in mils

w = line width, in mils

t = thickness of the conductor, in mils

For example, if the total dielectric thickness, b, is twice the line width, b = 2w, corresponding to roughly a 50-Ohm line, the capacitance per length is $C_L = 3.8$ pF/inch.

In both geometries, we see that the capacitance per length of a 50-Ohm line is about 3.5 pF/inch. This is a good rule of thumb to keep in mind.

TIP As a rough rule of thumb, the capacitance per length of a 50-Ohm transmission line in FR4 is about 3.5 pF/inch.

For example, in a multilayer BGA package, signal traces are designed as microstrip geometries and are roughly 50 Ohms. The dielectric material, Bismaleimide triazine (BT), has a dielectric constant of about 3.9. The capacitance per length of a signal trace is roughly 3.5 pF/inch. A trace that is 0.5 inch long will have a capacitance of about 3.5 pF/inch × 0.5 inch = 1.7 pF. The capacitive load of a receiver would be the roughly 2 pF of input-gate capacitance and the 1.7 pF of the lead capacitance, or about 3.7 pF.

TIP It is important to keep in mind that these approximations are APPROXIMATIONS. If it is important to have confidence in the accuracy, the capacitance per length should be calculated with a 2D field solver.

5.7 2D Field Solvers

When accuracy is important, the best numerical tool to use to calculate the capacitance per length between any two conductors is a 2D field solver. These tools assume the cross section of the conductors is constant down their length. When this is the case, the capacitance per length is also constant down the length.

A 2D field solver will solve LaPlace's Equation, one of Maxwell's Equations, using the geometry of the conductors as the boundary conditions. The process, which most tools use without the user really needing to know, is to set the voltage of one conductor to be 1 volt and solving for the electric fields everywhere in the space. The charge on the conductors is then calculated from the electric fields. The capacitance per length between the two conductors is directly calculated as the ratio of the charge on the conductors per the 1 v applied.

One way of evaluating the accuracy of a 2D field solver is to use the tool to calculate a geometry for which there is an exact expression, such as a coax or twin-rod structure. For example, Figure 5-7 shows the comparison of the capacitance per length calculated by a field solver for the twin-rod structure, compared to the exact expression above. The agreement is seen to be excellent. To quantify the residual error, the relative difference between the field-solver result and the

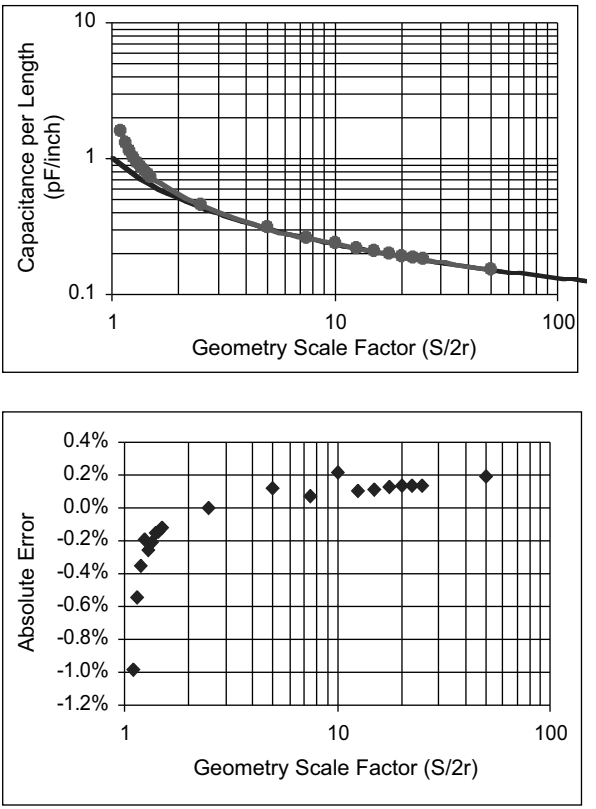


Figure 5-7 Calculated capacitance per length of twin rods comparing Ansoft's 2D Extractor field solver and the exact formula and the approximation. Top: The points are from the field solver; the line through them is the exact expression and the other line is the approximation. For s > 4r, the approximation is excellent. Bottom: Absolute error of the Ansoft 2D field solver is seen to be better than 1%.

analytic result is plotted. The worst case error is seen to be less than 1% for this specific tool.

TIP For any arbitrary geometry, it is possible to get better than 1% absolute error using a field solver to calculate the capacitance per length.

Using a field solver whose accuracy has been verified, it is interesting to evaluate the accuracy of some of the popular approximations, such as the capacitance per length for a microstrip or stripline structure.

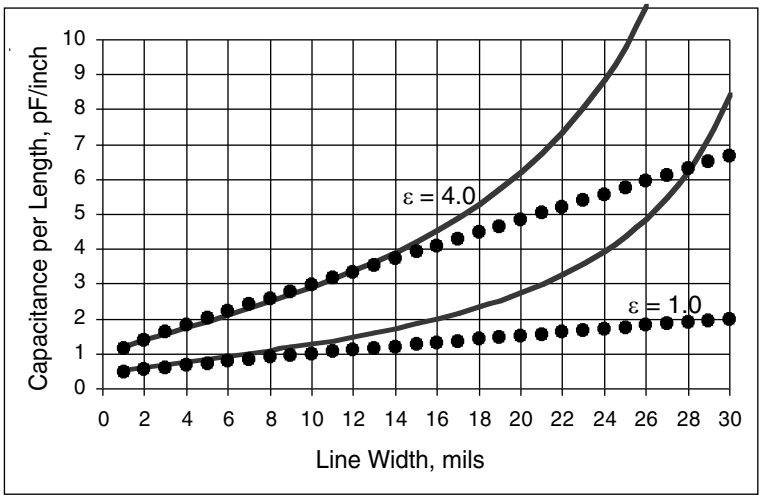


Figure 5-8 Calculated capacitance per length of a microstrip with 5-mil-thick dielectric, and dielectric constants of 4 and 1, comparing Ansoft's SI2D field solver and the IPC approximation above, as the line width is increased. Circles are the field-solver results; the lines are the IPC approximation.

Figure 5-8 compares the approximation above for a microstrip and the field-solver result. In some cases, the approximation can be as good as 5%, but in other cases, they differ by more than 20%. No approximation should be relied on for better than 10%–20% unless previously verified.

An additional advantage of a field solver is the ability to take into account second-order effects. One important effect is the impact from increasing trace thickness on the capacitance per length for a microstrip. We would expect that as the metal thickness increases, from very thin to very thick, the fringe fields between the signal line and the return plane will increase. In fact, as shown in Figure 5-9, the capacitance does increase with increasing trace thickness, but not very much. It increases only about 3% from very thin to 3 mils thick, or 2-ounce copper. There are no approximations that can accurately predict these sorts of effects. For comparison, we show how the IPC approximation for the capacitance matches the more accurate field-solver results. The poor agreement is the reason the approximation should not be used to evaluate the impact of second-order effects, such as trace thickness.

A 2D field solver is a very important tool to calculate the electrical properties of any uniform cross-section interconnect. It is especially important when the dielectric material is not homogeneously distributed around the conductors.

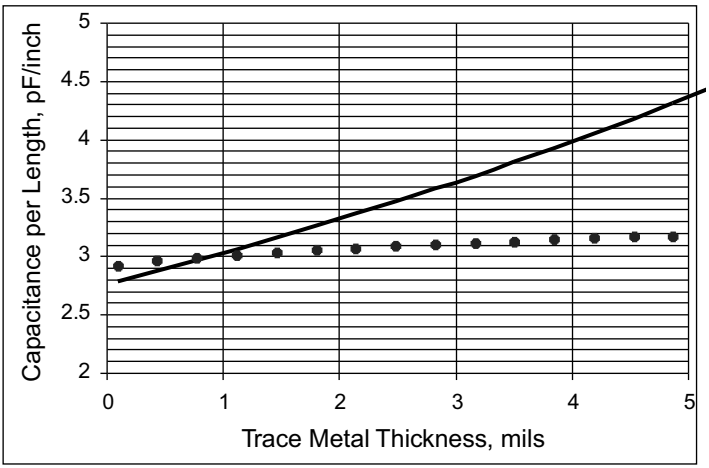


Figure 5-9 Effect on capacitance per length of a microstrip as the trace thickness is increased from 0.1 mil to 5 mils with 5-mil-thick dielectric and 10-mil-wide conductor. Circles are the field-solver results; the line is the IPC approximation.

5.8 Effective Dielectric Constant

In any cross section where the dielectric material completely surrounds the conductors, all the field lines between the conductors will see the same dielectric constant. This is the case for stripline, for example. However, if the material is not uniformly distributed around the conductors, such as with a microstrip or twisted pair or coplanar, some field lines may encounter air while other field lines encounter the dielectric. An example of the field lines in microstrip is shown in Figure 5-10.

The presence of the insulating material will increase the capacitance between the conductors compared to if the material was not there. When the material is homogeneously distributed between and around the conductors, such as in stripline, the material will increase the capacitance by a factor equal to the dielectric constant of the material. In microstrip, some of the field lines encounter air and some encounter the dielectric constant of the laminate. The capacitance between the signal path and the plane will be increased due to the material between them, but by how much?

The combination of air and partial filling of the dielectric creates an *effective dielectric constant*. Just as the dielectric constant is the ratio of the filled capacitance to the empty capacitance, the effective dielectric constant is the ratio of the capacitance when the material is in place with whatever distribution is really

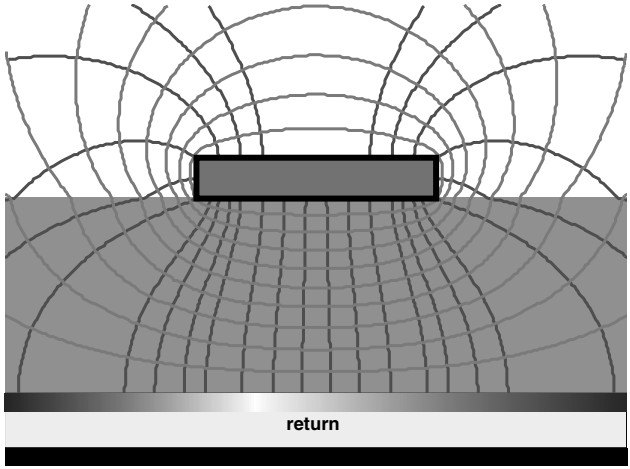


Figure 5-10 Field lines in a microstrip showing some in air and some in the bulk material. The effective dielectric constant is a combination of the dielectric constant of air and the bulk dielectric constant. Field lines calculated using Mentor Graphics Hyperlynx.

there, compared to the capacitance between the conductors when there is only air surrounding them.

The first step in calculating the effective dielectric constant, ε_{eff}, is to calculate the capacitance per length between the two conductors with empty space between them, C_0. The second step is to add the material as it is distributed and calculate the resulting capacitance per length, C_{filled}. The effective dielectric constant is just the ratio:

$$\varepsilon_{eff} = \frac{C_{filled}}{C_0} \tag{5-18}$$

where:

C_0 = empty space capacitance

C_{filled} = capacitance with the actual material distribution

ε_{eff} = the effective dielectric constant

Both the empty and filled capacitance can be accurately calculated using a 2D field solver. The only accurate way of calculating the effective dielectric constant of a transmission line is using a 2D field solver. As we show in a later chapter,

the effective dielectric constant is a very important performance term, as it directly determines the speed of a signal in a transmission line.

Figure 5-11 shows the calculated effective dielectric constant for a microstrip as the width of the trace is increased. The bulk dielectric constant in this case is 4. For a very wide trace, most of the field lines are in the bulk material and the effective dielectric constant approaches 4. When the line width is narrow, most of the field lines are in the air and the effective dielectric constant is below 3, reflecting the contribution from the lower dielectric constant of air.

TIP The intrinsic bulk dielectric constant of the laminate material is not changing. Only how it affects the capacitance changes as the fields between the conductors encounter a different mix of air and dielectric.

If dielectric material is added to the top surface of the microstrip, the fringe field lines that were in air will see a higher dielectric constant and the capacitance of the microstrip will increase. With dielectric above, we call this an embedded microstrip. When only some of the field lines are in the material it is called a partially embedded microstrip. This is the case for a soldermask coating,

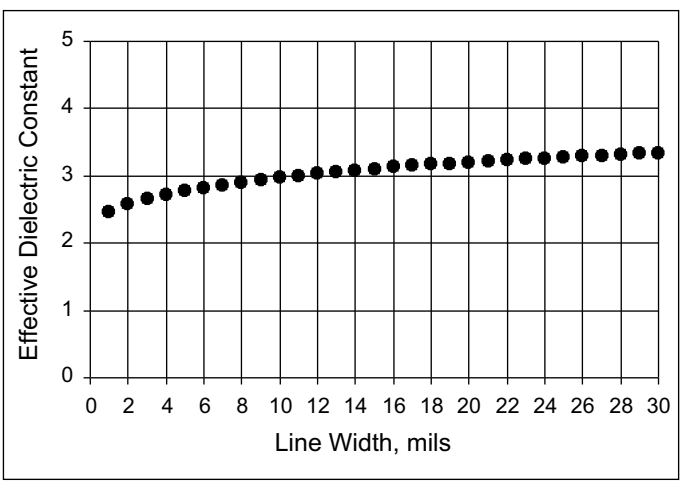


Figure 5-11 Effective dielectric constant of a microstrip with a 5-mil-thick dielectric, as the line width is increased. The dielectric constant of the bulk material is 4. Results are calculated with the Ansoft 2D Extractor.

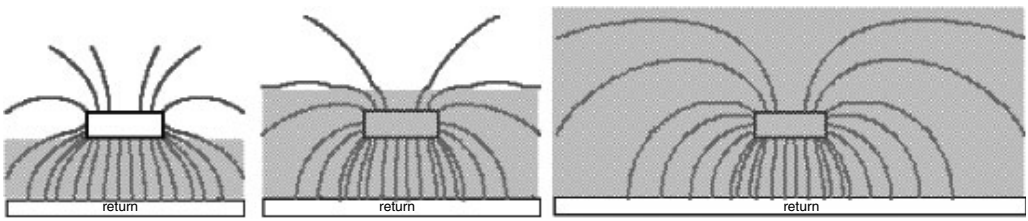


Figure 5-12 The electric field distribution around a microstrip with different thickness covering. For a thick enough layer, all the fields are confined in the bulk material and the capacitance will be independent of more thickness. Simulations performed with Mentor Graphics Hyperlynx.

for example. When all the field lines are covered by dielectric, it is called a fully embedded microstrip. Figure 5-12 shows the field-line distribution for three different degrees of embedded microstrip.

How much material would have to be added to a microstrip to completely cover all the field lines and have the bulk dielectric constant match the effective dielectric constant? This is an easy problem to solve with a 2D field solver. Figure 5-13 shows the calculated capacitance per length for a microstrip with 5-mil-thick dielectric and 10-mil-wide trace as a top layer of dielectric is added with the same bulk dielectric constant value as the laminate, 4.

In this example, a thickness of dielectric on top of the trace about equal to the line width is required to fully cover all the fringe fields.

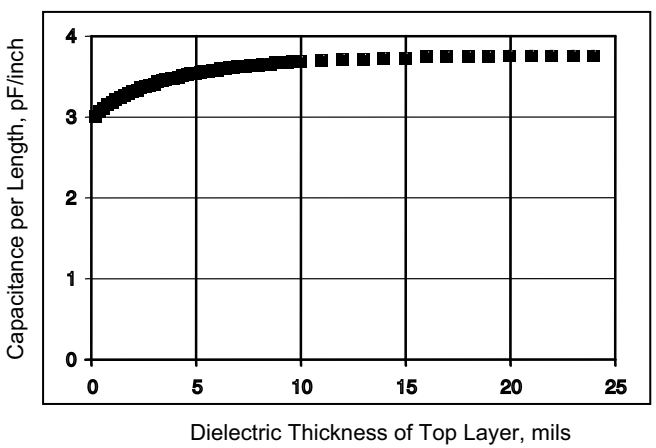


Figure 5-13 Capacitance per length of microstrip as top dielectric thickness is increased. Calculated using Ansoft's 2D Extractor.

5.9 The Bottom Line

1. Capacitance is a measure of the capacity to store charge between two conductors.
2. Current flows through a capacitor when the voltage between the conductors changes. Capacitance is a measure of how much current will flow.
3. With few exceptions, every formula relating the capacitance of two conductors is an approximation. If better than 10%–20% accuracy is required, approximations should not be used.
4. The only three exact expressions are for coax, twin rods, and rod over ground.
5. In general the farther apart the conductors, the lower the capacitance. The greater the overlap of their areas, the higher the capacitance.
6. Dielectric constant is an intrinsic bulk-material property that relates how much the capacitance is increased due to the presence of the material.
7. The power and ground planes in a circuit board have capacitance. However, the amount is so small as to be negligible. The value the planes provide is not for their decoupling capacitance, it is for their low loop inductance.
8. The IPC approximations for microstrip and stripline should not be used if accuracy better than 10% is required.
9. A 2D field solver, once verified, can be used to calculate the capacitance per length of a uniform transmission line structure to better than 1%.
10. Increasing conductor thickness in a microstrip will increase the capacitance per length, but only slightly. A change from very thin to 2-ounce copper only increases the capacitance by 3%.
11. Increasing the thickness of the dielectric coating on top of a microstrip will increase the capacitance. Completely enclosing all the fringe field lines happens when the coating is as thick as the trace is wide and the capacitance can increase by as much as 20%.
12. The effective dielectric constant is the composite dielectric constant when the material is not homogeneously distributed and some field lines see different materials, such as in a microstrip. It can easily be calculated with a 2D field solver.

CHAPTER 6

The Physical Basis of Inductance

Inductance is a critically important electrical property because it affects virtually all signal-integrity problems. Inductance plays a role in signal propagation for uniform transmission lines as a discontinuity, in the coupling between two signal lines, in the power-distribution network, and in EMI.

In many cases, the goal will be to decrease inductance, such as the mutual inductance between signal paths for reduced switching noise, the loop inductance in the power-distribution network, and the effective inductance of return planes for EMI. In other cases, the goal may be to optimize the inductance, as in achieving a target characteristic impedance.

By understanding the basic types of inductance and how the physical design influences the magnitude of the inductance, we will see how to optimize the physical design for acceptable signal integrity.

6.1 What Is Inductance?

There is not a single person involved with signal integrity and interconnect design who has not worried about inductance at one time or another. Yet, very few engineers use the term correctly. This is fundamentally due to the way we all learned about inductance in high school or college physics or electrical engineering.

Typically, we were taught about inductance and how it related to flux lines in coils. We were introduced to the inductance of a coil with figures of coiled wires,

or solenoids, with flux lines through them. Or, we were told inductance was mathematically an integral of magnetic field density through surfaces. For example, a commonly used definition of inductance, L, is:

$$L = \frac{1}{I} \int_{area} \vec{B} \bullet \hat{n} da \tag{6-1}$$

While all of these explanations may be perfectly true, it doesn't help us on a practical level. Where are the coils in a signal-return path? What does an integral of magnetic field density really mean? We have not been trained to apply the concepts of inductance to the applications we face daily—applications related to the interaction of signals in interconnects as packages, connectors, or boards.

While all this is true, we still need to understand inductance in a more fundamental way. Our understanding of inductance needs to feed our intuition and give us the tools we need to solve real interconnect problems. A practical approach to inductance is based on just three fundamental principles.

6.2 Inductance Principle #1: There Are Circular Rings of Magnetic-Field Lines Around All Currents

There is a new fundamental entity, called *magnetic-field lines*, that surrounds every current. If we have a straight wire, as shown in Figure 6-1 for example, and send a current of 1 Amp through it, there will be concentric circular magnetic-field lines in the shape of rings created around the wire. These rings exist up and down the length of the wire. Imagine walking along the wire and counting the specific number of field line rings that completely surround it; the farther from the surface of the current, the fewer the number of field line rings we will encounter. If we move far enough away from the surface, we will count very few magnetic-field line rings.

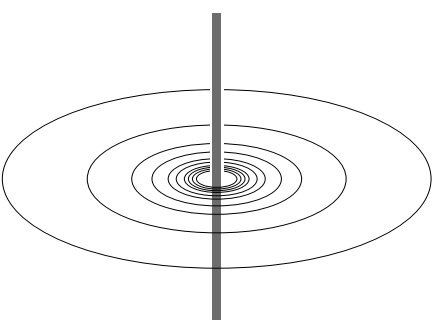


Figure 6-1 Some of the circular magnetic-field line rings around a current. The rings exist up and down the length of the wire.

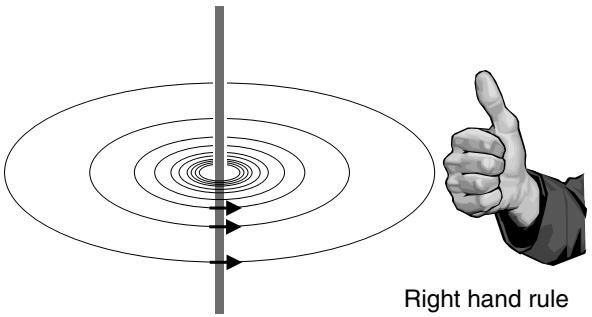


Figure 6-2 The direction of circulation of the magnetic-field line rings is based on the right-hand rule.

These field line rings have a specific direction, as though they are circulating around the wire. To determine their direction, use the familiar right-hand rule: Point the thumb of your right hand in the direction of the positive current and your fingers in the direction the field line rings circulate. This is illustrated in Figure 6-2.

TIP Magnetic-field line rings are always complete circles and always enclose some current. There must be some current encircled by the field line rings.

In what units do we count the field line rings? We count pens in units of gross. There are 144 pens in a gross. Paper is counted in units of reams, with 500 sheets to each ream. Apples are counted by the bushel. How many apples are there in a bushel? It's not clear exactly how many apples are in a bushel, but there is some number.

Likewise, we count the number of magnetic-field line rings around a current in units of Webers. Like the number of apples in a bushel, the number of magnetic-field line rings in a Weber of field lines is influenced by a number of factors. First is the amount of current in the conductor. If we double the current in the wire, we will double the number of Webers of magnetic-field line rings that are around it.

Second, the length of the wire around which the field line rings appear will affect the number of field line rings we count. The longer the wire length, the more field line rings there will be. Third, the cross section of the wire affects the total number of field line rings surrounding the current. This is a second-order effect and is more subtle. As we will see, if the cross-sectional area is increased,

for example, if the wire is made thicker, the number of field line rings will decrease a little bit.

Fourth, the presence of other currents nearby will affect the number of field line rings around the first current. A special current to watch for is the return current. As the return current is brought closer, some of its field line rings will be around the first current and will change the total number of field line rings. On the other hand, the presence of dielectric materials does not affect the number of field line rings around the current.

> **TIP** Magnetic fields do not interact with dielectric materials at all. There is no change in the number of magnetic-field line rings around a current if it is surrounded by Teflon or by barium titanate.

Fifth, the metal of which the wire is composed will affect the total number of magnetic-field lines around the current only if the conductor contains iron, nickel, or cobalt. These three metals are called the ferromagnetic metals. They, or alloys containing them, have a permeability greater than 1. If any magnetic-field line rings are completely contained within one of these metals, the metal will have the effect of dramatically increasing the number of field line rings. However, it is only those field line rings that are circulating internally in the conductor that will be affected. Two common interconnect alloys, Alloy 42 and Kovar, are ferromagnetic because they both contain iron, nickel, and cobalt.

The composition of a wire made of any other metal, such as copper, silver, tin, aluminum, gold, lead, or even carbon, will have absolutely no effect on the number of field line rings around the current.

6.3 Inductance Principle #2: Inductance Is the Number of Webers of Field Line Rings Around a Conductor per Amp of Current Through It

Inductance is fundamentally related to the number of magnetic-field line rings around a conductor, per Amp of current through it.

> **TIP** Inductance is about the number of rings of magnetic-field lines enclosing a current, not about the absolute value of the magnetic field at any one point. Don't worry about magnetic-field concentration, worry about the number of magnetic-field line rings.

The units we use to measure inductance are Webers of field line rings per Amp of current. One Weber/Amp is given the special name, Henry. The inductance of most interconnect structures is typically such a small fraction of a Henry, it is more common to use the units of nanoHenry. A nanoHenry, abbreviated as nH, is a measure of how many Webers of field line rings we would count around a conductor per Amp of current through it:

$$L = \frac{N}{I} \tag{6-2}$$

where:

L = the inductance, in Henrys

N = the number of magnetic-field line rings around the conductor, in Webers

I = the current through the conductor, in Amps

If the current through a conductor doubles, the number of field line rings doubles but the ratio stays the same. This ratio is completely independent of how much current is going through the conductor. A conductor has the same inductance if 0 Amps are flowing through it or 100 Amps. Yes, the number of field line rings changes, but the ratio doesn't, and inductance is the ratio.

TIP This means that inductance is really related to the geometry of the conductors. The only thing that influences inductance is the distribution of the conductors and in the case of ferromagnetic metals, their permeability.

In this respect, inductance is a measure of the efficiency of the conductors to create magnetic field line rings. A conductor that is not very efficient at generating magnetic field line rings has a low inductance. A conductor has an efficiency whether or not it has current through it. The amount of current in the conductor does not in any way affect its efficiency. Inductance is only about a conductor's geometry.

We can apply this simple definition to *all* cases involving inductance. What makes it complicated and confusing is having to keep track of how much of the current loop we are counting the field line rings around and which other currents are present, creating field line rings. This gives rise to many qualifiers for inductance.

To keep track of the source of the magnetic-field line rings, we will use terms of *self-inductance* and *mutual inductance*. To keep track of how much of the current loop around which we are counting field line rings, we will use the terms *loop inductance* and *partial inductance*. Finally, when referring to the magnetic-field line rings around just a section of an interconnect, while the current is flowing through the entire loop, we will use the terms *total inductance*, *net inductance*, or *effective inductance*. These last three terms are used interchangeably in the Industry.

Just using the term *inductance* is ambiguous. We must develop the discipline to always use the qualifier of the exact type of inductance to which we are referring. The most common source of confusion about inductance arises from mixing up the various types of inductance.

6.4 Self-Inductance and Mutual Inductance

If the only current that existed in the universe were the current in a single wire, the number of field line rings around it would be easy to count. However, when there are other currents nearby, their magnetic-field line rings can encircle multiple currents. Consider two adjacent wires, labeled a and b, as shown in Figure 6-3. If there were current in only one wire, a, it would have some number of field line rings around it and some inductance.

Suppose there were to be some current in the second wire, b. It would have some field line rings around it, and hence, it would have some inductance. And, some of the field line rings from this wire, b, would also encircle the first wire, a.

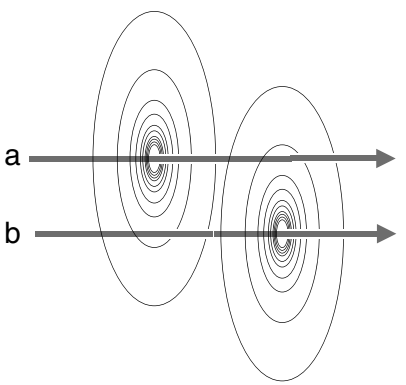


Figure 6-3 Magnetic-field line rings around one conductor can arise from its own currents and from another current.

Around the first wire, a, there are now some field line rings from its own current and some field line rings from the adjacent, second current, b.

When we count field line rings around one wire, we need a way of keeping track of the source of the field line rings. We do this by labeling those field line rings from a wire's own currents as the self-field line rings and those from an adjacent current's, the mutual-field line rings.

TIP The self-field line rings are those field line rings around a wire that arise from its own currents *only*. The mutual-field line rings are those magnetic-field line rings completely surrounding a wire that arise from another wire's current.

Every field line that is created by current in wire b and also goes around wire a must be around both of them. In this way, we say the mutual-field line rings "link" the two conductors.

If we have two adjacent wires and we put current in the second wire, it will have some of its field line rings around the first wire. As we might imagine, if we move the second wire farther away, the number of mutual-field line rings that are around both wires will decrease. Move them closer and the number of mutual-field line rings will increase.

But, what happens to the total number of field line rings around the first wire? If there is current through both wires, they will each have some self-field line rings. If the currents are in the same direction, the circulation direction of their self-field line rings will be the same. To count the total number of field line rings around the first wire, we would count its self-field line rings plus the mutual-field line rings from the second wire, because they are both circulating in the same direction.

However, if the direction of their currents is opposite, the circulation direction of the self- and mutual-field line rings around the first wire will be opposite. The mutual-field line rings will subtract from the self-field line rings. The total number of field line rings around the first wire will be decreased due to the presence of the adjacent opposite current.

Given this new perspective of keeping track of the source of the field line rings, we can make inductance more specific.

TIP We use the term *self-inductance* to refer to the number of field line rings around a wire, per Amp of current in its own wire. What we normally think of as *inductance* we now see is really specifically the self-inductance of a wire.

The self-inductance of a wire will be independent of the presence of another conductor's current. If we bring a second current near the first wire, the total number of field line rings around the first may change, but the number of field line rings from its own currents will not.

TIP Likewise, we use the term *mutual inductance* to refer to the number of field line rings around one wire, per Amp of current in another wire.

As we bring the two wires close together, their mutual inductance increases. Pull them farther apart and their mutual inductance decreases. Of course, the units we use to measure mutual inductance are also nH, since it is the ratio of a number of field line rings per Amp of current.

Mutual inductance has two very unusual and subtle properties. Mutual inductance is symmetric. Whether we send one Amp of current in one wire and count the number of field line rings around the other wire, or send one Amp of current through the other and count the field line rings it produces around the first, we get exactly the same ratio of field line rings per Amp of current. In this respect, mutual inductance is related to the field line rings that link two conductors and is tied to both of them equally. We sometimes refer to the "mutual inductance between two conductors," because it is a property shared equally by the two conductors.

This is true no matter what the size or shape of each individual wire. One can be a narrow strip and the other a wide plane. The number of field line rings around one conductor per Amp of current in the other conductor is the same whether we send the current in the wide conductor or the narrow one.

The second property is that the mutual inductance between any two conductors can never be greater than the self-inductance of either one. After all, every mutual-field line ring is coming from one conductor and must also be a self-field line ring around that conductor. Since the mutual inductance between two conductors is independent of which one has the source current, it must always be less than the smallest of the two conductors' self-inductance.

6.5 Inductance Principle #3: When the Number of Field Line Rings Around a Conductor Changes, There Will Be a Voltage Induced Across the Ends of the Conductor

So far, we have discussed what inductance is. Now we will look at the so what, why do we care about inductance? A special property of magnetic-field line rings is that when the actual, total number of field line rings around a section of a wire

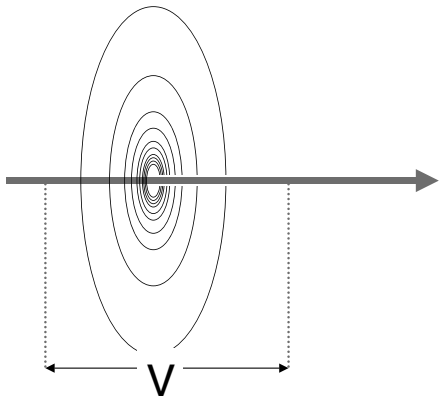


Figure 6-4 Voltage induced across a conductor due to the changing number of magnetic-field line rings around it.

changes, for whatever reason, there will be a voltage created across the length of the conductor. This is illustrated in Figure 6-4. The voltage created is directly related to how fast the total number of field line rings changes:

$$V = \frac{\Delta N}{\Delta t} \tag{6-3}$$

where:

V = the voltage induced across the ends of a conductor

ΔN = the number of field line rings that change

Δt = the time in which they change

If the current in a wire changes, the number of self-field line rings around it will change and there will be a voltage generated across the ends of the wire. The number of field line rings around the wire is $N = L \times I$, where L in this example is the self-inductance of the section of the wire. The voltage created, or induced, across a wire can be related to the inductance of the wire and how fast the current in it changes:

$$V = \frac{\Delta N}{\Delta t} = \frac{\Delta LI}{\Delta t} = L\frac{dI}{dt} \tag{6-4}$$

TIP Induced voltage is the fundamental reason why inductance plays such an important role in signal integrity. If there were no induced voltage when a current changed, there would be no impact from inductance on a signal. This induced voltage from a changing current gives rise to transmission line effects, discontinuities, cross talk, switching noise, rail collapse, ground bounce, and most sources of EMI.

This relationship is, after all, the definition of an inductor. If we change the current through an inductor, we get a voltage generated across it. The voltage polarity is created to drive an induced current that would counteract the change in current. This is why we say, "an inductor resists a change in current."

If we happened to have an adjacent current in another wire, near the first wire, some of the field line rings from this second wire may also go around the first wire. If the current in the second wire changes, the number of its field line rings around the first wire will change. This changing number of field line rings will cause a voltage to be created across the first wire. This is illustrated in Figure 6-5. These changes in the mutual-field line rings contribute to the induced voltage across the first wire. When it is another conductor in which the current changes, we typically use the term *cross talk* to describe the induced noise in the adjacent conductor. In this case, the voltage noise generated is:

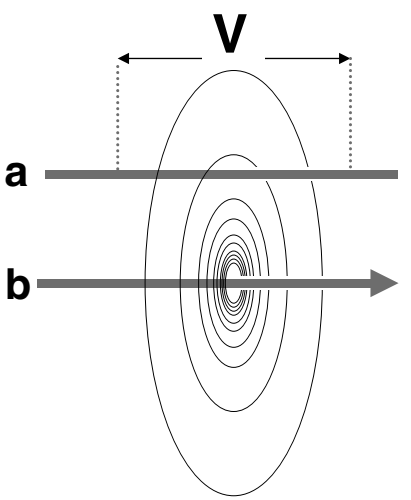


Figure 6-5 Induced voltage on one conductor due to a changing current in another and the subsequent changing mutual field lines between them are a form of cross talk.

$$V_{noise} = M\frac{dI}{dt} \tag{6-5}$$

where:

V_{noise} = the voltage noise induced in the first, quiet wire

M = the mutual inductance between the two wires, in Webers

I = the current in the second wire

Because the voltage induced depends on how fast the current changes, we sometimes use the term *switching noise* or *delta I noise* when describing the noise created when the current switches through an inductance.

To be able to analyze real-world problems that involve multiple conductors, we need to be able to keep track of all the various currents that are the sources of field line rings. The effects are the same; it's just more complicated when there are many conductors, each with possible currents and magnetic-field line rings.

6.6 Partial Inductance

Of course, real currents *only* flow in complete circuit rings. In the previous examples, we have been looking at just a part of a wire, where the only current that exists is the specific current in the part of wire we had drawn. As we counted field line rings, we ignored the rings from current in the rest of the current loop to which this wire segment belonged. When we look at the field line rings around part of a loop from currents in only part of the loop, we call this type of inductance the *partial inductance* of the wire.

It is important to keep in mind that when we speak of partial inductance, it is as though the *rest of the loop does not exist*. In the view of partial inductance, no other currents exist, except in the specific part of the conductor in which we are looking. The concept of partial inductance is a mathematical construct. It can never be measured, since an isolated current can never exist.

TIP In reality, we can never have a partial current—we must always have current loops. However, the concept of partial inductance is a very powerful tool to understand and calculate the other flavors of inductance, especially if we don't know what the rest of the loop looks like yet.

Partial inductance has two flavors: partial self-inductance and partial mutual inductance. What we have been discussing above has really been the partial

inductance of parts of two wires. More often than not, when referring to the inductance of a lead in a package, or a connector pin, or a surface trace, we are really referring to the partial self-inductance of this interconnect element.

The precise definition of partial self- and mutual inductance is based on a mathematical calculation of the number of field line rings around a section of a wire. Take a fixed-length section of conductor that might be part of a current loop. Rip it out of the loop so it is isolated in space but maintains its original geometry. On the ends, put large planes that are perpendicular to the length of the conductor. Now imagine injecting 1 A of current, appearing suddenly in one end of the wire, traveling through the wire, coming out the other end, and disappearing back into nothing when it exits.

The only current that exists in the universe is in the section of wire between the end planes. From this small section of current, count the number of field line rings that fit between the two end-cap planes. The number of field line rings counted, per the 1A of current in the wire section, is the partial self-inductance of that section of the conductor. Obviously, if we make the part of the wire longer, the total number of field line rings surrounding it will increase and its partial self-inductance will increase.

Now bring another short section of interconnect near this first section. Inject current from nowhere into the second wire and have it disappear out the opposite end. This partial current will create magnetic-field line rings throughout space, some of which will fit within the plane end caps of the first partial section of wire, completely enclosing the first wire segment. The number of field line rings around the first conductor segment, per Amp of current in the second wire, is the partial mutual inductance between the two sections.

Obviously, in the real world, we can't create a current into the end of the wire without it coming from some other part of a circuit. However, we can perform this operation mathematically. This term, *partial inductance*, is a very well-defined quantity, it just can't be measured. As we shall see, it is a very powerful concept to facilitate optimizing the design for reduced ground bounce and calculating the other, measurable terms of inductance.

There are only a few geometries of conductors with a reasonably good approximation for their partial self-inductance. The partial self-inductance of a straight, round conductor, illustrated in Figure 6-6, can be calculated to better than a few percent accuracy using a simple approximation. It is given by:

$$L = 5d\left\{\ln\left(\frac{2d}{r}\right) - \frac{3}{4}\right\} \tag{6-6}$$

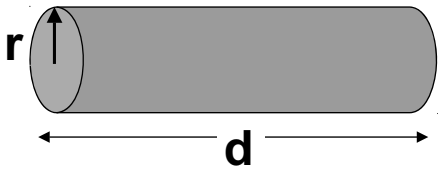


Figure 6-6 Geometry for the approximation of the partial self-inductance of a round rod.

where:

L = partial self-inductance of the wire, in nH

r = radius of the wire, in inches

d = length of the wire in inches

For example, an engineering change wire is typically 30-gauge wire, or roughly 10 mils in diameter. For a length that is 1 inch long, the partial self- inductance is:

$$L = 5 \times 1\left\{\ln\left(\frac{2 \times 1}{0.005}\right) - \frac{3}{4}\right\} = 26\text{nH} \tag{6-7}$$

TIP This gives rise to an important rule of thumb: The partial self-inductance of a wire is about 25 nH/inch or 1 nH/mm. Always keep in mind, this is a rule of thumb and sacrifices accuracy for ease of use.

We see that the partial self-inductance increases as we increase the length of the conductor. But, surprisingly, it increases faster than just linearly. If we double the length of the conductor, the partial self-inductance increases by more than a factor of two. This is because as we increase the length of the wire, there are more field line rings around this newly created section of the wire from the current in the new section, and some field line rings from the current in the other part of the wire are also around this new section.

The partial self-inductance decreases as the cross-sectional area increases. If we make the radius of the wire larger, the current will spread out more and the partial self-inductance will decrease. As we spread out the current distribution, the total number of field line rings will decrease.

TIP This points out a very important property of partial self-inductance: The more spread out the current distribution, the lower the partial self-inductance. The more dense we make the current distribution, the higher the partial self-inductance.

In this geometry of a round rod, the partial self-inductance varies only with the natural log of the radius, so it is only weakly dependent on the cross-sectional area. Other cross sections, such as wide planes, will have a partial self-inductance that is more sensitive to spreading out the current distribution.

Using the rule of thumb above, we can estimate the partial self-inductance of a number of interconnects. A surface trace from a capacitor to a via, 50 mils long, has a partial self-inductance of about 25 nH/inch × 0.05 inch = 1.2 nH. A via through a board, 064 mils thick, has a partial self-inductance of about 25 nH/inch × 0.064 inch = 1.6 nH.

Both the approximation and the rule of thumb are very good estimates of the partial self-inductance of a narrow rod. Figure 6-7 compares the estimated partial self-inductance for a 1-mil-diameter wire bond, based on the rule of thumb and the approximation given above, with the calculation from a 3D field solver. For lengths of typical wire bonds, about 100 mils long, the agreement is very good.

The partial mutual inductance between two conductor segments is the number of field line rings from one conductor that completely surrounds the other conductor's segment. In general, the partial mutual inductance between two wires is a

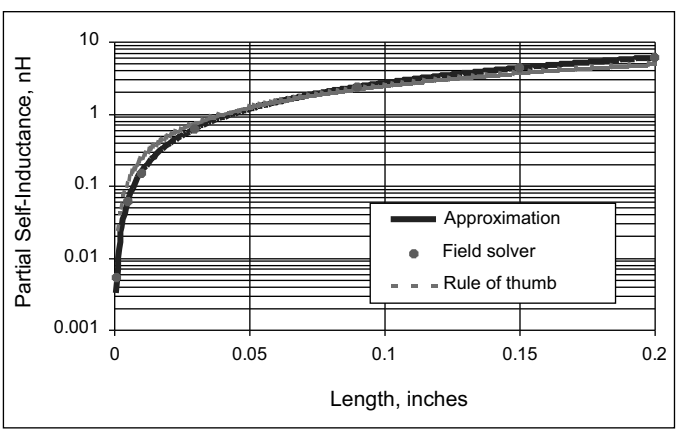


Figure 6-7 Partial self-inductance of a round rod, 1 mil in diameter, comparing the rule of thumb, the approximation, and the results from the Ansoft Q3D field solver.

small fraction of the partial self-inductance of either one and drops off very rapidly as the wires are pulled apart. The partial mutual inductance between two straight, round wires can be approximated by:

$$M = 5d\left\{\ln\left(\frac{2d}{s}\right) - 1 + \frac{s}{d} - \left(\frac{s}{2d}\right)^2\right\} \tag{6-8}$$

where:

M = partial mutual inductance, in nH

d = length of the two rods, in inches

s = center-to-center separation, in inches

This formidable approximation takes into account second-order effects and is often referred to as a second-order model. It can be simplified by a further approximation when the separation is small compared to the length of the rods ($s << d$) as:

$$M = 5d\left\{\ln\left(\frac{2d}{s}\right) - 1\right\} \tag{6-9}$$

This is a first-order model, as it ignores some of the details of the long-range coupling between the rods. What it sacrifices in accuracy is made up for in being slightly easier to use. Figure 6-8 shows the predictions for the mutual inductance between two rods, as the rods are moved farther apart, and how they compare with the predictions from a 3D field solver. We see that when the partial mutual inductance is greater than about 20% of the partial self-inductance (i.e., when mutual inductance is significant), the first-order approximation is pretty good. It is a good, practical approximation.

For example, two wire bonds may each have a partial self-inductance of 2.5 nH, if they are 100 mils long. If they are on a 5-mil pitch, they will have a partial mutual inductance of 1.3 nH. This means that if there were 1A of current in one wire bond, there would be 1.3 nH × 1 Amp = 1.3 nanoWebers of field line rings around the second wire bond. The ratio of the partial mutual inductance between the two wire bonds to the partial self-inductance of either one is about 50% with this separation.

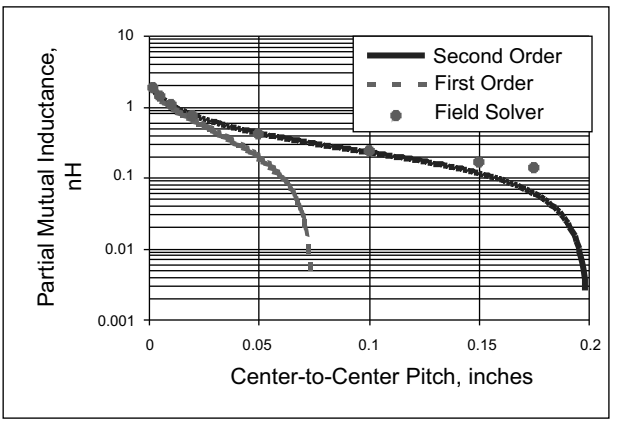


Figure 6-8 Partial mutual inductance between two round rods, 0.1 inch long, as their center-to-center separation increases, comparing the accurate approximation, the simplified approximation, and the result from Ansoft's Q3D field solver.

TIP From the graph of the mutual inductance and spacing, a good rule of thumb can be identified: If the spacing between two conductor segments is farther apart than their length, their partial mutual inductance is less than 10% of the partial self-inductance of either one and can often be ignored.

This says, the coupling between two sections of an interconnect are not important if they are farther apart than their length. For example, two vias 20 mils long have virtually no coupling between them if they are spaced more than 20 mils apart, center to center.

The concept of partial inductance is really the fundamental basis for all aspects of inductance. All the other forms of inductance can be described in terms of partial inductance. Package and connector models are really based on partial inductances. The output result of 3D static field solvers, when they calculate inductance values, is really in partial-inductance terms. SPICE models really use partial-inductance terms.

TIP If we can identify the performance design goals in terms of how we need to optimize each type of inductance, we can use our understanding of how the physical design will affect the partial self- and mutual inductances of a collection of conductors to optimize the physical design.

6.7 Effective, Total, or Net Inductance and Ground Bounce

Consider a wire that is straight for some length and then loops back on itself, as shown in Figure 6-9, making a complete loop. This sort of configuration is very common for all interconnects, including signal and return paths and power paths and their "ground"-return paths. For example, it is common to have adjacent power- and ground-return wire bonds in a package. The pair could be adjacent signal and return leads in an IC package, or they could be an adjacent signal and a return plane pair in a circuit board.

When current flows in the loop, magnetic-field line rings are created from each of the two legs. If the current in the loop changes, the number of field line rings around each half of the wire would change. Likewise, there would be a voltage created across each leg that would depend on how fast the total number of field line rings around each leg was changing.

The voltage noise created across one leg of the current loop depends on how fast the *total* number of field line rings around the leg changes, while it is part of this complete current loop. The total number of field line rings around one leg arises from the current in that leg (the partial self-field line rings) and the field line rings coming from the other leg (the partial mutual-field line rings). But, the field line rings from the two currents circulate around the leg in opposite directions, so the total number of field line rings around this section of the loop is the difference between the self- and mutual-field line rings around it. The total number of field line rings per Amp of current around this leg is given the special name *effective inductance*, *total inductance*, or *net inductance*.

TIP Effective, net, or total inductance of a section of a loop is the total number of field line rings around just this section, per Amp of current in the loop. This includes the contribution of field line rings from *all* the current segments in the complete loop.

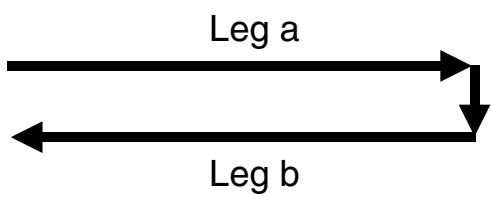


Figure 6-9 A current loop with two legs: an initial current and its return current.

We can calculate the effective inductance of one leg, based on the partial inductances of the two legs. Each of the two legs of the loop has a partial self-inductance associated with it, which we label as L_a and L_b. There is a partial mutual inductance between the two legs, which we label as L_{ab}. We label the current in the loop, I, which is the same current in each leg, but of course, going in opposite directions.

If we could distinguish the separate sources of the field line rings around leg b, we would see some of the field line rings come from the current in leg b and are self-field lines. Around leg b, there would be $N_b = I \times L_b$ field line rings from its own current. At the same time, some of the field line rings around leg b are mutual field line rings and come from the current in leg a. The number of mutual field line rings coming from leg a, surrounding leg b is $N_{ab} = L_{ab} \times I$.

What is the total number of field line rings around leg b? Since the current in leg a is moving in the opposite direction as leg b, the field line rings from leg a, around leg b, will be in the opposite direction as the field line rings from the current in leg b. When we count the total number of field line rings around leg b, we will be subtracting both sets of field line rings. The total number of field line rings around leg b is:

$$N_{total} = N_b - N_{ab} = (L_b - L_{ab}) \times I \tag{6-10}$$

We call $(L_b - L_{ab})$ the total, net, or effective inductance of leg b. It is the total number of field line rings around leg b, per Amp of current in the loop, including the effects of all the current segments in the entire loop. When the adjacent current is in the opposite direction, as when the two legs are part of the same loop and one leg has the return current for the other leg, the effective inductance will determine how much voltage is created across one leg when the current in the loop changes. When the second leg is the return path, we call this voltage generated across the return path *ground bounce*.

The ground-bounce voltage drop across the return path is:

$$V_{gb} = L_{total} \times \frac{dI}{dt} = (L_b - L_{ab}) \times \frac{dI}{dt} \tag{6-11}$$

where:

V_{gb} = the ground-bounce voltage

L_{total} = the total inductance of just the return path

I = the current in the loop

L_b = the partial self-inductance of the return-path leg

L_{ab} = the partial mutual inductance between the return path and the initial path

If the goal is to minimize the voltage drop in the return path (i.e., the ground-bounce voltage), there are only two approaches. First, we can do everything possible to decrease the rate of current change in the loop. This means slow down the edges, limit the number of signal paths that use the same, shared return path, and use differential signaling for the signals. Rarely do we have the luxury of affecting these terms much. However, we should always ask.

Secondly, we must find every way possible to decrease L_{total}. There are only two knobs to tweak to decrease the total inductance of the return path: *decrease* the partial self-inductance of the leg and *increase* the partial mutual inductance between the two legs. Decreasing the partial self-inductance of the leg means making the return path as short and wide as possible (i.e., use planes). Increasing the mutual inductance between the return path and the initial path means doing everything possible to bring the first leg and its return path as close together as possible.

TIP Ground bounce is the voltage between two points in the return path due to a changing current in a loop. Ground bounce is the primary cause of switching noise and EMI. It is primarily related to the total inductance of the return path and shared return current paths. To decrease ground-bounce voltage noise, there are two significant features to change: Decrease the partial self-inductance of the return path by using short lengths and wide interconnects, and increase the mutual inductance of the two legs by bringing the current and its return path closer together.

Surprisingly, decreasing ground bounce on the return path requires more than just doing something to the return path. It also requires thought about the placement of the signal current path and the resulting partial mutual inductance with the return path.

We can evaluate how much the total inductance of one wire bond can be reduced by bringing an adjacent wire bond closer, using the approximations above. Suppose one wire bond carries the power current and the other carries the ground-return current. They would have equal and opposite currents. In this case,

the partial mutual inductance between them will act to decrease the total inductance of one of them: $L_{total} = L_a - L_{ab}$. The closer we bring the wires, the greater the partial mutual inductance between them and the greater the reduction in the total inductance of one of the wire bonds.

If each wire bond were 1 mil in diameter and 100 mils in length, the partial self-inductance of each would be about 2.5 nH. Using the approximation above for the partial self-inductance and the partial mutual inductance, we can estimate the net or effective inductance of one wire bond as we change the center-to-center spacing, s.

When the spacing is greater than 100 mils, the partial mutual inductance is less than 10% of the partial self-inductance. The effective inductance of one wire bond is nearly the same as its isolated partial self-inductance. But, when we bring them as close as 5 mils center-to-center pitch, the mutual inductance increases considerably and we can reduce the effective inductance of one wire bond down to 1.3 nH. This is a reduction of more than 50%. The lower the effective inductance, the lower the voltage drop across this wire bond and the less ground-bounce voltage noise the chip will experience.

If the effective inductance of one wire bond, when the other current is far away, is 2.5 nH and there is 100 mA of current that switches in 1 nsec (typical of what goes into a transmission line), the ground-bounce voltage generated across the wire bond is $V_{gb} = 2.5\ nH \times 0.1A/1\ nsec = 250\ mV$. This is a lot of voltage noise. When the two wire bonds are routed close together, with a center pitch of 5 mils, the ground-bounce noise is reduced to $V_{gb} = 1.3\ nH \times 0.1A/1\ nsec = 130\ mV$—considerably less.

TIP This demonstrates the very important design rule: To decrease effective inductance, bring the return current as close as possible to the signal current.

Suppose we have the opposite case. Suppose that both wires will be carrying power current. This is often the case in many IC packages, because multiple leads are used to carry power and carry the ground return. If we look at the net inductance of one of the power wires and bring an adjacent power wire nearby, what happens?

In this case, both currents are in the same direction. The mutual-field line rings will be in the same direction as the self-field line rings. The mutual-field line rings from the second wire will add its field line rings to the field line rings

already around the first wire. The net inductance of one of the power wires will be $L_{total} = L_a + L_{ab}$.

If the goal is to minimize the net inductance of the power lead, our design goal, as always, is to do everything possible to decrease the partial self-inductance of the lead. However, in this case, because the field line rings from the adjacent wire are in the same direction, we must do everything possible to *decrease* their mutual inductance. This means space the wires out as much as possible.

We can estimate the net or effective inductance of one wire bond when the adjacent one carries the same current, for the two cases of current in the same direction or in opposite directions, as we pull them farther apart. This is plotted in Figure 6-10.

As long as the two wires are farther apart than their length, the net inductance is not much different from the partial self-inductance of one. As they are brought closer together, the net inductance decreases when the currents are in opposite directions, but increases when the currents are in the same direction.

TIP A good general design rule to minimize the net inductance of either leg in the power-distribution system is to keep similar parallel currents as far apart as their lengths.

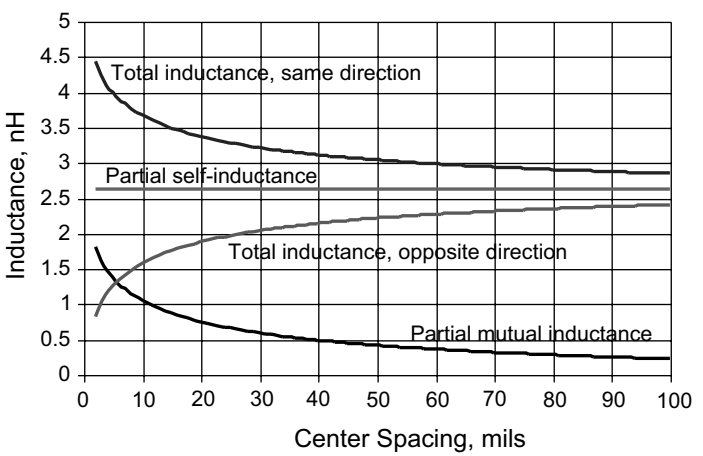


Figure 6-10 Total inductance of one 100-mil-long wire bond when an adjacent wire bond carries the same current, for the two cases of the currents in the same direction and in opposite directions. The wires are pulled apart, comparing the total inductances with the partial self- and mutual inductances.

In other words, if there are two adjacent wire bonds, each 100 mils long and both carrying power, they should be at least 100 mils center-to-center. Any closer and their mutual inductance will increase the effective inductance of each leg and increase the switching noise on one of the wires. This is not to say that there is no benefit for parallel currents in close proximity. It's just that their effectiveness is reduced over the maximum possible gain.

A common practice in high-power chips is double bonding, that is, using two wire bonds between the same pad on the die and the same pad on the package. The series resistance between the two pads will be decreased because of the two wires in parallel, and the equivalent inductance of the two wire bonds will be reduced compared to that when using only one wire bond. The closer the wires, the higher the mutual inductance, and the larger the effective inductance. But, since there are two in parallel, the equivalent inductance is half the total inductance of either one.

When double bonding, the wire bond loops should be created to keep the wires as far apart as possible. If they are 50 mils long and can be kept 5 mils apart, the partial self-inductance of either wire bond would be about 1.25 nH, and their partial mutual inductance would be about 0.5 nH. Their effective inductance would be 1.75 nH. With the two in parallel, their equivalent inductance would be $1/2 \times 1.75$ nH = 0.88 nH. This is reduced from the 1.25 nH expected with just one wire bond. Double bonding can, in fact, decrease the equivalent inductance between two pads.

As another example, consider the vias from a decoupling capacitor's pads to the power- and ground-return planes below, as illustrated in Figure 6-11. Suppose the distance to the plane below is 20 mils, and the vias are 10 mils in diameter. Is there any advantage in using multiple vias in parallel from each pad of the capacitor?

If the center-to-center spacing, s, between the vias is greater than the length of a via, 20 mils, their partial mutual inductance will be very small and they will not interact. The net inductance of either one will be just its partial self-inductance. Having multiple vias in parallel from one pad to the plane below will reduce the equivalent inductance to the plane below inversely with the number of vias. The more vias there are in parallel, the lower the equivalent inductance. In the previous figure, this means s_2 must be at least as large roughly as the distance to the planes, 20 mils. Likewise, if it is possible to bring vias with opposite currents closer together than their length, the effective inductance of each via will be reduced. If it is possible to place the vias with s_1 less than 20 mils apart, the net inductance of each via will be reduced, the equivalent inductance from the pad to the plane below will be reduced, and the rail-collapse voltage will be reduced.

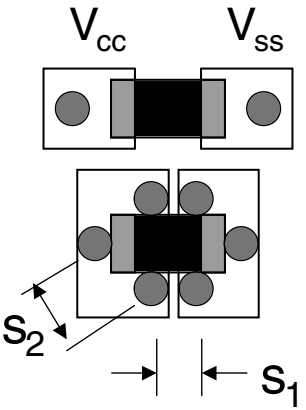


Figure 6-11 Via placement for decoupling capacitor pads between Vcc and Vss planes. Top: conventional placement. Bottom: optimized for low total inductance and lowest voltage-collapse noise: s_2 > via length, s_1 < via length.

TIP This suggests the following design rule for minimizing the total inductance of either path: Keep the center-to-center spacing between vias of the same current direction at least as far apart as the length of the via; keep the center-to-center spacing between vias with opposite direction current much closer than the length of the vias.

6.8 Loop Self- and Mutual Inductance

The general definition of inductance is the number of field line rings around a conductor per Amp of current through it. In the real world, current always flows in complete loops. When we measure the total inductance of the complete current loop, we call it the *loop inductance*. This loop inductance is really the self-inductance of the entire current loop, or the loop self-inductance.

TIP The loop self-inductance of a current loop is the total number of field line rings surrounding the entire loop, per Amp of current in the loop. With 1 Amp in the loop, we start at one end of the loop, walk along the wire, and count the total number of field line rings we encounter from all the current in the loop. This includes the effect of the current distribution of each section of the wire.

Let's look at the loop self-inductance of the wire loop with two straight legs as in Figure 6-9 on page 167. Leg a is like a signal path and leg b is like a return path. As we walk along leg a counting field line rings, we see the field line rings coming from the current in leg a (the partial self-inductance of leg a), and we see the field line rings around leg a coming from the current in leg b (or the partial mutual inductance between leg a and leg b).

Moving down leg a, the total number of field line rings we count is really the total inductance of leg a. When we move down leg b and count the total number of field line rings around leg b, we count the total inductance of leg b. The combination of these two is the loop self-inductance of the entire loop:

$$L_{loop} = L_a - L_{ab} + L_b - L_{ab} = L_a + L_b - 2L_{ab} \quad (6\text{-}12)$$

where:

L_{loop} = the loop self-inductance of the twin-lead loop

L_a = partial self-inductance of leg a

L_b = partial self-inductance of leg b

L_{ab} = partial mutual inductance between legs a and b

This may look familiar, as it appears in many textbooks. What is often not explicitly stated and what often makes inductance confusing is that the self- and mutual inductances in this relationship are really the partial self- and mutual inductances.

This relationship says that as we bring the two legs closer together, the loop inductance will decrease. The partial self-inductances will stay the same; it is the partial mutual inductance that increases. In this case, the larger mutual inductance between the two legs will act to decrease the total number of field line rings around each leg and to reduce the loop self-inductance.

TIP It is sometimes stated that the loop self-inductance depends on the "area of the loop." While this is roughly true in general, it doesn't help feed our intuition much. As we have seen, it is not so much the area that is important, it is the total number of field line rings encircling each leg that is important.

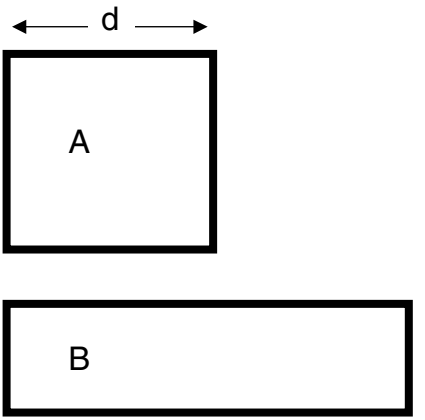


Figure 6-12 Two loops with exactly the same area but very different loop inductances. Bringing the return section of the loop closer to the other leg reduces the loop inductance by increasing the partial mutual inductance.

For example, in Figure 6-12 we show two different current loops, each with exactly the same area. They will have different loop inductances, since the partial mutual inductances are so different. The closer together we bring the legs with opposite current, the larger their partial mutual inductances and the smaller the resulting loop inductance. As we have seen, the underlying mechanism for decreasing loop self-inductance is the increase in the partial mutual inductance between the signal and return paths when the return path is brought closer to the other leg and the loop is made smaller.

There are three important special-case geometries that have good approximations for loop inductance: a circular loop; two long, parallel rods; and two wide planes.

For a circular loop, the loop inductance is given by

$$L_{loop} = 32 \times R \times \ln\left(\frac{4R}{D}\right) nH \tag{6-13}$$

where:

L_{loop} = loop inductance, in nH

R = radius of the loop, in inches

D = diameter of the wire making up the loop, in inches

For example, a 30-gauge wire, about 10 mils thick, that is bent in a circle with a 1-inch diameter, has a loop inductance of:

$$L_{loop} = 32 \times 0.5 \times \ln\left(\frac{4 \times 0.5}{0.01}\right) \text{nH} = 85 \text{ nH} \quad (6\text{-}14)$$

TIP This is a good rule of thumb to remember. Hold your index finger and thumb in a circle. A loop of 30-gauge wire this size has a loop inductance of about 85 nH.

The loop inductance is not really proportional to the area or to the circumference. It is proportional to the radius times ln (radius). The larger the circumference, the larger the partial self-inductance of each section, but also the farther away the opposite-direction currents in the loop and the lower their mutual inductance.

However, to first order, the loop inductance is roughly proportional to the radius. If the circumference is increased, the loop inductance will increase. For the 1-inch loop, the circumference is 1 inch × 3.14 or about 3.14 inches. This corresponds to a loop inductance per inch of circumference of 85 nH/3.14 inches ~ 25 nH/in. As a good rule of thumb, we see again that the loop inductance per length is about 25 nH/inch, for loops near 1 inch in diameter.

Figure 6-13 compares the loop inductance predicted by this approximation with the actual measured loop inductance of small copper wires. The accuracy is good to a few percent.

The loop inductance of two adjacent, straight-round wires, assuming one carries the return current of the other, is given by:

$$L_{loop} = 10 \times \text{len} \times \ln\left(\frac{s}{r}\right) \text{nH} \quad (6\text{-}15)$$

where:

L_{loop} = loop inductance, in nH

len = the length of the rods, in inches

r = radius of the rods, in mils

s = center-to-center separation of the rods, in mils

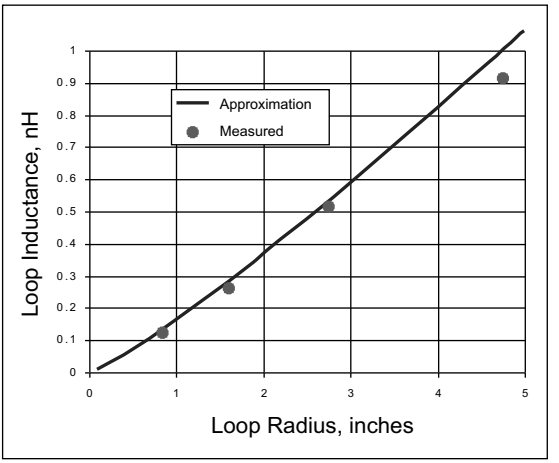


Figure 6-13 Comparison of the measured loop inductance of small loops made from 25-mil-thick wire with different loop radius and the predictions of the approximation. The approximation is seen to be good to a few percent.

For example, the loop inductance of two 1-mil-diameter wire bonds, 100 mils long and separated by 5 mils, is:

$$L_{loop} = 10 \times 0.1 \times \ln\left(\frac{5}{0.5}\right) \text{nH} = 2.3 \text{ nH} \tag{6-16}$$

This relationship points out that the loop inductance of two parallel wires is directly proportional to the length of the wires. It scales directly with the natural log of the separation. Increase the separation and the loop inductance increases, but only slowly, as the natural log of the separation.

The loop inductance of long, straight, parallel rods is directly proportional to the length of the rods. For example, if the rods represent the conductors in a ribbon cable, with a radius of 10 mils and center-to-center spacing of 50 mils, the loop inductance for adjacent wires carrying equal and opposite current is about 16 nH for a 1-inch-long section.

In the special case when the signal and return paths have a constant cross section down their length, the loop inductance scales directly with the length, and we refer to the loop inductance per length of the interconnect. Signal and return paths in ribbon cable have a constant loop inductance per length. In the example

above, the loop inductance per length of the ribbon-cable wires is about 16 nH/inch. In the case of two adjacent wire bonds, the loop inductance per length is 2.3 nH/0.1 inch, or 23 nH/inch.

As we shall see, any controlled-impedance interconnect will have a constant loop inductance per length.

6.9 The Power-Distribution Network (PDN) and Loop Inductance

When we think "signal integrity," we usually think about reflection problems and cross talk between signal nets. Though these are important problems, they represent only some of the signal-integrity problems. Another set of problems is due not to the signal paths, but to the power and ground paths. We call this the *power-distribution network* (PDN) and the design of the PDN we lump under the general heading of power integrity. Though power integrity is covered in more depth in a later chapter, some of the relevant aspects of the role inductance plays in the PDN are introduced here.

The purpose of the PDN is to deliver a constant voltage across the power and ground pads of each chip. Depending on the device technology, this voltage difference is typically 5 v to 0.8 v. Most noise budgets will allocate no more than 5% ripple on top of this. Isn't the regulator supposed to keep the voltage constant? If ripple is too high, why not just use a heftier regulator?

The expression "there's many a slip twixt cup and lip" pretty much summarizes the problem. Between the regulator and the chip are a lot of interconnects in the PDN—vias, planes, package leads, and wire bonds, for example. When current going to the chip changes (e.g., when the microcode causes more or fewer gates to switch, or at clock edges, where most gates tend to switch), the changing current, passing through the impedance of the PDN interconnects, will cause a voltage drop, called either *rail droop* or *rail collapse*.

To minimize this voltage drop from the changing current, the design goal is to keep the impedance of the PDN below a target value. There will still be changing current, but if the impedance can be kept low enough, the voltage drop across this impedance can be kept below the 5% allowed ripple.

TIP The impedance of the PDN is kept low by two design features: the addition of low-loop inductance decoupling capacitors to keep the impedance low at lower frequencies and minimized loop inductance between the decoupling capacitors and the chip's pads to keep the impedance low at higher frequencies.

How much decoupling capacitance is really needed? We can estimate roughly how much total capacitance is needed by assuming the decoupling capacitor provides all the charge that must flow for some time period, Δt.

During this time, the voltage across the capacitor, C, will drop as its charge is depleted by ΔQ, which has flowed through the chip. The voltage drop, ΔV, is given by:

$$\Delta V = \frac{\Delta Q}{C} \tag{6-17}$$

where:

ΔV = the change in voltage across the capacitor

ΔQ = the charge depletion from the capacitor

C = the capacitance

How much current, I, flows through the chip? Obviously this depends strongly on the specific chip and will change a lot depending on the code running through it. However, we can get a rough estimate by assuming the power dissipation of the chip, P, is related to the voltage, V, across it, and the average current, I, through it. Given the chip's average power dissipation, the average current through the chip is:

$$I = \frac{P}{V} \tag{6-18}$$

The total amount of decoupling needed for the chip to be able to decouple for a time, Δt, is related by:

$$\frac{P}{V} = \frac{\Delta Q}{\Delta t} = \frac{C\Delta V}{\Delta t} \tag{6-19}$$

From this relationship, the time a capacitor will decouple can be found as:

$$\Delta t = 0.05 \times C \times \frac{V^2}{P} \tag{6-20}$$

or the capacitance required to decouple for a given time can be found from:

$$C = \frac{1}{0.05} \times \frac{P}{V^2} \times \Delta t \tag{6-21}$$

where:

Δt = the time during which the charge flows from the capacitor, in sec

0.05 = the 5% voltage droop allowed

C = the capacitance of the decoupling capacitor, in Farads

V = the voltage of the rail, in Volts

P = the power dissipation of the chip, in watts

For example, if the chip power is 1 watt, typical for a memory chip or small ASIC running at 3.3 v and with a 5% ripple allowed, the amount of total decoupling capacitance needed is about:

$$C = \frac{1}{0.05} \times \frac{1}{3.3^2} \times \Delta t = 2 \times \Delta t \tag{6-22}$$

If the regulator cannot respond to a voltage change in less time than 10 microseconds, for example, then we need to provide at least 2 × 10 microsec = 20 microFarad capacitance for decoupling. Less than this and the voltage droop across the capacitor will exceed the 5% allowed ripple.

Why not just use a single 20-microFarad capacitor to provide all the decoupling required? The impedance of an ideal capacitor will decrease with increasing frequency. At first glance, it would seem that if a capacitor has low enough impedance (e.g., in the 1-MHz range) where the regulator cannot respond, then it should have even lower impedance at higher frequency.

Unfortunately, in real capacitors, there is a loop associated with the connection between the terminals of the capacitor and the rest of the connections to the pads on the chip. This loop inductance, in series with the ideal capacitance of the component, causes the impedance of a real capacitor to increase with increasing frequency.

Figure 6-14 is a plot of the measured impedance of an 0603 decoupling capacitor. This is the measured loop impedance between one end of the capacitor and the other, through a plane below the component. At low frequency, the imped-

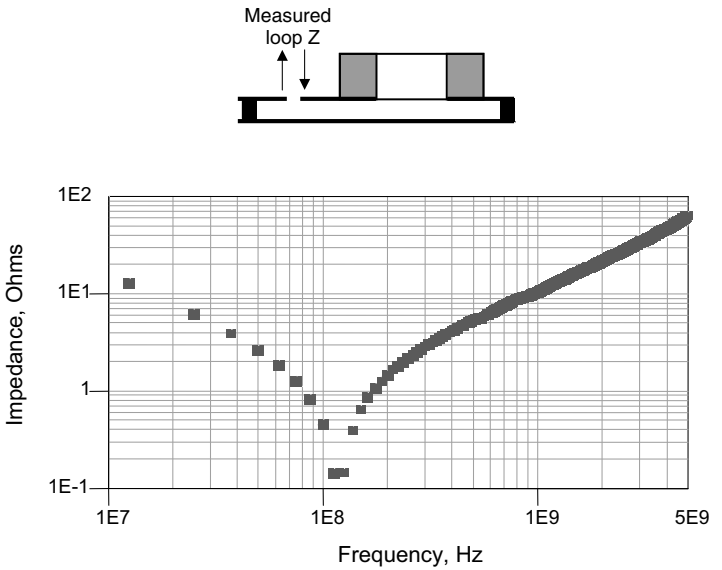


Figure 6-14 Measured loop impedance of a 1 nF 0603 decoupling capacitor, with a current loop configured as shown, measured with a GigaTest Labs Probe Station.

ance decreases, exactly as it does in an ideal capacitor. However, as the frequency increases, we reach a point where the series loop inductance of the capacitor begins to dominate the impedance. Above this frequency, called the *self-resonant frequency*, the impedance begins to increase. Above the self-resonant frequency, the impedance of the capacitor is completely independent of its capacitance. It is only related to its associated loop inductance. If we want to decrease the impedance of decoupling capacitors at the higher frequency end, we need to decrease their associated loop inductance, not increase the capacitance.

TIP A key feature of decoupling capacitors is that at high frequency, the impedance is solely related to their loop inductance, referred to as the equivalent series inductance (ESL). Decreasing the impedance of a decoupling capacitor at high frequency is all about decreasing the loop inductance of the complete path from the chip's pads to the decoupling capacitor.

The measured loop impedance of six 0603 decoupling capacitors with different values is shown in Figure 6-15. Their impedance at low frequency is radically different, since they have orders-of-magnitude different capacitance. However, at

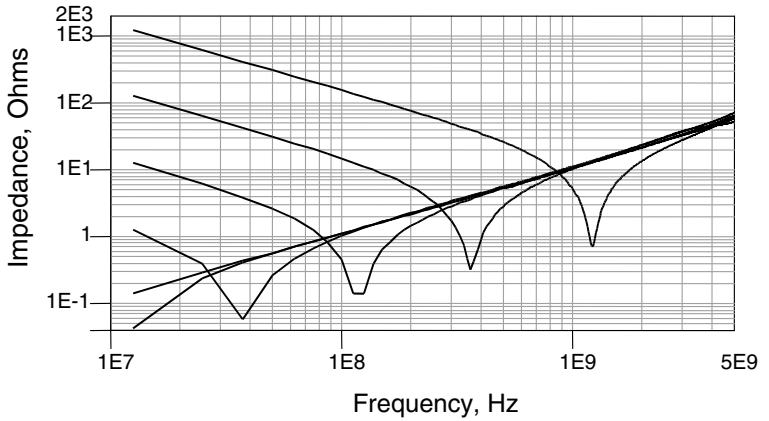


Figure 6-15 Measured loop impedance of six different 0603 capacitors with capacitances varying from 10 pF to 1 μF, but all with the same mounting geometry, measured with GigaTest Labs Probe Station.

high frequency, their impedances are identical because they have the same mounting geometry on the test board.

TIP The only way of decreasing the impedance of a decoupling capacitor at high frequency is by decreasing its loop self-inductance.

The best ways of decreasing the loop inductance of a decoupling capacitor are as follows:

1. Keep vias short by assigning the power and ground planes close to the surface.
2. Use small-body-size capacitors.
3. Use very short connections between the capacitor pads and the vias to the underlying planes.
4. Use multiple capacitors in parallel.

If the loop inductance associated with one decoupling capacitor and its mounting is 2 nH and the maximum allowed inductance is 0.1 nH, then there must be at least 20 capacitors in parallel for the equivalent loop inductance to meet the requirement.

From the decoupling capacitors to the chips' pads, the interconnect should be designed for the lowest loop inductance. In addition to short surface pads and short vias, planes are the interconnect geometry with the lowest loop inductance.

6.10 Loop Inductance per Square of Planes

The loop inductance for a current path going down one plane and back the other, as illustrated in Figure 6-16, depends on the partial self-inductance of each plane path and the partial mutual inductance between them. The wider the planes, the more spread out the current distribution, the lower the partial self-inductance of each plane, and the lower the loop inductance. The longer the planes, the larger their partial self-inductance and the larger the loop inductance. The closer we bring the planes, the larger their mutual inductance and the lower the loop inductance.

For the case of wide conductors, where the width, w, is much larger than their spacing, h, or w >> h, the loop inductance between two planes is to a very good approximation given by:

$$L_{loop} = \mu_0 h \frac{Len}{w} \tag{6-23}$$

where:

L_{loop} = the loop inductance, in nH

μ_0 = permeability of free space = 32 pH/mil

h = spacing between the planes, in mils

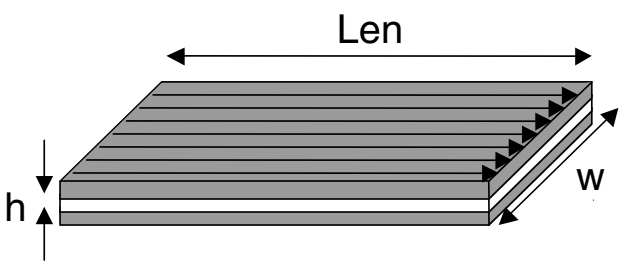


Figure 6-16 Geometrical configuration for the current flow in two planes forming a loop. The return current is in the opposite direction in the bottom plane.

Len = length of the planes, in mils

w = width of the planes, in mils

This assumes the current flows uniformly from one edge and back to the other edge.

If the section is in the shape of a square, the length equals the width, and the ratio is always 1, independent of the length of a side. It's startling that the loop inductance of a square section of a pair of planes that is 100 mils on a side is exactly the same as for a square section 1 inch on a side. Any square section of a pair of planes will have the same loop inductance. This is why we often use the term *loop inductance per square* of the planes, or the shortened version *inductance per square* or *sheet inductance* of the board. It makes life more confusing, but now we know this really refers to the loop inductance between two edges of a square section of planes, when the other edges are shorted together.

For the thinnest dielectric spacing currently in volume production, 2 mils, our approximation above estimates the loop inductance per square as about L_{loop} = 32 pH/mil × 2 mil = 64 pH. As the dielectric thickness increases, the loop inductance per square increases. A dielectric spacing of 5 mils has a loop inductance per square of L_{loop} = 32 pH/mil × 5 mil = 160 pH.

As the spacing between adjacent planes increases, the partial mutual inductance will decrease and there won't be as many field line rings from one plane around the other to cancel out the total number of field line rings. With greater dielectric spacing, the loop inductance increases and the rail-collapse noise will increase. This will make the PDN noise worse and will also increase the ground-bounce noise that drives common currents on external cables and causes EMI problems.

TIP Spacing the power and ground planes as close together as possible will decrease the loop inductance in the planes, decreasing rail collapse, ground bounce in the planes, and EMI.

6.11 Loop Inductance of Planes and Via Contacts

Current doesn't flow from one edge to another edge in planes. Between a discrete decoupling capacitor and the package leads, the connection to the planes is more like point contacts. In the above analysis we assumed the current was flowing uni-

formly down the plane. However, in actual practice, the current is not uniform. If the current is restricted due to point contacts, the loop inductance will increase.

The only reason we assumed the current was uniform was because this was the only case for which we had a simple approximation to help estimate the loop inductance. When balancing accuracy versus little effort to get an answer, we chose the path of little effort. The only way to get a better estimate of the loop inductance associated between two planes with real contacts is by using a 3D field solver.

We can gain some insight into how the geometry affects the associated loop inductance between two contact points between the planes. By using a 3D field solver, we can calculate the specific current distribution between the contact points and from the specific current distributions, the resulting loop inductance.

The good news about field solvers is that the accuracy can be very good, and they can include many real-world effects for which there are no good approximations. The bad news is that we can't generalize an answer from a field solver. It can run only one specific problem at a time.

As an example of the impact of the via contacts on the loop inductance between two planes, two special cases are compared. In both cases, a 1-inch-square section of two planes, 2 mils apart, was set up. In the first case, one edge of the top plane and the adjacent edge of the bottom plane were used as the current source and sink. The far ends of the planes were shorted together.

In the second case, two small via contacts were used as the current source and sink at one end, and a similar pair of contacts shorted the planes together at the other end. The contacts were 10 mils in diameter, spaced on 25-mil centers, similar to how a pair of vias would contact the planes in an actual board.

Figure 6-17 shows the current distribution in one of the planes for each case. When the edge contact is used, the current distribution is uniform, as expected. The loop inductance between the two planes is extracted as 62 pH. The approximation above, for the same geometry, was 64 pH. We see that for this special case, the approximation is pretty good.

When the current flows from one via contact, down the board to a second via contact, through the via to the bottom plane, back through the board and up through the end vias, as would be the case in a real board, the loop inductance extracted by the field solver is 252 pH. This is an increase by about a factor of four. The increase in loop inductance between the planes is due to the higher current density where the current is restricted to flow up the via. The more constricted the current flow, the higher the partial self-inductance and loop inductance. This

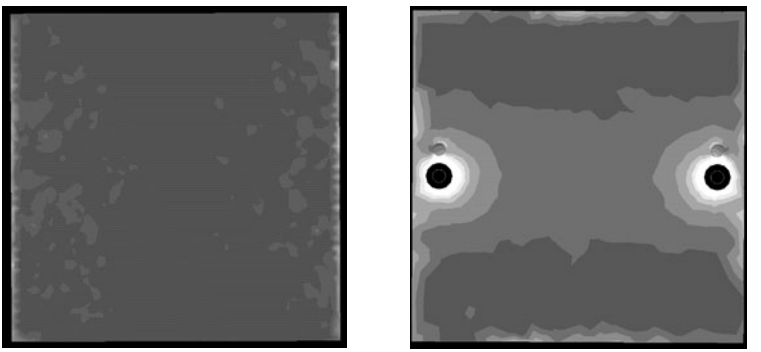


Figure 6-17 Current distribution in the top plane of a pair of planes with edge contact and a pair with point contacts. The lighter shade corresponds to higher current density. With edge contacts, the current distribution is uniform. With via point contacts, the current is crowded near the contact points. The higher current density creates higher inductance. Simulation with Ansoft's Q3D field solver, courtesy of Charles Grasso.

increase in loop inductance is sometimes referred to as *spreading inductance*. If the contact area were increased, the current density would decrease and the spreading inductance would decrease.

The loop inductance between the planes, even with the spreading inductance, will still scale with the plane-to-plane separation. The thinner the dielectric between the power and ground planes, the lower the sheet inductance and the lower the spreading inductance. Likewise, when the dielectric spacing between planes is large, the spreading inductance is large.

TIP The spreading inductance associated with the via contact points to a pair of power and ground planes is usually larger than the associated sheet loop inductance and must be taken into account to accurately estimate the loop inductance of the planes.

When many pairs of vias contribute current to the planes connecting many capacitors and many package leads, closely spaced planes will minimize the voltage drop from all the simultaneous dI/dt's.

The loop inductance in the planes associated with the decoupling capacitor, is dominated by the spreading inductance, not by the distance between the chip and the capacitor. To a first order, the total loop inductance of the decoupling capacitor is only weakly dependent on its proximity to the chip. However, the closer the capacitor is to the chip, the more the high-frequency power and return

currents will be confined to the proximity of the chip and the lower the ground-bounce voltage will be on the return plane.

> **TIP** By keeping decoupling capacitors close to the high-power chips, the high-frequency currents in the return plane can be localized to the chip and kept away from I/O regions of the board. This will minimize the ground-bounce voltage noise that might drive common currents on external cables and contribute to EMI.

6.12 Loop Inductance of Planes with a Field of Clearance Holes

A field solver is a useful tool to explore the impact of a field of via clearance holes on the loop inductance between two planes. Arrays of vias occur all the time. They happen under BGA packages, under connectors, and in very high-density regions of the board.

Often, there will be clearance holes in the power and ground planes of the via field. What impact will the holes have on the loop inductance of the planes? To a first order, we would expect the inductance to increase. But by how much? We often hear that an array of clearance holes under a package—the Swiss cheese effect—will dramatically increase the inductance of the planes. The only way to know by how much is to put in the numbers, that is, use a field solver.

Two identical pairs of planes were created, each 0.25 inch on a side and separated by 2 mils. Two via contacts were connected to each end of the planes. At one end, the vias were shorted together. At the other end, current was injected in one via and taken out from the other. This simulates one end connected to a decoupling capacitor and the other connected to the power and ground leads of a package through vias to the top surface.

In one pair of planes, a field of clearance holes was created in each plane. Each hole was 20 mils in diameter on 25-mil centers. This is an open area of about 50%. For each of the two cases, the current distribution was calculated and the loop inductance extracted with a 3D static field solver. Figure 6-18 shows the current distribution with and without the field of holes. The clearance holes constrict the current to the narrow channels between the holes, so we would expect to see the loop inductance increase.

The field solver calculates a loop inductance of 192 pH with no holes and 243 pH with this high density of holes. This is an increase of about 25% in loop inductance for an open area of 50%. We see that holes do increase the loop inductance of planes but not nearly as dramatically as we might have expected. To minimize the impact from the holes, it is important to make the holes as small as possible.

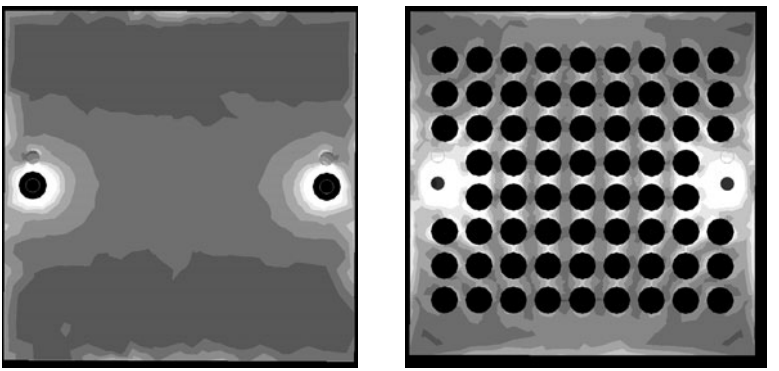


Figure 6-18 Current distribution on closely spaced planes with contact via points with and without a field of clearance holes. The lighter colors correspond to higher current densities. The holes cause the current to constrict, increasing the loop inductance. Simulated with Ansoft's Q3D field solver, courtesy of Charles Grasso.

Of course, bringing the planes closer together will also decrease the loop inductance of the planes, with or without clearance holes.

It is important to note that while a field of clearance holes in the power and ground planes will increase the loop inductance, the loop inductance can be kept well below a factor of two increase. It is not nearly as much of a disaster as commonly believed.

T I P The optimum power and ground interconnects, for lowest loop inductance, are planes, as wide as possible and as closely spaced as possible. When we use a very thin dielectric between the planes, we reduce the loop inductance between the bulk decoupling capacitors and the chip's pads. This will decrease the rail collapse and EMI.

Both of these problems, rail collapse and EMI, will get worse as rise times decrease. As we move into the future of higher and higher clock frequencies, thin dielectrics in the power-distribution network will play an increasingly important role.

6.13 Loop Mutual Inductance

If there are two independent current loops, there will be a mutual inductance between them. The loop mutual inductance is the number of field line rings from the current in one loop that completely surrounds the current of the second loop, per Amp of current in the first loop.

When current in one loop changes, it will change the number of field line rings around the second current loop and induce noise in the second current loop. The amount of noise created is:

$$V_{noise} = L_m \frac{dI}{dt} \tag{6-24}$$

where:

V_{noise} = the voltage noise induced on one loop

L_m = the loop mutual inductance between the two rings

dI/dt = how fast the current in the second loop changes

The noise in the quiet loop will happen only when there is a dI/dt in the active loop, which is only during the switching transitions. This is why this sort of noise is often called *switching noise*, *simultaneous switching noise (SSN)*, or *delta I noise*.

TIP The most important way of reducing switching noise is reducing the mutual inductance between the signal- and return path loops. This can be accomplished by moving the loops farther apart from each other. Since the mutual inductance between two loops can never be greater than the self-inductance of the smallest loop, another way of decreasing the loop mutual inductance is to decrease the loop self-inductance of both loops.

Loop mutual inductance also contributes to the cross talk between two uniform transmission lines and is discussed in a later chapter.

6.14 Equivalent Inductance of Multiple Inductors

So far, we have been considering just the partial inductance associated with a single interconnect element that has two terminals, one on either end, and then the resulting loop inductance consisting of the two elements in series. For two separate interconnect elements, there are two ways of being connected together: end to end (in series) or with each end together (in parallel). These two circuit configurations are illustrated in Figure 6-19.

When connected together, the resulting combination has two terminals, and there is an equivalent inductance for the combination. We are used to thinking

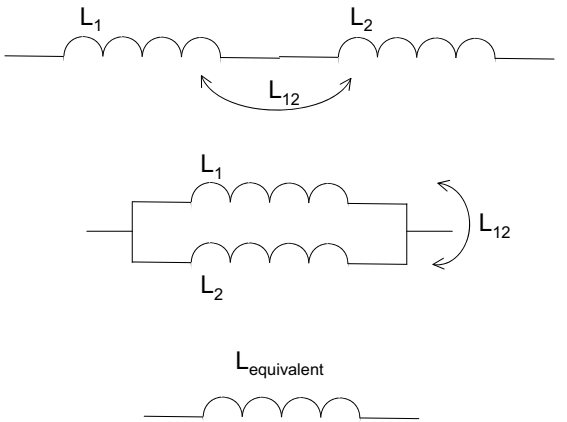


Figure 6-19 Circuit topologies for combining partial inductors in series (top) and in parallel (middle) into an equivalent inductance (bottom).

about the equivalent inductance of a series combination of two inductors simply as the sum of each individual partial self-inductance. But what about the impact from their mutual inductance?

The inclusion of the mutual inductance between the interconnect elements makes the equivalent inductance a bit more complicated. For the series combination of two partial inductances, the resulting equivalent partial self-inductance of the series combination is:

$$L_{series} = L_1 + L_2 + 2L_{12} \tag{6-25}$$

The equivalent partial self-inductance when the elements are connected in parallel is given by:

$$L_{parallel} = \frac{L_1 L_2 - L_{12}^2}{L_1 + L_2 - 2L_{12}} \tag{6-26}$$

where:

L_{series} = the equivalent partial self-inductance of the series combination

$L_{parallel}$ = the equivalent partial self-inductance of the parallel combination

L_1 = partial self-inductance of one element

L_2 = partial self-inductance of the other element

L_{12} = the partial mutual inductance between the two elements

When the partial mutual inductance is zero, and the partial self-inductances are the same, both of these reduce to the familiar expressions of the series combination. That is, they are just twice the self-inductance of one of them and the equivalent parallel inductance is just half the partial self-inductance of one of them.

In the special case when the partial self-inductance of the two inductors is the same, the series combination of the two inductors is just twice the sum of the self- and mutual inductance. For the parallel combination, when the two partial self-inductances are equal, the equivalent inductance is:

$$L_{parallel} = \frac{1}{2}(L + M) \tag{6-27}$$

where:

$L_{parallel}$ = the equivalent partial self-inductance of the parallel combination

L = the partial self-inductance of either element

M = the partial mutual inductance of either element

This says that if the goal is to reduce the equivalent inductance of two current paths in parallel, we do get a reduction, as long as the mutual inductance between the elements is kept small.

6.15 Summary of Inductance

All the various flavors of inductance are directly related to the number of magnetic-field line rings around a conductor per Amp of current. The importance of inductance is due to the induced voltage across a conductor when currents change. Just referring to an inductance can be ambiguous.

To be unambiguous, we need to specify the source of the currents, as in the self-inductance or mutual inductance. Then we need to specify whether we are referring to part of the circuit using partial inductance or the whole circuit using loop inductance. When we look at the voltage noise generated across part of the circuit, because this depends on all the field line rings and how they change, we need to specify the total inductance of just that section of the circuit. Finally, if we

have multiple inductors in some combination, such as multiple parallel leads in a package or vias in parallel, we need to use the equivalent inductance.

The greatest source of confusion arises when we misuse the term *inductance*. As long as we include the correct qualifier, we will never go wrong. The various flavors of inductance are listed here:

1. **Inductance:** The number of magnetic-field line rings around a conductor per Amp of current through it
2. **Self-inductance:** The number of field line rings around the conductor per Amp of current through the same conductor
3. **Mutual inductance:** The number of field line rings around a conductor per Amp of current through another conductor
4. **Loop inductance:** The total number of field line rings around the complete current loop per Amp of current
5. **Loop self-inductance:** The total number of field line rings around the complete current loop per Amp of current in the same loop
6. **Loop mutual inductance:** The total number of field line rings around a complete current loop per Amp of current in another loop
7. **Partial inductance:** The number of field line rings around a section of a conductor as though no other currents exist anywhere
8. **Partial self-inductance:** The number of field line rings around a section of a conductor per Amp of current in that section as though no other currents exist anywhere
9. **Partial mutual inductance:** The number of field line rings around a section of a conductor per Amp of current in another section as though no other currents exist anywhere
10. **Effective, net, or total inductance:** The total number of field line rings around a section of a conductor per Amp of current in the entire loop, taking into account the presence of field line rings from current in every part of the loop
11. **Equivalent inductance:** The single self-inductance corresponding to the series or parallel combination of multiple inductors including the effect of their mutual inductance

6.16 Current Distributions and Skin Depth

In evaluating the resistance and inductance of conductors, we have been assuming the currents were uniformly distributed through the conductors. While this is the case for DC currents, it is not always true when currents change. AC currents can have a radically different current distribution, which can dramatically affect the resistance and to some extent the inductance of conductors.

It is easiest to calculate the current distributions in the frequency domain where currents are sine waves. This is a situation where moving to the frequency domain gets us to an answer faster than staying in the time domain.

At DC frequency, the current distribution in a solid rod is uniform throughout the rod. When we counted magnetic-field line rings previously, we focused on the field line rings that were around the outside of the conductor. In fact, there are some magnetic-field line rings inside the conductor as well. They are also counted in the self-inductance. This is illustrated in Figure 6-20.

To distinguish the magnetic-field line rings contributing to the conductor's self-inductance that are enclosed inside the conductor from those that are outside the conductor, we separate the self-inductance into internal self-inductance and external self-inductance.

The internal magnetic-field line rings are the only field line rings that actually see the conductor metal and can be affected by the metal. For a round wire, the external-field line rings never see any conductor and they will never change with frequency. However, the internal-field line rings do see the conductor and they can change as the current distribution inside the conductor changes with frequency.

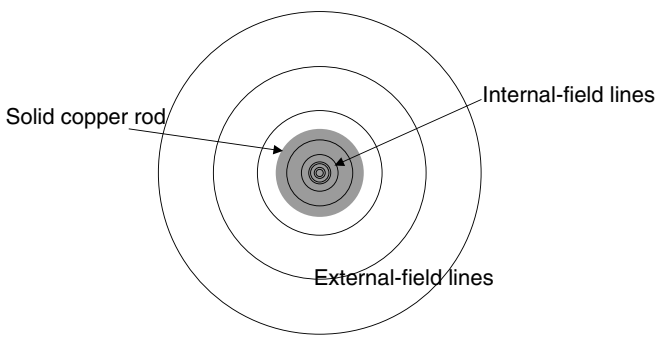


Figure 6-20 Magnetic-field line rings from the DC current in a uniform solid rod of copper. Some of the field line rings are internal and some are external.

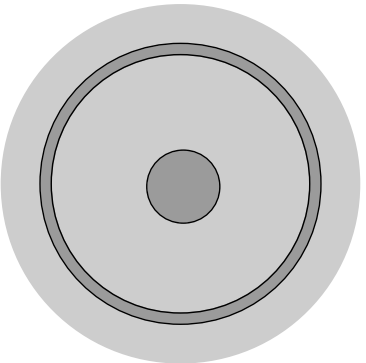


Figure 6-21 Two cylinders of current, traveling into the paper, singled out in a solid rod of copper, both of exactly the same cross-sectional area.

Consider two cylinders of current with exactly the same cross-sectional area in a solid rod of copper, as shown in Figure 6-21. If the cross-sectional area of each is exactly the same, each cylinder has the same current. Which cylinder has more field line rings around it?

The outer cylinder and the inner cylinder have exactly the same number of field line rings in the region outside the outer cylinder. The number of field line rings outside a current is only related to the amount of total current enclosed by the field line rings. There are no field line rings from currents in the outer cylinder inside this cylinder, since field line rings must encircle a current.

The current from the inner cylinder has more internal self-field line rings, since it has more distance from its current to the outer wall of the rod. The closer to the center of the rod the current is located, the greater the total number of field line rings around that current.

TIP Currents closer to the center of the rod will have more field line rings per Amp of current and a higher self-inductance than currents toward the outside of the conductor.

Now, we turn on the AC current. Currents are sine waves. Each frequency component will travel the path of lowest impedance. The current paths that have the highest inductance will have the highest impedance. As the sine-wave frequency increases, the impedance of the higher inductance paths get even larger. The higher the frequency, the greater the tendency for current to want to take the lower inductance path. This is the path toward the outer surface of the rod.

In general, the higher the frequency, the greater the tendency for current to travel on the outside surface of the conductor. At a given frequency, there will be some distribution of current from the center to the outside surface. This will depend on the relative amount of resistive impedance and inductive impedance. The higher the current density the higher the voltage drop from the resistive impedance. But the higher the frequency, the bigger the difference between the inductive impedance of the inner and outer paths. This balance means the current distribution changes with frequency, tending toward all the current in a thin layer around the outer surface, at high frequency.

TIP As the frequency of the sine waves of current through the rod increases, the current will redistribute so most of the current is taking the path of lowest impedance, that is, toward the outside of the conductor. At high frequency, it appears as though all the current is traveling in a thin shell at the conductor's surface.

There are only a few simple geometries where there is a good approximation for the actual current distribution of the current in a conductor. One of these is for a round cylinder. At each frequency, the current distribution will drop off exponentially toward the center of the conductor. This is illustrated in Figure 6-22.

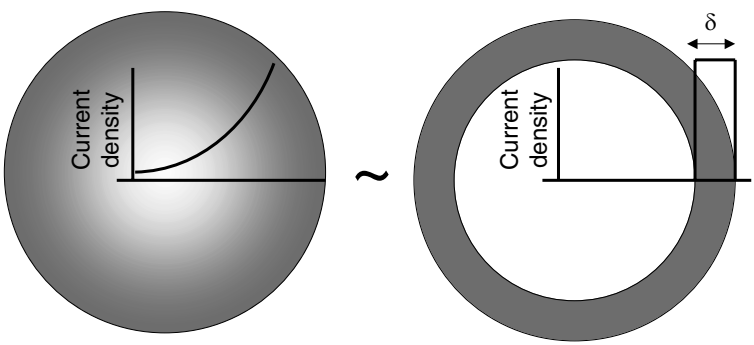


Figure 6-22 Left: Current distribution in a solid rod of copper at some frequency, showing the concentration of current near the outer surface. Darker color is high-current density. Right: Approximating the current in the rod in terms of a uniform current distribution with a thickness equal to the skin depth.

In this geometry, we can approximate the shell of current as a uniform distribution with a fixed thickness, a distance δ, from the outer surface. We call this equivalent thickness of the current shell, the skin depth. It depends on the frequency, the conductivity of the metal, and the permeability of the metal:

$$\delta = \sqrt{\frac{1}{\sigma \pi \mu_0 \mu_r f}} \tag{6-28}$$

where:

δ = skin depth, in meters

σ = conductivity of conductor, in Siemens/m

μ_0 = permeability of free space, $4 \times \pi \times 10^{-7}$ H/m

μ_r = relative permeability of the conductor

f = sine-wave frequency, in Hz

For the case of copper with a conductivity of 5.6×10^7 Siemens/m and relative permeability of 1, the skin depth is approximately:

$$\delta = 66 \text{ microns} \sqrt{\frac{1}{f}} \tag{6-29}$$

where:

δ = the skin depth, in microns

f = the sine-wave frequency, in MHz

At 1 MHz, the skin depth in copper is 66 microns. In Figure 6-23, the skin depth for copper is plotted and compared to the geometrical thickness of 1-ounce and 1/2-ounce copper. This points out that for 1-ounce copper traces, when the sine-wave frequency of the current is higher than about 10 MHz, the current distribution is determined by skin depth; not the geometrical cross section. Below 10 MHz, the current distribution will be uniform and independent of frequency. When the skin depth is thinner than the geometrical cross section, the current distribution, the resistance, and the loop inductance will be frequency dependent.

This is a handy rule of thumb to keep in mind. This means that if we have a board with 1-ounce copper traces or a geometrical thickness of 34 microns, cur-

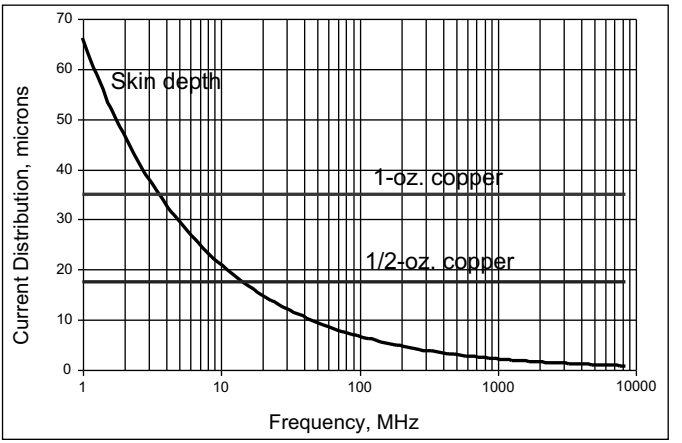


Figure 6-23 Current distribution in copper when skin depth is limited and compared with the geometrical thickness of 1-ounce and 1/2-ounce copper.

rent that is moving through the conductor at 10-MHz or higher sine-wave frequencies would not use the entire cross section of the trace. It would be dominated by these skin-depth effects.

In real interconnects, there is always a signal path and a return path. As the current loop propagates down the signal and return path, it is the loop self-inductance that influences the impedance the current sees. As frequency increases, the impedance from the loop self-inductance will increase and the current in both conductors will redistribute to take the path of lowest impedance or lowest loop self-inductance. What is the current distribution for lowest loop self-inductance?

There are two ways the loop self-inductance can be decreased: by spreading out the current within each conductor and by bringing the return current closer to the signal current. This will decrease the partial self-inductance of each conductor and increase the partial mutual inductance between them. Both effects will happen: Current will spread out within each conductor, and the current distribution in both conductors will redistribute so the two currents get closer to each other. The precise current distribution in each conductor is determined by the balance of two forces. Within each conductor, the current wants to spread apart to decrease its partial self-inductance. At the same time, the current in one conductor wants to get as close as possible to the current in the other conductor to increase their partial mutual inductance. The resulting current distribution can only be calculated with a 2D field solver. Figure 6-24 shows the simulated current distribution in a pair of

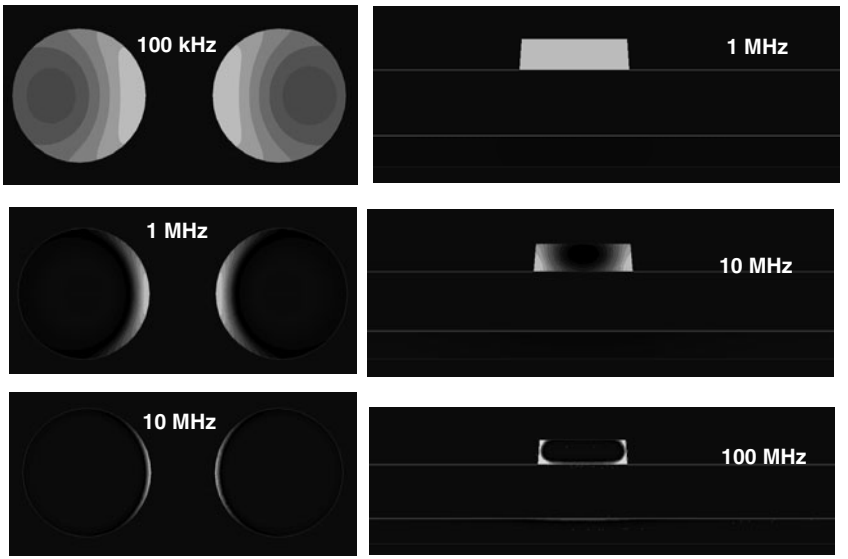


Figure 6-24 Current distribution in 20-mil-diameter wire at three different frequencies and 1-ounce copper microstrip. The lighter the color the higher the current density. Simulated with Ansoft's 2D Extractor.

20-mil (500-micron) diameter ribbon wires. At low frequency, the skin depth is large compared with the geometrical cross section and the current is uniformly distributed in each conductor. At 100 kHz, the skin depth in copper is about 10 mils (250 microns) and is comparable with the cross section, so it will begin to redistribute. At 1 MHz, the skin depth is 2.5 mils (66 microns), small compared to the diameter. The current distribution will be dominated by the skin depth. As the frequency increases, the current redistributes to minimize the loop impedance.

Also shown is the current distribution for a 1-ounce copper microstrip. At 1 MHz, the current is mostly uniformly distributed. At 10 MHz, it begins to redistribute. Above 10 MHz, the skin depth is much smaller than the cross section and will dominate the current distribution. In both examples, the current redistributes at higher frequency to minimize the impedance.

As the frequency increases, the bulk resistivity of the conductor does not change. For copper, the bulk resistivity doesn't begin to change until above 100 GHz. However, if the current travels through a thinner cross section due to skin-depth effects, the resistance of the interconnect will increase.

TIP In the skin-depth regime, when the skin depth is thinner than the geometrical cross section, as frequency increases, the cross-sectional area the current travels through decreases proportionally to the square root of the frequency. This will cause the resistance per length of the line to increase with the square root of the frequency as well.

Consider the case of a simple microstrip line, made from 1-ounce copper, 5 mils wide. The resistance per length of the signal path at DC frequency is:

$$R_{DC} = \frac{\rho}{wt} \tag{6-30}$$

where:

R_{DC} = the resistance per length for DC currents

ρ = the bulk resistivity of copper

w = the line width of the signal trace

t = the geometrical thickness of the signal trace

At frequencies above about 10 MHz, the current is skin-depth limited and the resistance will be frequency dependent. Above this frequency, the thickness of the conductor that is actually being used by the current is roughly the skin depth, so the high-frequency resistance is really:

$$R_{HF} = \frac{\rho}{w\delta} \tag{6-31}$$

where:

R_{HF} = the resistance per length for high-frequency currents

ρ = the bulk resistivity of copper

w = the line width of the signal trace

δ = the skin depth of copper at the high frequency

The ratio of resistance at high frequency to the resistance at DC is roughly $R_{HF}/R_{DC} = t/\delta$. At 1 GHz, where the skin depth of copper is 2 microns, the high-

frequency resistance will be 30 microns/2 microns = 15 × higher than the low-frequency resistance for 1-ounce copper. The series resistance of a signal trace will only get larger at higher frequencies.

Figure 6-25 shows the measured resistance of a small loop of 22-gauge copper wire, 25 mils in thickness. The loop was about 1 inch in diameter. The skin depth is comparable to the geometrical thickness at 10 kHz. At higher frequencies, the resistance should increase with roughly the square root of frequency.

One consequence of the frequency-dependent current distribution is the frequency-dependent resistance. In addition, the inductance will change. Since the driving force for the current redistribution is decreasing loop self-inductance, the loop self-inductance will decrease with higher frequency.

At DC frequency, the self-inductance of a wire will be composed of the external self-inductance and the internal self-inductance. The external self-inductance will not change as the current redistributes within the conductor, but the internal self-inductance will decrease as more current moves to the outside. At frequencies far above where the skin depth is comparable to the geometrical thickness, there will be very little current inside the conductor and there will be no internal self-inductance.

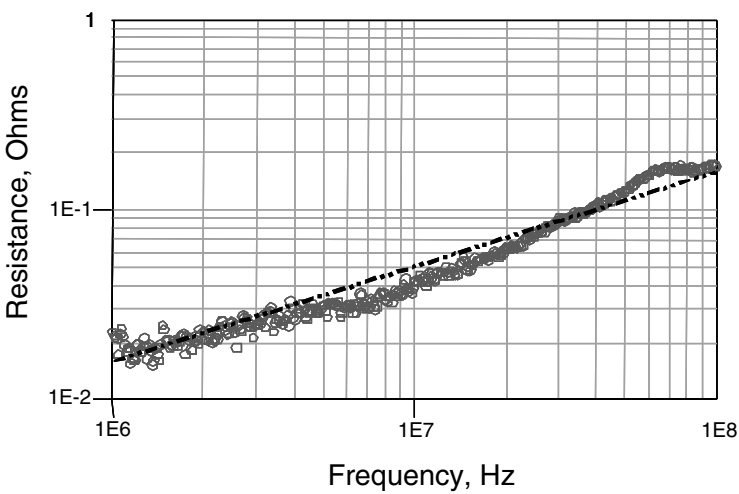


Figure 6-25 Measured series resistance of a 22-gauge copper wire loop, 1 inch in diameter, showing the resistance increasing with the square root of frequency. Circles are measured resistance; the line is resistance increasing with the square root of frequency.

We would expect the self-inductance of a wire to be frequency dependent. At low frequency, it will be $L_{internal} + L_{external}$. At high frequency, it should be simply $L_{external}$. The transition should start at about where the skin depth is comparable to the geometrical thickness and reach a steady value above a frequency where the skin depth is just a small fraction of the geometrical thickness.

The precise current distribution and the contribution of internal and external self-inductance are hard to estimate analytically, especially for rectangular cross sections. However, they can be calculated very easily using a 2D field solver.

Figure 6-26 shows an example of the loop self-inductance per length for a microstrip as the current redistributes itself. This illustrates that due to skin-depth effects, the low-frequency inductance is higher than the high-frequency inductance by an amount equal to the internal self-inductance. Above about 100 MHz, the current travels in a thin shell, and the inductance is constant with further increase in frequency.

T I P When we refer to the loop self-inductance of a microstrip, for example, we are more often referring to this high-frequency limit, assuming all the current is in the outside surface. *High frequency* refers to above the skin-depth limit, where current is close to the surface and not dependent on the geometrical thickness.

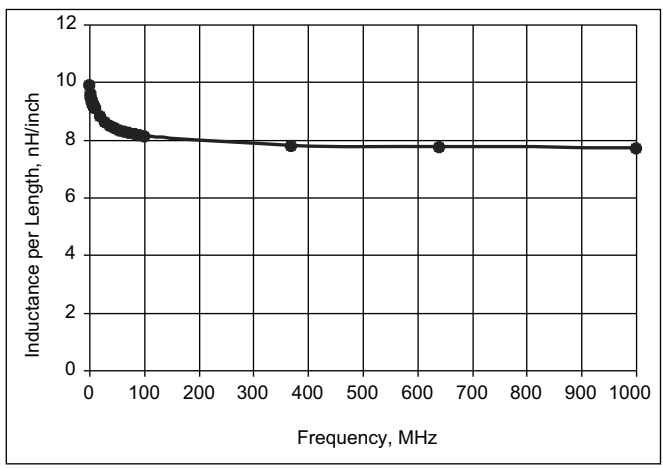


Figure 6-26 Loop self-inductance of a microstrip as the current redistributes due to skin-depth effects, calculated with Ansoft's 2D Extractor.

6.17 High-Permeability Materials

There is an important term affecting skin depth that applies to only a few special materials. This term is the *permeability of the conductor.* Permeability refers to how the conductor interacts with magnetic-field line rings. Most metals have a permeability of 1 and they do not interact with magnetic-field line rings.

However, when the permeability is greater than 1, the number of field line rings in the metal is amplified over what would be there if the permeability were 1. There are only three metals that have a permeability other than 1. These are the ferromagnetic metals: iron, nickel, and cobalt. Most alloys that contain any combination of these metals have a permeability much greater than 1. We are most familiar with ferrites, which usually contain iron and cobalt and can have permeabilities of over 1,000. Two important interconnect metals are ferromagnetic: Alloy 42 and Kovar. The permeability of these metals can be 100–500. This high permeability can have a significant impact on the frequency dependence of the resistance and inductance of an interconnect made from these materials.

For a ferromagnetic wire, at DC frequency, the self-inductance of the wire will be related to the internal and external self-inductance. All the field line rings that contribute to the external self-inductance see only air, which has a permeability of 1. The external self-inductance of a ferromagnetic wire will be exactly the same as if the wire were made of copper. After all, for the same current in the wire, there would be the same external-field line rings per amp of current.

However, the internal-field line rings in the ferromagnetic wire will see a high permeability, and these magnetic-field line rings will be amplified. At low frequency, the inductance of a ferromagnetic wire is very high, but above about 1 MHz, all the field line rings are external and the loop self-inductance is comparable to a copper loop of the same dimension.

> **TIP** The loop inductance that high-speed signals would encounter in a ferromagnetic conductor is comparable to the loop inductance if the conductor were made from copper, since above the skin-depth limit the loop inductance is composed almost solely of external-field line rings.

The skin depth for a ferromagnetic conductor can be much smaller than for a copper conductor, due to its high permeability. For example, nickel, having a bulk conductivity of about 1.4×10^7 Siemens/m and permeability of about 100, has a skin depth of roughly:

$$\delta = 13 \text{ microns} \sqrt{\frac{1}{f}} \tag{6-32}$$

At the same frequency, the cross section for current in a nickel conductor will be much thinner than for an equivalent-geometry copper conductor. In addition, the bulk resistivity is higher. This means the series resistance will be much higher. Figure 6-27 shows the measured series resistance of 1-inch-diameter rings of copper wire and nickel wire of roughly comparable cross section. The resistance of the nickel wire is more than 10 times higher than that of the copper wire and is clearly increasing with the square root of frequency, which is characteristic of a skin-depth-limited current distribution. This is why the high-frequency resistance of Kovar or Alloy 42 leads can be very high compared to nonferromagnetic leads.

Surface microstrip traces on circuit boards are often plated on their top surface with a nickel/gold layer to facilitate solder-attach of components. This nickel layer has virtually no impact on the electrical properties of the trace because it is on the other side of the trace from the return path. The current will travel the path of lowest impedance, which is not through the nickel layer. If the conductor were solid nickel, the resistance and inductance would be strongly frequency dependent. With the thick copper on one side, all the current can flow through the lower-impedance path in the copper.

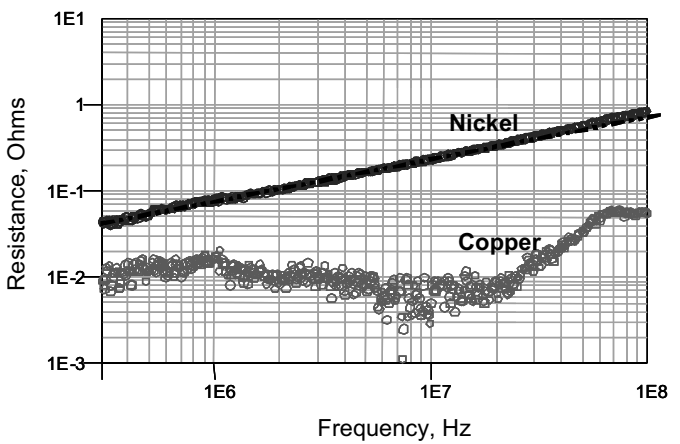


Figure 6-27 Measured resistance of a 1-inch-diameter loop of copper wire and nickel wire, of roughly the same cross section, showing much higher resistance of the nickel conductor, due to skin-depth effects. The resistance increases with the square root of frequency, shown by the superimposed line. The noise floor of the measurement was about 10 milliOhms.

This is why a plating of silver is sometimes applied to Alloy 42 leads to limit their resistance at high frequency. It provides a nonferromagnetic conductor on the outer surface for high-frequency currents to travel. The highest frequency components will experience a larger skin-depth and the higher conductivity material.

The precise resistance of a conductor will depend on the frequency-dependent current distribution, which, for arbitrary shapes, may be difficult to calculate. This is one of the values of a good 2D field solver that allows the calculation of frequency-dependent current distributions and the resulting inductance and resistance.

6.18 Eddy Currents

As mentioned previously, if there are two conductors and the current in one changes, there will be a voltage created across the second, due to the changing mutual magnetic-field line rings around it. This voltage induced in the second conductor can drive currents in the second conductor. In other words, changing the current in one conductor can induce current in the second conductor. We call the induced currents in the second conductor *eddy currents*.

There is an important geometry where eddy currents can significantly affect the partial self-inductance of a conductor and the loop self-inductance of a current loop. This geometry occurs when a loop is near a large conducting surface, such as a plane in a circuit board or the sides of a metal enclosure.

As the simplest example, consider a round wire above a metal plane. It is important to keep in mind that this metal plane can be any conductor and can float at any voltage. It does not matter what its voltage is, or what else it is connected to. All that is important is that it is conductive and that it is continuous.

When there is current in the wire, some of the field line rings will pass through the conducting plane. There will be some mutual inductance between the wire and the plane. When current in the wire changes, some of the magnetic-field line rings going through the plane will change and voltages will be induced in the plane. These voltages will drive eddy currents in the plane. These eddy currents will, in turn, produce their own magnetic fields.

By solving Maxwell's Equations, it can be shown that the pattern of the magnetic-field line rings generated by the eddy currents looks exactly like the magnetic-field lines from another current that would be located below the surface of the plane (i.e., a distance below equal to the height above the plane of the real current). This is illustrated in Figure 6-28. This fictitious current is called an *image current*. The direction of the image current is opposite the direction of the

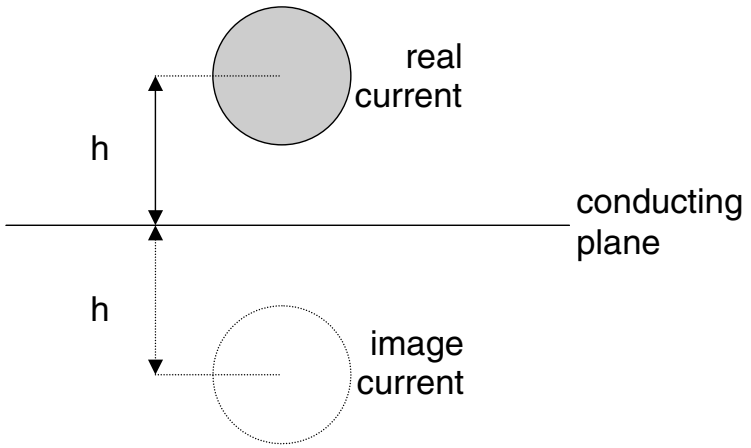


Figure 6-28 Generation of an image current in a plane exactly opposite from the initial current.

real current that induced it. The net magnetic-field lines from the real current and the induced eddy currents will have the same field line distribution as from the real current and the image current, as though the plane were not there. To understand the actual magnetic-field line rings present with the combination of the real current and the induced eddy currents, we can throw out the plane and the real eddy currents and replace them with the image currents. The field line rings from the real current and the image current will be exactly the same as the field line rings from real current and the eddy currents.

The current in the image current is exactly the same magnitude as the current in the real current, but in the opposite direction. Some of the field line rings from the image current will go around the real current. But, because the direction of the image current is opposite the direction of the source current, the field line rings from the image current will subtract from the field line rings around the real current.

This has the unusual effect of decreasing the partial self-inductance of the wire. If a current loop is over the surface of a floating, conducting plane, which has absolutely no electrical contact to the loop, the mere presence of the plane will decrease the loop inductance of the loop. The closer the wire is to the plane, the closer is the image current, and the more mutual-field line rings there are from the image current, and the lower the partial self-inductance of the real current. The closer the floating plane below, the larger the induced eddy currents in the plane and the greater the reduction of the self-inductance of the signal path. Figure 6-29 shows the eddy current distributed in an adjacent plane when the signal path is in proximity to the floating plane.

Figure 6-29 Current distribution at 1 MHz in a round conductor near a floating plane showing the eddy currents induced in the plane.

When two long, rectangular coplanar conductors make up a signal and return path loop, they will have some loop self-inductance per length. If a uniform, floating conducting plane is brought in proximity, their loop self-inductance will decrease due to the field line rings from the eddy currents in the plane below. The closer the plane, the lower the loop inductance. Figure 6-30 shows this reduction in loop self-inductance per length for a simple case.

In this example, the line width is 5 mils and the spacing is 10 mils. The span between the outer edges of each conductor is 20 mils. As a rough rule of thumb, the induced eddy currents will play a role if the spacing to the floating plane is closer than the total span of the conductors.

T I P Induced eddy currents will be created in any plane conductor when a current loop is as close as the span of the conductors. The presence of adjacent planes will always decrease the loop self-inductance of an interconnect.

6.19 The Bottom Line

1. Inductance is a critically important term that affects all aspects of signal integrity.
2. The basic definition of inductance is that it is the number of magnetic-field line rings around a conductor per Amp of current through the conductor.
3. All the various types of inductance are special qualifiers that specify from which conductor the field line rings arise (self and mutual), around how

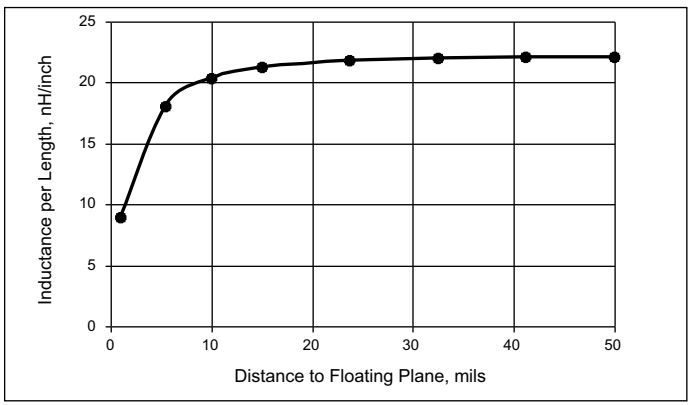


Figure 6-30 Top: current distribution in the two legs of a long rectangular coplanar loop, showing the induced eddy currents in the floating plane. Bottom: the change in loop self-inductance per length as the distance to the floating plane changes, calculated with Ansoft's 2D Extractor.

much of the conductor the field line rings are counted (partial and loop), and if all the field line rings are included, even those from the rest of the loop (total).

4. The only reason inductance is important is because of induced voltage: If the number of field line rings around a conductor changes, a voltage is induced across the ends of the conductor related to how fast the field line rings change.
5. Ground bounce is the voltage induced between different parts of the ground-return path due to a total inductance of the return path and a dI/dt through it.
6. Reducing ground bounce is about reducing the total inductance of the return path: wide conductors, short lengths, and the signal path as close as possible to the return path.
7. Lowest rail-collapse noise is obtained when the loop inductance from the chip pad to the decoupling capacitor is as low as possible. The lowest loop inductance interconnect is two wide planes as close together as possible.
8. The loop inductance between two planes is increased by the presence of a field of via holes. For the case of about 50% open area, the loop inductance is increased by about 25%.

9. As the sine-wave frequency components of currents increase, they will take the path of lowest impedance, which translates to a distribution toward the outside surface of the conductors and signal and return currents as close together as possible. This causes the inductance to be slightly frequency dependent, decreasing toward higher frequency, and the resistance to be strongly frequency dependent, increasing with the square root of the frequency.
10. When a current is in proximity to a uniform plane, even if it is floating, induced eddy currents will cause the self-inductance of the current to decrease.

CHAPTER 7

The Physical Basis of Transmission Lines

We hear the words *transmission line* all the time and we probably use them ever day, yet what really is a transmission line? A coax cable is a transmission line. A PCB trace in a multilayer board is a transmission line.

> **TIP** Fundamentally, a transmission line is composed of any two conductors that have length. This is all it takes to make a transmission line.

As we will see, a transmission line is used to transport a signal from one point to another. Figure 7-1 illustrates the general features of all transmission lines. To distinguish the two conductors, we refer to one as the signal path and the other as the return path.

A transmission line is a new ideal circuit element with very different properties of the three previously introduced ideal circuit elements: resistors, capacitors, and inductors. It has two very important parameters: a characteristic impedance and a time delay. The way a signal interacts with an ideal transmission line is radically different from the way it interacts with the other three ideal elements.

Though in some cases we can approximate the electrical properties of an ideal transmission line with combinations of L's and C's, the behavior of an ideal transmission line matches the actual, measured behavior of real interconnects

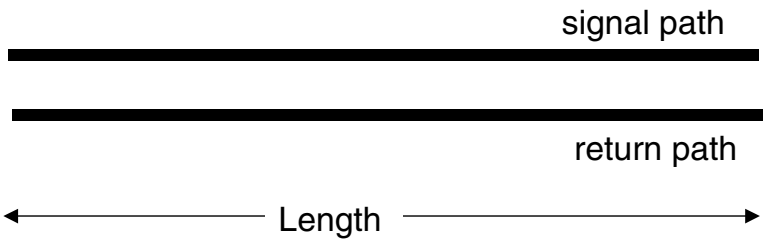


Figure 7-1 A transmission line is any two conductors with length. We label one of the conductors as the signal path, the other as the return path.

much better and to much higher bandwidth than does an LC approximation. Adding an ideal transmission line circuit element to our toolbox will dramatically increase our ability to describe the interactions of signals and interconnects.

7.1 Forget the Word *Ground*

Too often, the other line in a transmission line is referred to as the *ground* line.

> **TIP** Far more problems are created than solved by referring to the second line as the *ground*. It is a good habit to use the term *return path* instead.

One of the most common ways of getting in trouble with signal-integrity design is overusing the term *ground*. It is much healthier to get in the habit of calling the other conductor and thinking about it as the return path.

Many of the problems related to signal integrity are due to poorly designed return paths. If we are always consciously aware that the other path plays the important role of being the return path for the signal current, we will take as much care in designing its geometry as we take for the signal path.

When we label the other path as ground, we typically think it is a universal sink for current. Return current goes into this connection and comes out wherever there is another ground connection. *This is totally wrong.* The return current will closely follow the signal current. As we saw in the last chapter, at higher frequencies, the loop inductance of the signal and return paths will be minimized, which means the return path will distribute as close to the signal path as the conductors will allow.

Figure 7-2 Forget the word *ground* and more problems will be avoided than created.

Further, the return current has no idea what the absolute voltage level is of the return conductor. The actual return conductor may in fact be a voltage plane such as the Vcc or Vdd plane. Other times, it might be a low-voltage plane. That it is labeled as a *ground* connection in the schematic is totally irrelevant to the signal, which is propagating on the transmission line. Start calling the return path the return path and problems will be reduced in the future. This guideline is illustrated in Figure 7-2.

7.2 The Signal

As we will see, when a signal moves down a transmission line, it simultaneously uses the signal and the return path. Both conductors are equally important in determining how the signal interacts with the interconnect.

When both lines look the same, as in a twisted pair, it is inconsequential which we call the signal path and which we call the return path. When one is different from the other, such as in a microstrip, we usually refer to the narrow conductor as the signal path and the plane as the return path.

When a signal is launched into a transmission line, it will propagate down the line at the speed of light in the material. After the signal is launched in the transmission line, we can freeze time a moment later, move along the line, and measure the signal. The signal is always the voltage difference between two adjacent points on the signal and return paths. This is illustrated in Figure 7-3.

TIP If we know the impedance the signal sees, we can always calculate the current associated with the signal voltage. In this respect, a signal is equally well defined as a voltage or a current.

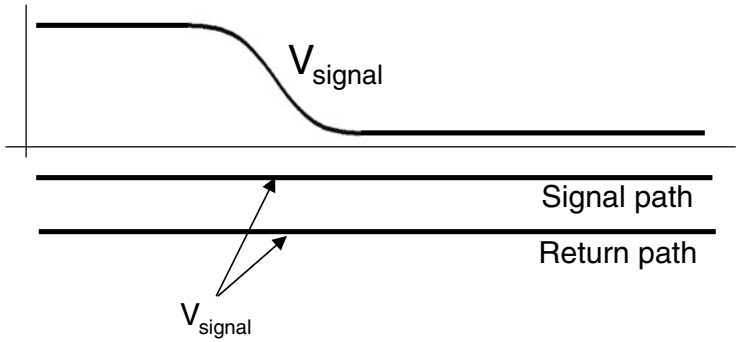


Figure 7-3 Map of the signal, frozen in time on a transmission line. The signal is the voltage between two adjacent points on the signal and the return paths.

These general principles apply to all transmission lines, single ended, and as we will see, differential transmission lines.

7.3 Uniform Transmission Lines

We classify transmission lines by their geometry. The two general features of geometry that strongly determine the electrical properties of a transmission line are how uniform the cross section is down its length and how identical each of its two conductors is.

When the cross section is the same down the length, as in a coax cable, the transmission line is called *uniform*. Examples of various uniform transmission lines are illustrated in Figure 7-4.

As we will see, uniform transmission lines are also called controlled impedance lines. There are a great variety of uniform transmission lines, such as twin leads, microstrips, striplines, and coplanar lines.

TIP Reflections will be minimized and signal quality optimized, if the transmission lines are uniform, or have controlled impedance. All high-speed interconnects should be designed as uniform transmission lines.

Nonuniform transmission lines exist when some geometry or material property changes as we move down the length of the line. For example, if the spacing between two wires is not controlled but varies, this is a nonuniform line. A pair of leads in a dual in-line package (DIP) or quad flat pack (QFP) are nonuniform

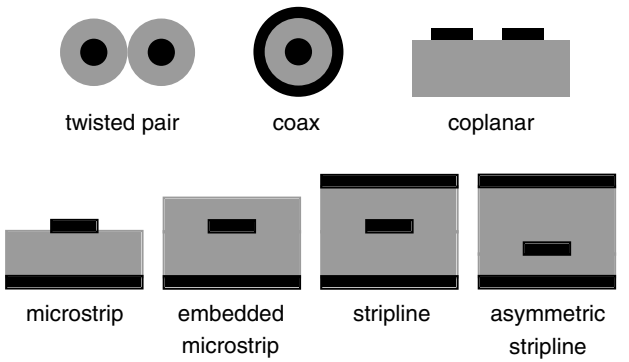


Figure 7-4 Examples of cross sections of uniform transmission lines commonly used as interconnects.

lines. Adjacent traces in a connector are often nonuniform transmission lines. Traces on a PCB that do not have a return-path plane are often nonuniform lines. Nonuniform transmission lines will lead to signal-integrity problems and should be avoided, unless they are kept short enough.

TIP One of the goals in designing for optimized signal integrity is to design all interconnects as uniform transmission lines and to minimize the length of all nonuniform transmission lines.

Another quality of the geometry affecting transmission lines is how similar the two conductors are. When each conductor is the same shape and size as the other one, there is symmetry and we call this line a balanced line. A pair of twisted wires is symmetric since each conductor looks identical. A coplanar line has two narrow strips side by side on the same layer and is balanced.

A coax cable is unbalanced since the center conductor is much smaller than the outer conductor. A microstrip is unbalanced, since one wire is a narrow trace while the second is a wide trace. A stripline is unbalanced for the same reason.

TIP In general, for most transmission lines, the signal quality and cross-talk effects will be completely unaffected by whether the line is balanced or unbalanced. However, ground-bounce and EMI issues will be strongly affected by the specific geometry of the return path.

Whether the transmission line is uniform or nonuniform, balanced or unbalanced, it has just one role to play: to transmit a signal from one end to the other with an acceptable level of distortion.

7.4 The Speed of Electrons in Copper

How fast do signals travel down a transmission line? If is often erroneously believed that the speed of a signal down a transmission line depends on the speed of the electrons in the wire. With this false intuition, we might imagine that lowering the resistance of the interconnect will increase the speed of a signal. In fact, the speed of the electrons in a typical copper wire is actually about 10-billion-times slower than the speed of the signal.

It is easy to estimate the speed of an electron in a copper wire. Suppose we have a roughly 18-gauge round wire, 1 mm in diameter, with 1 Amp of current. We can calculate the speed of the electrons in the wire based on how many electrons pass by one section of the wire per second, the density of the electrons in the wire, and the cross-sectional area of the wire. This is illustrated in Figure 7-5. The current in the wire is related to:

$$I = \frac{\Delta Q}{\Delta t} = \frac{q \times n \times A \times v \times \Delta t}{\Delta t} = q \times n \times A \times v \tag{7-1}$$

from which we can calculate the velocity of the electrons as:

$$v = \frac{I}{q \times n \times A} \tag{7-2}$$

where:

I = the current passing one point, in Amps

ΔQ = the charge flowing in a time interval, in Coulombs

Δt = the time interval

q = the charge of one electron, = 1.6×10^{-19} Coulombs

n = the density of free electrons, in #/m^3

A = the cross-sectional area of the wire, in m^2

v = the speed of the electrons in the wire, in m/sec

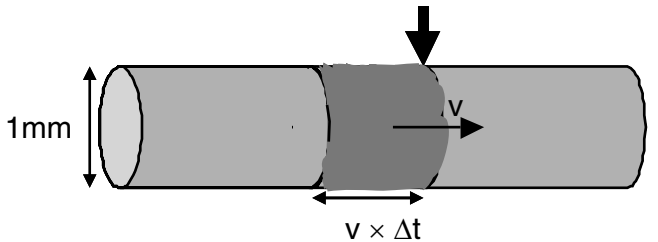


Figure 7-5 The electrons moving down a wire. The number passing the arrow per second is the current and is related to their velocity and number density.

Each copper atom contributes roughly two free electrons that can move through the wire. Atoms of copper are about 1 nm apart. This makes the density of free electrons, n, about $n \sim 10^{27}/m^3$.

For a wire that is 1 mm in diameter, the cross-sectional area is about $A \sim 10^{-6}\ m^2$. Combining these terms and using a 1-Amp current in the wire, the speed of an electron in the wire can be estimated to be roughly:

$$v = \frac{I}{q \times n \times A} = \frac{1\,Amp}{10^{-19} \times 10^{27} \times 10^{-6}} = 10^{-2}\frac{m}{sec} = 1\frac{cm}{sec} \tag{7-3}$$

TIP An electron travels at a speed of about 1 cm/sec. This is about as fast as an ant scurries on the ground.

With this simple analysis, we see that the speed of an electron in a wire is incredibly slow compared to the speed of light in air. The speed of an electron in a wire really has virtually nothing to do with the speed of a signal. Likewise, as we will see, the resistance of the wire has only a very small, almost irrelevant effect on the speed of a signal in a transmission line. It is only in extreme cases that the resistance of an interconnect affects the signal speed, and then only very slightly. We must recalibrate our intuition from the erroneous notion that lower resistance will mean faster signals.

7.5 The Speed of a Signal in a Transmission Line

If it's not the speed of the electrons that determines the speed of the signal, what does?

TIP The speed of a signal depends on the materials that surround the conductors and how quickly the changing electric and magnetic fields associated with the signal can build up and propagate in the space around the transmission line conductors.

The simplest way to think of a signal propagating down a transmission line is illustrated in Figure 7-6. The signal, after all, is a voltage difference between the signal path and the return path. As the signal propagates, a voltage difference must be created between the two conductors. Accompanying the voltage difference is an electric field between the conductors.

In addition to the voltage, a current must be flowing in the signal conductor, and in the return conductor, to provide the charge that charges up the conductors that generates the voltage difference that creates the electric field. This current loop moving through the conductors will produce a magnetic field.

A signal can be launched into a transmission line simply by touching the leads of a battery to the signal and return paths. The sudden voltage change creates a sudden electric and magnetic-field change. This *kink of field* will propagate through the dielectric material surrounding the transmission line at the speed of a changing electric and magnetic field, which is the speed of light in the material.

We usually think of light as the electromagnetic radiation we can see. However, all changing electromagnetic fields are exactly the same and are described by exactly the same set of equations, Maxwell's Equations. The only difference is the frequency of the waves. For visible light, the frequency is about 1,000,000 GHz. For the signals typically found in high-speed digital products, the frequency is about 1–10 GHz.

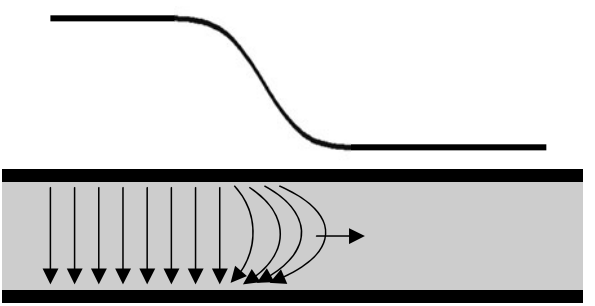


Figure 7-6 The electric field building up in a transmission line as the signal propagates down the line. The speed of the signal depends on how fast the changing electric and magnetic fields can build up and propagate in the materials surrounding the signal- and return-path conductors.

How quickly the electric and magnetic fields can build up is what really determines the speed of the signal. The propagation and interaction of these fields is described by Maxwell's Equations. These say that if the electric and magnetic fields ever change, the kink they make will propagate outward at a speed that depends on some constants and material properties.

The speed of the change, or the kink, v, is given by:

$$v = \frac{1}{\sqrt{\varepsilon_0 \varepsilon_r \mu_0 \mu_r}} \tag{7-4}$$

where:

ε_0 = permittivity of free space = 8.89×10^{-12} F/m

ε_r = relative dielectric constant of the material

μ_0 = permeability of free space = $4\pi \times 10^{-7}$ H/m

μ_r = relative permeability of the material

Putting in the numbers, we find,

$$v = \frac{2.99 \times 10^8}{\sqrt{\varepsilon_r \mu_r}} \frac{\text{m}}{\text{secs}} = \frac{(11.9)}{\sqrt{\varepsilon_r \mu_r}} \frac{\text{inches}}{\text{nsecs}} \tag{7-5}$$

TIP In air, where the relative dielectric constant and relative permeability are both 1, the speed of light is about 12 inches/nsec. This is a really good rule of thumb to keep in mind.

For virtually all interconnect materials, the magnetic permeability of the dielectrics, μ_r, is 1. All polymers that do not contain a ferromagnetic material have a magnetic permeability of 1. Therefore, this term can be ignored.

In comparison, the relative dielectric constant of materials, ε_r, is never less than or equal to 1, except for in the case of air. In all real interconnect materials, the dielectric constant is greater than 1. This means the speed of light in interconnects will always be less than 12 inches/nsec. The speed is:

$$v = \frac{12}{\sqrt{\varepsilon_r}} \frac{\text{inches}}{\text{nsecs}} \tag{7-6}$$

For brevity, we usually refer to the relative dielectric constant as just the "dielectric constant." This number characterizes some of the electrical properties of an insulator. It is an important electrical property. For most polymers, it is roughly 4. For glass, it is about 6 and for ceramics, it is about 10.

It is possible in some materials for the dielectric constant to vary with frequency. In other words, the speed of light in a material may be frequency dependent. In general, the dielectric constant decreases with higher frequency. This makes the speed of light in the material increase as we go toward higher frequency.

In most common materials, such as FR4, the dielectric constant varies very little from 500 MHz to 10 GHz. Depending on the ratio of epoxy resin to fiber glass, the dielectric constant of FR4 can be from 3.5 to 4.5. Most interconnect laminate materials have a dielectric constant of about 4. This suggests a simple, easy-to-remember generalization.

TIP A good rule of thumb to remember is that the speed of light in most interconnects is about 12 inches/nsec / sqrt(4) = 6 inches/nsec. When evaluating the speed of signals in a board-level interconnect, we can assume they are about 6 inches/nsec.

As pointed out in an earlier chapter, when the field lines see a combination of dielectric materials, as in a microstrip where there are some field lines in the bulk material and some in the air above, the effective dielectric constant that affects the signal speed is a combination of the different materials. The only way to predict the effective dielectric constant when the materials are inhomogeneous throughout the cross section is with a 2D field solver. In the case of stripline, for example, all the fields see the same material and the effective dielectric constant is the bulk dielectric constant.

The time delay, TD, and length of an interconnect are related by:

$$TD = \frac{Len}{v} \tag{7-7}$$

where:

TD = the time delay, in nsec

Len = the interconnect length, in inches

v = the speed of the signal, in inches/nsec

This means that to travel down a 6-inch length of interconnect in FR4, for example, the time delay is about 6 inches / 6 inches/nsec or about 1 nsec. To travel 12 inches takes about 2 nsec.

The wiring delay, the number of psec of delay per inch of interconnect, is also a useful metric. It is just the inverse of the velocity: 1/v. For FR4, the wiring delay is about 1/6 inches/nsec = 0.166 nsec/inch, or 170 psec/inch. This is the delay of a signal propagating down a transmission line in FR4. Every inch of interconnect has a propagation delay of 170 psec. The wiring delay through the 0.5 inch of a BGA lead is 170 psec/inch × 0.5 inches = 85 psec.

7.6 Spatial Extent of the Leading Edge

Every signal has a rise time, RT, usually measured from the 10% to 90% voltage levels. As a signal moves down a transmission line, the leading edge spreads out on the transmission line and has a spatial extent. If we could freeze time and look at the size of the voltage distribution as it moves out, we would find something like Figure 7-7.

The length of the rise time, d, on the transmission line depends on the speed of the signal and the rise time:

$$d = RT \times v \tag{7-8}$$

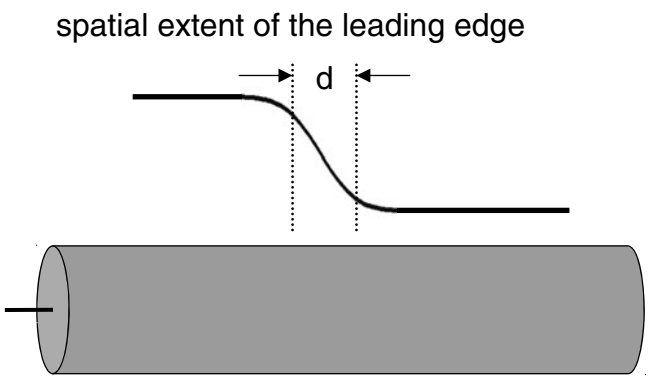


Figure 7-7 Spatial extent of the leading edge of the signal as it propagates down a transmission line.

where:

d = the spatial extent of the rise time, in inches

RT = the rise time of the signal, in nsec

v = the speed of the signal, in inches/nsec

For example, if the speed is 6 inches/nsec and the rise time is 1 nsec, the spatial extent of the leading edge is 1 nsec × 6 inches/nsec = 6 inches. As the leading edge moves down the circuit board, it is really a 6-inch section of rising voltage moving down the board. A rise time of 0.1 nsec has a spatial extent of 0.6 inch.

TIP Many of the signal-integrity problems related to imperfections in the transmission line depend on the size of the discontinuity compared to the spatial extent of the leading edge. It is always a good idea to be aware of the spatial extent of the leading edge for all signals.

7.7 "Be the Signal"

All a signal cares about is how fast it moves down the line and what impedance it sees. As we saw previously, speed is based on the material properties of the dielectric and their distribution. To evaluate the impedance a signal sees in propagating down the line, we will take a microstrip transmission line and launch a signal into one end. A microstrip is a uniform, but unbalanced transmission line. It has a narrow-width signal line and a wide return path.

The analysis we will do of this line is identical to that for any transmission line. We'll make it 10 feet long, so that we can actually walk down it and, in a Zen way, "be the signal" to observe what the signal would see. With each step along the way we will ask, what impedance do we see? We will answer this question by determining the ratio of the voltage applied, 1 v, and the current coming out of our foot to drive the signal down the transmission line.

In this case, we launch a signal into one end by connecting a 1-v battery between the two conductors at the front end. At the initial instant we have launched the signal into the line, there has not been enough time for the signal to travel very far down the line.

Just to make it easier, let's assume that we have air between the signal and return paths, so the speed of propagation is 1 foot per nsec. After the first nsec, the voltage on the far end of the line is still zero, as the signal hasn't had enough time

to get very far. The signal along the line would be about 1 volt for the first foot and zero for the remaining length of the line.

Let's freeze time after the first nsec and look at the charges on the line. What we see is illustrated in Figure 7-8. Between the signal- and return-path conductors in the first 12 inches, there will be a 1-volt difference. This is, after all, the signal. We know that because the signal and return paths are two separated conductors, there will be some capacitance between the conductors in this region. If there is a 1-v difference between them, there must also be some charge on the signal conductor and an equal and opposite amount on the return-path conductor.

In the next 1 nsec, we, being the signal, will move ahead another 12 inches. Let's stop time again. We now have the first two feet of line charged up. We see that having made this last step, we have brought the signal to the second foot-long section and created a voltage difference between the signal and return conductors. There is now a charge difference between the two conductors, at the point of each footstep, where there was none a nsec ago.

As we walk down the line, we are bringing a voltage difference to the two conductors and charging them up. In each nsec, we take another 1-foot step and charge up this new section of the line. Each step of the signal will leave another foot of charged transmission line in our wake.

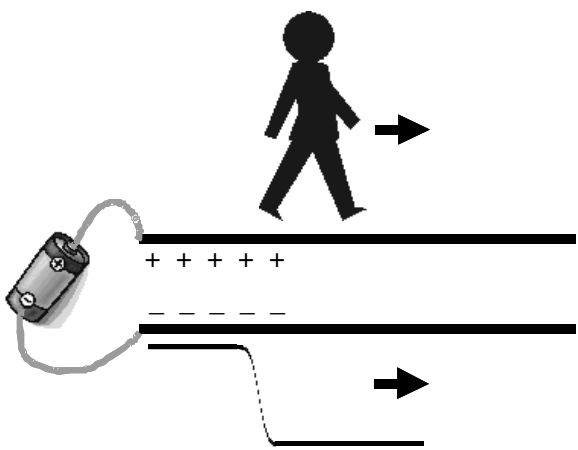


Figure 7-8 Charge distribution on the transmission line after a 1-v signal has propagated for 1 nsec. There is no charge ahead of us (the signal) at this instant of time.

The charge that flowed to charge up each footstep came from the signal as each foot came down and ultimately from the battery. The fact that the signal is propagating down the line means that the capacitance between the signal and return paths is getting charged up. How much charge has to flow from our foot into the line, in each footstep? In other words, what is the current that must flow as the signal propagates?

If the signal is moving down the line at a steady speed and the line is uniform, that is, it has the same capacitance per foot of length, then we are injecting the same amount of charge into the line with each footstep. Each step charges up the same amount of capacitance to the same voltage. If we are always taking the same time per step, then there is the same charge per unit of time required for the signal to charge up the line. The same amount of charge per nsec flowing into the line means that there is a constant current flowing into the line from our foot.

TIP From the signal's perspective, as we walk down the line at our speed of 1 foot/nsec, we are charging up each foot of line in the same amount of time. Coming out of the bottom of our foot is the charge that is added to the line to charge it up. An equal charge out of our foot in an equal time interval means we are injecting a constant current into the line.

What affects the current coming out of our foot to charge up the line? If we are moving down the line at a constant speed, and if we were to increase the width of the signal path, the capacitance we need to charge up will increase and the charge that must come out of our foot in the time we have till the next step will also increase. Likewise, if anything is done to decrease the capacitance per length, the current coming out of our foot will decrease. For the same reason, if the capacitance per length stays the same, but our speed increases, we will be charging up more length per nsec and the current needed will increase.

In this way, we can deduce that the current coming out of our foot will scale directly with both the capacitance per length and the speed of the signal. If either increases, the current out of our foot with each step will increase. If either decreases, the current from the signal to charge up the line will decrease. We have deduced a simple relationship between the current coming out of our foot and the properties of the line:

$$I \sim v \times C_L \tag{7-9}$$

where:

I = the current out of our foot

v = the speed with which we move down the line, charging up regions

C_L = the capacitance per length of the line

As we, the signal, move down the transmission line, we will constantly be asking, "what is the impedance of the line?" The basic definition of impedance of any element is the ratio of voltage applied to current through it. So as we move down the line, we will constantly be asking with each footstep, what is the ratio of the voltage applied to the current being injected into the line?

The voltage of the signal is fixed at the signal voltage. The current into the line depends on the capacitance of each footstep and how long it takes to charge up each one. As long as the speed of the signal is constant and the capacitance of each footstep is constant, the current injected into the line by our foot will be constant and the impedance of the line the signal sees will be constant.

Suppose the line width were to suddenly widen. The capacitance of each footstep would be larger and the current coming out of each footstep to charge up this capacitance would be larger. A higher current for the same voltage means we see the transmission line having a lower impedance. Likewise, if the line were to suddenly narrow, the capacitance of each footstep would be less, the current required to charge it up would be less, and the impedance the signal would see for the line would be higher.

TIP We call the impedance the signal sees with each step, the instantaneous impedance of the transmission line. As the signal propagates down the line, it is constantly probing the instantaneous impedance of each footstep, as the ratio of the voltage applied to the current required to charge the line and propagate toward the next step.

The instantaneous impedance depends on the speed of the signal, which is a material property, and the capacitance per length. For a uniform transmission line, the cross-sectional geometry is constant down the line and the instantaneous impedance the signal sees will be constant down the line. As we will see, an important behavior of a signal interacting with a transmission line is that whenever the signal sees a change in the instantaneous impedance some of the signal will reflect and some will continue distorted. Additionally, signal integrity may be

affected. This is the primary reason why controlling the instantaneous impedance the signal sees is so important.

TIP The chief way of minimizing reflection problems is to keep the instantaneous impedance the signal sees constant by keeping the geometry constant. This is what is meant by a controlled-impedance interconnect, or a constant instantaneous impedance down the length.

7.8 The Instantaneous Impedance of a Transmission Line

We can quantify this analysis by building a simple physical model of the transmission line. We have modeled the line as an array of little capacitor buckets, each one equal to the capacitance in the transmission line that spans a footstep and separated by the distance we, the signal, move in each footstep. We call this model (the simplest model we can come up with that provides engineering insight) the zeroth-order model of a transmission line. It is a physics model and *not* an equivalent circuit model. Circuit models do not have lengths in them. This model is outlined in Figure 7-9.

In this model, the size of each footstep is Δx. The magnitude of each little capacitor bucket is the capacitance per length, C_L, of the line times the length of each footstep:

$$C = C_L \times \Delta x \tag{7-10}$$

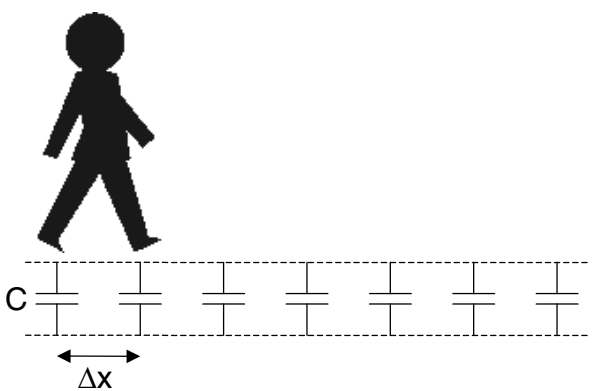


Figure 7-9 Zeroth-order model of a transmission line composed of an array of capacitors. With each footstep another capacitor is charged up. The spacing is the size of our footstep.

We can calculate the current out of our foot, I, using this model. The current is the charge that flows out of our foot to charge up each bucket in the time interval between each step. The charge we dump in each capacitor bucket, Q, is the capacitance of the bucket times the voltage applied, V. For each step we take, we are dumping the charge Q into the line, in a time interval of a step. The time between steps, Δt, is the length of our step, Δx, divided by our speed down the line, v. Of course, as the real signal propagates, each footstep is really small, but the time interval gets really small as well. The ratio of the charge required to each time interval is a constant value, which is the current that flows into the line as the signal propagates:

$$I = \frac{Q}{\Delta t} = \frac{CV}{\left(\frac{\Delta x}{v}\right)} = \frac{C_L \Delta x v V}{\Delta x} = C_L v V \tag{7-11}$$

where:

I = the current from the signal

Q = the charge in each footstep

C = the capacitance of each footstep

Δt = the time to step from capacitor to capacitor

C_L = the capacitance per length of the transmission line

Δx = the distance between the capacitors or each footstep

v = the speed of walking down the line

V = the voltage of the signal

This says the current coming out of our foot and going into the line is simply related to the capacitance per length, the speed of propagation, and the voltage of the signal—exactly as we reasoned.

This is the defining relationship for the current-voltage (I-V) behavior of a transmission line. It says the instantaneous current of a signal everywhere on a transmission line is directly proportional to the voltage. Double the voltage applied and the current into the transmission line will double. This is exactly how a resistor behaves. With each step down the transmission line, the signal sees an impedance that behaves like a resistive load.

From this relationship, we can calculate the instantaneous impedance a signal would see as it propagates down a transmission line. The instantaneous impedance is the ratio of the voltage applied to the current through the device:

$$Z = \frac{V}{I} = \frac{V}{C_L v V} = \frac{1}{C_L v} = \frac{83}{C_L}\sqrt{\varepsilon_r} \tag{7-12}$$

where:

Z = the instantaneous impedance of the transmission line, in Ohms

C_L = the capacitance per length of the line, in pF/inch

v = the speed of light in the material

ε_r = dielectric constant of the material

The instantaneous impedance a signal sees depends only on two terms, both of which are intrinsic to the line. It doesn't depend on the length of the line. The instantaneous impedance of the line depends on the cross section of the line and the material properties. As long as these two terms are constant as we move down the line, a signal would see the same constant instantaneous impedance. And, of course, the units we use to measure the instantaneous impedance of the line are in units of Ohms, as with any impedance.

Since the speed of the signal depends on a material property, we can relate the capacitance per length of the transmission line to the instantaneous impedance. For example, if the dielectric constant is 4 and the capacitance per length of the line is 3.3 pF/inch, the instantaneous impedance of the transmission line is:

$$Z = \frac{83}{C_L}\sqrt{\varepsilon_r} = \frac{83}{3.3}\sqrt{4} = 50\Omega \tag{7-13}$$

Right about now, we might be asking, what about the inductance of the line? Where does that come into play in this model? The answer is that this zeroth-order model is not an electrical model; it is a physical model. Rather than approximating the transmission line with L's and C's, we added the observation that the speed of the signal is the speed of light in the material.

In reality, the finite speed of the signal arises partly because of the series loop inductance of the signal and return paths. If we used a first-order, equivalent

circuit model, which included the inductance per length, it would derive for us the current into the transmission line and the finite propagation speed, but the model would be more complicated mathematically.

These two models are really equivalent considering the connection between the propagation speed and the inductance per length. As we shall see, the propagation delay is directly related to the capacitance per length combined with the inductance per length. The speed of the signal has in it some assumptions about the inductance of the conductors.

7.9 Characteristic Impedance and Controlled Impedance

For a uniform line, anywhere we choose to look, we will see the same instantaneous impedance as we propagate down the line. There is one instantaneous impedance that characterizes the transmission line and we give it the special name, characteristic impedance.

> **TIP** There is one instantaneous impedance that is *characteristic* for a uniform line. We refer to this constant, instantaneous impedance as the impedance that is characteristic of the line, or the *characteristic impedance* of the line.

To distinguish that we are speaking of the special term the *characteristic impedance* of the line, which is an intrinsic property of the line, we give it the special symbol, Z_0 (Z with a subscript zero). Characteristic impedance is given in units of Ohms. Every uniform transmission line has a characteristic impedance, which is one of the most important terms describing its electrical properties and how signals will interact with it.

> **TIP** The characteristic impedance of the line will tell us the instantaneous impedance a signal will see as it propagates down the line. As we will see, this is the chief factor that influences signal integrity in transmission line circuits.

Characteristic impedance is numerically equal to the instantaneous impedance of the line and is intrinsic to the line. It depends only on the material properties, the dielectric constant, and the capacitance per length of the line. It does not depend on the length of the line.

For a uniform line, the characteristic impedance is:

$$Z_0 = \frac{83}{C_L}\sqrt{\varepsilon_r} \tag{7-14}$$

If the line is uniform, it has only one instantaneous impedance, which we call the characteristic impedance. One measure of the uniformity of the line is how constant the instantaneous impedance is down the length. If the line width varies down the line, there is no one single value of instantaneous impedance for the whole line. By definition, a nonuniform line has no characteristic impedance. When the cross section is uniform, the impedance a signal sees as it propagates down the interconnect will be constant and we say the impedance is controlled. For this reason, we call uniform cross-section transmission lines *controlled-impedance* lines.

TIP We call a line that has a constant instantaneous impedance down its length a controlled-impedance line. A board that is fabricated with all its interconnects as controlled-impedance lines, all with the same characteristic impedance, is called a controlled-impedance board. All high-speed digital products, with boards larger than about 6 inches and clock frequencies greater than 100 MHz, are built with controlled-impedance boards.

When the geometry and material properties are constant down the line, the characteristic impedance of the line is uniform and one number fully characterizes the line.

A controlled-impedance line can be fabricated with virtually any uniform cross section. There are many standard cross-sectional shapes that can have controlled impedance, and many of these families of shapes have special names. For example, two round wires twisted together are called a *twisted pair.* A center conductor surrounded by an outer conductor is a *coaxial* or *coax transmission line*. A narrow strip of signal line over a wide plane is a *microstrip*. When the return path is two planes and the signal line is a narrow strip between them, we call this a *stripline*. The only requirement for a controlled-impedance interconnect is that the cross section be constant.

With this connection between the capacitance per length and the characteristic impedance, we can now relate our intuition about capacitance to our new intuition about characteristic impedance. We generally have a pretty good intuitive

feel for capacitance and capacitance per length for the two conductors in a transmission line. If we make the two conductors wider, we increase the capacitance per length. If we move them farther apart, we make the capacitance per length lower.

For a microstrip using FR4 dielectric, when the line width is twice the dielectric thickness, the characteristic impedance is about 50 Ohms. What happens to the characteristic impedance when we make the dielectric separation larger? It's not obvious initially. However, we now know that the characteristic impedance of a transmission line is inversely proportional to the capacitance per length between the conductors.

Therefore, if we move the conductors farther apart, the capacitance will decrease, and the characteristic impedance will increase. Making the microstrip signal-trace wider will increase the capacitance per length and decrease the characteristic impedance. This is illustrated in Figure 7-10.

In general, a wide conductor with a thin dielectric will have a low characteristic impedance. For example, the characteristic impedance of the transmission line formed from the power and ground planes in a PCB will have a low characteristic impedance, generally less than 1 Ohm. Narrow conductors with a thick dielectric will have a high characteristic impedance. Signal traces, with narrow lines, will have high characteristic impedance, typically between 60 Ohms and 90 Ohms.

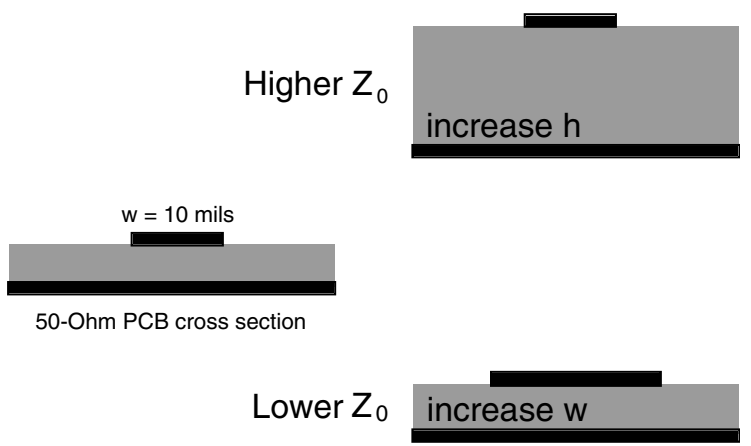


Figure 7-10 If line width increases, capacitance per length increases and characteristic impedance decreases. If dielectric spacing increases, capacitance per length decreases and characteristic impedance increases.

7.10 Famous Characteristic Impedances

Over the years, various specs have been established for specialized controlled-impedance interconnects. A number of these are listed in Figure 7-11. One of the most common ones is RG58. Virtually all general-purpose coax cables used in the lab, with Berkeley Nuclear Corp. (BNC) type bayonet connectors, are made with RG58 cable. This spec defines an inner- and outer-conductor diameter and a dielectric constant. In addition, when the spec is followed, the characteristic impedance is about 52 Ohms. Look on the side of the cable and you will see "RG58" stamped.

There are other cable specs as well. RG174 is useful to know about. It is a thinner cable than RG58 and is much more flexible. When trying to snake a cable around in tight spaces or when low stress is required, the flexibility of RG174 is useful. It is specified with a characteristic impedance of 50 Ohms.

The coax cable used in Cable TV systems is specified at 75 Ohms. This cable will have a lower capacitance per length than a 50-Ohm cable and in general is thicker than a comparable 50-Ohm cable. For example, RG59 is thicker than RG58.

Twisted pairs, typically used in high-speed serial links, small computer system interconnect (SCSI) applications, and telecommunications applications, are made with 18- to 26-gauge wire. With the typical insulation thickness commonly used, the characteristic impedance is about 100 Ohms to 130 Ohms. This is typi-

RG174	50Ω
RG58	52Ω
RG59	75Ω
RG62	93Ω
TV Antenna	300Ω
Cable TV	75Ω
Twisted pairs	100–130Ω

Figure 7-11 Some famous controlled-impedance interconnects based on their specified characteristic impedance.

cally a higher impedance than used in circuit boards, but matches the differential impedance of typical board traces. Differential impedance is introduced in a later chapter.

There is one characteristic impedance that has special, fundamental significance. This is the characteristic impedance of free space. As described previously, a signal propagating in a transmission line is really light, the electric and magnetic fields trapped and guided by the signal- and return-path conductors. As a propagating field, it travels at the speed of light in the composite dielectric medium.

Without the conductors to guide the fields, light will propagate in free space as waves. These are waves of electric and magnetic fields. As the wave propagates through space, the electric and magnetic fields will see an impedance. The impedance a wave sees is related to two fundamental constants: the permeability of free space and the permittivity of free space:

$$Z_0 = \sqrt{\frac{\mu_0}{\varepsilon_0}} = 120\pi = 376.99 \sim 377\Omega \tag{7-15}$$

The combination of these two constants is the instantaneous impedance a propagating wave will see. We call this the characteristic impedance of free space, approximately 377 Ohms. This is a fundamental number. The amount of radiated energy from an antenna is optimized when its impedance matches the 377 Ohms of free space.

There is only one characteristic-impedance value that has fundamental significance, and it is 377 Ohms. All other impedances are arbitrary. The characteristic impedance of an interconnect can be almost any value, limited by manufacturability constraints.

But what about 50 Ohms? Why is it so commonly used? What's so special about 50 Ohms? Its use became popular in the early 1930s when radio communications and radar systems became important and drove the first requirements for using high performance transmission lines. The application was to transmit the radio signal from the not very efficient generator to the radio antenna, with the minimum attenuation.

As we show in a later chapter, the attenuation of a coax cable is related to the series resistance of the inner conductor and outer conductor divided by the characteristic impedance. If the outer diameter of the cable is fixed, using the largest diameter cable possible, there is an optimum inner radius that results in minimum impedance.

With too large an inner radius, the resistance is lower, but the characteristic impedance is lower as well and the attenuation is higher. Too small a diameter of the inner conductor and the resistance is high and the attenuation is high. When you explore the optimum value of inner radius, you find the value for the lowest attenuation is also the value that creates 50 Ohms.

The reason 50 Ohms was chosen almost 100 years ago was to minimize the attenuation in coax cables for a fixed outer diameter. It was adopted as a standard to improve radio and radar system efficiency, and it was easily manufactured. Once adopted, the more systems using this value of impedance, the better their compatibility. If all test and measurement systems matched to this standard 50 Ohms, then reflections between instruments would be minimized and signal quality would be optimized.

In FR4, a microstrip can be fabricated if the line width is twice as wide as the dielectric is thick. There is a broad range of characteristic impedances around 50 Ohms that can also be fabricated, so it is a soft optimum.

In high-speed digital systems, there are a number of trade-offs that determine the optimum characteristic impedance of the entire system. Some of these are illustrated in Figure 7-12. A good starting place is 50 Ohms. Using a higher characteristic impedance with the same pitch will mean more cross talk; however,

Property	Lower Z_0	Higher Z_0
Board costs	better	worse
Delay adders	better	worse
Cross talk	better	worse
Attenuation	better	worse
Connector costs	worse	better
Twisted pair/ cable costs	worse	better
Driver design	worse	better
Power dissipation	worse	better

Figure 7-12 Trade-offs in various system issues based on changing the characteristic impedance of the interconnects. Deciding on the optimum characteristic impedance, balancing performance and cost, is a difficult process. 50 Ohms is a good compromise in most systems.

higher-characteristic impedance connectors or twisted-pair cables will be lower cost since they are easier to fabricate. Lower characteristic impedance will mean lower cross talk and lower sensitivity to delay adders caused by connectors, components, and vias, but it also means higher power dissipation. This is important in high-speed systems.

Every system will have its own balance for the optimum characteristic impedance. In general, it is a very soft optimum and the exact value chosen is not critically important, as long as the same impedance is used throughout the system. Unless there is a strong driving force otherwise, 50 Ohms is usually used. In the case of Rambus memory, timing was critically important and a low impedance of 28 Ohms was chosen to minimize the impact from delay adders. To manufacture such a low impedance requires wide lines. But since the interconnect density in Rambus modules is low, the wider lines have only a small impact.

7.11 *The* Impedance of a Transmission Line

What impedance does a battery see connected to the front of a transmission line? As the signal propagates down the line, the current from the battery into the line is the same as the current coming out of the signal. The impedance the battery will see is thus the same as the instantaneous impedance of the line, *as long as the signal is propagating down the line*. The current into the line is directly proportional to the voltage applied.

A circuit element that has a constant current through it, for a constant voltage applied, is an ideal resistor. From the battery's perspective, when the terminals are attached to the front end of the line and the signal propagates down the line, the transmission line will draw a constant current and act like a resistor to the battery. The impedance of the transmission line, as seen by the battery, is a constant resistance, as long as the signal is propagating down the line. There is no test the battery can do that will distinguish the transmission line from a resistor, at least while the signal is propagating down and back on the line.

TIP We've introduced the idea of the characteristic impedance of an interconnect. We often use the terms the *characteristic impedance* and the *impedance* of a line interchangeably. But they are not always the same, and it's important to emphasize this distinction.

What does it mean to speak of the *impedance* of a cable? An RG58 cable is often referred to as a 50-Ohm cable. What does this really mean? Suppose we

were to take a three-foot length of RG58 cable and measure the impedance at the front end, between the signal and return paths. What impedance will we measure? Of course, we can measure the impedance with an ohmmeter. If we connect an ohmmeter to the front end of the three-foot transmission line, between the central signal line and the outer shield, as shown in Figure 7-13, what will it read? Will it be an open, a short, or 50 Ohms?

Let's be more specific. Suppose we were to measure the impedance using a Radio Shack ohmmeter with a liquid crystal display (LCD) that updates in half a second. What impedance will be measured?

Of course, if we wait long enough, the short length of cable will look like an open, and we would measure infinite as the input impedance. So, if the input impedance of this short cable is infinite, what does it mean to have a 50-Ohm cable? Where does the characteristic-impedance attribute come in?

To explore this further, consider a more extreme, very long length of RG58 cable. It's so long, in fact, that the line stretches from the earth to the moon. This is about 240,000 miles long. As we might recall from high school, the speed of light in a vacuum is about 186,000 miles per second, or close to 130,000 miles per second in the dielectric of the RG58 cable. It would take light about 2 seconds to go from one end of the cable to the far end and another 2 seconds to come back. If we attach our ohmmeter to the front end of this long line, what will it see as the impedance? Remember, the ohmmeter finds the resistance by connecting a 1-volt source to the device under test and measuring the ratio of the voltage applied to the current draw.

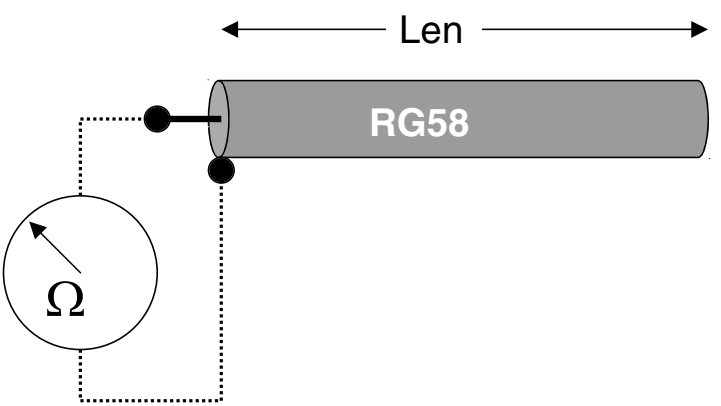


Figure 7-13 Measuring the input impedance of a length of RG58 cable with an ohmmeter.

This is exactly the case of driving a transmission line, providing we do the impedance measurement in less than the 4-second round-trip time of flight. During the first 4 seconds, while the signal is propagating out to the end of the interconnect and back, the current into the front of the line will be a constant amount equal to the current required by the signal to charge up successive sections of the cable as it propagates outward.

The impedance the source sees looking into the front of the line, the "input" impedance, will be the same as the instantaneous impedance the signal sees, which is the characteristic impedance of the line. In fact, the source won't know that there is an end to the line until a round-trip time of flight, or 4 seconds, later. During the first 4 seconds of measurement, the ohmmeter should read the characteristic impedance of the line, or 50 Ohms, in this case.

TIP As long as the measurement time is less than the round-trip time of flight, the input impedance of the line, measured by the ohmmeter, will be the characteristic impedance of the line.

But, we know that if we wait a day with the ohmmeter connected, we will eventually measure the input impedance of the cable to be on open. Here are the two extremes: Initially, we measure 50 Ohms, but after a long time, we measure an open. So, what is *the* input impedance of the line?

The answer is that there is no one value. It is time dependent; it changes. What this example illustrates is that the input impedance of a transmission line is time dependent; it depends on how long we are doing the measurement, compared to the round-trip time of flight. This is illustrated in Figure 7-14. Within the round-trip time of flight, the impedance looking into the front end of the transmission line is the characteristic impedance of the line. After the round-trip time of flight, the input impedance can be anywhere from infinite to zero, depending on what is at the far end of the transmission line.

When we refer to the impedance of a cable or a line as being 50 Ohms, we are really saying the instantaneous impedance a signal would see propagating down the line is 50 Ohms. Or, the characteristic impedance of the line is 50 Ohms. Or, initially, if we look for a time short compared to a round-trip time of flight, we will see 50 Ohms as the input impedance of the interconnect.

Even though these words sound similar, there is an important difference between *the* impedance, the *input* impedance, the *instantaneous* impedance, and

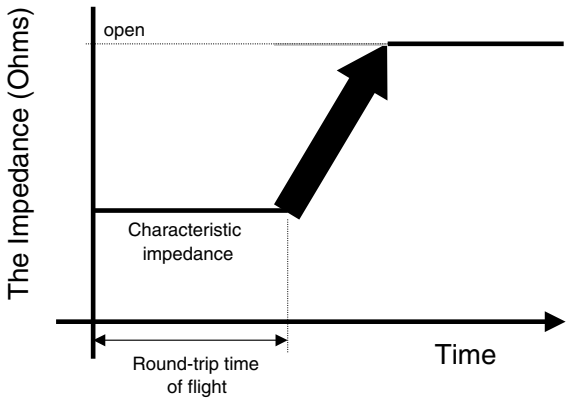


Figure 7-14 The input impedance, looking into a transmission line, is time dependent. During the round-trip time of flight, the measured input impedance will be the characteristic impedance. If we wait long enough, the input impedance will look like an open.

the *characteristic* impedance of an interconnect. Just saying "the impedance" is ambiguous.

> **TIP** The input impedance of an interconnect is what a driver would measure launching a signal into the front end of the interconnect. It is time dependent. It can be an open, it can be a short, or it can be anywhere in between, all for the same transmission line, depending on how the far end is connected, how long the transmission line is, and how the impedance is measured.

The instantaneous impedance of the transmission line is the impedance the signal sees as it propagates down the line. If the cross section is uniform, the instantaneous impedance will be the same down the line. However, it may change where there are discontinuities, for example, at the end. If the end is open, the signal will see an infinite instantaneous impedance when it hits the end of the line. If there is a branch, it will see a drop in the instantaneous impedance at the branch point.

The characteristic impedance of the interconnect is a physical quality of the line that characterizes the transmission line due to its geometry and material properties. It is equal to the instantaneous impedance the signal would see as it propagates down the uniform cross section.

Everyone who works in the signal-integrity field gets lazy sometimes and just uses the term *impedance*. We, therefore, must ask the qualifying question of which impedance we mean, or look at the context of its use to know which of

these three impedances we are referring to. Knowing the distinction, we can all try to use the right one and be less ambiguous.

When the rise time is shorter than the round-trip time of flight of an interconnect, a driver will see the interconnect with a resistive input impedance equal to the characteristic impedance of the line. Even though the line may be open at the far end, during the transition time, the front of the line will behave like a resistor.

The round-trip time of flight is related to the dielectric constant of the material and the length of the line. With rise times for most drivers in the sub-nanosecond regime, any interconnect longer than a few inches will look long and behave like a resistive load to the driver during the transition. This is one of the important reasons that the transmission line behavior of all interconnects must be considered.

TIP In the high-speed regime, an interconnect longer than a few inches does not behave like an open to the driver. It behaves like a resistor during the transition. When the length is long enough to show transmission line behavior, the input impedance the driver sees may be time dependent. This property will strongly affect the behavior of the signals that propagate on the interconnects.

Given this criterion, virtually all interconnects in high-speed digital systems will behave like transmission lines and these properties will dominate the signal-integrity effects. For a transmission line on a board that is 3 inches long, the round-trip time of flight is about 1 nsec. If the integrated circuit (IC) driving the line has a rise time less than 1 nsec, the impedance it will see looking into the front of the line during the rising or falling edge will be the characteristic impedance of the line. The driver IC will see an impedance that acts resistive. If the rise time is very much longer than 1 nsec, it will see the impedance of the line as an open. What it sees during the transition period, as the initial edge of the signal bounces back and forth, is very complicated and can often be analyzed only using simulation tools. These tools are described later.

The round-trip time of flight is a very important parameter of a transmission line. To a driver, the line will seem resistive for this time. Figure 7-15 shows the round-trip time of flight for various length transmission lines made with air ($\varepsilon_r = 1$), FR4 ($\varepsilon_r = 4$), and ceramic ($\varepsilon_r = 10$) dielectric materials. In most systems with clock frequencies higher than 200 MHz, the rise time is less than 0.5 nsec. For these systems, all transmission lines longer than about 1.5 inches will appear resistive during the rise time. This means for virtually all high-speed drivers, when they drive a transmission line, during the transition the input impedance they see will act like a resistor.

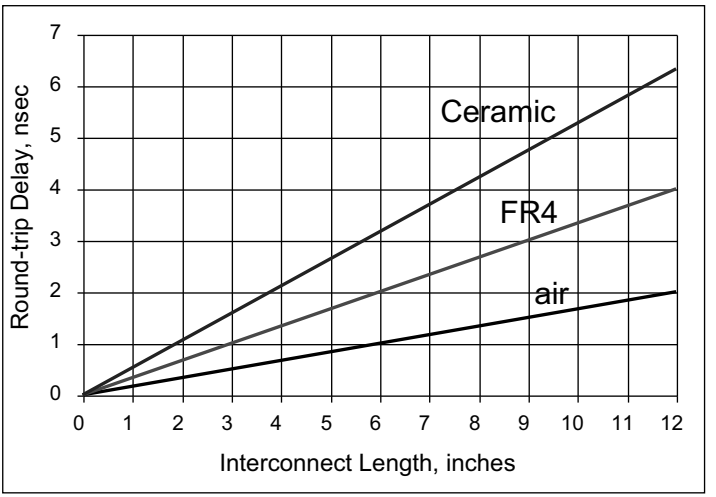


Figure 7-15 For a time equal to the round-trip time of flight, a driver will see the input impedance of the interconnect as a resistive load with a resistance equal to the characteristic impedance of the line.

7.12 Driving a Transmission Line

For a high-speed driver launching a signal into a transmission line, the input impedance of the transmission line during the transition time will behave like a resistance that is equivalent to the characteristic impedance of the line. Given this equivalent circuit model, we can build a circuit of the driver and transmission line and calculate the voltage launched into the transmission line. The equivalent circuit is shown in Figure 7-16.

The driver can be modeled as a voltage source element that switches on fast and as a source resistance. The voltage source has a voltage that is specified depending on the transistor technology. For CMOS, it ranges from 5 v to 1.5 v depending on the transistor generation. Older CMOS use 5 v, while PCI and some memory buses use 3.3 v. The fastest processors use 2.4 v and lower for their output rails and 1.5 v and lower for their core. These voltages are the supply voltages and are very close to the output voltage when the device is driving an open circuit.

The value of the source resistance also depends on the device technology. It is typically in the 5-Ohm to 60-Ohm range. When the driver suddenly turns on, some current flows through the source impedance to the transmission line and there is a voltage drop internal to the gate, before the signal comes out the pin. This means that the full-drive voltage does not appear across the output pins of the driver.

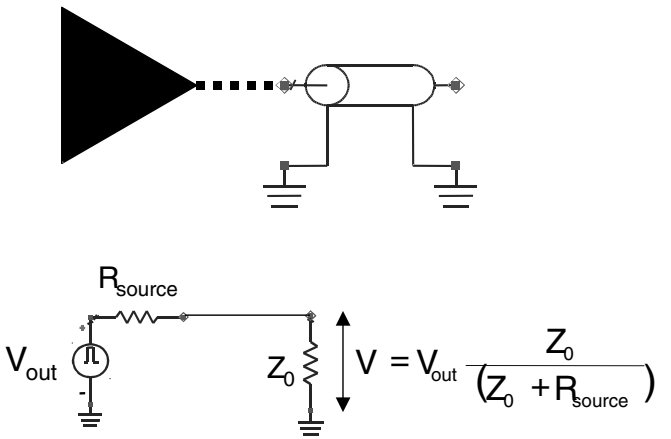


Figure 7-16 Top: Output gate driving a transmission line. Bottom: equivalent circuit model showing the voltage source, which is the driver, the output-source impedance of the driver gate itself, and the transmission line modeled as a resistor, which is valid during the round-trip time of flight of the transmission line.

The actual voltage launched into the transmission line can be calculated by modeling this circuit as a resistive-voltage divider. The signal sees a voltage divider composed of the source resistance and the transmission line's impedance. The magnitude of the voltage initially launched into the line is the ratio of the impedance of the line to the series combination of the line and source resistance. It is given by:

$$V_{launched} = V_{ouput}\left(\frac{Z_0}{R_{source} + Z_0}\right) \tag{7-16}$$

where:

$V_{launched}$ = the voltage launched into the transmission line

V_{output} = the voltage from the driver when driving an open circuit

R_{source} = the output-source impedance of the driver

Z_0 = the characteristic impedance of the transmission line

When the source resistance is high, the voltage launched into the line will be low—usually not a good thing. In Figure 7-17, we plot the percentage of the source voltage that actually gets launched into the transmission line and propagates down

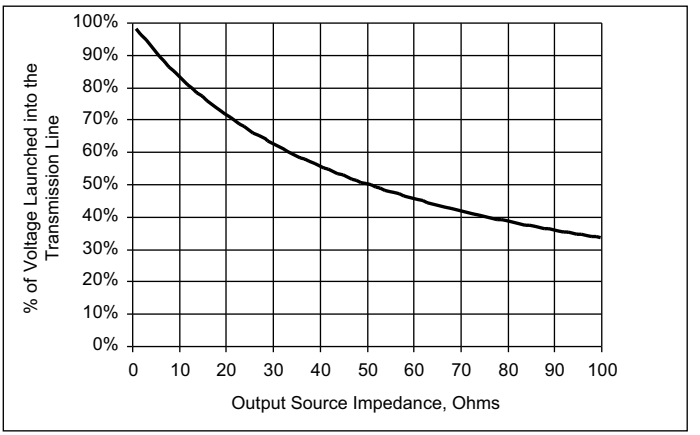


Figure 7-17 Amount of voltage launched into a 50-Ohm transmission line as the output source impedance of the driver varies.

it, for a characteristic impedance of 50 Ohms. When the output-source impedance is also 50 Ohms, we see that only half the open-circuit voltage is actually launched into the line. If the output is 3.3 volts, the signal launched into the line is only 1.65 volts. This is probably not enough to reliably trigger a gate that may be connected to the line. However, as the output resistance of the driver decreases, the signal voltage into the line increases.

T I P In order for the voltage, initially launched into the line, to be close to the source voltage, the driver's output source resistance must be small—significantly less than the characteristic impedance of the line.

We say that in order to "drive a transmission line," in other words, launch a voltage into the line that is close to the source voltage, we need an output impedance of the driver that is very small compared to the characteristic impedance of the line. If the line is 50 Ohms, then we need a source impedance less than 10 Ohms, for example.

Output devices that have exceptionally low output impedances, 10 Ohms or less, are often called *line drivers*, because they will be able to inject a large percentage of their voltage into the line. Older technology CMOS devices were not able to drive a line since their output impedances were in the 90-Ohm to 130-Ohm range. Since most interconnects behave like transmission lines, current-generation,

high-speed CMOS devices must all be able to drive a line and are designed with low-output impedance gates.

7.13 Return Paths

In the beginning of this chapter, we emphasized that the second trace is not the ground but the return path. We should always remember that *all* current, without exception, travels in loops.

> **TIP** There is *always* a current loop, and if some current goes out to somewhere, it will always come back to the source.

Where is the current loop in a transmission line? Suppose we have a microstrip that is very long. In this first case, we'll make it so long that the one-way time delay, TD, is 1 second. This is about the distance from the earth to the moon. To make it easier to think about for now, we will short the far end. We launch a signal into the line. This is shown in Figure 7-18. We said at the beginning, this meant that we had a constant current going into the signal path, related to the voltage applied and the characteristic impedance of the line.

If current travels in loops and must return to the source, eventually we'd expect to see the current loop back at the end and flow back down the return path. But, how long does this take? The current flow in a transmission line is very subtle. When do we see the current come out the return path? Does it take 2 seconds—1 second to go down and 1 second to come back? What would happen then if the far end were really open? If there is insulating dielectric material between the signal and return conductors, how could the current possibly get from the signal to the return conductor, except at the far end?

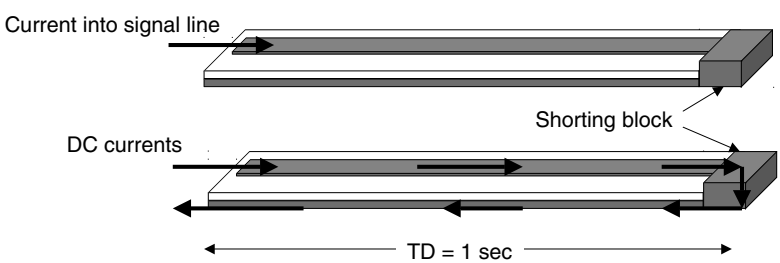


Figure 7-18 Current injected into the signal path of a transmission line and the current distribution after a long time. When does the current actually exit the return path?

The best way of thinking about it is by going back to the zeroth-order model, which describes the line as a bunch of tiny capacitors. This is shown in Figure 7-19. Consider the current flow initially. As the signal launches into the line it sees the first capacitor. As we described in an earlier chapter, if the voltage is constant, there will be no current flow through the capacitor. The only way current flows through a capacitor is if the voltage across it changes. As the signal is launched into the transmission line, the voltage across the signal- and return-path conductors ramps up. It is during this transition time, as the edge passes by, that the voltage is changing and current flows through the initial capacitor. As current flows into the signal path to charge up the capacitor, the exact same amount of current flows out of the return path, having gone through the capacitor.

In the first picosecond, the signal has not gotten very far down the line and it has no idea how the rest of the line is configured, whether it is open, shorted, or has some radically different impedance. The current flow back to the source, through the return path, depends only on the immediate environment and the region of the line where the voltage is changing, that is, where the signal edge is.

We can extend the transmission line model to include the rest of the signal and return paths with all the various distributed capacitors between them. As the signal propagates down the line, there is current—the return current—flowing through the capacitance to the return-path conductor and looping back to the source. However, this current from the signal path, through to the return path, flows between them only where the signal voltage is changing.

A few nsec after the signal launch, near the front end, the signal voltage is constant and there is no current flow between the signal to the return path. There is just constant current flowing into the signal conductor and back out the return

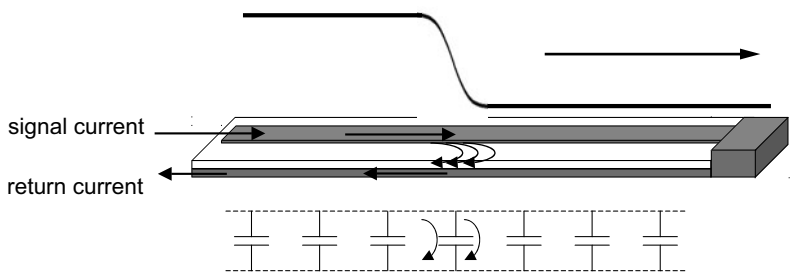


Figure 7-19 Signal current gets to the return path through the distributed capacitance of the transmission line. Current is flowing only from the signal conductor to the return conductor where the signal voltage is changing—where there is a dV/dt.

conductor. Likewise, in front of the signal edge, before the edge has gotten to that region of the line, the voltage is constant and there is no current flow between the signal and return paths. It is only at the signal edge that current flows through the distributed capacitance.

Once the signal is launched into the line, it will propagate down the line as a wave, at the speed of light. Current will flow down the signal line, pass through the capacitance of the line, and travel back through the return path as a loop. The front of this current loop propagates outward coincident with the voltage edge. We see that the signal is not only the voltage wave front, but it is also the current loop, which is propagating down the line. The instantaneous impedance the signal sees is the ratio of the signal voltage to the signal current.

Anything that disturbs the current loop will disturb the signal and cause a distortion, compromising signal integrity. To maintain good signal integrity, it is important to control both the current wave front and the voltage wave front. The most important way of doing this is to keep the impedance the signal sees constant.

TIP Anything that affects the signal current *or* the return-current path will affect the impedance the signal sees. This is why the return path should be designed just as carefully as the signal path, whether it is on a PCB, a connector, or an IC package.

When the return path is a plane, it is appropriate to ask where does the return current flow? What is its distribution in the plane? The precise distribution is slightly frequency dependent and is not easy to calculate with pencil and paper. This is where a good 2D field solver comes in handy.

An example of the current distribution in a microstrip and a stripline for 10-MHz and 100-MHz sine waves of current is shown in Figure 7-20. We can see two important features. First, the signal current is only along the outer edge of the signal trace. This is due to skin depth. Second, the current distribution in the return path is concentrated underneath the signal line. The higher the sine-wave frequency, the more tightly concentrated this current distribution will be.

As the frequency increases, the current in the return path will take the path of lowest impedance. This translates into the path of lowest loop inductance, which means the return current will move as close to the signal current as possible. The higher the frequency, the greater the tendency for the return current to flow directly under the signal current. Even at 10 MHz, the return current is highly localized.

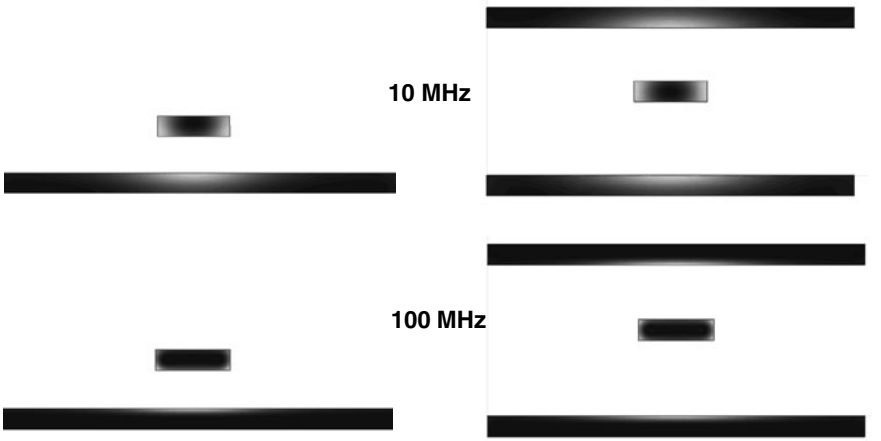


Figure 7-20 Current distribution in the signal and return paths for a microstrip and a stripline at 10 MHz and 100 MHz. In both cases, the line width is 5 mils and it is 1-ounce copper. Lighter color indicates higher current density. Results calculated with Ansoft's 2D Extractor.

In general, for frequencies above about 100 kHz, most of the return current flows directly under the signal trace. Even if the trace snakes around a curvy path or makes a right-angle bend, the return current in the plane follows it. By taking this path, the loop inductance of the signal and return will be kept to a minimum.

TIP Anything that prevents the return current from closely following the signal current, such as a gap in the return path, will increase the loop inductance and increase the instantaneous impedance the signal sees, causing a distortion.

We see that the way to engineer the return path is to control the signal path. Routing the signal path around the board will also route the return current path around the board. This is a very important principle of circuit board routing.

7.14 When Return Paths Switch Reference Planes

Cables are specifically designed with a return path adjacent to the signal path. This is true for coax and twisted pair cables. The return path is easy to follow. In planar interconnects found in circuit boards, the return paths are usually designed as planes, as in multilayer boards. For a microstrip, there is one plane directly underneath the signal path and the return current is easy to identify. But what if

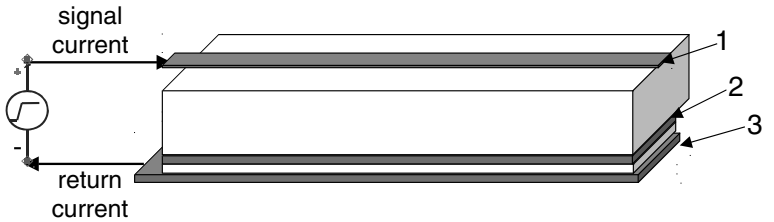


Figure 7-21 Driving a transmission line with the adjacent plane not being the driven return-path plane. What is the return-current distribution?

the plane adjacent to the signal path is not the plane that is being driven? What if the signal is introduced between the signal path and another plane, as shown in Figure 7-21. What will the return path do?

Current will always distribute so as to minimize the impedance of the signal-return loop. Right at the start of the line, the return path will couple between the bottom plane on layer 3 to the middle plane on layer 2 and then back to the signal path on layer 1.

One way to think about the currents is in the following way: The signal current in the signal path will induce eddy currents in the upper surface of the floating middle plane, and the return current in the bottom plane will induce eddy currents in the underside surface of the middle plane. These induced eddy currents will connect at the edge of the plane where the signal and return currents are being injected in the transmission line. The current flows are diagrammed in Figure 7-22.

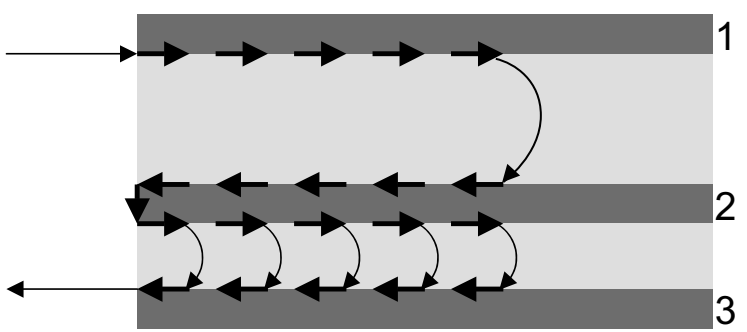


Figure 7-22 Side view of the current flow when driving a transmission line with the adjacent plane not being the driven return-path plane.

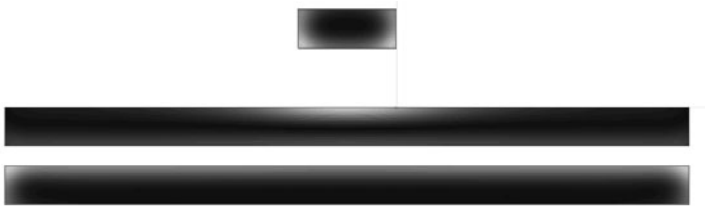


Figure 7-23 Current distribution, viewed end on, in the three conductors when the top trace and bottom plane are driven with the middle plane floating. Induced eddy currents are generated in the floating plane. Lighter colors are higher-current density. Calculated with Ansoft's 2D Extractor.

The precise current distribution in the planes will be frequency dependent, driven by skin-depth effects. In general, they will distribute in each plane to decrease the total loop inductance in the signal-return path. They can only be accurately calculated with a field solver. Figure 7-23 shows an example of the calculated current distribution, viewed from the end, for this sort of configuration for 2-mil-thick conductors at 20 MHz.

What impedance will the driver see looking into the transmission line between the signal line and the bottom plane? The driver, driving a signal into the signal and return paths with the floating plane between them, will see two transmission lines in series. There will be the transmission line composed of the signal and plane on layer 2 and the transmission line composed of the two planes on layers 2 and 3. This is illustrated in Figure 7-24. The series impedance the driver will see is:

$$Z_{driver} = Z_{1-2} + Z_{2-3} \tag{7-17}$$

The smaller the impedance between the two planes, $Z_{2–3}$, the closer to $Z_{1–2}$ the driver will see as the impedance looking into this transmission line.

This means that even though the driver is connected to the signal line and the bottom plane, the impedance the driver sees is really dominated by the impedance of the transmission line formed by the signal path and its most adjacent plane. *This is true no matter what the voltage of the adjacent plane is.* This is a startling result.

TIP The impedance a driver will see for a transmission line in a multilayer board will be dominated by the impedance between the signal line and the most adjacent planes, independent of which plane is actually connected to the driver's return.

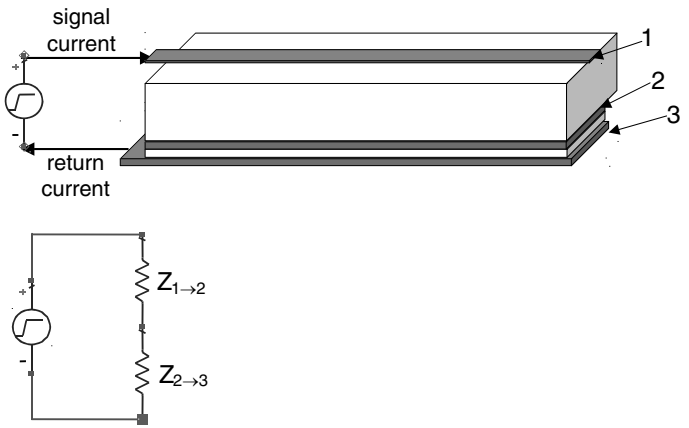


Figure 7-24 Top: Physical configuration of driver driving the transmission line with a floating plane between them. Bottom: Equivalent circuit model showing the impedance the driver sees as the sum of the impedance from the signal to the floating plane and between the two planes.

The smaller the impedance between the two planes compared to the impedance between the signal and its adjacent plane, the closer the driver will just see the impedance between the signal and floating plane.

The characteristic impedance between two long, wide planes, providing h << w, can be approximated by:

$$Z_0 = \frac{377\Omega}{\sqrt{\varepsilon_r}} \frac{h}{w} \tag{7-18}$$

where:

Z_0 = the characteristic impedance of the planes

h = the dielectric thickness between the planes

w = the width of the planes

ε_r = the dielectric constant of the material between the planes

For example, in FR4, for planes that are 2 inches wide with a 10-mil separation, the characteristic impedance a signal will see between the planes is about 377 Ohms × 2 × 0.01/2 = 3.8 Ohms. If the separation were 2 mils, the impedance between the planes would be 377 Ohms × 2 × 0.002/2 = 0.75 Ohm. When the plane-to-plane impedance is much less than 50 Ohms, it is less important which

plane is actually DC-connected to the driver and the more likely the most adjacent planes will dominate the impedance.

TIP The most important way to minimize the impedance between adjacent planes is to use as thin a dielectric between the planes as possible. This will keep the impedance between the planes low and provide tight coupling between the planes.

TIP When planes are tightly coupled and there is a low impedance between them, as there should be for low rail collapse anyway, it doesn't matter to which plane the driver actually connects. The coupling between the planes will provide a low-impedance path for the return current to get as close as possible to the signal current.

What if the signal path changes layers in mid-trace? How will the return current follow? Figure 7-25 shows a four-layer board with the signal path starting on layer 1 and transitioning through a via to layer 4. In the first half of the board, the return current must be on the plane directly underneath the signal path, on layer 2. In addition, for sine-wave-frequency components of the current above 10 MHz, the current is actually flowing on the top surface of the layer 2 plane.

In the bottom half of the board, where the signal is on layer 4, where will the return current be? It has to be on the plane adjacent to the signal layer, or on plane layer 3. And, it will be on the undersurface of the plane. In the regions of uniform transmission line, the return currents are easy to follow. The via is clearly the path the signal current takes to go from layer 1 to layer 4. How does the return current make the transition from layer 2 to layer 3?

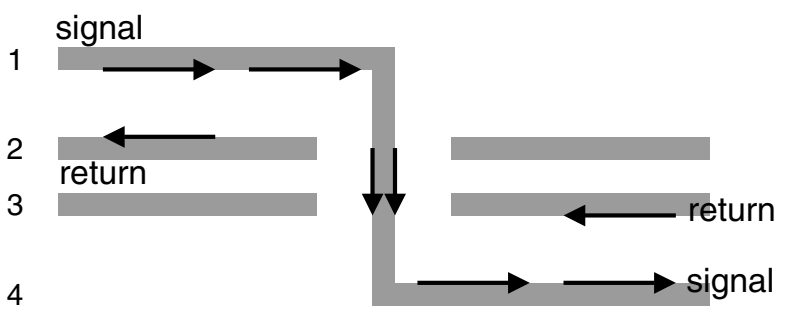


Figure 7-25 Cross section of a four-layer board with signal current transitioning from layer 1 to layer 4 through a via. How does the return current transition from layer 2 to layer 3?

If the two planes are at the same potential and have a via shorting them, the return current would take this path of low impedance. There might be a small jog for the return current, but it would be through a short length of plane, and a plane has low total inductance so would not have a large impedance discontinuity. This would be the preferred stack-up. If there were no other constraints such as cost, keeping the nearest reference planes at the same voltage and shorting them together close to the signal via is the optimum design rule. To reduce the voltage drop in the return path, always consider adding a return via adjacent to the signal via.

However, to minimize the total number of layers, it is sometimes necessary to use adjacent reference layers with different voltages. If plane 2 is a 5-v plane and plane 3 is a 0-v plane, there would be no DC path between them. How will the return current flow from plane 3 to plane 2?

It can only go through the capacitance between them. At the clearance hole, the return current will snake around and change surfaces on the same plane. The return current will then spread out on the inner surfaces of the planes and couple through the plane-to-plane capacitance. The current will spread out between the planes at the speed of light in the dielectric. The current flows in the return path are shown in Figure 7-26. The two return-path planes create a transmission line, and the return current will see an impedance that is the instantaneous impedance of the two planes.

TIP Whenever the return-path current switches planes and the planes are DC isolated, the return current will couple through the planes and see an impedance equal to the instantaneous impedance of the transmission line created by the planes.

The return current will have to go through this impedance and there will be a voltage drop in the return path. We call this voltage drop in the return path *ground bounce*. The larger the impedance of the return path, the larger the voltage drop, and the larger the ground-bounce noise generated. All other signal lines that are also changing reference planes will contribute to this ground-bounce voltage noise and encounter the ground-bounce noise created by the other signals.

TIP The goal in designing the return path is to minimize the impedance of the return path to minimize the ground-bounce noise generated in the return path. As we will see, this is primarily implemented by keeping the impedance between the return planes as low as possible, usually by keeping them on adjacent layers with as thin a dielectric as possible between them.

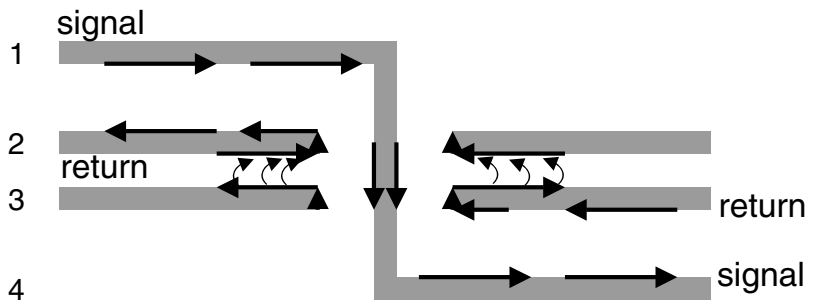


Figure 7-26 The return current transitions from layer 2 to layer 3 by capacitive coupling between the layers.

As the return current spreads out between the two return-path planes in an ever-expanding circle from the signal via, it will see an ever-decreasing instantaneous impedance. The capacitance per length that the signal sees gets larger as the radius gets larger. This makes the analysis complicated except for specific cases and in general, requires a field solver.

However, we can build a simple model to estimate the instantaneous impedance between the two planes and to provide insight into how to optimize the design of a stack-up and minimize this form of ground bounce.

To calculate the instantaneous impedance the signal sees as it propagates radially outward between the two planes, we need to calculate the capacitance per length of the radial-transmission line and the speed of the signal. The capacitance per length the signal will see is the change in capacitance per incremental change in radius. The total capacitance the return current sees is:

$$C = \varepsilon_0 \varepsilon_r \frac{A}{h} \tag{7-19}$$

and the area between the planes is:

$$A = \pi r^2 \tag{7-20}$$

Combining these relationships shows the capacitance increasing with distance as:

$$C = \varepsilon_0 \varepsilon_r \frac{A}{h} = \varepsilon_0 \varepsilon_r \frac{\pi r^2}{h} \tag{7-21}$$

where:

C = the coupling capacitance between the planes

ε_0 = permittivity of free space, 0.225 pF/in

ε_r = the dielectric constant of the material between the planes

A = the area of overlap of the return current in the planes

h = the spacing between the planes

r = the increasing radius of the coupling circle, expanding at the speed of light

The incremental increase in capacitance as the radius increases (i.e., the capacitance per length) is:

$$C_L = 2\pi\varepsilon_0 \varepsilon_r \frac{r}{h} \tag{7-22}$$

As expected, as the return current moves farther from the via, the capacitance per length increases. The instantaneous impedance this current will see is:

$$Z = \frac{1}{vC_L} = \frac{\sqrt{\varepsilon_r}}{c} \times \frac{h}{2\pi r \varepsilon_0 \varepsilon_r} = \frac{377\Omega}{2\pi} \frac{h}{r\sqrt{\varepsilon_r}} = 60\Omega \frac{h}{r\sqrt{\varepsilon_r}} \tag{7-23}$$

where:

Z = the instantaneous impedance the return current sees between the planes

C_L = the coupling capacitance per length between the planes

v = the speed of light in the dielectric

ε_0 = permittivity of free space, 0.225 pF/in

ε_r = the dielectric constant of the material between the planes

h = the spacing between the planes

r = the increasing radius of the coupling circle, expanding at the speed of light

c = the speed of light in a vacuum

For example, if the dielectric thickness between the planes is 10 mils, the impedance the return current will see by the time it is 1 inch away from the via is $Z = 60 \times 0.01/(1 \times 2) = 0.3$ Ohm. And, this impedance will get smaller as the return current propagates outward. This says that the farther from the via, the lower the impedance the return current sees, and the lower the ground-bounce voltage will be across this impedance.

We can relate the impedance the return current sees (which is the impedance in series with the signal current) to time, since the return current spreads out at the speed of light in the material and $r = v \times t$:

$$Z = 60\Omega\frac{h}{r\sqrt{\varepsilon_r}} = 60\Omega\frac{h}{vt\sqrt{\varepsilon_r}} = 60\Omega\frac{h\sqrt{\varepsilon_r}}{ct\sqrt{\varepsilon_r}} = 5\Omega\frac{h}{t} \quad (7\text{-}24)$$

where:

Z = the instantaneous impedance the return current sees between the planes, in Ohms

v = the speed of light in the dielectric

ε_r = the dielectric constant of the material between the planes

c = the speed of light in a vacuum

h = the spacing between the planes, in inches

t = the time the return current is propagating, in nsec

For example, for a spacing of 0.01 inch, the impedance the return current sees after the first 0.1 nsec is $Z = 5 \times 0.01/0.1 = 0.5$ Ohm. The very first leading edge of the signal will see an initial impedance that can be as large as 0.5 Ohm. If the signal current were 20 mA in the first 100 psec, corresponding to a 1-v signal in a 50-Ohm impedance, the ground-bounce voltage drop across the switching planes that are in series with the signal voltage, for the first 0.1 nsec, would be 20 mA × 0.5 Ohm = 10 mV.

This may not seem like a lot compared to the 1-v signal, but if 10 signals transition simultaneously between the same reference planes, all within less than 0.6 inch of each other, they will each see the same common 0.5-Ohm impedance and there will be 20 mA × 10 = 200 mA of current through the return-path impedance. This will generate a ground-bounce noise of 200 mA × 0.5 Ohm = 100 mV, which is now 10% of the signal voltage and can be significant. All signal paths

that also transition through this path will see the 100 mV of ground-bounce noise, even if they are not switching.

The initial impedance the return current will see can be high, if there is substantial current flowing in a very short time initially. All current flowing during this initial short time will see a high impedance and generate ground-bounce voltages. Figure 7-27 shows the impedance the return current sees over time. From this plot, it is clear that the impedance of the return path is significant only for very fast rise times, substantially less than 0.5 nsec.

When the impedance is significant—about 5% of 50 Ohms—for one signal line switching, the impedance of the return path is important. When there are n signal paths switching through these planes, the maximum acceptable impedance of the return path is 2.5 Ohms/n.

TIP This analysis indicates that for fast edges and multiple signals switching simultaneously between reference planes, significant ground-bounce voltage can be generated across the return path. The only way to minimize this ground-bounce voltage is by decreasing the impedance of the return path.

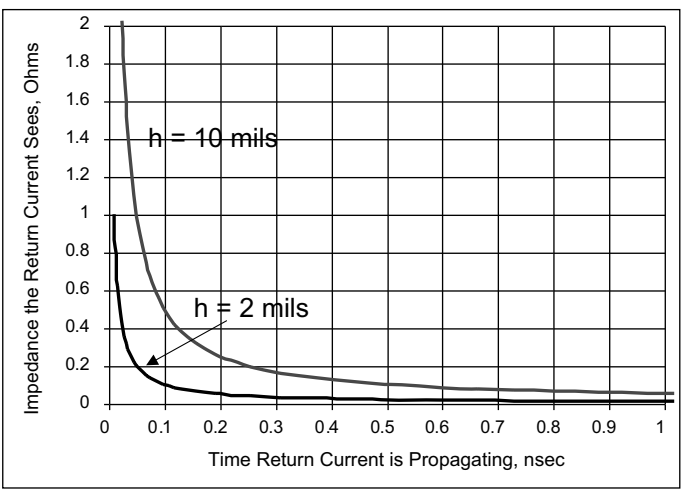


Figure 7-27 The impedance the return current sees between the planes for 2-mil spacing and 10-mil spacing, as the signal propagates outward from the via.

This is primarily accomplished by:

1. Make sure that as the signal path transitions layers, it always has an adjacent plane with the same voltage level for its reference and there is a shorting via between the switching planes in close proximity to any signal vias.
2. Keep the spacing between different DC-voltage-level reference planes as thin as possible.
3. Space out adjacent switching vias so that the return currents do not overlap during the initial transients when the impedance of the return path can be high.

It is sometimes believed that adding a decoupling capacitor between two reference planes for which there is a switching return current will help to decrease the impedance of the return path. It is hoped the added discrete capacitor will provide a low-impedance path for the return current to flow from one reference plane to the other.

TIP To provide any effectiveness, the real capacitor must keep the impedance between the planes less than 5% × 50 Ohms or 2.5 Ohms, at the bandwidth of the rise time's frequency components.

A real capacitor has some loop inductance and some equivalent series resistance associated with it. This will limit the usefulness of discrete decoupling capacitors for very short rise-time signals. After all, the impedance between the planes will already be low after long periods or for the low-frequency components.

For the higher-frequency components that could use a discrete component, it is not the capacitance that determines the real capacitor's impedance; it is its equivalent series inductance. Figure 7-28 shows the expected impedance of a real 1-nF capacitor with a loop inductance of 0.5 nH. This is an extremely optimistic loop inductance and can only be achieved by either a multiple via in-pad configuration or with the use of interdigitated capacitors (IDCs).

This real capacitor will provide a low impedance for the return path for signal bandwidths only up to 1 GHz. Higher capacitance than 1 nF will provide no added value, since the low-frequency components will see a low impedance between the planes anyway.

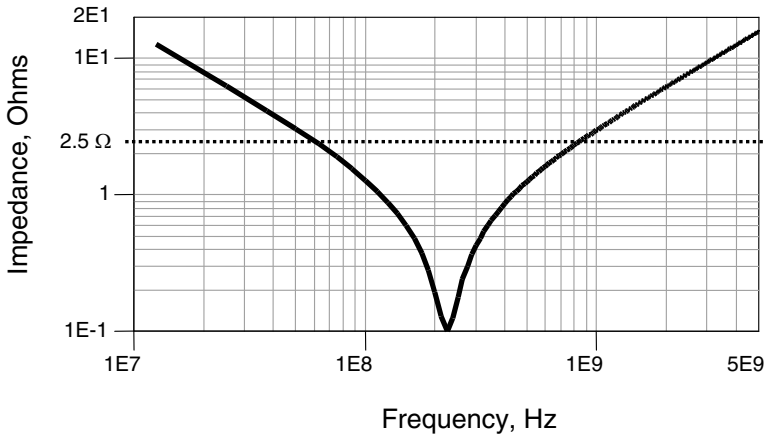


Figure 7-28 Impedance of a capacitor with 1-nF capacitance and only 0.5 nH of loop inductance.

TIP When using a discrete capacitor to decrease the impedance of the return path, it is far more important to keep the series inductance low than to use more than 1 nF of capacitance.

Unfortunately, even the 0.5 nH of loop inductance of a well-designed capacitor will still have a large impedance at frequencies above 1 GHz, where the planes do not have low impedance and their ground bounce may be a problem.

TIP A capacitor between different DC-voltage planes will not be very effective at managing the ground bounce from switching planes. It may, however, provide additional decoupling for lower-frequency noise but will not solve the ground-bounce problem as rise times continue to decrease.

TIP In a multilayer board, when signal paths must change different voltage-level reference layers, the only way of minimizing the ground-bounce voltage generated is by using a dielectric between the reference planes that is as thin as possible.

There is one additional problem created when a signal changes reference planes and current is driven in the transmission line formed by the two adjacent planes. Where does this current end up? As the current propagates outward, eventually it will hit the edges of the board. The current injected between the planes

when the signal current switches will rattle around between the planes, causing transient voltages between the two planes.

The impedance between the planes is very low, much less than 1 Ohm, so the transient voltages created will be low. However, with multiple signals switching between planes, each one injects some noise between the planes. This noise will get larger with more signals switching. The current injected into the planes is determined by the signal impedance, roughly 50 Ohms. The voltage noise generated between the planes depends on the impedance of the planes. To minimize this rattling voltage, the impedance of the planes should be kept to a minimum by keeping their spacing thin.

Signals transitioning reference planes is a dominant source of noise injected between the planes. This voltage noise will rattle around and contribute to the noise in the PDN. In low noise systems, it can be a major source of cross talk to very sensitive lines, such as rf receivers or analog to digital converter inputs or voltage references. To minimize the noise in these systems, care should be taken to minimize the return current injected between the planes by the careful selection of return plane voltages, return vias, and low inductance decoupling capacitors.

We sometimes call this voltage, which is rattling between the edges of the board between the adjacent plane layers, resonances in the planes. These resonances will eventually die out from conductor and dielectric attenuation. They will have frequency components that match the round-trip time of flight between the board dimensions. For boards with a dimension of 10–20 inches on a side, the resonant frequencies will be in the 150-MHz to 300-MHz range. This is why capacitors between the different voltage planes can provide some value. They help to maintain a low impedance between the planes (in the board resonant-frequency range) and keep the voltages between the planes low. However, as we saw, they do not affect the transient ground-bounce voltage during the very fast transition.

TIP To minimize the resonant rattling-around voltage, especially in small multilayer packages, it is important to avoid having any return currents switch between different reference voltage layers. The adjacent return layers should be at the same DC-voltage level, and a return via should connect the return path in proximity to the signal via. This will avoid injecting any currents between planes and avoid driving any plane resonances.

These problems will get more important as rise times decrease, especially below 100 psec.

7.15 A First-Order Model of a Transmission Line

An ideal transmission line is a new ideal circuit element that has the two special properties of a constant instantaneous impedance and a time delay associated with it. This ideal model is a "distributed" model in the sense that the properties of an ideal transmission line are distributed over its length rather than being concentrated in a single lumped point.

Physically, a controlled-impedance transmission line is composed of just two conductors with length and having a uniform cross section for the whole length. Earlier in this chapter, we introduced a zeroth-order model, which described a transmission line as a collection of capacitors spaced some distance apart. This was a physical model, not an equivalent electrical model.

We can further *approximate* the physical transmission line by describing the sections of signal- and return-path conductors as a loop inductance. The simplest equivalent circuit model for a transmission line would have small capacitors separated by small loop inductors, as shown in Figure 7-29. The C is the capacitance between the conductors and the L is the loop inductance between the sections.

Each segment of the signal path has some partial self-inductance associated with it, and each segment of the return path has some partial self-inductance associated with it. There is some partial mutual inductance between each of the signal- and return-path segments between the discrete capacitors. For an unbalanced transmission line, such as a microstrip, the partial self-inductances of the signal- and return-path segments are different. In fact, the partial self-inductance of the signal path can be more than 10 times larger than the partial self-inductance of the return path.

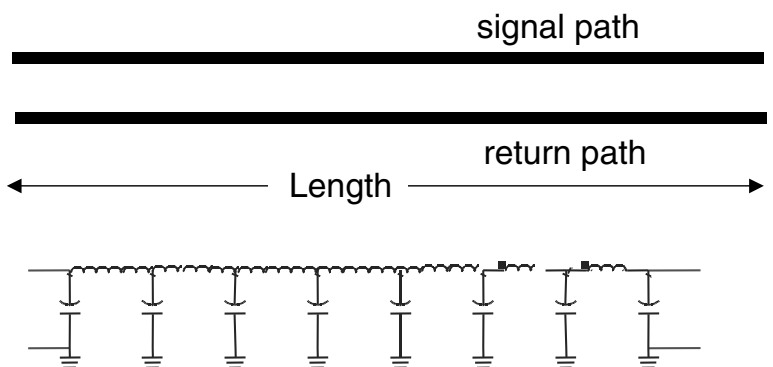


Figure 7-29 Top: physical structure of a uniform transmission line. Bottom: first-order equivalent circuit model approximation for a transmission line based on combinations of L's and C's.

However, from a signal's perspective as it propagates down the line, it is a current loop, from the signal to the return path, that propagates. In this sense, all the signal current sees is the loop inductance down the signal-path segment and returning back through the return-path segment. For signal propagation in a transmission line and most cross-talk effects, the partial inductances of the signal and return paths do not play a role. It is only the loop inductance that is important. When approximating an ideal distributed-transmission line as a series of LC segments, the inductance the signal sees, as represented in the model, is really a loop inductance.

TIP It is important to keep in mind that this lumped circuit model is an approximation of an ideal transmission line. In the extreme case, as the size of the capacitors and inductors gets diminishingly small and the number of each increases, the approximation gets better and better.

In the extreme, when each capacitor and inductor is infinitesimal and there is an infinite number of elements, there is a uniform capacitance per length, C_L, and a uniform loop inductance per length, L_L. These are often called the line parameters of a transmission line. Given the total length of the line, Len, the total capacitance is given by:

$$C_{total} = C_L \times Len \tag{7-25}$$

and the total inductance is given by:

$$L_{total} = L_L \times Len \tag{7-26}$$

where:

C_L = capacitance per length

L_L = loop inductance per length

Len = length of the transmission line

Just by looking at this LC circuit, it is difficult to get a feel for how a signal will interact with it. At first glance, we might think there will be a lot of oscillations and resonances. But what happens when the size of each element is infinitely small?

The only way to really know how a signal will interact with this circuit model is to apply network theory and solve the differential equations represented by this LC network. The results indicate that a signal traveling down the network will see a constant instantaneous impedance at each node. This constant instantaneous impedance is the same as the instantaneous impedance we discovered for an ideal distributed-transmission-line element. The instantaneous impedance is numerically the characteristic impedance of the line. Likewise, there will be a finite delay between the time the signal is introduced into the front of the LC network and the time it comes out.

Using network theory, we can calculate how the characteristic impedance and time delay depend on the line parameters and total length of the line:

$$Z_0 = \sqrt{\frac{L_L}{C_L}} \tag{7-27}$$

$$TD = \sqrt{C_{total}L_{total}} = Len \times \sqrt{C_L L_L} = \frac{Len}{v} \tag{7-28}$$

$$v = \frac{Len}{TD} = \frac{1}{\sqrt{C_L L_L}} \tag{7-29}$$

where:

Z_0 = the characteristic impedance, in Ohms

L_L = the loop inductance per length of the transmission line

C_L = the capacitance per length of the transmission line

TD = the time delay of the transmission line

L_{total} = the total loop inductance in the transmission line

C_{total} = the total capacitance in the transmission line

v = the speed of the signal in the transmission line

Without having to invoke a finite speed for the propagation of a signal down a transmission line, the electrical properties of an LC network predict this behavior.

Likewise, though it is hard to tell looking at the circuit model, network theory predicts there is a constant impedance the signal sees at each node along the circuit.

These two predicted properties (the characteristic impedance and the time delay) must match the same values we derived based on the zeroth-order physics model of finite speed and collection of capacitor buckets. By combining the results from these two models, a number of very important relationships can be derived.

Because the speed of a signal depends on the dielectric constant of the material and on the capacitance per length and inductance per length, we can relate the capacitance per length to the inductance per length:

$$v = \frac{c}{\sqrt{\varepsilon_r}} = \frac{1}{\sqrt{C_L L_L}} \tag{7-30}$$

$$L_L = 7\frac{\varepsilon_r}{C_L}\frac{nH}{inch} \tag{7-31}$$

$$C_L = 7\frac{\varepsilon_r}{L_L}\frac{pF}{inch} \tag{7-32}$$

From the relationship between the characteristic impedance and velocity, the following relationships can be derived:

$$C_L = \frac{1}{vZ_0} = \frac{1}{cZ_0}\sqrt{\varepsilon_r} = \frac{83}{Z_0}\sqrt{\varepsilon_r}\frac{pF}{inch} \tag{7-33}$$

$$L_L = \frac{Z_0}{v} = 0.083Z_0\sqrt{\varepsilon_r}\frac{nH}{inch} \tag{7-34}$$

And, from the time delay and the characteristic impedance of the transmission line, the following can be derived:

$$C_{total} = \frac{TD}{Z_0} \tag{7-35}$$

$$L_{total} = TD \times Z_0 \tag{7-36}$$

where:

Z_0 = the characteristic impedance, in Ohms

L_L = the loop inductance per length of the transmission line, in nH/inch

C_L = the capacitance per length of the transmission line, in pF/inch

TD = the time delay of the transmission line, in nsec

L_{total} = the total loop inductance in the transmission line, in nH

C_{total} = the total capacitance in the transmission line, in pF

v = the speed of the signal in the transmission line, in in/nsec

For example, a line that is 50 Ohms and has a dielectric constant of 4 has a capacitance per length of $C_L = 83/50 \times 2 = 3.3$ pF/inch. This is a startling conclusion.

TIP *All* 50-Ohm transmission lines with a dielectric constant of 4 have the same capacitance per length—about 3.3 pF/inch. This is a very useful rule of thumb to remember.

If the line width doubles, the dielectric spacing would have to double to maintain the same characteristic impedance, and the capacitance would stay the same. An interconnect in a BGA package that has been designed as a 50-Ohm controlled-impedance line, 0.5 inches long, has a capacitance of 3.3 pF/inch × 0.5 inch = 1.6 pF.

Likewise, the inductance per length of a 50-Ohm line made with FR4 is $L_L = 0.083 \times 50 \times 2 = 8.3$ nH/inch.

TIP *All* 50-Ohm transmission lines with a dielectric constant of 4 have the same loop inductance per length of about 8.3 nH/inch. This is a very useful rule of thumb to remember.

If the time delay of a line is 1 nsec and it is 50 Ohms, the total capacitance in the line is C_{total} = 1 nsec/50 = 20 pF. If the line is 6 inches long, this 20 pF of capacitance is distributed over 6 inches and the capacitance per length is 20 pF/6 inches = 3.3 pF/inch. For this same line, the total loop inductance going down the signal line and looping back through the return path is L_{total} = 1 nsec × 50 Ohms = 50 nH. Distributed over the 6 inches of length, this is an inductance per length of 50 nH/6 inches = 8.3 nH/inch.

These relationships that relate the capacitance, inductance, characteristic impedance, and dielectric constant, which are associated with a transmission line, apply to any transmission line. They make no assumptions on the cross-section geometry. They are very powerful tools to help us estimate one or more of these terms using existing approximations or field solvers. If we know any two, we can always find the others.

7.16 Calculating Characteristic Impedance with Approximations

Engineering a particular target characteristic impedance is really a matter of adjusting the line widths, dielectric thickness, and dielectric constants. If we know the length of the transmission line and the dielectric constant of the material around the conductors and we can calculate the characteristic impedance, we can use the relationships above to calculate all other parameters.

Of course, every different type of cross-sectional geometry will have a different relationship between the geometrical features and the characteristic impedance. In general, there are three types of analysis we can use to calculate the characteristic impedance from the geometry:

1. Rules of thumb

2. Approximations

3. 2D field solvers

The two most important rules of thumb relate the characteristic impedance of a microstrip and stripline fabricated with FR4. The cross sections for 50-Ohm transmission lines are illustrated in Figure 7-30.

T I P As a good rule of thumb, a 50-Ohm microstrip in FR4 has a line width twice the dielectric thickness. A 50-Ohm stripline has a total dielectric spacing between the planes equal to twice the line width.

There are only three cross-section geometries that have exact equations. All others are approximations. These three are coax, twin-round-wire, and round-wire-over-plane geometries. They are illustrated in Figure 7-31. The relationship between characteristic impedance and geometry for a coax is:

$$Z_0 = \frac{377\Omega}{2\pi\sqrt{\varepsilon_r}}\ln\left(\frac{b}{a}\right) = \frac{60\Omega}{\sqrt{\varepsilon_r}}\ln\left(\frac{b}{a}\right) \tag{7-37}$$

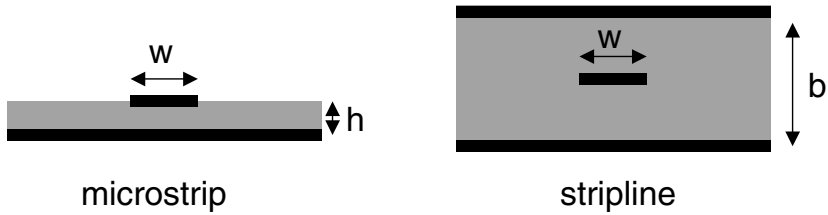


Figure 7-30 Scaled cross sections for 50-Ohm transmission lines in FR4. Left: 50-Ohm microstrip has w = 2 × h. Right: 50-Ohm stripline has b = 2 × w.

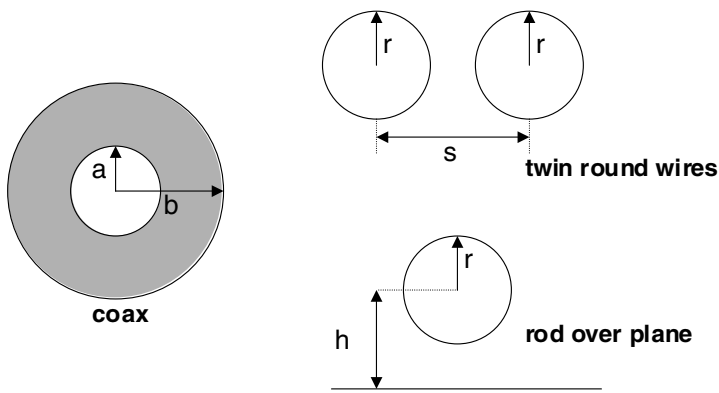


Figure 7-31 The only three cross sections for which there are exact equations for the characteristic impedance. All others are approximations.

For twin parallel round wires, the characteristic impedance is:

$$Z_0 = \frac{120\Omega}{\sqrt{\varepsilon_r}} \ln\left(\frac{s}{2r} + \sqrt{\left(\frac{s}{2r}\right)^2 - 1}\right) \tag{7-38}$$

For the case of a round rod over a plane, the characteristic impedance is:

$$Z_0 = \frac{60\Omega}{\sqrt{\varepsilon_r}} \ln\left(\frac{h}{r} + \sqrt{\left(\frac{h}{r}\right)^2 - 1}\right) \tag{7-39}$$

where:

Z_0 = the characteristic impedance, in Ohms

a = inner radius of the coax, in inches

b = outer radius of the coax, in inches

r = radius of the round rod, in inches

s = center-to-center spacing of the round rod, in inches

h = height of the center of the rod over the plane, in inches

ε_r = dielectric constant of the materials

These relationships assume the dielectric material completely and uniformly fills all space wherever there are electric fields. If this is not the case, the effective dielectric constant that affects the speed of the signal down the conductor will be some complicated combination of the mixture of dielectrics. This can often only be calculated with a field solver.

For homogeneous dielectric distributions, these are useful relationships to use to calibrate 2D field solvers, since they are exact.

TIP With very few exceptions, virtually all other equations used to relate the characteristic impedance to the geometry are approximations. When being off by more than 5% may increase design-time and costs excessively, approximations should not be used for final sign-off of the design of transmission lines. When accuracy counts, a verified 2D field solver should be used.

The value of approximations is that they show the relationship between the geometrical terms and can be used for sensitivity analysis in a spreadsheet. The most popular approximation for microstrip recommended by the IPC is:

$$Z_0 = \frac{87\Omega}{\sqrt{1.41 + \varepsilon_r}} \ln\left(\frac{5.98h}{0.8w + t}\right) \tag{7-40}$$

For stripline, the IPC recommended approximation is:

$$Z_0 = \frac{60\Omega}{\sqrt{\varepsilon_r}} \ln\left(\frac{2b + t}{0.8w + t}\right) \tag{7-41}$$

where:

Z_0 = characteristic impedance, in Ohms

h = dielectric thickness below the signal trace to the plane, in mils

w = line width, in mils

b = plane-to-plane spacing, in mils

t = metal thickness, in mils

ε_r = dielectric constant

If we ignore the impact from the trace thickness, t, then the characteristic impedance for both structures depends only on the ratio of the dielectric thickness to the line width. This is a very important relationship.

TIP To first order, the characteristic impedance of stripline and microstrip will scale with the ratio of the dielectric thickness to the line width. As long as this ratio is constant, the characteristic impedance will be constant.

For example, if the line width is doubled and the dielectric spacing is doubled, the characteristic impedance will be unchanged to first order.

Though these equations look complicated, this is no measure of accuracy. The only way to know the accuracy of an approximation is to compare it to the results from a verified field solver.

7.17 Calculating the Characteristic Impedance with a 2D Field Solver

If we need better than 10% accuracy or are worrying about second-order effects, such as the effect of trace thickness, solder-mask coverage, or side-wall shape, we should not be using one of these approximations. The most important tool that should be in every engineer's toolbox for doing impedance calculations is a 2D field solver.

The basic assumption made by all 2D field solvers is that the geometry is uniform. That is, as we move down the length of the transmission line, the cross-sectional geometry stays the same. This is the basic definition of a controlled-impedance line—that the cross section is uniform. In such a case, there is only one characteristic impedance that describes the line. This is why a 2D field solver is the right tool to provide an accurate prediction of the characteristic impedance of uniform transmission lines. Only the 2D cross-sectional information is important.

Whenever accuracy is important, a 2D field solver should be used. This means that before sign-off on a design and committing to hardware, the stack-up should be designed with a 2D field solver. Sometimes the argument is made, if the

manufacturing tolerance is 10%, that we should not care if the accuracy of the prediction is even 5%.

Accuracy is important because of yield. Any inaccuracy in the predicted impedance will shift the center position of the distribution of the manufactured impedances. The better we can center the distribution on the required target, the better we can increase yield. Even an inaccuracy of 1% will shift the distribution off center. Any product close to the spec limits could be shifted outside the limits and contribute to a yield hit. As we showed in an earlier chapter, the accuracy of a field solver can be better than 0.5%.

Using a 2D field solver, the characteristic impedance of a microstrip can be calculated. These results are compared to the predictions based on the IPC approximations. Figure 7-32 shows the predicted characteristic impedance of a microstrip with dielectric constant of 4, with 1/2-ounce copper and dielectric thickness of 10 mils, as the line width is varied. Near and above 50 Ohms, the agreement is very good. However, for lower impedances, the IPC approximation can be off by as much as 25%.

The same comparison is made for stripline and is shown in Figure 7-33. The agreement close to 50 Ohms is very good, but for low impedances, the approximation can be off by more than 25%. Approximations should not be used when accuracy counts.

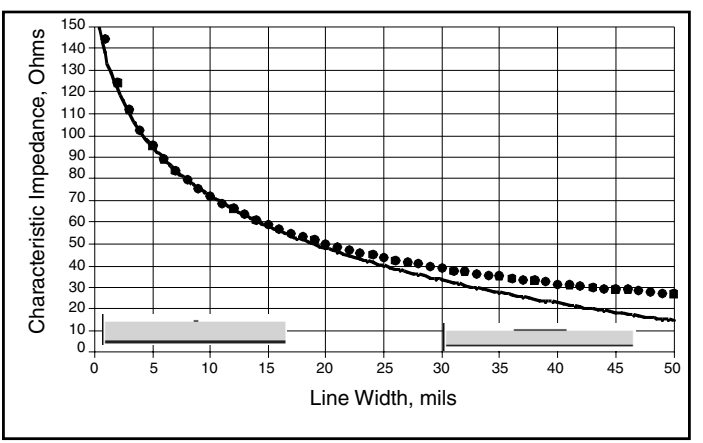


Figure 7-32 Comparison of field-solver results (circles) and the IPC approximation (line) for a microstrip with 10-mil-thick FR4 dielectric and 1/2-ounce copper. Field-solver results from the Ansoft 2D Extractor.

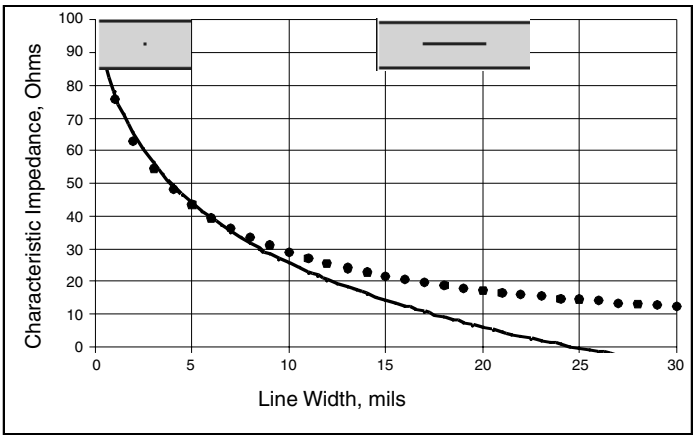


Figure 7-33 Comparison of field-solver results (circles) and the IPC approximation (line) for a stripline with 10-mil-thick FR4 dielectric and 1/2-ounce copper. Field-solver results from the Ansoft 2D Extractor.

In addition to providing an accurate estimate for the characteristic impedance, a 2D field solver can also provide insight into the impact from second-order effects, such as:

1. The width of the return path

2. The thickness of the signal-trace conductor

3. The presence of solder mask over a surface trace

4. The effective dielectric constant

How does the width of the return path influence the characteristic impedance? If the width of the return path is narrow, the capacitance should be smaller and the characteristic impedance would be higher. Using a field solver we can calculate how wide the return path has to be so that it doesn't matter anymore.

Figure 7-34 shows the calculated characteristic impedance of a microstrip with dielectric constant of 4 and trace thickness of 0.7 mil, corresponding to 1/2-ounce copper, dielectric thickness of 5 mils, and line width of 10 mils. This is nominally a 50-Ohm line. The width of the return path is varied and the characteristic impedance calculated. If the width of the extension on each side of the signal trace is greater than 15 mils, the characteristic impedance is less than 1% off from its infinite width value. Because the fringe fields along the edge of the trace scale with the dielectric thickness, this is the important scale factor. The extent of the

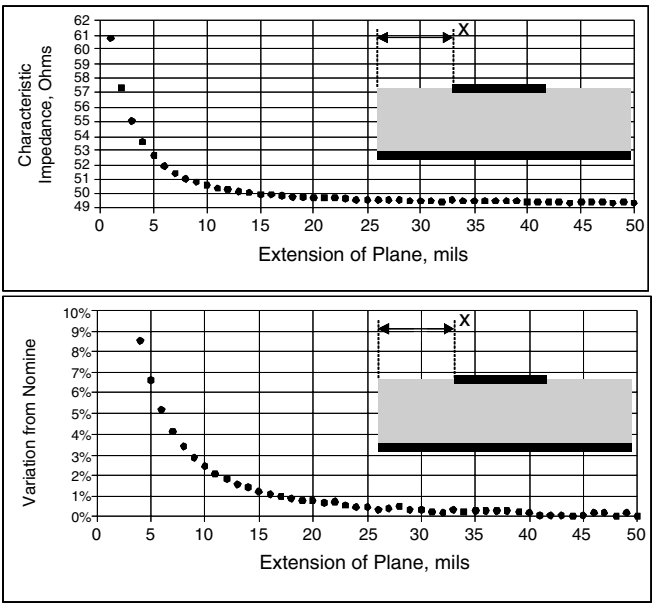


Figure 7-34 Top: calculated characteristic impedance of a microstrip with 10-mil-wide trace and 5-mil-thick dielectric, as the return-path plane is widened. Bottom: variation from the nominal value. Field solver results from the Ansoft 2D Extractor.

return path on either side of the signal path should be roughly 3 times the dielectric thickness, h.

> **TIP** As a rough rule of thumb, the return path should extend at least three times the dielectric thickness on either side of the signal trace for the characteristic impedance to not exceed more than 1% of the value when the return path is infinitely wide.

Using a 2D field solver, we can calculate the characteristic impedance as the trace thickness changes from 0.1 mil thick to 3 mils thick. This is plotted in Figure 7-35. Each point on the curve is the calculated characteristic impedance for a different trace thickness. It can be seen that as the metal thickness increases, the fringe field capacitance increases and the characteristic impedance decreases, as we would expect. Thicker metal means higher capacitance between the signal trace and the return path, which also means lower characteristic impedance. However, as we can see from the calculated results, this is not a large effect—it is second order.

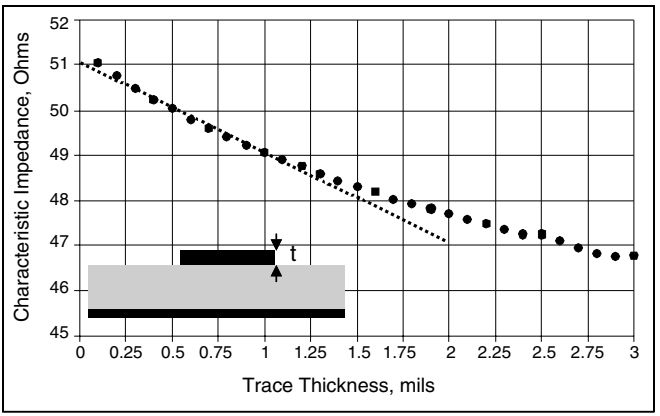


Figure 7-35 Calculated characteristic impedance of a nominal 50-Ohm microstrip as the trace thickness is changed. The circles are calculated results; the line is 2 Ohms/mil. Field solver results from the Ansoft 2D Extractor.

TIP As a rough rule of thumb, the decrease in characteristic impedance is about 2 Ohms per mil of signal-trace thickness.

If there is a thin solder-mask coating on top of the microstrip, the fringe field capacitance will increase and the characteristic impedance will decrease. For the same microstrip as above, but using a conductor only 0.1 mil thick, the characteristic impedance is calculated for an increasing solder-mask coating, using a dielectric constant of 4. Figure 7-36 shows the decrease in characteristic impedance. It is about 2 Ohms/mil of coating for thin coatings. After about 10-mil-thick coating, the characteristic impedance is no longer affected, as all the external fringe fields are contained in the first 10 mils of coating. This is a measure of how far the fringe fields extend above the surface.

Of course, solder mask is more typically 0.5 mil to 2 mils. In this regime, we see that the presence of the solder mask can lower the characteristic impedance by as much as 2 Ohms, a significant amount. To reach a target impedance with solder mask requires the line width to be narrower than nominal so the solder mask would bring the impedance back down to the target value.

This last example illustrates that the characteristic impedance depends on the dielectric distribution above the trace in a microstrip. Of course, in a stripline, all the fields are contained by the dielectric so solder mask on top of the upper plane has no impact on the characteristic impedance.

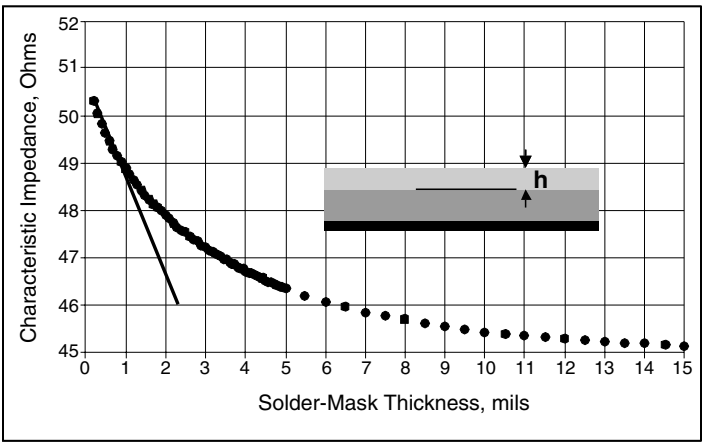


Figure 7-36 Calculated characteristic impedance of a nominal 50-Ohm microstrip as a solder mask with dielectric constant of 4 is added. Circles are the field solver results; the line is 2 Ohms/mil. Field solver results from the Ansoft 2D Extractor.

In addition to the characteristic impedance affected by the inhomogeneous dielectric distribution, the effective dielectric constant the fields experience is also affected by the dielectric distribution. In microstrip, the effective dielectric constant, which determines the speed of the signal, depends on the specific geometry of the dielectric. Often, the effective dielectric constant can only be accurately calculated with a field solver. This was discussed in detail in an earlier chapter.

7.18 An n-Section Lumped Circuit Model

The ideal transmission-line circuit element is a distributed element that very accurately predicts the measured performance of real interconnects. Figure 7-37 shows the comparison of the measured and simulated impedance of a 1-inch-long transmission line, in the frequency domain. We see the excellent agreement even up to 5 GHz, the bandwidth of the measurement.

T I P Real interconnects match the behavior of an ideal transmission line to high bandwidth. An ideal transmission line is a very good model for real interconnects.

We can approximate this ideal model by a combination of LC lumped-circuit sections. The question naturally arises, how do we know how many LC sections to use for a given level of accuracy and what happens if we use too few sections?

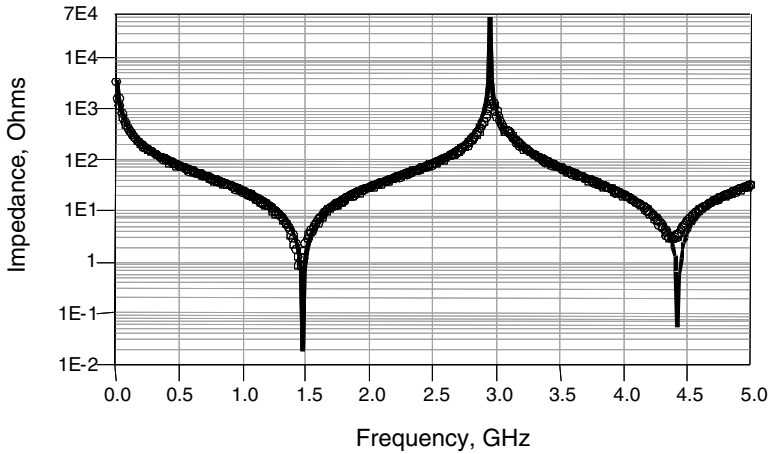


Figure 7-37 Measured (circles) and simulated (line) impedance of a 1-inch-long, 50-Ohm transmission line. The model is an ideal, lossless transmission line. The agreement is excellent up to the bandwidth of the measurement.

These questions can be explored using a simulation tool like SPICE. We will first work in the frequency domain to evaluate the impedance looking into the front end of a transmission line and then interpret this result in the time domain.

In the frequency domain, we can ask what is the impedance looking into the front end of a transmission line, with the far end open. In this example, we will use a 50-Ohm line that is 6 inches long with a dielectric constant of 4. Its time delay, TD, is 1 nsec.

The total capacitance is given by $C_{total} = TD/Z_0 = 1$ nsec/50 Ohms = 20 pF. The total loop inductance is given by $L_{total} = Z_0 \times TD = 50$ Ohms $\times$ 1 nsec = 50 nH.

The simplest approximation for a transmission line is a single LC model, with the L and C values as the total values of the transmission line. This is the simplest lumped circuit model for an ideal transmission line.

Figure 7-38 shows the predicted impedance for an ideal, distributed transmission line and the calculated impedance of a single-section LC lumped-circuit model using these values. In the low-frequency range, the LC model matches the performance really well. The bandwidth of this model is about 100 MHz. The limitation to the bandwidth is because, in fact, this ideal transmission line does not have all its capacitance in one place. Instead, it is distributed down the length, and between each capacitor is some loop inductance associated with the length of the sections. However, it is clear from this comparison that a transmission line, open at the far end, will look exactly like an ideal capacitor at low frequency.

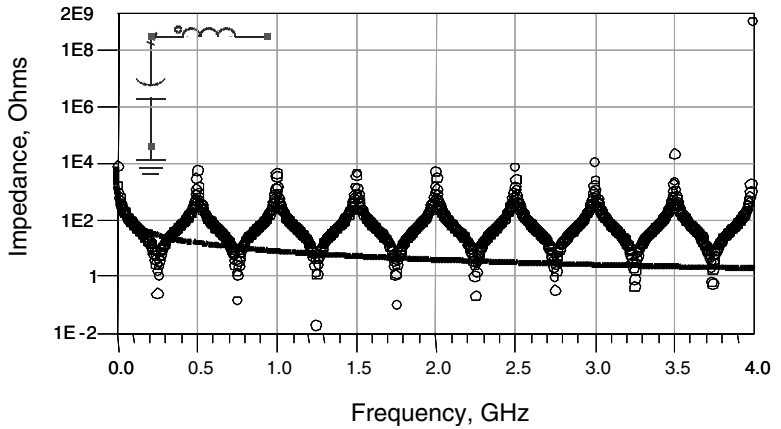


Figure 7-38 Simulated impedance of an ideal transmission line (circles) and simulated impedance of a single-section LC lumped-circuit model (line). The agreement is excellent up to the bandwidth of about 100 MHz.

The impedance of the ideal transmission line shows the resonance peaks occurring when the frequency matches another half wavelength that can fit in the length of the transmission line. The peak resonant frequencies, f_{res}, are given by:

$$f_{res} = m \times \frac{f_0}{2} = m \times \frac{1}{2TD} \tag{7-42}$$

where:

f_{res} = the frequency for the peaks in the impedance

m = the number of the peak, also the number of half waves that fit in the transmission line

TD = the time delay of the transmission line

f_0 = the frequency at which one complete wave fits in the transmission line

The first resonance for m = 1 is for 1 × 1 GHz/2 = 0.5 GHz. Here, just one half wave will fit evenly in the length of the transmission line, with a TD of 1 nsec. The second resonance for m = 2 is at 2 × 1 GHz/2 = 1 GHz. Here, exactly one wave will fit in the transmission line. These standing wave patterns are shown in Figure 7-39.

The bandwidth of the single-section LC model is about one quarter the frequency of the first resonance, up to about 125 MHz. We can improve the bandwidth of the model by breaking the transmission line into more sections. If we

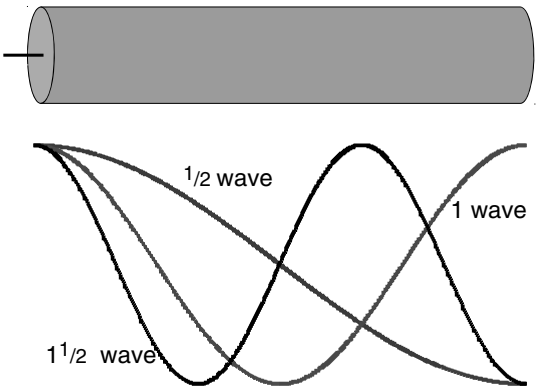


Figure 7-39 Voltage waves on the transmission line. Resonances are when an additional half wavelength can fit in the line.

break it up in two sections, then each section can be modeled as an identical LC model and the values of each L and C would be $L_{total}/2$ and $C_{total}/2$. The predicted impedance of this two-section LC model compared with the ideal T line is shown in Figure 7-40. The bandwidth of this model is roughly at half the first resonance peak. This is a frequency of about 250 MHz.

We can further increase the bandwidth of this lumped circuit model by breaking up the length into more LC sections. Figure 7-41 shows the comparison

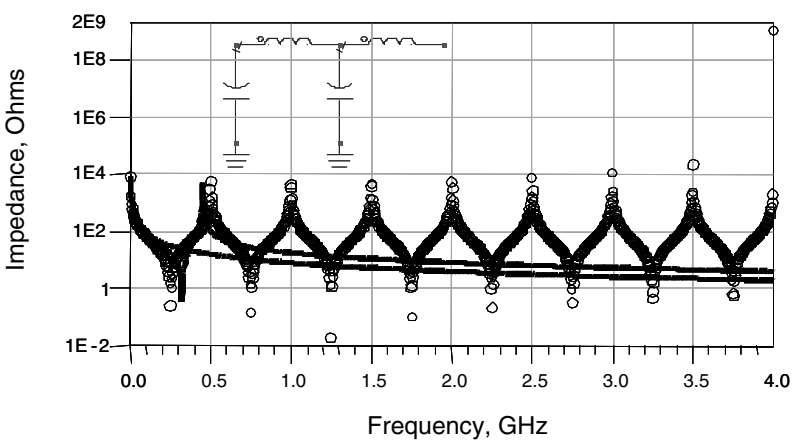


Figure 7-40 Simulated impedance of an ideal transmission line (circles) and simulated impedance of a one-section and two-section LC lumped-circuit model (lines).

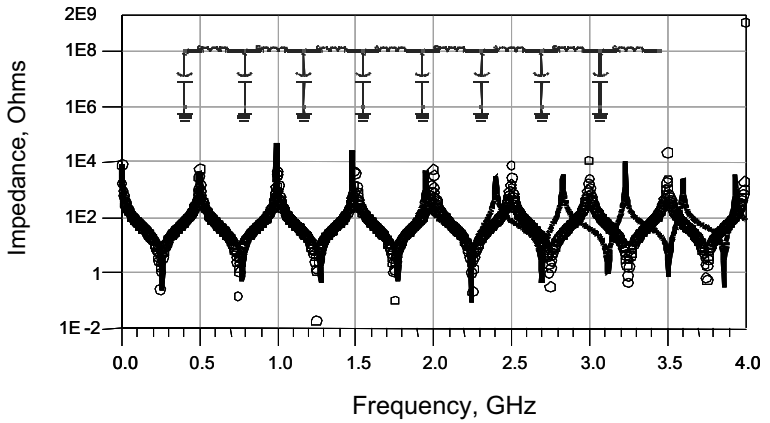


Figure 7-41 Simulated impedance of an ideal transmission line (circles) and simulated impedance of a 16-section LC lumped-circuit model (line).

with an ideal transmission line and using 16 different LC sections, with each L and C being $L_{total}/16$ and $C_{total}/16$. As we increase the number of sections, we are able to better approximate, to a higher bandwidth, the impedance behavior of an ideal transmission line. The bandwidth of this model is about up to the fourth resonance peak, at 2 GHz.

TIP An n-section LC circuit is an approximation of an ideal transmission line. The more sections, the higher the bandwidth of the approximation.

We can estimate the bandwidth of an n-section lumped-circuit model based on the time delay of this ideal transmission line. These last examples illustrate that the more segments in the LC model, the higher the bandwidth. One section has a bandwidth up to one quarter of the first resonant peak; two sections up to half the first resonance; and 16 sections up to the second resonant frequency. We can generalize that the highest frequency at which we have good agreement—the bandwidth of the model—is:

$$BW_{model} = \frac{n}{4} \times \frac{f_0}{2} \sim n \times \frac{f_0}{10} \tag{7-43}$$

or

$$n = 10 \times \frac{BW_{model}}{f_0} = 10 \times BW_{model} \times TD \tag{7-44}$$

where:

BW_{model} = the bandwidth of the n-section lumped-circuit model

n = the number of LC sections in the model

TD = the time delay of the transmission line

f_0 = the resonant frequency for one complete wavelength = 1/TD

We have approximated the relationship to be a little more conservative and a little easier to remember by using $n = 10 \times BW_{model} \times TD$, rather than $n = 8 \times BW_{model} \times TD$.

TIP This is a very important rule of thumb to keep in mind. It says, to reach a bandwidth of the model equal to the 1/TD, 10 LC sections are required. It also says that because this frequency corresponds to having one wavelength in the transmission line, for a good approximation, there should be one LC section for every 1/10th wavelength of the signal.

For example, if the interconnect has a TD = 1 nsec, and we would like an n-section LC model with a bandwidth of 5 GHz, then we need at least $n = 10 \times 5$ GHz $\times$ 1 nsec = 50 sections. At this highest frequency, there will be 5 GHz $\times$ 1 nsec = 5 wavelengths on the transmission line. For each wave, we need 10 sections; therefore, we need $5 \times 10 = 50$ LC sections for a good approximation.

If the TD of a line is 0.5 nsec and we need a bandwidth of 2 GHz, the number of sections required is $n = 10 \times 2$ GHz $\times$ 0.5 nsec = 10 sections.

We can also evaluate the frequency to which we can use a single LC section to model a transmission line. In other words, up to what frequency does a transmission line look like a simple LC circuit? The bandwidth of one section is:

$$BW = n \times \frac{1}{10 \times TD} = 1 \times \frac{1}{10 \times TD} = 0.1 \times \frac{1}{TD} \tag{7-45}$$

For the case of a transmission line with a TD = 1 nsec, the bandwidth of a single-section LC model for this line is 0.1 $\times$ 1/1 nsec = 100 MHz. If the TD = 0.16 nsec, (roughly 1 inch long), the bandwidth of a simple LC model for this line

is 0.1 × 1/0.16 nsec = 600 MHz. The longer the time delay of a transmission line, the lower the frequency at which we can approximate it as a simple LC model.

We have evaluated the number of sections we need to describe a transmission line for a required bandwidth. We found that we needed about 10 LC sections per wavelength of the highest frequency component of the signal and a total number of LC segments depending on the number of wavelengths of the highest frequency component of the signal that can fit in the transmission line.

If we have a signal with a rise time, RT, the bandwidth associated with the signal (the highest sine-wave frequency component that is significant) is BW_{sig} = 0.35/RT. If we have a transmission line that has a time delay of TD and we wish to approximate it with an n-section lumped-circuit model, we need to make sure the bandwidth of the model, BW_{model}, is at least > BW_{sig}:

$$BW_{model} > BW_{sig} \tag{7-46}$$

$$n \times \frac{1}{10 \times TD} > \frac{0.35}{RT} \tag{7-47}$$

$$n > 3.5\frac{TD}{RT} \tag{7-48}$$

where:

BW_{sig} = bandwidth of the signal

BW_{model} = the bandwidth of the model

RT = the rise time of the signal

TD = the time delay of the transmission line

n = the minimum number of LC sections needed for an accurate model

For example, with a rise time of 0.5 nsec and time delay of 1 nsec, we would need n > 3.5 × 1/0.5 = 7 sections for an accurate model.

When the rise time is equal to the TD of the line we want to model, we need at least 3.5 sections for an accurate model. In this case, the spatial extent of the

rise time is the length of the transmission line. This suggests a very important rule of thumb, given in the following TIP.

TIP To accurately describe an interconnect as an n-section LC model, we need to have at least 3.5 LC sections along the spatial extent of the leading edge. It is as though each third of the leading edge of the signal interacts with the interconnect approximately as a lumped-circuit element.

This rule of thumb is illustrated in Figure 7-42. In FR4, if the rise time is 1 nsec, the spatial extent of the leading edge is 6 inches. We need 3.5 LC sections for every 6 inches of length or about 1.7 inches per section. We can generalize this: If the rise time is RT, and the speed of the signal is v, then the length for each LC section is (RT × v)/3.5. In FR4, where the speed is about 6 inches/nsec, the length of each LC section required for a rise time, RT, is 1.7 × RT, with rise time in nsec.

TIP This suggests a very powerful rule of thumb: For an adequate bandwidth of the n-section lumped-circuit model at the given rise time, RT (in nsec), the size (in inches) of each LC section should be no longer than 1.7 × RT inches.

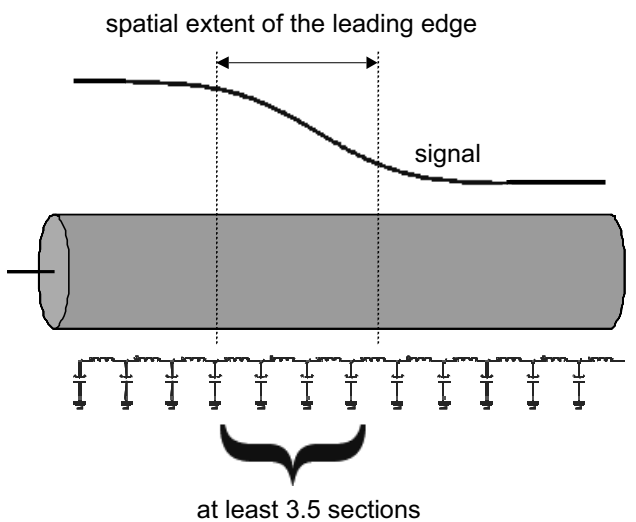


Figure 7-42 As a general rule of thumb, there should be at least 3.5 LC sections per spatial extent of the rise time for an accurate model of the interconnect at the bandwidth of the signal.

If the rise time is 1 nsec, the length of each single LC should be less than 1.7 inches. If the rise time is 0.5 nsec, the length of each LC section should be no longer than $0.5 \times 1.7 = 0.85$ inches.

TIP Of course, an ideal, distributed transmission line model is always a good model for a uniform interconnect, at low frequency and at high frequency.

7.19 Frequency Variation of the Characteristic Impedance

So far, we have been assuming the characteristic impedance of a transmission line is constant with frequency. As we have seen, the input impedance, looking into the front of a transmission line, is strongly frequency dependent. Does the characteristic impedance vary with frequency? In this section, we are assuming the transmission line is lossless. In a later chapter, we will look at the case where the transmission line has loss. We will see that the characteristic impedance does vary slightly due to the losses.

As we have seen, the characteristic impedance of an ideal lossless transmission line is related to the capacitance and inductance per length as:

$$Z_0 = \sqrt{\frac{L_L}{C_L}} \tag{7-49}$$

Providing the dielectric constant of the interconnect is constant with frequency, the capacitance per length will be constant. This is a reasonable assumption for most materials, though in some cases, the dielectric constant will vary slightly.

As we saw in the last chapter, the loop inductance per length of a line will vary with frequency due to skin-depth effects. In fact, the loop inductance will start out higher at low frequency and decrease as all the currents distribute to the outer surface. This would suggest that characteristic impedance will start out higher at low frequency and decrease to a constant value at higher frequency.

At frequencies well above the skin depth, we would expect all the currents to be on the outer surface of all the conductors and not to vary with frequency beyond this point. The loop inductance should be constant and the characteristic impedance should be constant. We can estimate this frequency for a 1-ounce copper conductor. The skin depth for copper is about 20 microns at 10 MHz. The thickness of 1-ounce copper is about 34 microns. We would expect to see the char-

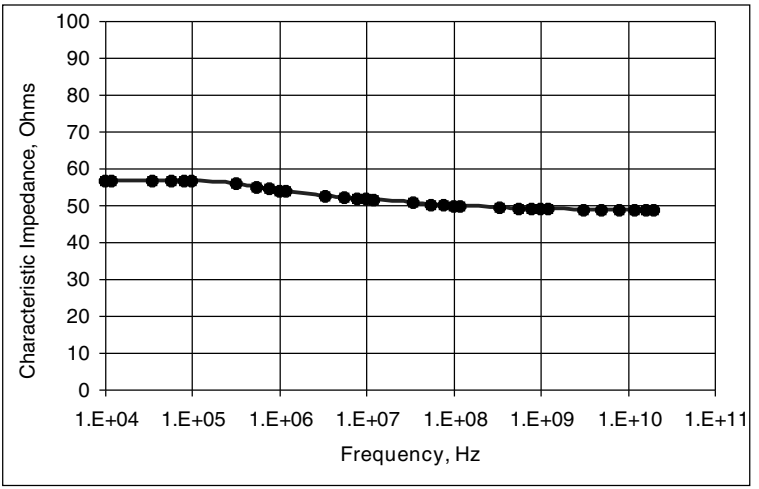


Figure 7-43 Calculated frequency variation of the characteristic impedance due to skin-depth effects, for a 50-Ohm line in FR4 with 1-ounce copper traces. Calculated with Ansoft's 2D Extractor.

acteristic impedance start to decrease around 1 MHz to 10 MHz and stop decreasing at about 100 MHz, where the skin depth is only 6 microns.

We can calculate the frequency dependence of the characteristic impedance for a 50-Ohm microstrip with 1-ounce copper traces using a 2D field solver. The result is shown in Figure 7-43.

This shows that at low frequency, the characteristic impedance is high, starts dropping at about 1 MHz, and continues to drop until about 50 MHz. The total drop from DC to high frequency is about 7 Ohms, or less than 15%.

TIP Above about 50 MHz, the characteristic impedance of a transmission line is constant with frequency. This is the "high-frequency" characteristic impedance and is the value typically used for all evaluation of the behavior of high-speed signals.

7.20 The Bottom Line

1. A transmission line is a fundamentally new ideal-circuit element that accurately describes all the electrical properties of a uniform cross-sectional interconnect.
2. Forget the word *ground*. Think *return path*.

3. Signals propagate down a transmission line at the speed of light in the material surrounding the conductors. This primarily depends on the dielectric constant of the insulation.
4. The characteristic impedance of a transmission line describes the instantaneous impedance a signal would see as it propagates down the line. It is independent of the length of the line.
5. The characteristic impedance of a line primarily depends inversely on the capacitance per length and the speed of the signal.
6. The input impedance looking into the front end of a transmission line changes with time. It is initially the characteristic impedance of the line during the round-trip time of flight, but can end up being anything depending on the termination, the length of the line, and how long we measure the impedance.
7. A controlled-impedance board has all its traces fabricated with the same characteristic impedance. This is essential for good signal integrity.
8. A signal propagates through a transmission line as a current loop with the current going down the signal path and looping back through the return path. Anything that disturbs the return path will increase the impedance of the return path and create a ground-bounce voltage noise.
9. An ideal transmission line can be approximated with an n-section LC lumped-circuit model. The higher the bandwidth required, the more LC sections required.
10. For good accuracy, there should be at least 3.5 LC sections along the spatial extent of the leading edge.
11. An ideal transmission line is always a good model for a uniform interconnect, independent of the rise time and interconnect length.

CHAPTER 8

Transmission Lines and Reflections

TIP If a signal is traveling down an interconnect and the instantaneous impedance the signal encounters at each step ever changes, some of the signal will reflect and some of the signal will continue down the line distorted. This principle is the driving force that creates most signal-quality problems on a single net.

Reflections and distortions from impedance discontinuities give rise to a degradation in signal quality. In some cases, this looks like ringing. The undershoot, when the signal level drops, can eat into the noise budget and contribute to false triggering. One example of the reflection noise generated from impedance discontinuities at the ends of a short-length transmission line is shown in Figure 8-1.

Reflections occur whenever the instantaneous impedance the signal sees changes. This can be at the ends of lines or wherever the topology of the line changes, such as at corners, vias, branches, connectors, and packages. By understanding the origin of these reflections and arming ourselves with the tools to predict their magnitude, we can engineer a design with acceptable system performance.

TIP For optimal signal quality, the goal in interconnect design is to keep the instantaneous impedance the signal sees as constant as possible.

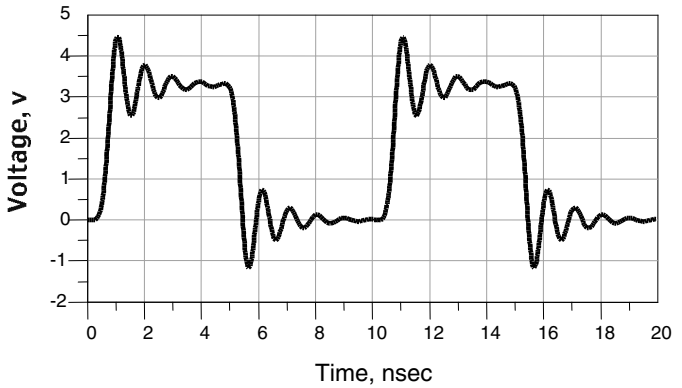


Figure 8-1 "Ringing" noise at the receiver end of a 1-inch-long controlled-impedance interconnect created because of impedance mismatches and multiple reflections at the ends of the line.

First, this means keeping the instantaneous impedance of the line constant. Hence, the growing importance in manufacturing controlled-impedance boards. All the various design guidelines, such as minimizing stub lengths, using daisy chains rather than branches, and using point-to-point topology, are all methods to keep the instantaneous impedance constant.

Second, this means managing the impedance changes at the ends of the line with a termination strategy. No matter how efficient we may be at constructing a uniform transmission line, the impedance will always vary at the ends of the line. The rattling around from the reflections at the ends will result in ringing noise, unless it is controlled. That's what a termination strategy does.

Third, even with controlled impedance interconnects, well terminated, the specific routing topology can influence the reflections. When a signal line splits in two branches, there is an impedance discontinuity. Maintaining a linear routing topology with no branches or stubs is an important strategy to minimize impedance changes and reflection noise.

8.1 Reflections at Impedance Changes

As a signal propagates down a transmission line, it sees an instantaneous impedance for each step along the way. If the interconnect is a controlled impedance, then the instantaneous impedance will be equal to the characteristic impedance of the line. If the instantaneous impedance should ever change, for whatever reason, some of the signal will reflect back in the opposite direction and some of it will

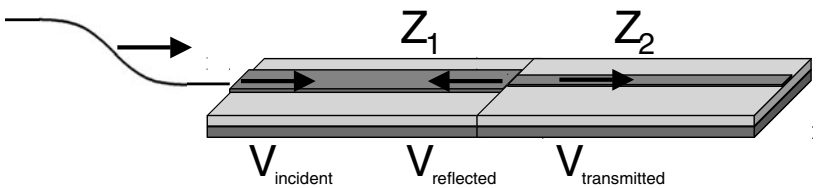


Figure 8-2 Whenever a signal sees a change in the instantaneous impedance, there will be some reflected signal, and the transmitted signal will be distorted.

continue with a different amplitude. We call all locations where the instantaneous impedance changes *impedance discontinuities* or just *discontinuities*.

The amount of signal that reflects depends on the magnitude of the change in the instantaneous impedance. This is illustrated in Figure 8-2. If the instantaneous impedance in the first region is Z_1 and the instantaneous impedance in the second region is Z_2, the magnitude of the reflected signal compared to the incident signal will be given by:

$$\frac{V_{reflected}}{V_{incident}} = \frac{Z_2 - Z_1}{Z_2 + Z_1} = \rho \tag{8-1}$$

where:

$V_{reflected}$ = the reflected voltage

$V_{incident}$ = the incident voltage

Z_1 = the instantaneous impedance of the region where the signal is initially

Z_2 = the instantaneous impedance of the section where the signal just enters

ρ = the reflection coefficient, in units of rho

The greater the difference in the impedances in the two regions, the greater the amount of reflected signal. For example, if a 1-v signal is moving on a 50-Ohm characteristic-impedance transmission line, it will see an instantaneous impedance of 50 Ohms. If it hits a region where the instantaneous impedance changes to 75 Ohms, the reflection coefficient will be (75 – 50)/(75 + 50) = 20%, and the amount of reflected voltage will be 20% × 1 v = 0.2 v.

For every part of the waveform that hits the interface, exactly 20% of it will reflect back. This is true no matter the shape of the waveform. In the time domain,

it can be a sharp edge, a sloping edge, or even a Gaussian edge. Likewise, in the frequency domain, where all waveforms are sine waves, each sine wave will reflect and the amplitude and phase of the reflected wave can be calculated from this relationship.

It is often the reflection coefficient, ρ (or rho), that is of interest. The reflection coefficient is the ratio of the reflected voltage to the incident voltage.

TIP The most important thing to remember about the reflection coefficient is that it is equal to the ratio of the second impedance minus the first, divided by their sum. This distinction is particularly important in determining the sign of the reflection coefficient.

When considering signals on interconnects, keeping track of their direction of travel on the interconnect is critically important. If a signal is traveling down a transmission line and hits a discontinuity, a second wave will be generated at the discontinuity. This second wave will be superimposed on the first wave, but will be traveling back toward the source. The amplitude of this second wave will be the incident voltage times rho.

8.2 Why Are There Reflections?

The reflection coefficient describes the fraction of the voltage that reflects back to the source. In addition, there is a transmission coefficient that describes the fraction of the incident voltage that is transmitted through the interface into the second region. This property of signals to reflect whenever the instantaneous impedance changes is the source of all signal-quality problems related to signal propagation on one net.

To minimize signal-integrity problems that arise because of this fundamental property, we must implement the following four important design features for all high-speed circuit boards:

1. Use controlled impedance interconnects.

2. Provide at least one termination at the ends of a transmission line.

3. Use a routing topology that minimizes the impact from multiple branches.

4. Minimize any geometry discontinuities.

But what causes the reflection? Why does a signal reflect when there is a change in the instantaneous impedance it sees? The reflected signal is created to match two important boundary conditions.

Consider the interface between two regions, labeled as region 1 and region 2, each with a different instantaneous impedance. As the signal hits the interface, we must see only one voltage between the signal- and return-path conductors and one current loop flowing between the signal- and return-path conductors. Whether we look from the region 1 side or change our perspective and look from the region 2 side, we must see the same voltage and the same current on either side of the interface. We must not have a voltage discontinuity across the boundary, because this would mean an infinitely large electric field in the boundary. We must not have a current discontinuity, as this would mean we are building up a net charge at the interface.

Without the creation of a reflected voltage heading back to the source and while maintaining the same voltage and current across the interface, we would have the condition of $V_1 = V_2$ and $I_1 = I_2$. But, $I_1 = V_1/Z_1$ and $I_2 = V_2/Z_2$. If the impedances of the two regions are not the same, there is no way all four conditions can be met.

To keep harmony in the universe, literally, to keep the Universe from exploding, a new voltage is created in the first region that reflects back to its source. Its sole purpose is to take up the mismatched current and voltage between the incident and transmitted signals. Figure 8-3 illustrates what happens at the interface.

The incident voltage, V_{inc}, moves toward the interface, while the transmitted voltage, V_{trans}, moves away from the interface. A new voltage is created as the incident signal tries to pass through the interface. This new wave is traveling only in region 1, back to the source. At any point in region 1, the total voltage between the signal and return conductors is the sum of the voltages traveling in the two directions, the incident voltage plus the reflected voltage.

The condition of the same voltages on both sides of the interface requires:

$$V_{inc} + V_{refl} = V_{trans} \tag{8-2}$$

The condition on the currents is a little more subtle. The total current at the interface, in region 1, is due to two current loops, each traveling in opposite directions and circulating in opposite directions. At the interface, the direction of the incident current loop is clockwise. The direction of the reflected current loop is counterclockwise. If we define the positive direction as clockwise, then the net

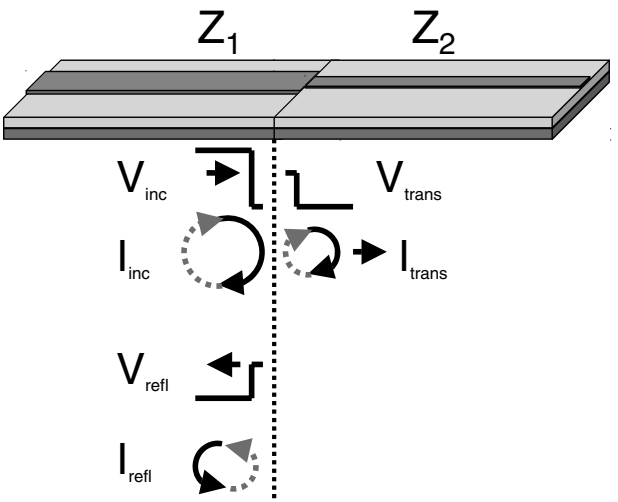


Figure 8-3 As the incident signal tries to pass through the interface, a reflected voltage and current are created to match the voltage and current loops on both sides of the interface.

current at the interface in region 1 is $I_{inc} - I_{refl}$. In region 2, the current loop is clockwise and is just I_{trans}. The condition of the same current viewed from either side of the interface is:

$$I_{inc} - I_{refl} = I_{trans} \tag{8-3}$$

The final condition is that the ratio of voltage to the current in each region is the impedance of each region:

$$\frac{V_{inc}}{I_{inc}} = Z_1 \tag{8-4}$$

$$\frac{V_{refl}}{I_{refl}} = Z_1 \tag{8-5}$$

$$\frac{V_{trans}}{I_{trans}} = Z_2 \tag{8-6}$$

Using these last relationships, we can rewrite the condition of the current as:

$$\frac{V_{inc}}{Z_1} - \frac{V_{refl}}{Z_1} = \frac{V_{trans}}{Z_2} \tag{8-7}$$

With a little bit of algebra, we get:

$$\frac{V_{inc}}{Z_1} - \frac{V_{refl}}{Z_1} = \frac{V_{inc} + V_{refl}}{Z_2} \tag{8-8}$$

and

$$V_{inc}\left(\frac{Z_2 - Z_1}{Z_2 Z_1}\right) = V_{refl}\left(\frac{Z_2 + Z_1}{Z_2 Z_1}\right) \tag{8-9}$$

and finally,

$$\frac{V_{refl}}{V_{inc}} = \frac{Z_2 - Z_1}{Z_2 + Z_1} = \rho \tag{8-10}$$

which is the definition of the reflection coefficient. Using the same approach, we can derive the transmission coefficient as:

$$t = \frac{V_{trans}}{V_{inc}} = \frac{2 \times Z_2}{Z_2 + Z_1} \tag{8-11}$$

Dynamically, what actually creates the reflected voltage? No one knows. We only know that if it is created, we are able to match the same voltage on one side of the interface with that on the other side. The voltage is continuous across the interface. Likewise, the current loop is exactly the same on both sides of the interface. The current is continuous across the interface. The universe is in balance.

8.3 Reflections from Resistive Loads

There are three important special cases to consider for transmission line terminations. In each case, the transmission line characteristic impedance is 50 Ohms. The signal will be traveling in this transmission line from the source and hit the far end with a particular termination impedance.

TIP It is important to keep in mind that in the time domain, the signal is sensitive to the instantaneous impedance. It is not necessary that the second region be a transmission line. It might also be a discrete device that has some impedance associated with it, such as a resistor, capacitor, inductor, or some combination thereof.

When the second impedance is an open, as is the case when the signal hits the end of a transmission line with no termination, the instantaneous impedance at the end is infinite. The reflection coefficient is (infinite – 50)/(infinite + 50) = 1. This means a second wave of equal size to the incident wave will be generated at the open but will be traveling in the opposite direction—back to the source.

If we look at the total voltage appearing at the far end, where the open is, we will see the superposition of two waves. The incident wave, with an amplitude of 1 v, will be traveling toward the open end. The other, the reflected wave, also with an amplitude of 1 v, will be traveling in the opposite direction. When we measure the voltage at the far end, we measure the sum of these two voltages, or 2 v. This is illustrated in Figure 8-4.

TIP It is often said that when a signal hits the end of a transmission line, it doubles. While this is technically true, it is not what is really happening. The total voltage, the sum of the two traveling waves, is twice the incident voltage. However, if we think of it as a doubling, we mis-calibrate our intuition. It is better to think of the voltage at the far end as the sum of the incident and the reflected voltage.

The second special case is when the far end of the line is shorted to the return path. The impedance at the end is 0 in this case. The reflection coefficient is (0 – 50)/(0 + 50) = –1. When a 1-v signal is incident to the far end, a –1-v signal is generated by the reflection. This second wave propagates back through the transmission line to the source.

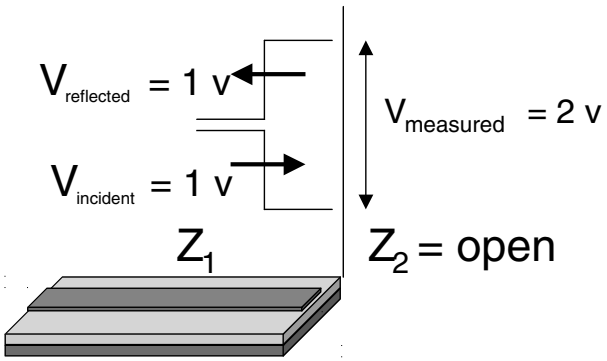


Figure 8-4 If the second impedance is an open, the reflection coefficient is 1. At the open, there will be two oppositely traveling waves superimposed.

The voltage that would be measured at the shorting discontinuity is the sum of the incident voltage and the reflected voltage, or 1 v + –1 v = 0. This is reasonable, because if we really have a short at the far end, by definition, we can't have a voltage across a short. We see now that the reason it is 0 v is that it is the sum of two waveforms, a positive one traveling in the direction from the source and a negative one traveling back toward it.

The third important impedance at the far end to consider is when the impedance matches the characteristic impedance of the transmission line. In this example, it could be created by adding a 50-Ohm resistor to the end. The reflection coefficient would be (50 – 50)/(50 + 50) = 0. There would be no reflected voltage from the end. The voltage appearing across the 50-Ohm termination resistor would just be the incident-voltage wave.

If the instantaneous impedance the signal sees does not change, there will be no reflection. By placing the 50-Ohm resistor at the far end, we have matched the termination impedance to the characteristic impedance of the line and reduced the reflection to zero.

For any resistive load at the far end, the instantaneous impedance the signal will see will be between 0 and infinity. Likewise, the reflection coefficient will be between –1 and +1. Figure 8-5 shows the relationship between terminating resistance and reflection coefficient for a 50-Ohm transmission line.

When the second impedance is less than the first impedance, the reflection coefficient is negative. The reflected voltage from the termination will be a negative voltage. This negative-voltage wave propagates back to the source.

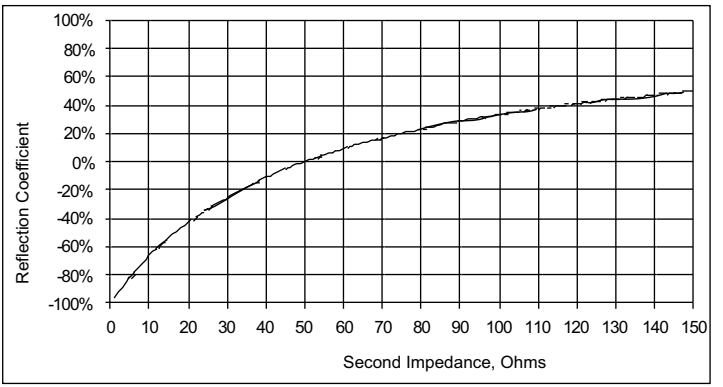


Figure 8-5 Reflection coefficient for the case of the first impedance being 50 Ohms and a variable impedance for the second region.

TIP When the second impedance is less than the first impedance, the voltage appearing across the resistor will always be less than the incident voltage.

For example, if the transmission line characteristic impedance is 50 Ohms and the termination is 25 Ohms, the reflection coefficient is (25 – 50)/(25 + 50) = –1/3. If 1 v were incident to the termination, a –0.33 v will reflect back to the source. The actual voltage that would appear at the termination is the sum of these two waves, or 1 v + –0.33 v = 0. 67 v.

Figure 8-6 shows the measured voltage that would appear across the termination for a 1-v incident voltage and a 50-Ohm transmission line. As the termination impedance increases from 0 Ohms, the actual voltage measured across the termination increases from 0 v up to 2 v when the termination is open.

8.4 Source Impedance

When a signal is launched into a transmission line, there is always some impedance of the source. For typical CMOS devices, this can be about 5 Ohms to 20 Ohms. For older-generation transistor-transistor logic (TTL) gates, this can be as high as 100 Ohms. The source impedance will have a dramatic impact on both the initial voltage launched into the transmission line and the multiple reflections. When the reflected wave finally reaches the source, it will see the output-source resistance as the instantaneous impedance right at the driver. The value of this output-

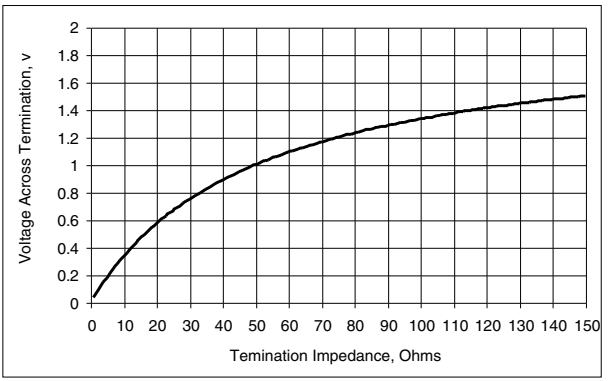


Figure 8-6 Voltage across the termination, for a 1-v incident signal. This voltage is the sum of the incident going and reflected wave.

source impedance will determine how the reflected wave reflects again from the driver.

If we have a model for the driver, either SPICE or IBIS based, a good estimate of the output impedance of the driver can be extracted with a few simple simulations. We assume an equivalent circuit model for the driver is an ideal voltage source in series with a source resistor, illustrated in Figure 8-7. We can extract the output voltage of the ideal source when driving a high-output impedance. If we connect a low impedance like 10 Ohms to the output, and measure the output voltage across this terminating resistor, we can back out the internal-source resistance from:

$$R_s = R_t\left(\frac{V_o}{V_t} - 1\right) \tag{8-12}$$

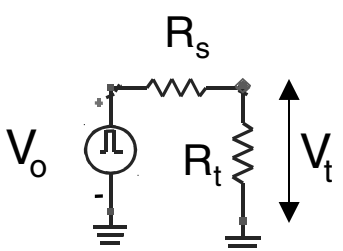


Figure 8-7 Simple, implicit model of an output driver with a terminating resistor connected.

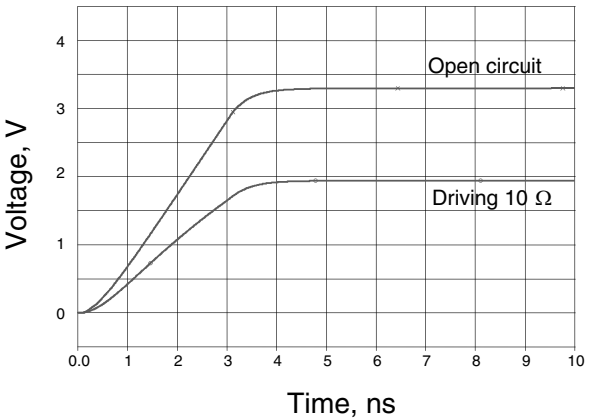


Figure 8-8 Simulated output voltage from a CMOS driver with a 10K-Ohm and then a 10-Ohm resistor connected. From the two voltage levels, the source impedance of the driver can be calculated. Simulation of a typical CMOS IBIS-driver model using the HyperLynx simulator.

where:

R_s = the source resistance of the driver

R_t = the terminating resistance connected to the output

V_o = the open-circuit output voltage from the driver

V_t = the voltage across the terminating resistor

To calculate the source resistance, we simulate the output voltage from the driver in two cases: when a very high resistance is attached (e.g., 10 kOhms) and when a low impedance is attached (e.g., 10 Ohms). An example of the simulated voltage using a behavioral model for a common CMOS driver, is shown in Figure 8-8. The open-circuit voltage is 3.3 v, and the voltage with the 10-Ohm resistor attached is 1.9 v. From the equation above, the output-source impedance can be calculated as 10 Ohms × (3.3 v/1.9 v – 1) = 7.3 Ohms.

8.5 Bounce Diagrams

As shown in the previous chapter, the actual voltage launched into the transmission line, or the incident voltage to the transmission line, is determined by the combination of the source voltage and the voltage divider made up of the source impedance and the transmission line.

Knowing the time delay of the transmission lines, TD, and the impedances of each region where the signal will propagate, and knowing the initial voltage

from the driver, we can calculate all the reflections at all the interfaces and predict the voltages that would be measured at any point in time.

For example, if the source voltage, driving an open termination, were 1 v and the source impedance were 10 Ohms, the actual voltage launched into a 1-nsec-long, 50-Ohm transmission line would be 1 v × 50/(10 + 50) = 0.84 v. This 0.84-v signal would be the initial incident voltage propagating down the transmission line.

Suppose the end of the transmission line was an open termination. The 0.84-v signal would hit the end of the line 1 nsec later and a +0.84-v signal would be generated by the reflection, traveling back to the source. At the end of the line, the total voltage measured across the open will be the sum of the two waves, or 0.84 v + 0.84 v = 1.68 v.

When the 0.84-v reflected wave hits the source end, another 1 nsec later, it will see an impedance discontinuity. The reflection coefficient at the source is (10 – 50)/(10 + 50) = –0.67. With a 0.84-v signal incident on the driver, a total of 0.84 v × –0.67 = –0.56 v will reflect back to the end of the line. Of course, this new wave will reflect again from the far end. A step voltage change of –0.56 v will reflect back. Measured at the far end, across the open, will be four simultaneous waves: 2 × 0.84 v or 1.68 v from the first wave, and 2 × –0.56 v or –1.12 v from the second reflection for a total voltage of 0.56 v.

The –0.56-v wave will hit the source impedance and reflect yet again. The reflected voltage will be +0.37 v. At the far end, there will be the 0.56 v from the first two waves plus the new set of incident and reflected 0.37-v waves, for a total of 0.56 v + 0.37 v + 0.37 v = 1.3 v. Keeping track of these multiple reflections is straightforward but tedious. Before the days of simple, easy-to-use simulation tools, these reflections were diagrammed using *bounce* or *lattice* diagrams. An example is shown in Figure 8-9.

When the source impedance is less than the characteristic impedance of the transmission line, we see that the reflection at the source end will be negative. This will contribute to the effect we normally call ringing. Figure 8-10 shows the voltage waveform at the far end of the transmission line for the previous example, when the rise time of the signal is very short compared to the time delay of the transmission line. This analysis was done using a SPICE simulator to predict the waveforms at the far end, taking into account all the multiple reflections and impedance discontinuities.

Two important features should be apparent. First, the voltage at the far end eventually approaches the source voltage, or 1 v. This must be the case, because we have an open and we must ultimately see the source voltage across an open.

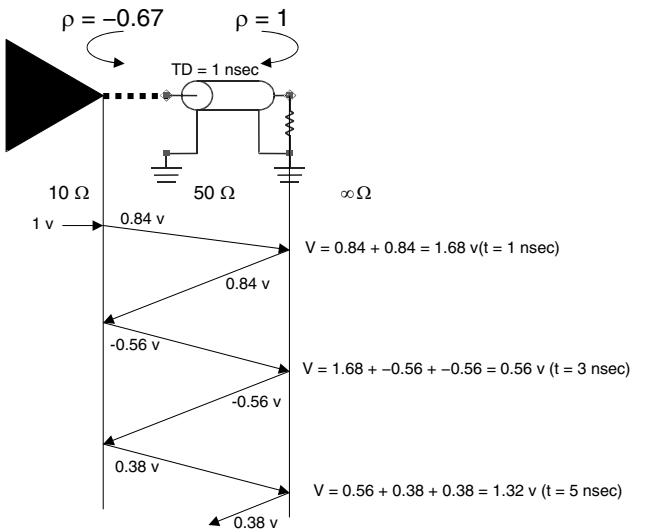


Figure 8-9 Lattice or bounce diagram used to keep track of all the multiple reflections and the time-varying voltages at the receiver on the far end.

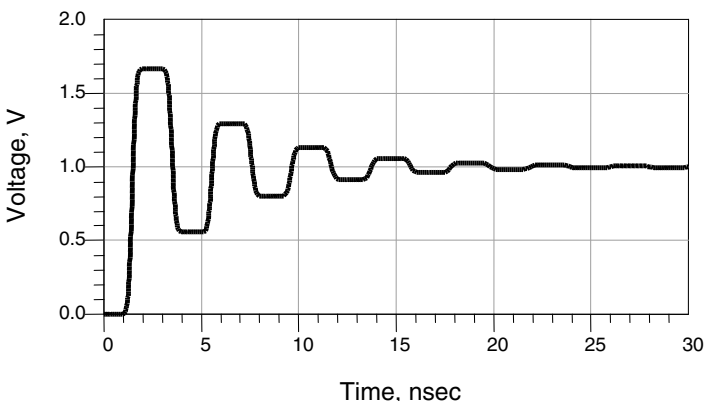


Figure 8-10 Simulated voltage at the far end of the transmission line described with the previous lattice diagram. Simulation performed with SPICE.

The second important effect is that the actual voltage across the open exceeds the voltage coming out of the source. The source has only 1 v, yet we would measure as much as 1.68 v at the far end. How was this higher voltage generated? The higher voltage is really generated by the resonance of the distributed L and C of the transmission line.

8.6 Simulating Reflected Waveforms

Using the definition of the reflection coefficient above, the reflected signal from any arbitrary impedance can be calculated. When the terminating impedance is a resistive element, the impedance is constant and the reflected voltages are easy to calculate. When the termination has a more complicated impedance behavior (such as a capacitive or inductive termination, or some combination of the two), calculating the reflection coefficient and how it changes as the incident waveform changes is difficult and tedious if done by hand. Luckily, there are simple, easy-to-use circuit-simulation tools that make this calculation much easier.

The reflection coefficients and the resulting reflected waveforms from arbitrary impedances and for arbitrary waveforms can be calculated using SPICE or other circuit simulators. With such tools, sources are created, ideal transmission lines are added, and terminations are connected. The voltage appearing across the termination and any other nodes can be calculated as the incident waveform hits the end and reflects from all the various discontinuities.

There are many possible combinations of source impedance, transmission line characteristic impedance, time delay, and end termination. Each of these can be easily varied using a simulation tool. Figure 8-11 shows the simulated voltage across the termination as the rise time of the signal is increased from a 0.1 nsec to 1 nsec and, in a separate simulation, the signal waveform as the source terminating resistance is varied from 0 Ohms to 90 Ohms.

TIP Using a SPICE or other circuit simulator, the performance of any arbitrary transmission line circuit can be simulated taking into account all specific features.

8.7 Measuring Reflections with a TDR

In addition to simulating the waveforms associated with transmission line circuits, it is possible to measure the reflected waveforms from physical interconnects using a specialized instrument usually referred to as a *time domain reflectometer*, (*TDR*). This is the appropriate instrument to use when characterizing a passive interconnect that does not have its own voltage source. Of course, when measuring the actual voltages in an active circuit, a fast oscilloscope with a high-impedance probe is the best tool.

A TDR will generate a short rise-time step edge, typically between 35 psec and 150 psec, and measure the voltage at an internal point of the instrument. Figure 8-12 is a schematic of the workings inside a TDR. It is important to keep in

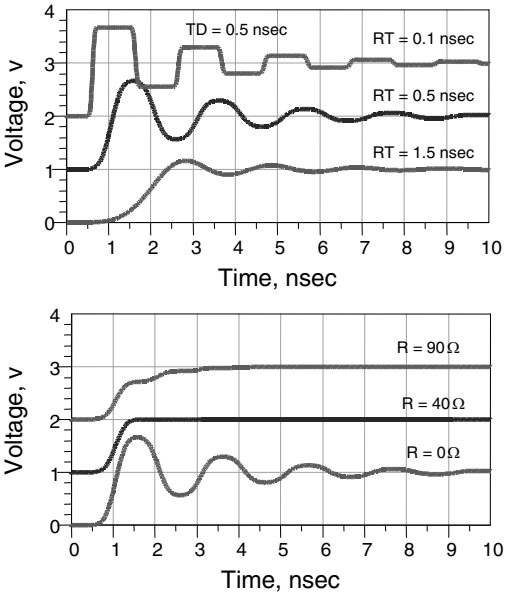


Figure 8-11 Examples of the variety of simulations possible with SPICE. Top: for a 10-Ohm driver and 50-Ohm characteristic-impedance line, showing the far-end voltage with different rise times for the signal. Bottom: changing the series-source terminating resistor and displaying the voltage at the far end.

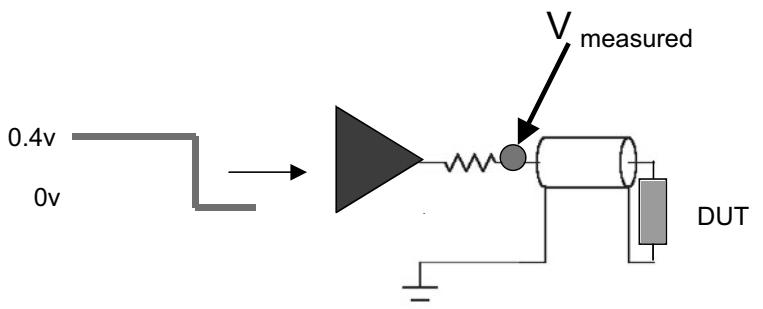


Figure 8-12 Schematic of the inside of a TDR. A very fast pulse generator creates a fast rising-voltage pulse. It travels through a precision 50-Ohm resistor in series with a short length of 50-Ohm coax cable to the front panel where the DUT is connected. The total voltage at an internal point is measured with a very fast sampling oscilloscope and displayed on the front screen.

mind that a TDR is nothing more than a fast step generator and a very fast sampling scope.

The voltage source is a very fast step generator, which outputs a step amplitude of about 400 mV. Right after the voltage source is a calibrated resistor of 50 Ohms. This assures that the source impedance of the TDR is a precision 50 Ohms. After the resistor is the actual detection point where the voltage is measured by a fast sampling amplifier. Connected to this point is a short-length coax cable that brings the signal to the front-panel SMA connector. This is where the DUT is connected. The signal from the source enters the DUT, and any reflected voltage is detected at the sampling point.

Before the step signal is generated, the voltage measured at the internal point will be 0 v. Where the voltage is actually measured, the signal encounters a voltage divider. The first resistor is the internal calibration resistor. The second resistor is the transmission line internal to the TDR. As the 400-mV step reaches the calibration resistor, the actual voltage measured at the detection point will be the result after going through the voltage divider.

The voltage detected is 400 mV × 50 Ohms/(50 Ohms + 50 Ohms) = 200 mV. This voltage is measured initially and is displayed by the fast sampling scope. The 200-mV signal continues moving down the internal coax cable to the DUT.

If the DUT is a 50-Ohm termination, there is no reflected signal and the only voltage present at the sampling point is the forward-traveling wave of 200 mV, which is constant. If the DUT is an open, the reflected signal from the DUT is +200 mV. A short time after launch, this 200-mV reflected-wave signal comes back to the sampling point, and what is measured and displayed is the 200-mV incident voltage + the 200-mV reflected wave. The total voltage displayed is 400 mV.

If the DUT is a short, the reflected signal from the DUT will be –200 mV. Initially, the 200-mV incident voltage is measured. After a short time, the reflected –200-mV signal comes back to the source and is measured by the sampling head. What is measured at this point is the 200-mV incident plus the –200-mV reflected signal, or 0 voltage. The measured-TDR plots of these cases are shown in Figure 8-13.

TIP The TDR will measure the reflected voltage from any interconnect attached to the front SMA connector of the instrument and how this voltage changes with time as the signal propagates down the interconnect, reflecting from all discontinuities.

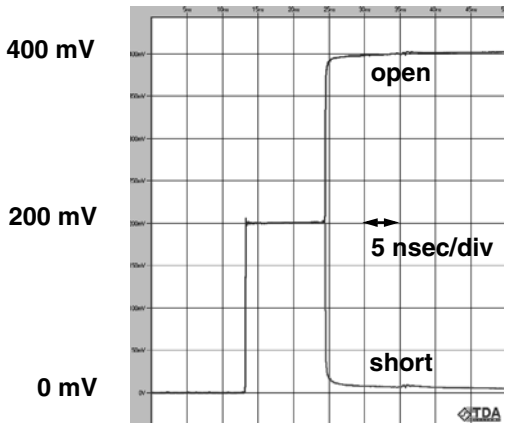


Figure 8-13 Measured-TDR response when the DUT is an open and a short. Data measured with an Agilent 86100 DCA and displayed with TDA Systems IConnect software.

As the transmitted signal continues down the DUT, if there are other regions where the instantaneous impedance changes, a new reflected voltage will be generated and will travel back to the internal measurement point to be displayed. In this sense, the TDR is really indicating changes in the instantaneous impedance the signal encounters and when the signal encounters it.

Since the incident signal must travel down the interconnect and the reflected signal must travel back down the interconnect to the detection point, the time delay measured on the front screen is really a round-trip delay to any discontinuity.

For example, if a uniform 4-inch-long, 50-Ohm transmission line is the DUT, we will see an initial small reflected voltage at the entrance to the DUT because it is not exactly 50 Ohms, and then a larger reflected signal when the incident signal reaches the open at the far end and reflects back to the detection point. This time delay is the round-trip time delay of the transmission line. If the transmission line is not 50 Ohms, then multiple reflections will take place at both ends of the line. The TDR will display the superposition of all the voltage waves that make it back to the internal measurement point. An example of the TDR response from a 50-Ohm transmission line and a 15-Ohm transmission line, both open at the far end, is shown in Figure 8-14.

The combination of understanding the principles and leveraging simulation and measurement tools will allow us to evaluate many of the important impedance discontinuities a signal might encounter. We will see that many of them are impor-

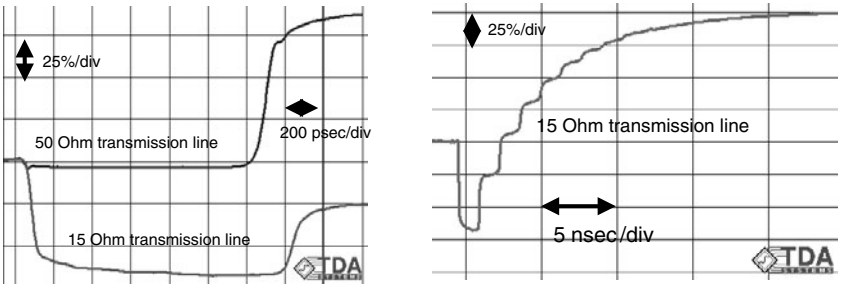


Figure 8-14 Measured TDR response of 4-inch-long transmission lines open at the far end: 50 Ohms and 15 Ohms. Left: time base of 200 psec/div. Right: reflections from the 15-Ohm line on expanded time base, 5 nsec/div. Measured with Agilent 86100 DCA and a GigaTest Labs Probe Station and displayed with TDA Systems IConnect.

tant and must be carefully engineered or avoided, while some of them are not important and can be ignored in some cases.

TIP It is only by applying engineering discipline and "putting in the numbers" that the important effects can be identified and managed while the unimportant ones are identified and ignored.

8.8 Transmission Lines and Unintentional Discontinuities

Wherever a signal sees an impedance change, there will be a reflection. Reflections can have a serious impact on signal quality. Predicting the impact on the signals from the discontinuities and engineering acceptable design alternatives is an important part of signal-integrity engineering.

Even if a circuit board is designed with controlled-impedance interconnects, there is still the opportunity for a signal to see an impedance discontinuity from such features as:

1. The ends of the line

2. A package lead

3. An input-gate capacitance

4. A via between signal layers

5. A corner

6. A stub

7. A branch

8. A test pad

9. A gap in the return path

10. A neck down in a via field

11. A crossover

When we model these effects, there are three common equivalent-circuit models we can use to electrically describe the unintentional discontinuity: an ideal capacitor, an ideal inductor, or a short-length ideal transmission line (either in series or in shunt). These possible equivalent circuit models are shown in Figure 8-15. These circuit elements can occur at the ends of the line or in the middle.

The two most important parameters that influence the distortion of the signal from the discontinuity are the rise time of the signal and how large the discontinuity is. For an inductor and capacitor, their instantaneous impedance depends on the instantaneous rate at which either the current is changing or the voltage is changing. As the signal passes across the element, the slope of the current and the voltage will change with time and the impedance of the element will change in time. This means the reflection coefficient will change with time and with the specific features of the rise or fall time. The peak reflected voltage will scale with the rise time of the signal.

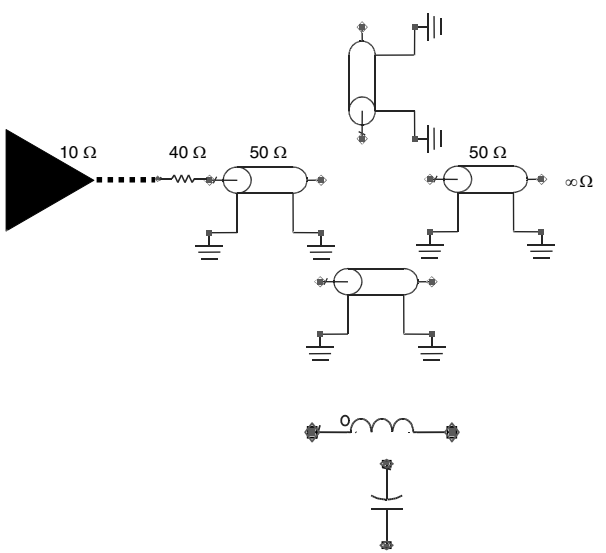


Figure 8-15 Transmission line circuit used to illustrate the specific impedance from the three types of discontinuities, short transmission line in series and shunt, shunt capacitive, and series inductive.

In general, the impact of a discontinuity is further complicated by the impedance of the driver and the characteristic impedance of the initial transmission line influencing the multiple bounces.

TIP These factors as well as the impact from the discontinuity itself can only be fully taken into account by converting the physical structure that creates the discontinuity into its equivalent electrical-circuit model and performing a simulation. Rule-of-thumb estimates can only provide engineering insight and offer rough guidelines for when a problem might arise.

Any impedance discontinuity will cause some reflection and distortion of the signal. It is not impossible to design an interconnect with absolutely no reflections. How much noise can we live with and how much noise is too much? This depends very strongly on the noise budget and how much noise voltage has been allocated to each source of noise.

TIP Unless otherwise specified, as a rough rule of thumb, the reflection noise level should be kept to less than 10% of the voltage swing. For a 3.3-v signal, this is 330 mV of noise. Some noise budgets might be more conservative and allocate no more than 5% to reflection noise. The tighter the noise budget, typically, the more expensive the solution. Often, the noise allocated to one source may be tightened up because the fix to correct it is less expensive to implement, while another might be loosened because it is more expensive to fix. As a rough rule of thumb, we should definitely worry about the reflection noise if it approaches or exceeds 10% of the signal swing. In some designs, less than 5% may be too much.

By evaluating a few simple cases, we can see what physical factors influence the signal distortions and how to engineer them out of the design before they become problems. Ultimately, the final evaluation of whether a design is acceptable or not must come from a simulation. This is why it is so important that every practicing engineer with a concern about signal integrity have easy access to a simulator to be able to evaluate specific cases.

8.9 When to Terminate

The simplest transmission line circuit has a driver at one end, a short length of controlled-impedance line, and a receiver at the far end. As we saw previously, the

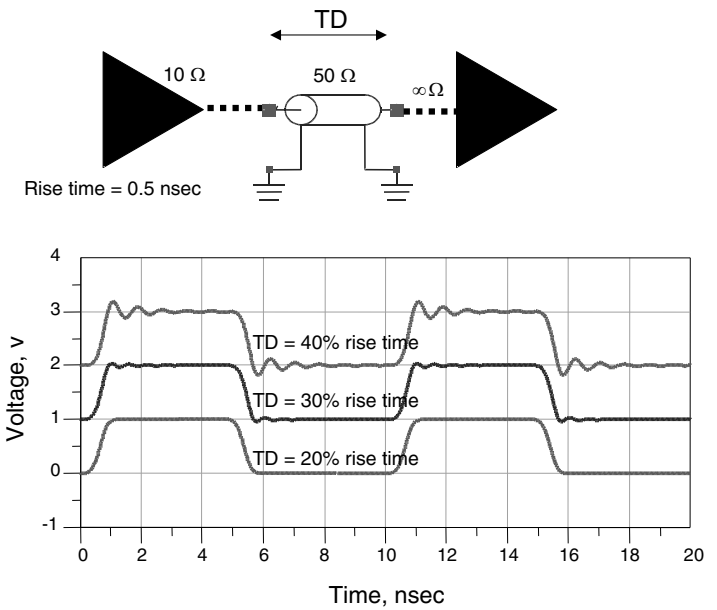


Figure 8-16 100-MHz clock signal at the far end of an unterminated transmission line as its length is changed from 20% of the rise time, to 30% and 40% of the rise time. When the time delay of the transmission line is greater than 20% of the rise time, ringing noise may cause a problem.

signal will bounce around between the high-impedance open at the far end and the low impedance of the driver at the near end. When the length of the line is long, these multiple bounces will cause signal-quality problems, which we lump in the general category of ringing. But if the line is short enough, though the reflections will still happen, they may be smeared out with the rising or falling edge and may not pose a problem. Figure 8-16 illustrates how the received waveform changes when the time delay increases from 20% of the rise time, to 30% of the rise time, and 40% of the rise time.

When the TD of the interconnect is 0.1 nsec long, all the reflections take place but they rattle around back and forth every 0.2 nsec, the round-trip time of flight. If this is short compared with the rise time, the multiple bounces will be smeared over the rise time and will be barely discernible, not posing a potential problem. In the previous figure, as a rough estimate, it appears that when the TD is less than 20%, the reflections are virtually invisible, but if the TD is greater than 20% of the rise time, the ringing begins to play a significant role.

TIP We use, as a rough rule of thumb, the threshold of TD > 20% of the rise time as the boundary of when to start worrying about ringing noise due to an unterminated line. If the TD of the transmission line is greater than 20% of the rise time, ringing will play a role and must be managed; otherwise, it will potentially cause a signal-integrity problem. If the TD < 20% of the rise time, the ringing noise may not be a problem and the line may not require termination.

For example, if the rise time is 1 nsec, the maximum TD for a transmission line that might be used unterminated is ~ 20% × 1 nsec = 0.2 nsec. In FR4, the speed of a signal is about 6 inches/nsec, so the maximum physical length of an unterminated line is roughly 6 inches/nsec × 0.2 nsec = 1.2 inches.

This allows us to generalize a very useful rule of thumb that the maximum length for an unterminated line before signal-integrity problems arise is roughly:

$$Len_{max} < RT \tag{8-13}$$

where:

Len_{max} = the maximum length for an unterminated line, in inches

RT = the rise time, in nsec

TIP This is a very useful and easy to remember result. As a rough rule of thumb, the maximum length of an unterminated line (in inches) is the rise time (in nsec).

If the rise time is 1 nsec, the maximum unterminated length is about 1 inch. If the rise time is 0.1 nsec, the maximum unterminated length is 0.1 inch. As we will see, this is the most important general rule of thumb to identify when ringing noise will play a significant role. This is also why signal integrity is becoming a significant problem in recent years and might have been avoided in older generation technologies.

When clock frequencies were 10 MHz, the clock periods were 100 nsec and the rise times were about 10 nsec. The maximum unterminated line would be 10 inches. This is longer than virtually all traces on a typical mother board. Back in the days of 10-MHz clocks, though the interconnects always behaved like transmission lines, the reflection noise never caused a problem and the interconnects

were "transparent" to the signals. We never had to worry about impedance matching, terminations, or transmission line effects.

However, the form factor of products is staying the same, and lengths of interconnects are staying fixed, but rise times are decreasing. Therefore, it is inevitable that we reach a high enough clock frequency with a short enough rise time that virtually *all* the interconnects on a board will be longer than the maximum possible unterminated length and termination will be important.

These days, with rise times of signals as short as 0.1 nsec, the maximum unterminated length of a transmission line before ringing noise becomes important is about 0.1 inch. Virtually 100% of all interconnects are longer than this. A termination strategy is a must in all of today's and future-generations' products.

8.10 The Most Common Termination Strategy for Point-to-Point Topology

We have identified the origin of the ringing as the impedance discontinuities at the source and the far end and the multiple reflections back and forth. If we eliminate the reflections from at least one end, we can minimize the ringing.

> **TIP** Engineering the impedance at one or both ends of a transmission line to minimize reflections is called *terminating the line*. Typically one or more resistors are added at strategic locations.

When one driver drives one receiver, we call this a *point-to-point topology*. There are four techniques to terminate a point-to-point topology illustrated in Figure 8-17. The most common method is using a resistor in series at the driver. This is called *source-series termination*. The sum of the terminating resistor and the source impedance of the driver should add up to the characteristic impedance of the line.

If the driver-source impedance is 10 Ohms and the characteristic impedance of the line is 50 Ohms, then the terminating resistor should be about 40 Ohms. With this terminating resistor in place, the 1-v signal coming out of the driver will now encounter a voltage divider composed of the 50-Ohm total resistance and the 50-Ohm transmission line. In this case, 0.5 v will be launched into the transmission line.

At first glance, it might seem that half the source voltage will not be enough to affect any triggering. However, when this 0.5-v signal hits the open, far end of the transmission line, it will again see an impedance discontinuity. The reflection coefficient at the open is 1, and the 0.5-v incident signal will reflect back to the source with an amplitude of 0.5 v. At the far end, the total voltage across the open termination will be the 0.5-v incident voltage and the 0.5-v reflected voltage, or a total of 1 v.

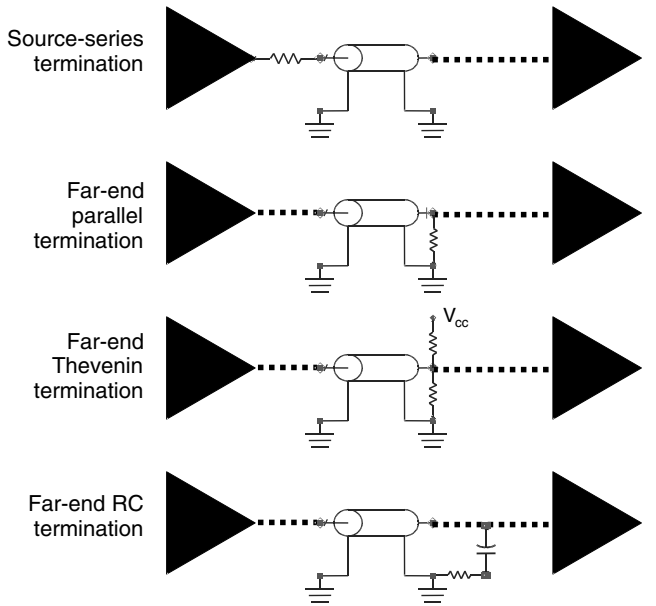


Figure 8-17 Four common termination schemes for point-to-point topologies. The top one, source-series termination, is the most commonly used approach.

The 0.5-v reflected signal travels back to the source. When it reaches the series terminating resistance, the impedance it sees looking into the source is the 40 Ohms of the added series resistor plus the 10 Ohms of the source, or 50 Ohms. It is already in a 50-Ohm transmission line; therefore the signal will encounter no impedance change and there will be no reflection. The signal will merely be absorbed by the terminating resistor and the source resistor.

At the far end, all that is seen is a 1-v signal and no ringing. Figure 8-18 shows the waveform at the far end for the case of no terminating resistor and the 40-Ohm source-series terminating resistor.

TIP Understanding the origin of the reflections allows us to engineer a solution to eliminate reflections at one end and prevent ringing. The resulting waveform is very clean and free of signal-quality problems.

At the near end, coming out of the source, right after the source-series terminating resistor, the initial voltage that would be measured is just the incident voltage launched into the transmission line. This is about half the signal voltage.

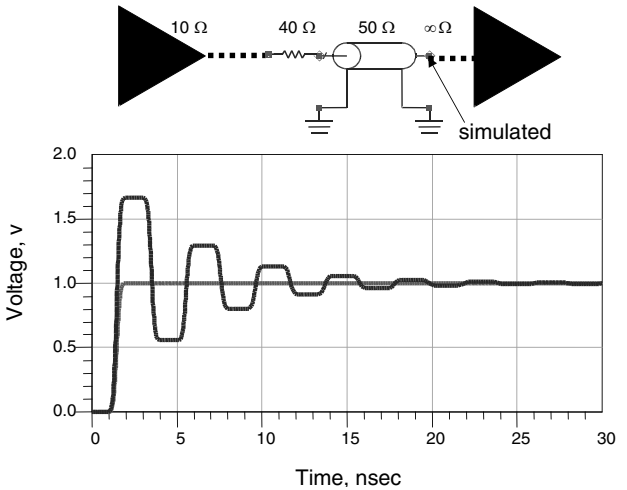


Figure 8-18 Voltage signal of a fast edge, at the far end of the transmission line with and without the source series terminating resistor.

Looking at the source end, we would have to wait for the reflected wave to come by to bring the total voltage up to the full voltage swing. For a time equal to the round-trip time of flight, the voltage at the source end, after the series resistor, will encounter a shelf. The longer the round-trip time delay of the transmission line compared to the rise time, the longer the shelf will last. This is a fundamental characteristic of source-series terminated lines. An example of the measured voltage at the source end is shown in Figure 8-19.

As long as there are no other receivers near the source to see this shelf, it will not cause a problem. It is when other devices are connected near the source, that the shelf may cause a problem and other topologies and termination schemes might be required.

In the following examples, we are always assuming the source impedance has been matched to the 50-Ohm characteristic impedance of the initial transmission line.

8.11 Reflections from Short Series Transmission Lines

Many times the line width of a trace on a board must neck down, as it might in going through a via field or routing around a congested region of the board. If the line width changes for a short length of the line, the characteristic impedance will change, typically increasing. How much change in impedance and over what length might start to cause a problem?

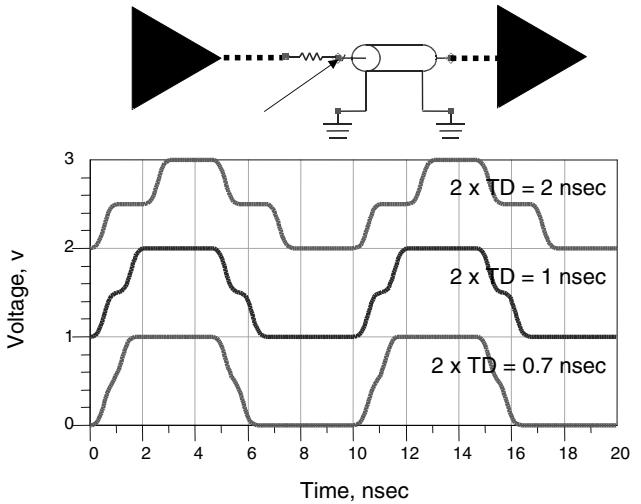


Figure 8-19 100-MHz clock signal, measured at the source end of the transmission line with a source-series resistor as the length of the line is increased. Rise time of the signal is 0.5 nsec.

There are three features that determine the impact from a short transmission line segment: the time delay (TD) of the discontinuity, the characteristic impedance of the discontinuity (Z_0), and the rise time of the signal (RT). When the time delay is long compared to the rise time, in other words, the discontinuity is electrically long, the reflection coefficient will saturate. The maximum value of the reflection coefficient will be related to the reflection from the front of the discontinuity:

$$\rho = \frac{Z_2 - Z_1}{Z_2 + Z_1} \tag{8-14}$$

For example, if the neck down causes an impedance change from 50 Ohms to 75 Ohms, the reflection coefficient would be 0.2. Some examples of the reflected and transmitted signals from electrically long transmission line discontinuities are shown in Figure 8-20.

These impedance discontinuities cause the signal to rattle around, contributing to reflection noise. This is why it is so important to design interconnects with one, uniform characteristic impedance. To keep the reflection noise to less than 5% of the voltage swing requires keeping the characteristic-impedance change to

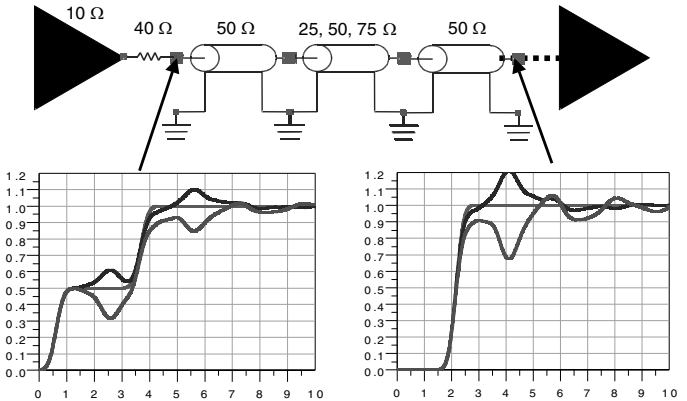


Figure 8-20 Reflected and transmitted signal in a transmission line circuit with an electrically long but uniform discontinuity, as the impedance of the discontinuity is changed.

less than 10%. This is why the typical spec for the control of the impedance in a board is for +/– 10%.

Note that whatever the reflection is from the first interface, it will be equal but of opposite sign from the second interface, since Z_1 and Z_2 would be reversed. If the discontinuity can be kept short, the reflections from the two ends might cancel and the impact on signal quality can be kept negligible. Figure 8-21 shows the reflected and transmitted signal with a short-length discontinuity that is 25 Ohms. If the TD of the discontinuity is shorter than 20% the rise time, the discontinuity may not cause problems. This gives rise to the same general rule of thumb as before; that is, the maximum acceptable length for an impedance discontinuity is:

$$\text{Len}_{max} < \text{RT} \tag{8-15}$$

where:

Len_{max} = the maximum length for a discontinuity, in inches

RT = the rise time, in nsec

For example, if the rise time is 0.5 nsec, neck downs shorter than 0.5 inch may not cause a problem.

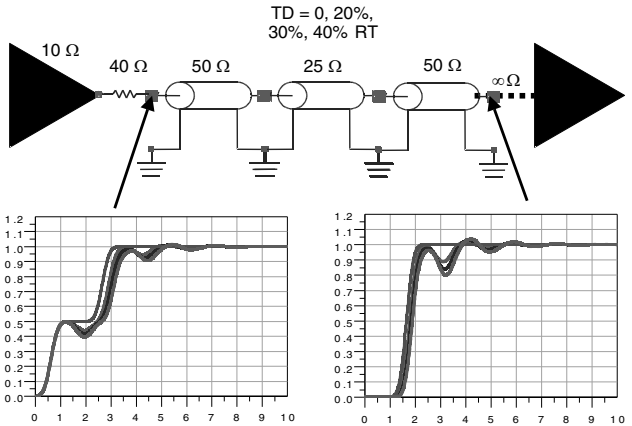


Figure 8-21 Reflected and transmitted signal in a transmission line circuit with an electrically short but uniform discontinuity, as the time delay of the discontinuity is increased from 0% to 40% of the rise time.

TIP If the TD of the discontinuity can be kept shorter than 20% of the rise time, the impact from the discontinuity may be negligible. This is the same rule of thumb that the length of the discontinuity, in inches, should be less than the rise time of the signal, in nsec.

8.12 Reflections from Short-Stub Transmission Lines

Often, a branch is added to a uniform transmission line to allow the signal to reach multiple fan outs. When the branch is short, it is called a stub. A stub is commonly found on BGA packages to allow busing all the pins together so the bonding pads can be gold plated. The buses are broken off during manufacturing, leaving small, short-length stubs attached to each signal line.

The impact from a stub is complicated to analyze because of all the many reflections that must be taken into account. As the signal leaves the driver, it will first encounter the branch point. Here it will see a low impedance from the parallel combination of the two transmission line segments, and a negative reflection will head back to the source. A fraction of the signal will continue down both branches. When the signal in the stub hits the end of the stub, it will reflect back to the branch point, again reflecting back to the end and rattling around in the stub. At the same time, at each interaction with the branch point, some fraction of the signal in the stub will head back to the source and to the far end. Each interface acts as a point of reflection.

The only practical way of evaluating the impact of a stub on signal quality is by using a SPICE or a behavioral simulator. The two important factors that determine the impact of the stub on signal quality are the rise time of the signal and the length of the stub. In this example, we assume the stub is located in the middle of the transmission line and has the same characteristic impedance as the main line. Figure 8-22 shows the simulated reflected signal and transmitted signal as the stub length is increased from 20% of the rise time to 60% of the rise time.

TIP As a rough rule of thumb, if the stub length is kept shorter than 20% of the spatial extent of the rise time, the impact from the stub may not be important. Likewise, if the stub is longer than 20% of the rise time, it may have an important impact on the signal and must be simulated to evaluate whether it will be acceptable.

For example, if the rise time of the driver is 1 nsec, a stub with a time delay shorter than 0.2 nsec might be acceptable. The length of the stub would be about 1 inch. Once again, the rule of thumb is:

$$Lstub_{max} < RT \tag{8-16}$$

where:

$Lstub_{max}$ = the maximum acceptable stub length, in inches

RT = the signal rise time, in nsec

This is a simple, easy-to-remember rule of thumb. For example, for a 1-nsec rise time, keep stubs shorter than 1 inch. If the rise time is 0.5 nsec, keep stubs shorter than 0.5 inch. It is clear that as rise times decrease, it gets harder and harder to engineer stubs short enough to not impact signal integrity.

In BGA packages, it is often not possible to avoid plating stubs used in the manufacture of the packages. These stubs are typically less than 0.25 inch long. When the rise time of the signals is longer than 0.25 nsec, these plating stubs may not cause a problem, but as rise times drop below 0.25 nsec, they definitely will cause problems, and it may be necessary to pay extra for packages that are manufactured without plating stubs.

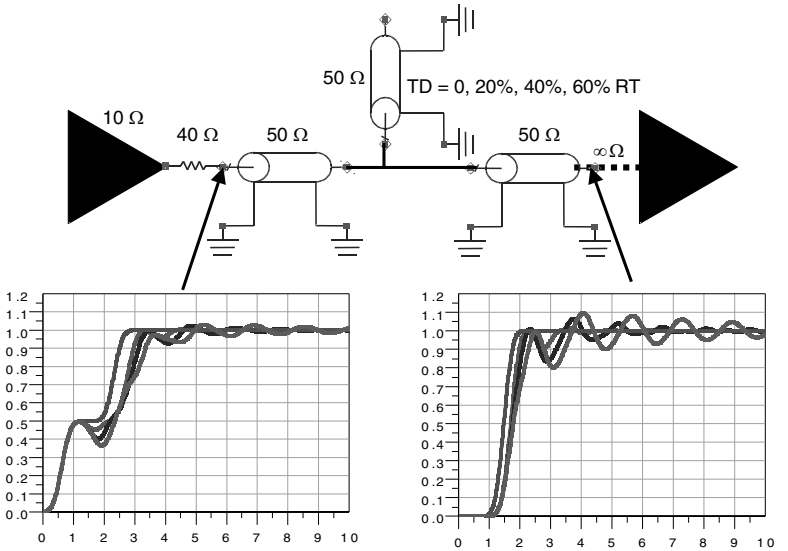


Figure 8-22 Reflected and transmitted signal in a transmission line circuit with a short stub in the middle, while the time delay of the stub is increased from 20% to 60% of the rise time.

8.13 Reflections from Capacitive End Terminations

All real receivers have some input-gate capacitance. This is typically on the order of 2 pF. In addition, the receiver's package-signal lead might have a capacitance to the return path of about 1 pF. If there is a bank of three memory devices at the end of a transmission line, there might be as much as a 10-pF load at the end of the transmission line.

When a signal travels down a transmission line and hits an ideal capacitor at the end, the actual instantaneous impedance the signal sees, which determines the reflection coefficient, will change with time. After all, the impedance of a capacitor, in the time domain, is related to:

$$Z = \frac{V}{C\frac{dV}{dt}} \tag{8-17}$$

where:

Z = the instantaneous impedance of the capacitor

C = the capacitance of the capacitor

V = the instantaneous voltage in the signal

If the rise time is short compared with the charging time of the capacitor, then initially, the voltage will rise up very fast and the impedance will be low. But as the capacitor charges, the voltage across it gets smaller and the dV/dt slows down. As the capacitor charges up, the rate at which the voltage across it changes will slow down. This will cause the impedance of the capacitor to increase dramatically. If we wait long enough, the impedance of a capacitor, after it has charged fully, is open.

This means the reflection coefficient will change with time. The reflected signal should suffer a dip and then move up to look like an open. The exact behavior will depend on the characteristic impedance of the line (Z_0), the capacitance of the capacitor, and the rise time of the signal. The simulated reflected and transmitted voltage behavior for a 2-pF, 5-pF, and 10-pF capacitance is shown in Figure 8-23.

The long-term response of the transmitted voltage pattern looks like the charging of a capacitor by a resistor. The presence of the capacitor filters the rise time and acts as a "delay adder" for the signal at the receiver. We can estimate the new rise time and the increase in time delay for the signal to transition through the midpoint (i.e., the delay adder), because it is very similar to the charging of an RC circuit, where the voltage increases with an exponential time constant:

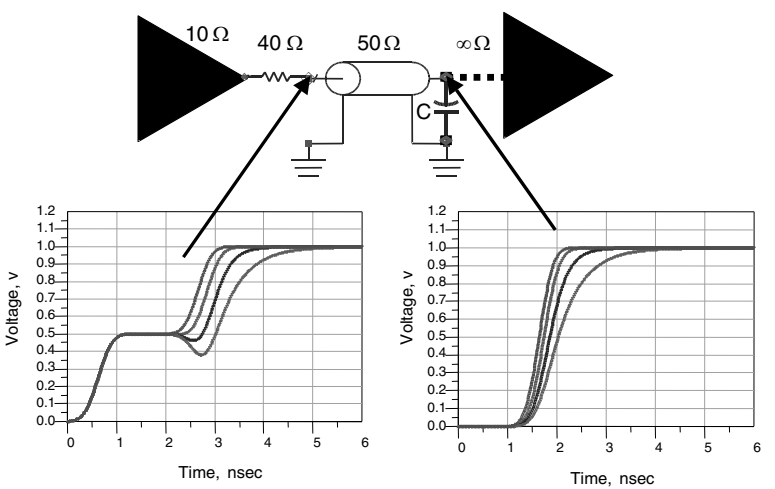


Figure 8-23 Reflected and transmitted signal in a transmission line circuit with a capacitive load at the far end, for a 0.5-nsec rise time and capacitances of 0, 2 pF, 5 pF, and 10 pF.

$$\tau_e = R \times C \tag{8-18}$$

This time constant is how long it takes for the voltage to rise up to 1/e or 37% of the final voltage. The 10%–90% rise time is related to the RC time constant by:

$$\tau_{10-90} = 2.2 \times \tau_e = 2.2 \times R \times C \tag{8-19}$$

At the end of a transmission line with a capacitive load, it looks like the voltage is charging up with an RC behavior. The C is the capacitance of the load. The R is the characteristic impedance of the transmission line, Z_0. The 10–90 rise time for the transmitted signal, if dominated by the RC charging, is roughly:

$$\tau_{10-90} = 2.2 \times Z_0 \times C \tag{8-20}$$

For example, if the transmission line has a characteristic impedance of 50 Ohms and the capacitance is 10 pF, the 10–90 charging time will be 2.2 × 50 Ohms × 10 pF = 1.1 nsec. If the initial-signal rise time is short compared with this 1.1-nsec charging time, the presence of the capacitive load at the end of the line will dominate and will now determine the rise time at the receiver. If the initial rise time of the signal is long compared to the 10–90 charging rise time, the capacitor at the end will add a delay to the rise time, roughly equal to the 10–90 rise time.

TIP Always be aware of the 10–90 RC rise time, which is based on the characteristic impedance of the line and the typical capacitive load of the input receiver. When the 10–90 rise time is comparable to the initial-signal rise time, the capacitive load at the far end will affect the timing.

A typical case is with a capacitance of 2 pF and characteristic impedance of 50 Ohms. The 10–90 rise time is about 2.2 × 50 × 2 = 0.2 nsec. When rise times are 1 nsec, this additional 0.2-nsec delay adder is barely discernible and may not be important. But when rise times are 0.1 nsec, the 0.2-nsec RC delay can be a significant delay adder. When driving multiple loads grouped at the far end, it is important to include the RC delay adder in all timing analysis.

8.14 Reflections from Capacitive Loads in the Middle of a Trace

A test pad, a via, a package lead, or even a small stub attached to the middle of a trace will act like a lumped capacitor. Figure 8-24 shows the reflected voltage and the transmitted voltage when a capacitor is added to the middle of a trace. Since the capacitor has a low impedance initially, the signal reflected back to the source will have a slight negative dip. If there were a receiver connected near the front end of the trace, this dip may cause problems as it would look like a nonmonotonic edge.

The transmitted signal initially might not experience a large impact on the first pass of the signal, but after the signal reflects from the end of the line, it will head back to the source. It will then hit the capacitor again and some of the signal, now with a negative sign, will reflect back to the far end. This reflection back to the receiver will be a negative voltage and will pull the received signal down, causing undershoot.

The impact of an ideal capacitor in the middle of a transmission line will depend on the rise time of the signal and the size of the capacitance. The larger the capacitor, the lower the impedance it will have and the larger the negative reflected voltage, which will contribute to a larger undershoot at the receiver. Likewise, the shorter the rise time, the lower the impedance of the capacitor and

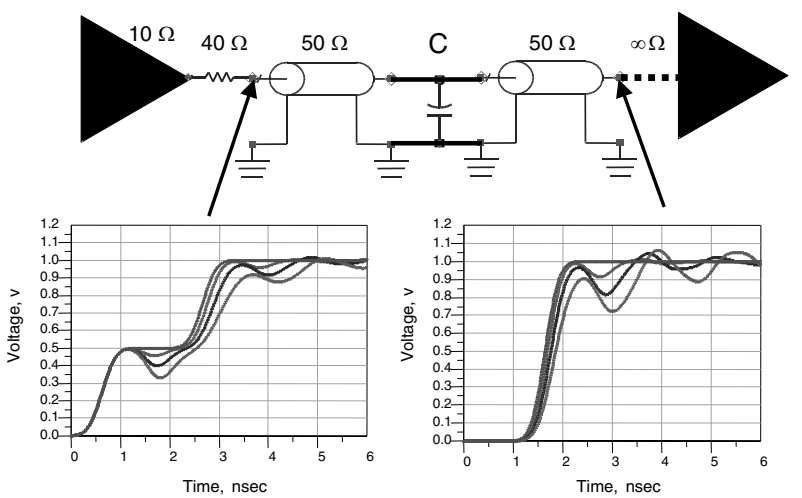


Figure 8-24 Reflected and transmitted signal in a transmission line circuit with a small capacitive discontinuity in the middle of the trace for a 0.5-nsec rise time and capacitances of 0, 2 pF, 5 pF, and 10 pF.

the greater the undershoot. If a certain capacitance, C_{max}, is barely acceptable for a certain rise time, RT, and if the rise time were to decrease, the maximum allowable capacitance would have to decrease as well. It is as though the ratio of RT/C_{max} must be greater than some value to be acceptable.

This ratio of the rise time to the capacitance has units of Ohms. But what is it the impedance of? The impedance of a capacitor, in the time domain, is:

$$Z_{cap} = \frac{V}{C\frac{dV}{dt}} \tag{8-21}$$

If the signal were a linear ramp, with a rise time, RT, then the dV/dt would be V/RT and the impedance of a capacitor would be:

$$Z_{cap} = \frac{V}{C\frac{dV}{dt}} = \frac{V}{C\frac{V}{RT}} = \frac{RT}{C} \tag{8-22}$$

where:

Z_{cap} = the impedance of the capacitor, in Ohms

C = the capacitance of the discontinuity, in nF

RT = the rise time of the signal, in nsec

It is as though the capacitor between the signal and return paths is a shunting impedance, of Z_{cap}, during the time interval of the rise time. This shunting impedance across the transmission line causes the reflections. This is illustrated in Figure 8-25. In order for this impedance to not cause serious problems, we would want its impedance to be large compared to the impedance of the transmission line. In other words, we would want $Z_{cap} >> Z_0$. As a starting place, this can be translated as $Z_{cap} > 5 \times Z_0$. The limit on the capacitance and rise time is translated as:

$$Z_{cap} > 5 \times Z_0 \tag{8-23}$$

$$\frac{RT}{C_{max}} > 5 \times Z_0 \tag{8-24}$$

$$C_{max} < \frac{RT}{5 \times Z_0} \tag{8-25}$$

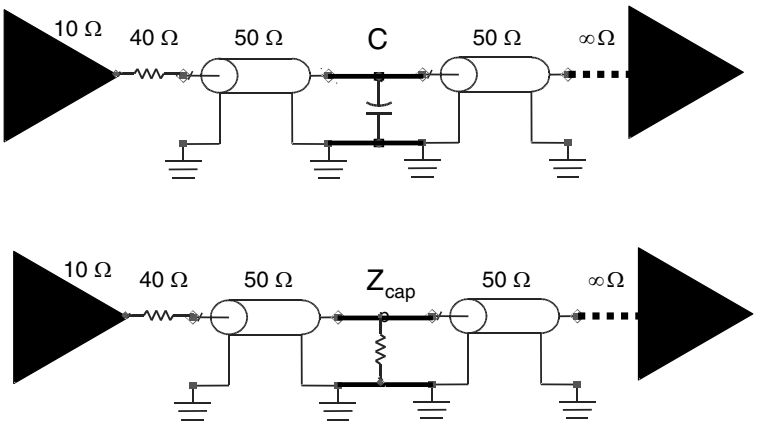


Figure 8-25 Describing the capacitive discontinuity shunting a transmission line as a shunt impedance, during the time interval when the edge passes by.

where:

Z_{cap} = the impedance of the capacitor during the rise time

Z_0 = the characteristic impedance of the transmission line, in Ohms

RT = the rise time, in nsec

C_{max} = the maximum acceptable capacitance, in nF, before reflection noise may be a problem

For example, if the characteristic impedance is 50 Ohms, the maximum allowable capacitance is:

$$C_{max} < \frac{RT}{5 \times 50} = 4 \times RT \tag{8-26}$$

where:

RT = the rise time in nsec

C_{max} = the maximum acceptable capacitance, in pF, before reflection noise may be a problem

This is the origin of a very simple rule of thumb.

TIP To keep capacitive discontinuities from causing excessive undershoot noise, keep the capacitance, in pF, less than four times the rise time, with the rise time in nsec.

If the rise time is 1 nsec, the maximum allowable capacitance would be 4 pF. If the rise time is 0.25 nsec, the maximum allowable capacitance discontinuity before undershoot problems might arise would be $0.25 \times 4 = 1$ pF. Likewise, if the capacitive discontinuity is 2 pF, the shortest rise time we might be able to get away with is 2 pF/4 = 0.5 nsec.

This rough limit suggests that if the rise time of a system is 1 nsec, it may be possible to get away with capacitive discontinuities on the order of 4 pF. Likewise, if the capacitance of an empty connector is 2 pF, for example, then this might be acceptable if rise times were longer than 0.5 nsec. However, if the rise time is 0.2 nsec, there may be a problem and it is critical to simulate the performance before committing to hardware. It may be worth the effort to look for alternative connectors or designs.

8.15 Capacitive Delay Adders

A capacitive load will cause the first-order problem of undershoot noise at the receiver. There is a second, more subtle impact from a capacitive discontinuity: The received time of the signal at the far end will be delayed. The capacitor combined with the transmission line acts as an RC filter. The 10–90 rise time of the transmitted signal will be increased, and the time for the signal to pass the 50% voltage threshold will be increased. The 10–90 rise time of the transmitted signal is roughly:

$$RT_{10-90} = 2.2 \times RC = 2.2 \times \frac{1}{2} Z_0 C = Z_0 C \tag{8-27}$$

The increase in delay time for the 50% point is referred to as the delay adder and is roughly:

$$\Delta TD = RC = \frac{1}{2} Z_0 C \tag{8-28}$$

where:

$RT_{10–90}$ = the 10% to 90% rise time, in nsec

ΔTD = the increase in time delay, in nsec for the 50% threshold

Z_0 = the characteristic impedance of the line, in Ohms

C = the capacitive discontinuity, in nF

The factor of 1/2 is there because the first half of the line charges the capacitor while the back half is discharging the capacitor, so the effective impedance charging the capacitor is really half the characteristic impedance of the line.

For example, in a 50-Ohm line, the increase in the 10–90 rise time of the transmitted signal for a 2-pF discontinuity will be about 50×2 pF = 100 psec. The 50% threshold delay adder will be about $0.5 \times 50 \times 2$ pF = 50 psec. Figure 8-26 shows the simulated rise time and delay in the received signal reaching the 50% threshold for three different capacitive discontinuities. The capacitor values of 2 pF, 5 pF, and 10 pF have an expected delay adder of 50 psec, 125 psec, and 250 psec. This estimate is very close to the actual simulated values.

It is very difficult to keep some capacitive discontinuities, which are created by test pads, connector pads, and via holes, to less than 1 pF. Every 1-pF pad will add about $0.5 \times 50 \times 1$ pF = 25 psec of added delay and increase the rise time of the signal. In very high-speed serial links, such as OC-48 data rates and above,

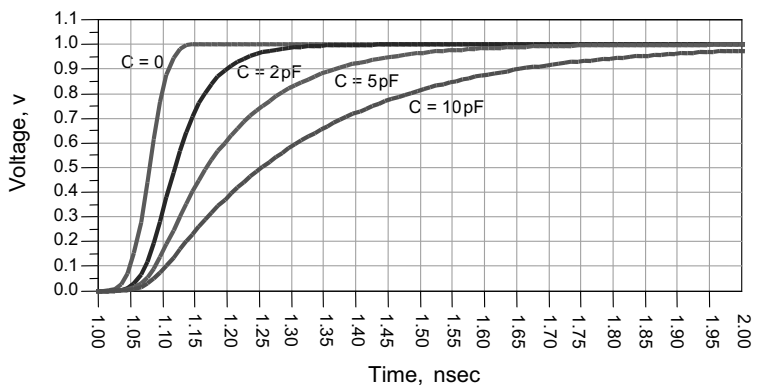


Figure 8-26 The resulting increase in delay time at the receiver for different values of a capacitive discontinuity in the middle of a 50-Ohm trace with a rise time of the signal of 50 psec. The estimates of the delay adders based on the simple rule of thumb are 50 psec, 125 psec, and 250 psec.

where the rise time of the signal may be 50 psec, every via pad or connector has the potential of adding 25 psec of delay and increasing the rise time of the signal by 50 psec. One via can easily double the rise time causing significant timing problems.

One way of minimizing the impact of delay adders is to use a low characteristic impedance. The lower the characteristic impedance, the less the delay adder for the same capacitive discontinuity.

8.16 Effects of Corners and Vias

As a signal passes down a uniform interconnect, there is no reflection and no distortion of the transmitted signal. If the uniform interconnect has a 90-degree bend, there will be an impedance change and some reflection and distortion of the signal. It is absolutely true that a 90-degree corner in an otherwise uniform trace will be an impedance discontinuity and affect the signal quality. Figure 8-27 is the measured TDR response of a 50-psec rise-time signal, reflecting off the impedance discontinuity of two 90-degree bends in close proximity. This is an easily measured effect.

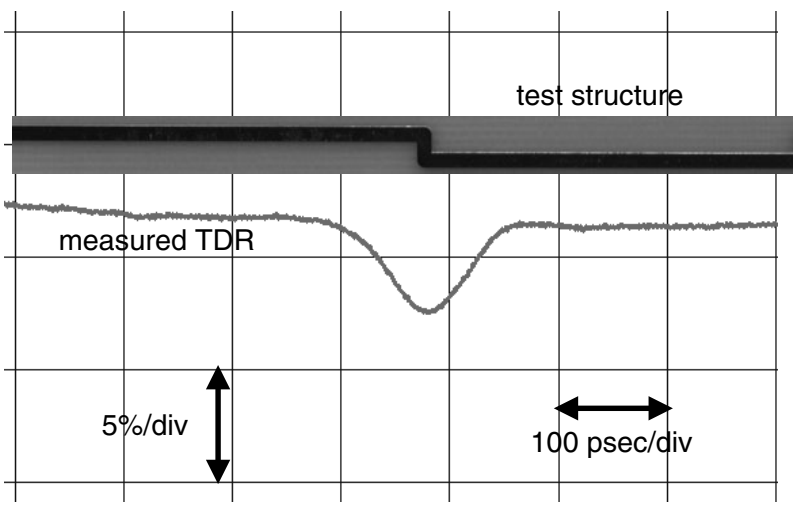


Figure 8-27 Measured TDR response of a uniform 50-Ohm line, 65 mils wide, with two 90-degree corners in close proximity. The rise time of the source is about 50 psec. Measured with an Agilent DCA 86100 and GigaTest Labs Probe Station.

Converting the 90-degree turn into two 45-degree bends will reduce this effect and using a rounded bend of constant width will reduce the impact from a corner even more. But, is the distortion from a corner a problem? Is the magnitude of the discontinuity large enough to worry about, and in what cases might a corner pose a problem? The only way to know the answer is to "put in the numbers," and the only way to do this is to understand the root cause of why a corner affects signal integrity.

It is sometimes thought that a 90-degree bend causes the electrons to accelerate around the bend, resulting in excess radiation and distortion. As we saw in an earlier chapter, the electrons in a wire are actually moving at the slow speed of about 1 cm/sec. In fact, the presence of the corner has no impact at all on the speed of the electrons. It is true that there will be high electric fields at the sharp point of the corner, but this is a DC effect and is due to the sharp radius of the outside edge. These high DC fields might cause enhanced filament growth and lead to a long-term reliability problem, but they will not affect the signal quality.

TIP The only impact a corner has on the signal transmission is due to the extra width of the line at the bend. This extra width acts like a capacitive discontinuity. It is this capacitive discontinuity that causes a reflection and a delay adder for the transmitted signal.

If the trace were to make the turn with a constant width, the width of the line would not change and the signal would encounter the same instantaneous impedance at each point around the turn and there would be no reflection. We can roughly estimate the extra metal a corner represents; Figure 8-28 illustrates that a corner will represent a small fraction of a square of extra metal. It is definitely less than one square, and as a very rough approximation, might be on the order of half a square of metal.

The capacitance of a corner can be estimated from the capacitance of a square and the capacitance per length of the trace:

$$C_{corner} = 0.5 \times C_{sq} = 0.5 \times C_L \times w \tag{8-29}$$

The capacitance per length of the line is related to the characteristic impedance of the line:

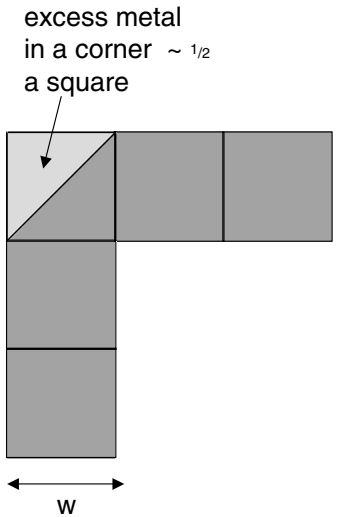


Figure 8-28 Simple estimate of the extra metal associated with a corner is about half of a square.

$$C_L = \frac{83}{Z_0}\sqrt{\varepsilon_r} \tag{8-30}$$

So an estimate of the capacitance of a corner is roughly:

$$C_{corner} = 0.5 \times C_L \times w = 0.5 \times w \times \frac{83}{Z_0}\sqrt{\varepsilon_r} \sim \frac{40}{Z_0} \times \sqrt{\varepsilon_r} \times w \tag{8-31}$$

where:

C_{corner} = the capacitance per corner in pF

C_L = the capacitance per length, in pF/inch

w = the line width of the line, in inches

Z_0 = the characteristic impedance of the line, in Ohms

ε_r = the dielectric constant of the dielectric

For example, in the 0.065-inch-wide trace measured above, the estimated capacitance in each of the two 90-degree bends is about $40/50 \times 2 \times 0.065 = 0.1$ pF

= 100 fF. In the above example, there are two corners in close proximity, so the total capacitance of the discontinuity is estimated as about 200 fF. Using the measured TDR response, we can estimate the excess capacitance caused by the discontinuity. Figure 8-29 shows the comparison of the measured response and the simulated response of a uniform line with a lumped capacitance of 200 fF in the middle. The excellent agreement suggests the discontinuity caused by the two corners can be modeled as a 200-fF capacitor. This is very close to the 200-fF capacitance suggested by the simple model.

We can generalize this estimate into a simple, easy-to-remember rule of thumb:

TIP In a 50-Ohm transmission line, the capacitance associated with a corner, in fF, is equal to 2 × the line width in mils.

As the line width gets narrower, while keeping the 50-Ohm impedance, the capacitance of a corner will decrease and its impact will be less significant. For a typical signal line in a high-density board, 5 mils wide, the capacitance of a corner is about 10 fF. The reflection noise from a 10-fF capacitor will be important for rise times on the order of 0.010 pF/4 ~ 3 psec. The delay adder from a 10-fF capacitance will be about 0.5 × 50 × 0.01 pF = 0.25 psec. It is unlikely the capacitance of a corner will play a significant signal-integrity role for 5-mil-wide lines.

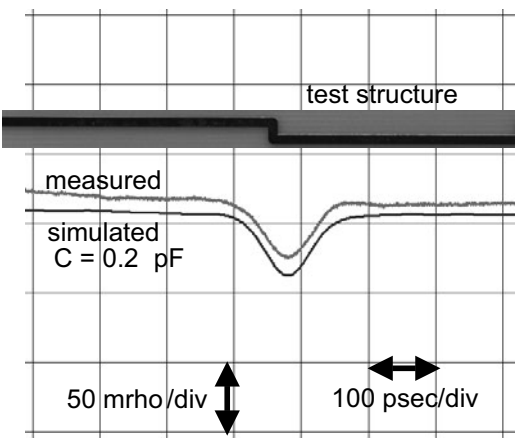


Figure 8-29 Measured and simulated TDR response of a uniform 50-Ohm line, 65 mils wide, with two 90-degree corners in close proximity. The rise time of the source is about 50 psec. The simulated response, based on a capacitance of 0.2 pF, is shifted down slightly for clarity. Measured with an Agilent DCA 86100 and GigaTest Labs Probe Station and simulated with TDA Systems IConnect software.

If a via connects a signal line to a test point or connects signal lines on adjacent layers but continues through the entire board, the barrel of the via may have excess capacitance to the various planes in the board. The residual via stub will cause a via often to look like a lumped capacitive load to a signal. The capacitance of a via stub will depend very strongly on the barrel size, the clearance holes, the size of the pads on the top and bottom of the board, and of course, the length of the stub. The residual capacitance can range from 0.1 pF to more than 1 pF. Any vias touching the signal line will probably look like a capacitive discontinuity, and in high-speed serial connections, are a chief limitation to the signal quality of the line.

The residual capacitance of a via can be estimated from a simple approximation. Unless special care is taken, the effective characteristic impedance of a via, including the return path through the various planes, is lower than 50 Ohms, on the order of 35 Ohms. The capacitance per length of a 50 Ohm line is 3.3 pF/inch. The capacitance per length of a via stub would be about 5 pF/inch, or about 5 fF/mil. This rough rule of thumb can be used to estimate the capacitive load of a via stub.

For example, a via stub 20 mils long, would have a capacitance of 20 mils × 5 fF/mil = 100 fF. A via stub in a thick board, 100 mils long, would have a capacitance of 100 mils × 5 fF/mil = 500 fF or 0.5 pF.

Figure 8-30 shows the measured TDR response of a uniform line with and without a single through-hole via in the middle of a 15-inch-long line in a 10-layer board. The trace impedance is about 58 Ohms, and the line width is nominally 8 mils. The rise time of the signal is about 50 psec. In this trace, the capacitance associated with both the SMA connector's via and the through-hole via in the middle of the line is about 0.4 pF. The difference in reflected voltage between the two vias is due to the rise-time degradation of the signal from dielectric losses as the signal propagates to the middle of the board and back. The changing reflected voltage along the line is a measure of the impedance variation due to manufacturing process variations.

This particular via can be approximated as a capacitance of about 0.4 pF. We would expect a delay adder from this single via to be about 0.5 × 50 × 0.4 pF = 10 psec. The transmitted signal in Figure 8-31 shows a 9-psec increase in delay time compared to a signal traveling through an identical line with no via. This is very close to the estimate based on the simple rule of thumb.

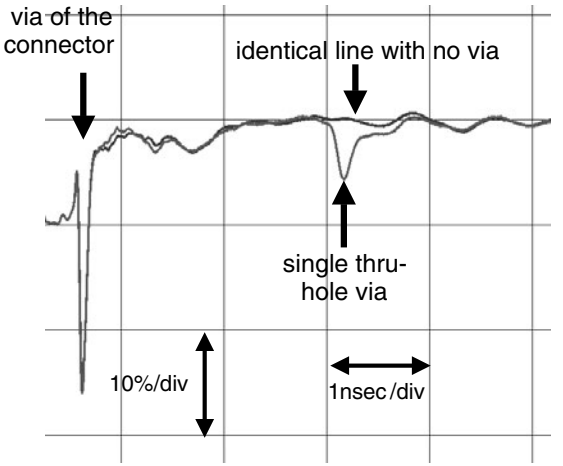


Figure 8-30 Measured TDR response of a uniform transmission line with and without a through-hole via in the middle, creating a capacitive discontinuity. The connector via at the front of both lines is also a capacitive discontinuity. Sample provided courtesy of Doug Brooks, UltraCAD. Measured with an Agilent DCA 86100, GigaTest Labs Probe Station, and TDA Systems IConnect software.

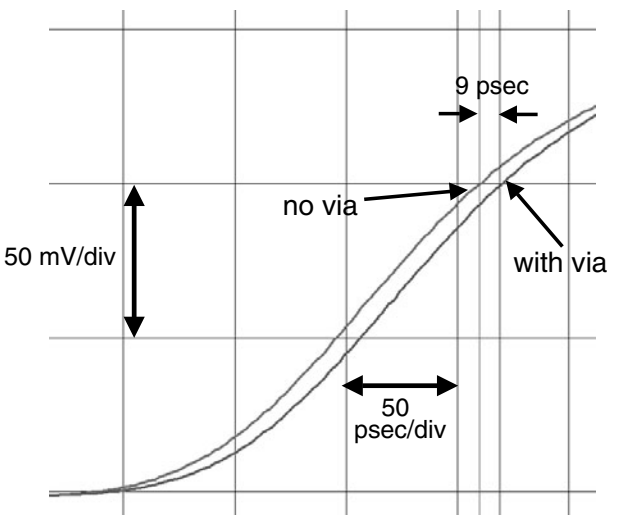


Figure 8-31 Measured transmitted signal after traveling 15 inches in a uniform transmission line and an identical line with a single through-hole via, showing a delay adder of 9 psec. Sample courtesy of Doug Brooks, UltraCAD. Measured with an Agilent DCA 86100, GigaTest Labs Probe Station, and TDA Systems IConnect software.

8.17 Loaded Lines

When there is one small capacitive load on a transmission line, the signal will be distorted and the rise time degraded. Each discrete capacitance acts to lower the impedance the signal sees in its proximity. If there are multiple capacitive loads distributed on the line (for example, 2-pF connector stubs spaced every 1.2 inches for a bus, or 3-pF package and input-gate capacitances distributed every 0.8 inch for a memory bus) and if the spacing is short compared to the spatial extent of the rise time, the reflections from each capacitive discontinuity may smear out. In this case, it appears as though the characteristic impedance of the line has been reduced. The transmission line with a distribution of uniformly spaced capacitive loads is called a *loaded line*.

Each discontinuity will look like a region of lower impedance. When the rise time is short compared to the time delay between the capacitances, each discontinuity acts like a discrete discontinuity to the signal. When the rise time is long compared to the time delay between them, the lower impedance regions overlap and the average impedance of the line looks lower.

An example of the reflected signal from a loaded line with three different rise times is shown in Figure 8-32. In this example, there are five 3-pF capacitors

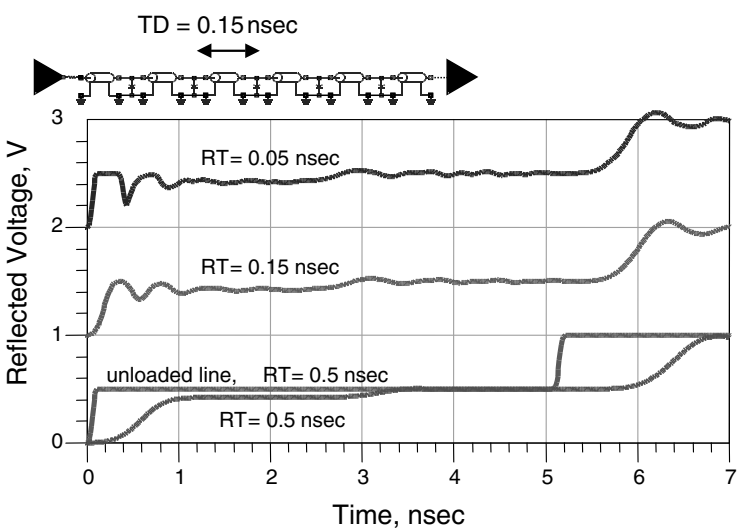


Figure 8-32 Reflected signal from a loaded line with the time delay between the 3-pF capacitors of 0.15 nsec. As the rise time increases, the reflection from each capacitor smears out.

distributed every 1 inch on a nominal 50-Ohm line. The last 10 inches of the line are unloaded. Each capacitor has an intrinsic 10–90 rise time of about 2.2 × 0.5 × 50 Ohms × 3 pF = 150 psec. Even though the initial rise time in the first example is 50 psec, after the first capacitor, the rise time is increased to 150 psec and longer after each capacitor.

The first few capacitors are visible as discrete discontinuities, but the later ones are smeared out by the long rise time of the transmitted signal. When the rise time of the signal is long compared to the time delay between the capacitive discontinuities, the uniformly distributed capacitive loads act to lower the apparent characteristic impedance of the line. In such a loaded line, it is as though the capacitance per length of the line has been increased by the added board features. The higher capacitance per length means a lower characteristic impedance and a longer time delay.

In a uniform, unloaded transmission line, the characteristic impedance and time delay are related to the capacitance per length and inductance per length as:

$$Z_0 = \sqrt{\frac{L_L}{C_{0L}}} \tag{8-32}$$

$$TD_0 = Len\sqrt{L_L C_{0L}} \tag{8-33}$$

where:

Z_0 = the characteristic impedance of the unloaded line, in Ohms

L_L = the inductance per length of the line, in pH/inch

C_{0L} = the capacitance per length of the unloaded line, in pF/inch

Len = the length of the line, in inches

TD_0 = the time delay of the unloaded line, in psec

When there are uniformly distributed capacitive loads, each of C_1 and spaced a distance of d_1, the distributed capacitance per length of the line is increased from C_{0L} to (C_{0L} +C_l/d_1). This changes the characteristic impedance of the line and its time delay to:

$$Z_{Load0} = \sqrt{\frac{L_L}{C_{0L} + \frac{C_1}{d_1}}} = Z_0 \sqrt{\frac{C_{0L}}{C_{0L} + \frac{C_1}{d_1}}} = Z_0 \sqrt{\frac{1}{1 + \frac{C_1}{C_{0L} d_1}}} \quad (8\text{-}34)$$

$$TD_{Load} = Len \sqrt{L_L \left(C_{0L} + \frac{C_1}{d_1} \right)} = TD_0 \sqrt{\left(1 + \frac{C_1}{C_{0L} d_1} \right)} \quad (8\text{-}35)$$

where:

Z_0 = the characteristic impedance of the unloaded line, in Ohms

Z_{Load0} = the characteristic impedance of the loaded line, in Ohms

L_L = the inductance per length of the line, in pH/inch

C_{0L} = the capacitance per length of the unloaded line, in pF/inch

C_1 = the capacitance of each discrete capacitor, in pF

d_1 = the distance between each discrete capacitor, in inches

Len = the length of the line, in inches

TD_0 = the time delay of the unloaded line, in psec

TD_{Load} = the time delay of the loaded-line region

In a 50-Ohm line, the capacitance per length is about 3.4 pF/inch. When the added distributed capacitive load is comparable to this, the characteristic impedance and time delay can be changed significantly. For example, if 3-pF loads from the input-gate capacitance of a memory bank are spaced every 1 inch in a multidrop bus, the additional loaded capacitance per length is 3 pF/inch. The loaded characteristic impedance is decreased to $0.73 \times Z_0$ and the time delay is increased to $1.37 \times TD_0$.

Since the characteristic impedance of the line is reduced, the terminating resistance should be reduced as well. Alternatively, in the region where the distributed capacitive loads are positioned, the line width could be reduced so the unloaded impedance is higher. When the loads are attached, the loaded line impedance would be closer to the target value of the impedance.

8.18 Reflections from Inductive Discontinuities

Just about every series connection added to a transmission line will have some series loop inductance associated with it. Every via used to change signal layers,

every series terminating resistor, every connector, and every engineering change wire will have some extra loop inductance. The signal will see this loop inductance as an additional discontinuity above and beyond what is in the transmission line.

If the discontinuity is in the signal path, the loop inductance will be dominated by the partial self-inductance of the signal path, though there will still be some partial mutual inductance with the return path. If the discontinuity is in the return path, the partial self-inductance of the return path will dominate the loop inductance. In either case, it is the loop inductance to which the signal is sensitive, because the signal is a current loop propagating down the line between the signal and return paths.

A series loop inductance will initially look like a higher impedance to an incident, fast rise-time signal. This will cause a positive reflection back to the source. Figure 8-33 shows the measured reflected signal from a uniform transmission line that travels over a small gap in the return path.

Figure 8-34 shows the signal at the receiver and source for different values of an inductive discontinuity. The shape of the signal at the near end, going up and then back down, is called *nonmonotonicity*. The signal is not increasing at a steady upward pace, monotonically. This feature, by itself, may not cause a signal-integrity problem. However, if there were a receiver located at the near end, and it received a signal increasing past the 50% point and then dropping down below 50%, it might be falsely triggered. Nonmonotonic behavior is to be avoided wherever possible. At the far end, the transmitted signal will show overshoot and a delay adder.

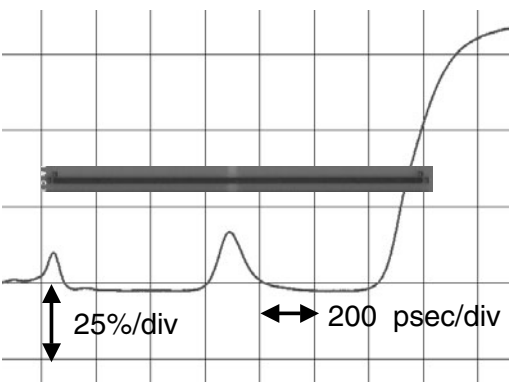


Figure 8-33 TDR-reflected signal from a uniform transmission line with an inductive discontinuity in the middle caused by a gap in the return path. The rise time is about 50 psec and was measured with an Agilent DCA 86100 and a GigaTest Labs Probe Station, analyzed with TDA Systems IConnect software.

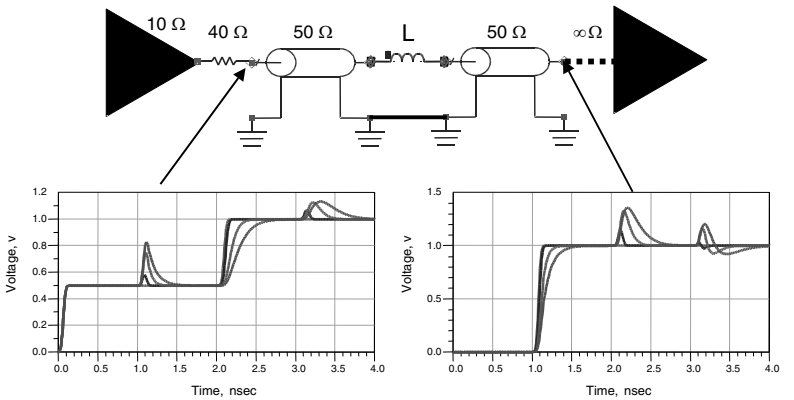


Figure 8-34 Signal at the source and the receiver with a 50-psec rise time interacting with an inductive discontinuity. Inductance values are L = 0, 1 nH, 5 nH, and 10 nH.

In general, the maximum amount of inductance that might be acceptable in a circuit depends on the noise margin and other features of the circuit. This means that each case must be simulated to evaluate what might be acceptable. However, as a rough measure of how much inductance might be too much, we can use the limit of when the series-impedance discontinuity of the discrete inductor increases to more than 20% of the characteristic impedance of the line. At this point, the reflected signal will be about 10% of the signal swing, usually the maximum acceptable noise allocated to reflection noise.

We can approximate the impedance of the inductor if it is small compared to the characteristic impedance and if the rise time is a linear ramp, while the rise time is passing through it, by:

$$Z_{inductor} = \frac{V}{I} = \frac{L\frac{dI}{dt}}{I} = \frac{L}{RT} \tag{8-36}$$

where:

$Z_{inductor}$ = the impedance of the inductor, in Ohms

L = the inductance, in nH

RT = the rise time of the signal, in nsec

The estimate of the maximum acceptable inductive discontinuity is set by keeping the impedance of the inductor less than 20% of the impedance of the line:

$$Z_{inductor} < 0.2 \times Z_0 \tag{8-37}$$

$$\frac{L_{max}}{RT} < 0.2 \times Z_0 \tag{8-38}$$

$$L_{max} < 0.2 \times Z_0 \times RT \tag{8-39}$$

where:

L_{max} = the maximum allowable series inductance, in nH

Z_0 = the characteristic impedance of the line, in Ohms

RT = the rise time of the signal, in nsec

For example, if the characteristic impedance of the line is 50 Ohms and the rise time is 1 nsec, the maximum acceptable series inductance would be about $L_{max} = 0.2 \times 50 \times 1$ nsec = 10 nH. This suggests a simple rule of thumb.

TIP As a rough approximation, in a 50-Ohm line, the maximum allowable excess loop inductance (in nH) is 10 times the rise time (in nsec). Likewise, if there is some loop inductance from a discontinuity, the shortest rise time that might be acceptable before reflection noise exceeds the noise budget (in nsec) is L/10 with the inductance (in nH).

If there is a residual loop inductance from a connector of 5 nH, the shortest usable rise time for that connector might be on the order of 5 nH/10 = 0.5 nsec. If the rise time of the signal is 0.1 nsec, all inductive discontinuities should be kept less than $10 \times RT = 10 \times 0.1 = 1$ nH.

Based on this estimate, we can evaluate the useful rise time for an axial lead and an SMT terminating resistor. The series loop inductance in an axial lead resistor is about 10 nH, while in an SMT resistor, it is about 2 nH.

TIP The shortest rise time at which an axial-lead resistor might be used, before reflection noise causes a problem, is about 10 nH/10 ~ 1 nsec, while for an SMT resistor, it is about 2 nH/10 ~ 0.2 nsec.

When the rise times are in the sub-nsec regime, axial-lead resistors are not suitable components and should be avoided. As rise times approach 100 psec, it will be increasingly important to engineer the use of SMT resistors with as low a loop inductance as possible. The two most important design features of a high-performance SMT resistor are a short length and a return plane as close to the surface as possible. Alternatively, resistors that are integrated into the board or package can have considerably lower loop inductance than 2 nH and might be required.

An inductive discontinuity contributes to reflection noise and also to a delay adder. When the rise time is very short and the transmitted rise time is dominated by the series inductor, the 10–90 rise time of the transmitted signal is approximately:

$$TD_{10-90} = 2.2 \times \frac{L}{2Z_0} = \frac{L}{Z_0} \tag{8-40}$$

$$TD_{adder} = 0.5 \times \frac{L}{Z_0} \tag{8-41}$$

where:

TD_{10-90} = the 10–90 rise time of the transmitted signal, in nsec

L = the series loop inductance of the discontinuity, in nH

Z_0 = the characteristic impedance of the line, in Ohms

TD_{adder} = the delay adder at the 50% point, in nsec

For example, a 10-nH discontinuity will increase the 10–90 rise time to about 10/50 = 0.2 nsec. The time delay added to the midpoint is roughly half of this or 0.1 nsec. Figure 8-35 shows the simulated time delay of the received signal with discontinuities of 1, 5, and 10 nH.

8.19 Compensation

Sometimes, it is unavoidable to have a series loop inductance in a circuit, as when a specific connector has already been designed in. Left to itself, it may cause excessive reflection noise. There is a technique called *compensation* that is used to cancel out some of this noise.

The idea is to try to trick the signal into not seeing a large inductive discontinuity, but seeing a section of transmission line that matches the characteristic impedance of the line. After all, an ideal transmission can be approximated to first

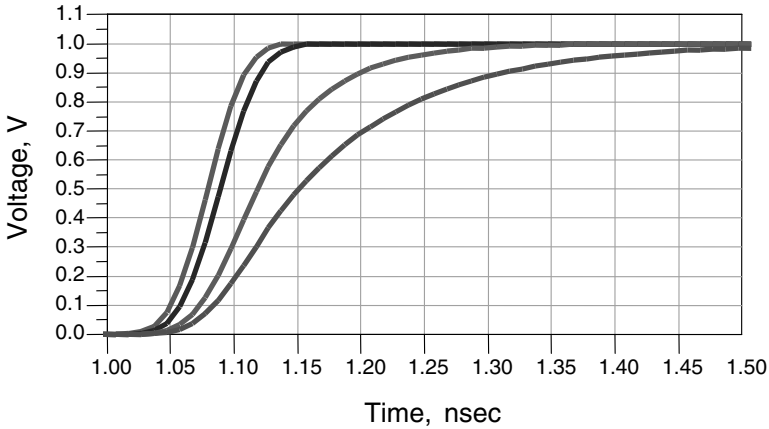


Figure 8-35 Delay adders in the received signal for a 50-psec rise time and inductive discontinuities of 0, 1 nH, 5 nH, and 10 nH. The estimated delay adders are 0, 10 psec, 50 psec, and 100 psec.

order, as a single section of an LC network. In this case, the characteristic impedance of any section of the line is given by:

$$Z_0 = \sqrt{\frac{L_L}{C_L}} = \sqrt{\frac{L}{C}} \tag{8-42}$$

where:

Z_0 = the characteristic impedance of the line, in Ohms

L_L = the inductance per length of the line, in nH/inch

L = the total inductance of any section of the line, in nH

C_L = the capacitance per length of the line, in nF/inch

C = the total capacitance of any section of the line, in nF

An inductive discontinuity can be turned into a transmission line segment by adding a small capacitor to either side. This is illustrated in Figure 8-36. In such a case, the apparent characteristic impedance of the inductor is:

$$Z_1 = \sqrt{\frac{L_1}{C_1}} \tag{8-43}$$

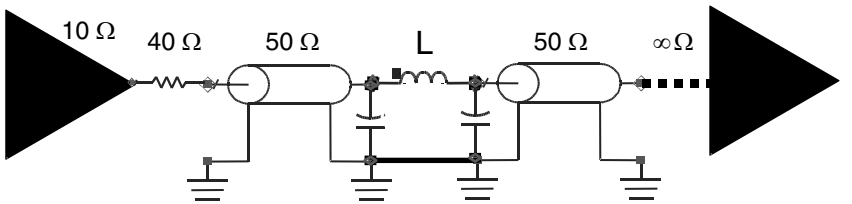


Figure 8-36 Compensation circuit for an inductive discontinuity. Add enough capacitance on either side to make the inductive discontinuity look like part of a 50-Ohm transmission line.

To minimize the reflected noise, the goal is to find values of the capacitors so that the apparent characteristic impedance of the connector, Z_1, is equal to the characteristic impedance of the rest of the circuit, Z_0. Using the relationship above, the capacitance to add is:

$$C_1 = \frac{L_1}{Z_0^2} \tag{8-44}$$

where:

C_1 = the total compensation capacitance to add, in nF

L_1 = the inductance of the discontinuity, in nH

Z_0 = the characteristic impedance of the line, in Ohms

For example, if the inductance of the connector is 10 nH and the characteristic impedance of the line is 50 Ohms, the total compensation capacitance to add is $10/50^2$ = 0.004 nF = 4 pF. For optimum compensation, the 4-pF capacitance should be distributed with 2 pF on each side of the inductor.

Figure 8-37 shows the reflected and transmitted signals for the three cases of no connector, an uncompensated connector, and the same connector compensated. Depending on the rise time of the system, the reflected noise can sometimes be reduced by more than 75%.

This technique applies to all inductive discontinuities such as vias and resistors as well. Depending on the relative amount of capacitance in the pads and the series inductance, a real discontinuity may look capacitive or inductive.

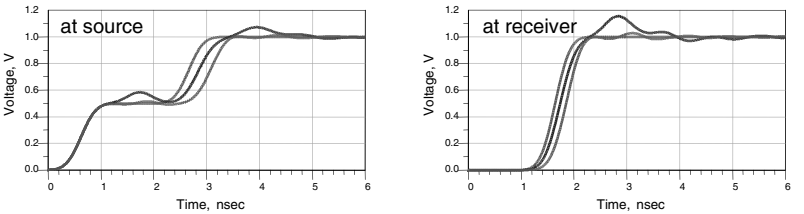


Figure 8-37 Source and transmitted signal for a 10-nH inductive discontinuity and 0.5-nsec rise time. The simulations are for no connector, with connector but no compensation, and for 2-pF capacitance added to each side of the inductor to compensate for the inductance.

TIP The goal in designing interconnects is to engineer the pads and other features to make the structure look like a section of uniform transmission line. In this way, some inductive discontinuities, such as connectors, can be made to nearly disappear.

8.20 The Bottom Line

1. Wherever the instantaneous impedance a signal sees changes, there will be a reflection and the transmitted signal will be distorted. This is the primary source of signal-quality problems associated with a single net.
2. As a rough rule of thumb, any transmission line longer, in inches, than the rise time, in nsec, will need a termination to prevent excessive ringing noise.
3. Source-series termination is the most common termination scheme for point-to-point connections. A series resistor should be added so that the sum of the resistor and the source impedance match the characteristic impedance of the line.
4. A SPICE or other circuit simulator is essential for every engineer involved with signal integrity. Many are low cost and easy to use. These tools allow simulation of all the multiple bounces due to impedance discontinuities.
5. As a rough rule of thumb, variations in the characteristic impedance of a line should be kept to less than 10% to keep reflection noise less than 5%.
6. As a rough rule of thumb, if the length of a short transmission line discontinuity can be kept shorter, in inches, than the rise time of the signal, in nsec, the reflections from the discontinuity might not cause a problem.

7. As a rough rule of thumb, if the length of a short stub can be kept shorter, in inches, than the rise time of the signal, in nsec, the reflections from the stub might not cause a problem.
8. A capacitive load at the far end of a line will cause a delay adder but no signal quality noise.
9. As a rough rule of thumb, a capacitive discontinuity in the middle of a line will cause excessive reflection noise if the capacitance, in pF, is greater than four times the rise time, in nsec.
10. The delay adder, in psec, for a capacitive discontinuity in the middle of a line will be about 25 times the capacitance, in pF.
11. A corner contributes a capacitance, in fF, of about two times the line width, in mils.
12. A uniformly spaced distribution of capacitive loads will lower the effective characteristic impedance of the line.
13. The magnitude of an acceptable inductive discontinuity, in nH, is about 10 times the rise time of the signal, in nsec.
14. The impact from an inductive discontinuity can be greatly minimized by adding capacitance on either side of the inductor to fool the signal into seeing a section of a uniform transmission line. In this way, vias can be engineered to be nearly invisible to high-speed signals.

This page intentionally left blank

CHAPTER 9

Lossy Lines, Rise-Time Degradation, and Material Properties

A signal with a very fast edge going into a real transmission line will come out with a longer rise time. As an example, the measured response of a 50-psec signal transported through a 36-inch-long, 50-Ohm backplane line in FR4 is shown in Figure 9-1. The rise time has been increased to almost 1 nsec. This rise-time degradation is due to the losses in the transmission line and is the dominant root cause of intersymbol interference (ISI) and the collapse of the eye diagram.

Loss in transmission lines is the principal signal-integrity problem for all signals with clocks higher than 1 GHz, transported for distances longer than 10 inches, such as in high-speed serial links and Gigabit ethernet.

TIP The reason the rise time is increased in propagating down a real transmission line is specifically because the higher-frequency components of the signal are preferentially attenuated more than the lower frequencies.

The simplest way to analyze the frequency-dependent losses is in the frequency domain. On the other hand, the problems created by lossy lines are time-domain dependent, and ultimately the final response must be analyzed in the time domain. In this chapter, we look first in the frequency domain to understand the

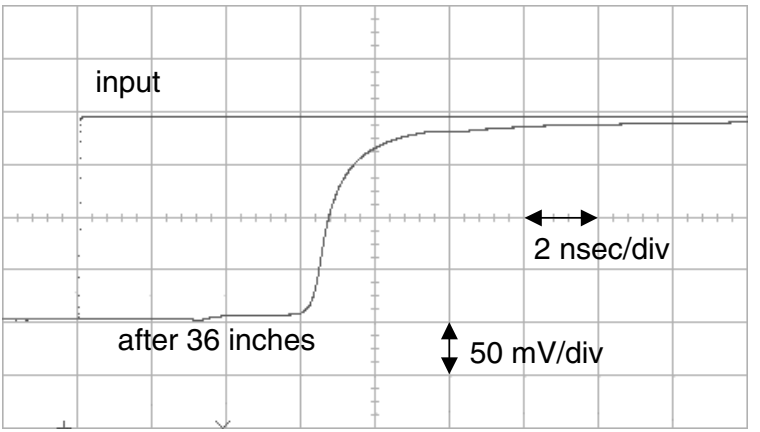


Figure 9-1 Measured input and output signals through a 50-Ohm line, 36 inches long. A 50-psec rise-time signal goes in, while a 1-nsec rise-time signal comes out.

loss mechanisms and then switch to the time domain to evaluate the impact on the signal-integrity response.

9.1 Why Worry About Lossy Lines?

If the losses were independent of frequency and low-frequency components were attenuated the same as high-frequency components, the entire signal waveform would uniformly decrease in amplitude, but the rise time would stay the same. This is illustrated in Figure 9-2. The effect of the constant attenuation could be compensated with some gain at the receiver. The rise time, the timing, and the jitter would remain unaffected by the constant attenuation.

It is not the loss that causes rise-time degradation, ISI, collapse of the eye, and deterministic jitter, it is the frequency dependence of the loss.

As a signal propagates down a real lossy transmission line, the amplitudes of the higher-frequency components are reduced and the low-frequency components stay about the same. The bandwidth of the signal is decreased because of this selective attenuation. As the bandwidth of the signal decreases, the rise time of the signal increases. It is the frequency dependence of the losses that specifically drives the rise-time degradation.

If the rise-time degradation were small compared to the period for one bit, or the unit interval, the bit pattern would be very constant and independent of what came previously. By the time one bit cycle ended, the signal would have stabilized and reached the final value. The voltage waveform for one bit in the stream would

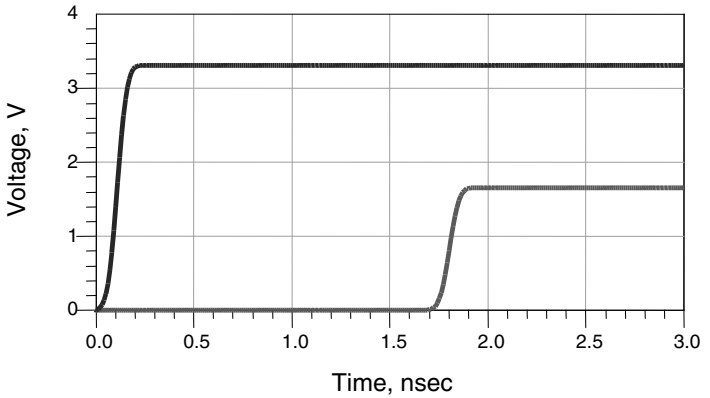


Figure 9-2 Simulated signal propagation of a 100-psec rise-time signal with losses that are independent of frequency. The only impact is a scaling of the signal.

be independent of what the previous bit was, whether it was a high or a low and for how long it was high or low. In this case, there would be no ISI.

However, if the rise-time degradation increases the received rise time significantly, comparable to the unit interval, the actual voltage level for a bit will depend on how long the signal was in a previous high or low state. If the bit pattern was high for a long time previously and the signal drops down for one bit and then up, the low voltage will not have had time to fall all the way to the lowest voltage. The precise voltage levels achieved by a single bit will depend on the previous bit pattern. This is called *intersymbol interference* (*ISI*) and is illustrated in Figure 9-3.

The important consequence of frequency-dependent loss and rise-time degradation is ISI: The precise waveform of the bit pattern will depend on the previous bits that have passed by. This will significantly affect the ability to tell the difference between a low and high level signal by the receiver, increasing the bit error rate.

TIP In addition, the time for the signal to reach the switching threshold will change depending on the previous data pattern. ISI is a significant contributor to jitter. If the rise time were short compared to the bit period, there would be no ISI.

One of the common metrics for describing the signal quality of a high-speed serial link, at the receiver, is an eye diagram. A pseudorandom bit stream, whose pattern represents all possible bit-stream patterns, is simulated or measured, using the clock reference as the trigger point. Each received cycle is taken out of the bit

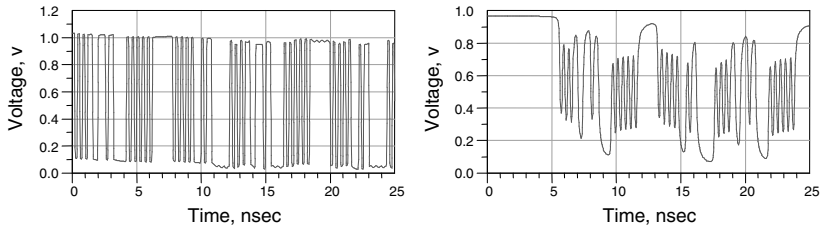


Figure 9-3 5-Gbps pseudorandom bit stream. Left: bit pattern when the rise time is much shorter than the bit period. Right: bit pattern when the rise time is comparable to the bit pattern, causing pattern-dependent voltage levels or intersymbol interference.

stream and overlaid on the previous cycle and hundreds of cycles are superimposed. This set of superimposed waveforms is called an eye diagram because it looks like an open eye.

The closing of the eye diagram is a measure of the bit error rate. The smaller the opening, the higher the bit error rate. Therefore, a large opening allows the possibility of a low bit error rate and good signal quality. The horizontal width of the cross-over region that separates eye openings is a measure of the jitter. A direct consequence of frequency-dependent lossy lines and an indirect measure of the ISI is the collapse of the opening of the eye. Figure 9-4 shows the collapse of the eye diagram for the same 5-Gbps (Giga-bit-per-second) waveform, with and without losses.

9.2 Losses in Transmission Lines

The first-order approximate model for a transmission line is an n-section LC model. This approximation is often referred to as the *lossless model*. It accounts for the two important features of a transmission line: characteristic impedance and

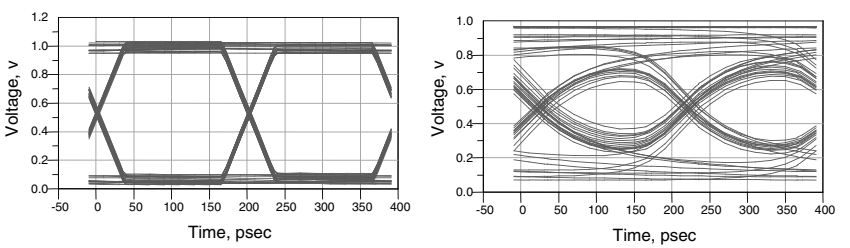


Figure 9-4 Eye diagrams of a 5-Gbps pseudorandom bit stream. Left: little loss. Right: same bit pattern when there is a lot of loss, showing the collapse of the eye diagram, and increased jitter, indicated by the widening of the cross-over regions.

time delay, but offers no mechanism to account for the loss of the voltage as the signal propagates.

The losses need to be added to this model so that the received waveform can be accurately predicted. There are five ways energy can be lost to the receiver while the signal is propagating down a transmission line:

1. Radiative loss

2. Coupling to adjacent traces

3. Impedance mismatches

4. Conductor loss

5. Dielectric loss

While radiative loss is important when it comes to EMI, the amount of energy typically lost to radiation is very small compared to the other loss processes, and this loss mechanism will have no impact on the received signal.

Coupling to adjacent traces is important and can cause rise-time degradation. This effect can be very accurately modeled and the resulting waveforms on both the active and the quiet lines predicted. In tightly coupled transmission lines, the signal on one trace will be affected by the energy coupled to the adjacent one and must be included in a critical net simulation to accurately predict performance and the transmitted signal. This topic is covered in the next chapter.

TIP Impedance discontinuities can have a dramatic impact in distorting the transmitted signals. This will have the direct consequence of degrading the rise time of the received signal. Even a line with no loss will show rise-time degradation from impedance discontinuities. This is why it is so important to have accurate models for the transmission lines, board vias, and connectors, to accurately predict the signal quality from simulations. And this is why it is so important to minimize discontinuities in the design of high-speed interconnects.

If the rise time is degraded due to the removal of high-frequency components, where did the high-frequency components go? After all, capacitive or inductive discontinuities will not in themselves absorb energy. The high-frequency components are reflected back to the source and ultimately get absorbed and dissipated in any terminating resistors or the source impedance of the driver.

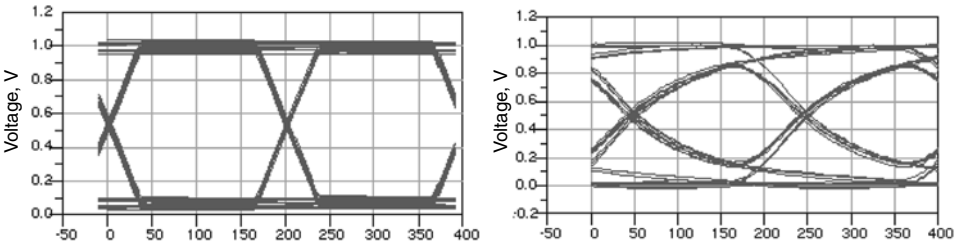


Figure 9-5 Eye diagrams of a 5-Gbps pseudorandom bit stream. Left: little loss. Right: same bit pattern still with no loss, but a 4-pF capacitive discontinuity from four through-hole vias.

Figure 9-5 is an example of the 5-Gbps signal from above, as transmitted through a short, ideal, lossless transmission line with four via pads in series, each of 1-pF load, contributing a total of 4-pF capacitive load. The resulting 50% point rise-time degradation expected is about $1/2 \times 50 \times 4$ pF = 100 psec, equal to half the bit period. Impedance discontinuities and their impact on rise-time degradation are covered in the previous chapter.

The last two loss mechanisms represent the primary causes of attenuation in transmission lines, not taken into account by other models. Conductor loss refers to energy lost in the conductors in both the signal and the return paths. This is ultimately due to the series resistance of the conductors. Dielectric loss refers to the energy lost in the dielectric due to a specific material property—the dissipation factor of the material.

TIP In general, with typical 8-mil-wide traces and 50-Ohm transmission lines in FR4, the dielectric losses are greater than the conductor losses at frequencies above about 1 GHz. For high-speed serial links with clock frequencies at or above 2.5 Gbps, the dielectric losses dominate. This is why the dissipation factor of the laminate materials is such an important property.

9.3 Sources of Loss: Conductor Resistance and Skin Depth

The series resistance a signal sees in propagating down the signal and return paths is related to the conductors' bulk resistivity and the cross section through which the current propagates. At DC, the current distribution in the signal conductor is uniform and the resistance is:

$$R = \rho \frac{Len}{w \times t} \tag{9-1}$$

where:

R = the resistance of the line, in Ohms

ρ = the bulk resistivity of the conductor, in Ohm-inches

Len = the length of the line, in inches

w = the line width, in inches

t = the thickness of the conductor, in inches

The DC current distribution in the return path, if it is a plane, is spread out through the cross section and the resistance is much smaller than the signal path's resistance and can be ignored.

For a typical 5-mil-wide trace, with 1.4-mil-thick copper (1-ounce copper), 1 inch in length, the resistance in the signal path at DC is about $R = 0.72 \times 10^{-6}$ Ohms-inches $\times$ 1 inch/(0.005×0.0014) = 0.1 Ohm.

The bulk resistivity of copper, and virtually all other metals, is absolutely constant with frequency until frequencies near 100 GHz. At first glance, we might expect the resistance of a trace to be constant with frequency. After all, this is how an ideal resistor behaves. However, as described in a previous chapter, the current will redistribute itself at higher frequencies due to skin-depth effects.

At a high frequency, the cross section through which current will be flowing in a copper conductor is in a thickness approximately equal to the skin depth, δ:

$$\delta = 2.1\sqrt{\frac{1}{f}} \tag{9-2}$$

where:

δ = the skin depth, in microns

f = the sine-wave frequency, in GHz

In copper, at 1 GHz, the current in the signal path of a microstrip, for example, is concentrated in a layer about 2.1 microns thick, on either side of the trace. At 10 MHz, or 0.01 GHz, it would be concentrated in a layer about 21 microns thick. This is a rough approximation. The actual current distribution in both the signal and return paths can always be calculated with a 2D field solver. An example of the current distribution in a microstrip and a stripline for sine waves at 10 MHz is shown in Figure 9-6.

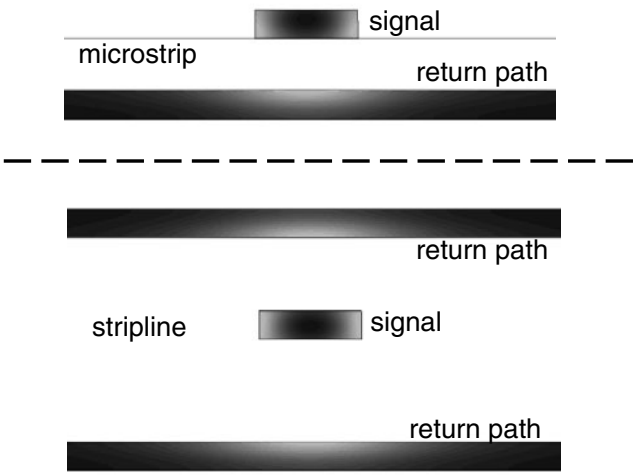


Figure 9-6 Current distribution in 1-ounce copper, for near 50-Ohm lines, at 10 MHz, showing onset of current redistribution due to skin-depth effects. Top: microstrip. Bottom: stripline. The lighter the color, the higher the current density. Simulated with Ansoft's 2D Extractor.

For 1-ounce copper, with a geometrical thickness of 35 microns, the skin depth is thinner for all frequencies above about 10 MHz. The current distribution is driven by the requirement of the current to find the path of least impedance, or at higher frequencies, the path of lowest loop inductance. This translates into two trends: The current in each conductor wants to spread out as far apart as possible to minimize the partial self-inductance of each conductor, and simultaneously, the oppositely directed current in each conductor will move as close together as possible to maximize the partial mutual inductance between the two currents.

TIP This means that for all important signal-frequency components, the current distribution in most PC-board interconnects is always going to be skin-depth limited, and the resistance will always be frequency dependent for frequency components above 10 MHz.

The resistance of a signal will depend on the actual cross section of the conductor transporting current. At higher frequencies, the current will be using a thinner section of the conductor and the resistance will increase with frequency. The frequency dependence of skin depth will cause a frequency dependence to resistance. It is important to note that the resistivity of copper, and most conductors, is

very constant with frequency. What is changing is the cross section through which the current flows. Above about 10 MHz, the resistance per length of the signal path will be frequency dependent.

In this regime of skin-depth-limited current, if the current were flowing in just the bottom half of the conductor, the resistance of a conductor would be approximated by:

$$R = \rho \frac{Len}{w \times \delta} \tag{9-3}$$

where:

R = the resistance of the line, in Ohms

ρ = the bulk resistivity of the conductor, in Ohm-inches

Len = the length of the line, in inches

w = the line width, in inches

δ = the skin depth of the conductor, in inches

As seen in the previous figure, the current flows in more than just the bottom half of the conductor, even for microstrip. There is substantial current also in the top of the conductor. These two regions are in parallel. To first order, the resistance of the signal path can be approximated as $0.5 \times R$, to take into account the two parallel paths in the signal path. The current distribution in the signal path of the microstrip and stripline are very similar.

Skin depth is driven by the need for the currents to take the path of the lowest impedance, which is dominated by the loop inductance at higher frequencies. This mechanism also drives the current in the return path to redistribute and change with frequency. At DC, the return current will be distributed all throughout the return plane. When in the skin-depth-limited regime, the current distribution in the return path will concentrate close to the signal path, near the surface, so as to minimize the loop inductance.

To first order, in a microstrip, the width of the current distribution in the return path is roughly three times the width of the signal path, as observed in the previous figure. This resistance in the return path is in series with the resistance of the signal path. At frequencies above about 10 MHz, we would expect the total series resistance of the transmission line to be $0.5R + 0.3R = 0.8R$. The total resistance of the signal path in a microstrip is expected to be about:

$$R = 0.8 \times \rho \frac{Len}{w \times \delta} \tag{9-4}$$

where:

R = the resistance of the line, in Ohms

ρ = the bulk resistivity of the conductor, in Ohm-inches

Len = the length of the line, in inches

w = the line width, in inches

δ = the skin depth of the conductor, in inches

0.8 = a factor due to the specific current distribution in the signal and return paths

Figure 9-7 compares this simple first-order model with the results from a 2D field solver, which calculates the precise current distribution at each frequency. The agreement at both low-frequency and skin-depth-limited frequencies is excellent for so simple a model. The total resistance per length of a stripline should be slightly lower. This is also compared in the figure.

So far, all we have pointed out is that the series resistance of the conductors in a transmission line will increase with increasing frequency. The question of how this frequency-dependent resistance affects loss is addressed later in this chapter.

9.4 Sources of Loss: The Dielectric

An ideal capacitor with air as the dielectric has an infinite DC resistance. Apply a DC voltage and no current will flow through it. However, if a sine wave of voltage, $V = V_0 \sin(wt)$, is applied, a cosine wave of current will flow through the capacitor, determined by the capacitance and the frequency. The current through an ideal capacitor is given by:

$$I = C_0 \frac{dV}{dt} = C_0 \omega V_0 \cos(\omega t) \tag{9-5}$$

where:

I = the current through the capacitor

C_0 = the capacitance of the capacitor

ω = the angular frequency, in radians/sec

V_0 = the amplitude of the voltage sine wave applied across the capacitor

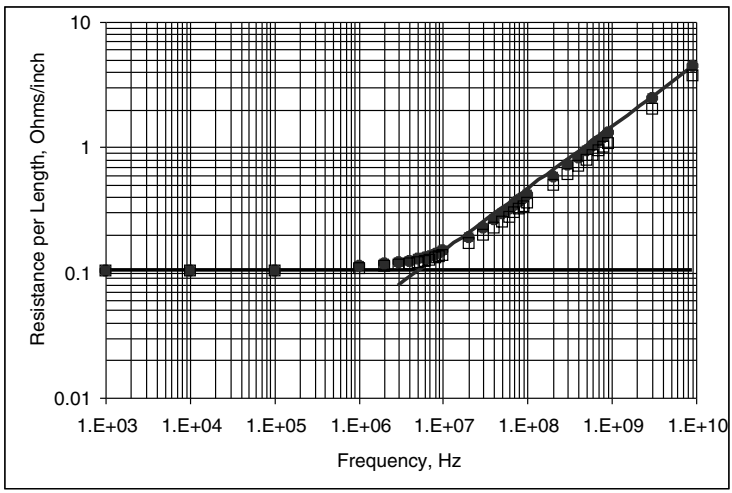


Figure 9-7 Calculated resistance vs. frequency of a 5-mil-wide, near 50-Ohm microstrip and stripline based on a field solver and the simple approximation of DC resistance and skin-depth-limited current. The circles are the microstrip and the squares are the stripline calculated with Ansoft's 2D Extractor field solver. The line is the simple first-order model of DC resistance and skin-depth-limited resistance.

There is no loss mechanism in an ideal capacitor. The current that flows through the capacitor is exactly 90 degrees out of phase with the voltage sine wave. If the ideal capacitor were filled with an insulator with a dielectric constant of ε_r, the capacitance would increase to $C = \varepsilon_r \times C_0$.

TIP The current through an ideal capacitor, when filled with an ideal lossless dielectric, will be increased by a factor equal to the dielectric constant. All of the current will be exactly 90 degrees out of phase with the voltage, no power will be dissipated in the material, and there is no dielectric loss.

However, real dielectric materials have some resistivity associated with them. When a real material is placed between the plates of a capacitor with a DC voltage across it, there will be some DC current that flows. This is usually referred to as *leakage current*. It can be modeled as an ideal resistor. The leakage resistance that is due to the material between two conductors making up a microstrip transmission line can be approximated with the parallel-plate approximation as:

$$R_{leakage} = \rho\frac{Len \times w}{h} = \frac{1}{\sigma}\frac{Len \times w}{h} \tag{9-6}$$

The amount of leakage current that flows through this resistance is:

$$I_{leakage} = \frac{V}{R_{leakage}} = V\frac{1}{\rho}\frac{h}{Len \times w} = V\sigma\frac{h}{Len \times w} \tag{9-7}$$

where:

$I_{leakage}$ = the leakage current through the dielectric

V = the applied DC voltage

$R_{leakage}$ = the leakage resistance associated with the dielectric

ρ = the bulk-leakage resistivity of the dielectric

σ = the bulk-leakage conductivity of the dielectric ($\rho = 1/\sigma$)

Len = the length of the transmission line

w = the line width of the signal path

h = the dielectric thickness between the signal and return paths

The leakage current, since it is going through a resistor, has current that is precisely in phase with the voltage. This current will dissipate power in the material and contribute to loss. The power dissipated by a resistor, with a constant voltage across it, is:

$$P = \frac{V^2}{R} \sim \sigma \tag{9-8}$$

where:

P = the power dissipated, in watts

V = the voltage across the resistor, in Volts

R = the resistance, in Ohms

σ = the conductivity of the material

For most dielectrics, the bulk resistivity is very high, typically 10^{12} Ohm-cm, so the leakage resistance through a typical 10-inch-long, 50-Ohm transmission

line, with w ~ 2 × h, is very high (on the order for 10^{11} Ohms). The resulting DC power loss through this resistance will be insignificant, or less than 1 nWatt.

However, for most materials, the bulk-leakage resistivity of the material is frequency dependent, getting smaller at higher frequencies. This is due to the origin of the leakage current. There are two ways of getting leakage current through a dielectric. The first is ionic motion. This is the dominant mechanism for DC currents. The reason the DC current is low in most insulators is that the density of mobile charge carriers (i.e., ions in the case of most insulators) is very small and their mobility is so low. This is compared with the high density of free electrons in a metal and their very high mobility.

The second mechanism for current flow in a dielectric is by the re-orientation of permanent electric dipoles in the material. When a voltage is applied across a capacitor, an electric field is generated. This field will cause some randomly oriented dipoles in the dielectric to align with the field. The motion of the negative end of the dipole toward the positive electrode and the positive end of the dipole to the negative electrode looks like a transient current through the material. This is illustrated in Figure 9-8.

Of course, the dipoles move only a very short distance and for only a very short time. If the voltage applied is a sine wave, then the dipoles will be sinusoidally rotated back and forth. This motion gives rise to an AC current. The higher the sine-wave frequency, the faster the charges will rotate back and forth and the higher the current. The higher the current, the lower the bulk resistivity at that frequency. The resistivity of the material is decreasing with increasing frequency.

The conductivity of the material is exactly the inverse of the resistivity, $\sigma = 1/\rho$. Just as the bulk resistivity relates to the ability of the material to resist the flow of current, the bulk conductivity relates to the ability of the material to conduct

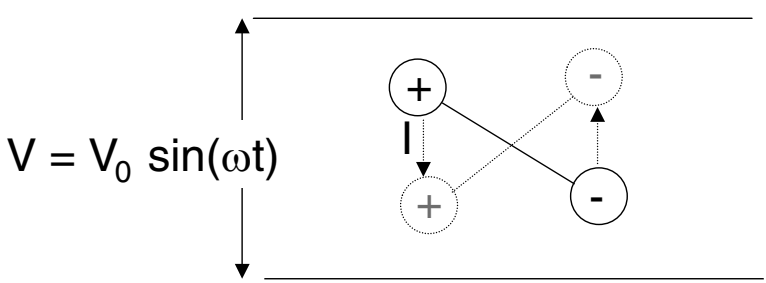


Figure 9-8 The re-orientation of the permanent dipoles in the dielectric, when the external field changes is an AC current through the dielectric.

current. A higher conductivity means the material conducts better. The bulk resistivity of a dielectric decreases with frequency, and the bulk conductivity increases with frequency. If the movements of the dipoles are able to follow the externally applied field and move the same distance for the same applied field, the current they create, and the bulk conductivity of the material, will increase linearly with frequency.

Most dielectrics behave in this way: Their conductivity is constant from DC until some frequency is reached and then it begins to increase and continues increasing, proportional to frequency. Figure 9-9 shows the bulk conductivity of FR4 material, with the transition frequency of about 10 Hz.

At frequencies above this transition frequency, where the motion of dipoles plays a significant role, there can be very high-leakage current through a real capacitor, as frequency is increased. This current will be in phase with the voltage and will look like a resistor. At higher frequencies, the leakage resistance will go down and the power dissipated will go up causing the dielectric to heat up.

TIP Effectively, the rotation of the dipoles translates electrical energy into mechanical energy. The friction of the motion of the dipoles with their neighbors and the rest of the polymer backbone causes the material to heat up, ever so slightly.

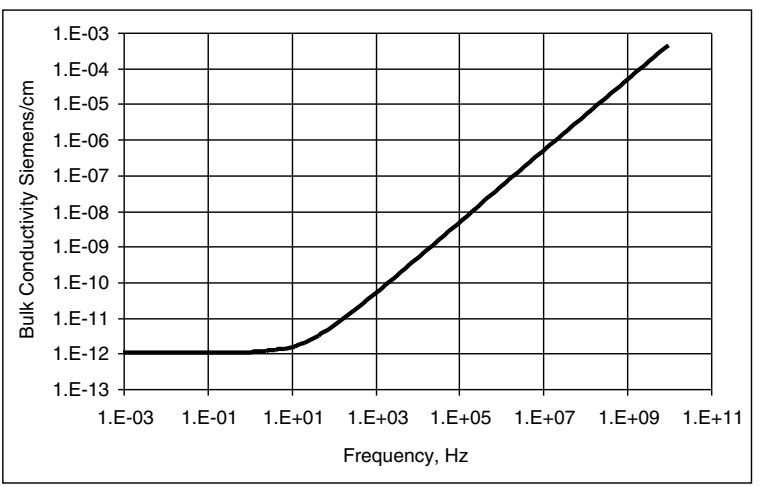


Figure 9-9 Simulated bulk conductivity of FR4 material from DC-leakage current and the AC-dipole motions.

The actual heat energy absorbed in normal cases is so small as to contribute to a negligible temperature rise. For the case of the previous 10-inch-long, 50-Ohm microstrip, even at 1 GHz, the leakage resistance of the dielectric is less than 1 kOhm, and the power dissipated is less than 10 mWatts. However, this is not always the case for dielectric loss. The notable exception is a microwave oven. The 2.45-GHz radiation is transferred from electrical energy in the microwaves to mechanical motion and heat by the rotation of water molecules, which strongly absorb the radiation.

In transmission lines, the dielectric's dipoles will suck energy out of the signal and cause less signal at the far end. It is not enough energy to heat the substrate very much, but it is enough to cause rise-time degradation. The higher the frequency, the higher the AC-leakage conductivity, and the higher the power dissipation in the dielectric.

9.5 Dissipation Factor

At low frequency, the leakage resistance of a dielectric material is constant and a bulk conductivity is used to describe the electrical properties of the material. This bulk conductivity is related to the density and mobility of ions in the material.

At high frequency, the conductivity increases with frequency due to the increasing motion of the dipoles. The more dipoles in the material that can rotate, the higher the bulk conductivity of the material. The farther the dipoles can move with an applied field, the higher the conductivity. To describe this new material property that measures the dipoles in a material, a new material electrical property must be introduced. This new material property, relating to dipole motion, is called the dissipation factor of the material:

$$\sigma = 2\pi f \times \varepsilon_0 \varepsilon_r \times \tan(\delta) \tag{9-9}$$

where:

σ = the bulk AC conductivity of the dielectric

f = the sine-wave frequency, in Hz

ε_0 = the permittivity of free space, 8.89×10^{-14} F/cm

ε_r = the relative dielectric constant, dimensionless

$\tan(\delta)$ = the dissipation of the material, dimensionless

The dissipation factor, usually written as the tangent of the loss angle, tan(δ), and also abbreviated sometimes as Df, is a measure of the number of dipoles in the material and how far each of them can rotate in the applied field:

$$\tan(\delta) \sim n \times p \times \theta_{max} \tag{9-10}$$

where:

tan(δ) = the dissipation factor, Df

n = the number density of dipoles in the dielectric

p = the dipole moment, a measure of the charge and separation of each dipole

θ_{max} = how far the dipoles rotate in the applied field

As the frequency increases, the dipoles move the same distance, but faster, so the current increases and the conductivity increases. This is taken into account by the frequency term. It is important to note that in the definition of the dissipation factor, as the tangent of an angle, δ, the angle, δ, is completely unrelated and separate from the skin-depth term, which, unfortunately, also uses the Greek letter δ to describe it. It is an unfortunate coincidence that both terms relate to two different, unrelated loss processes in transmission lines. Don't confuse them.

In real materials, the dipoles can't really move the same for all applied frequencies. This slight, second-order variation of the dipole motions with frequency is taken into account by the frequency variation of θ_{max}, which causes the dissipation factor to be slightly frequency dependent. The ability of the dipoles to move in the applied field is strongly dependent on how they are attached to the polymer-backbone chain and on the mechanical resonances of the other molecules nearby. At high enough frequency, the dipoles will not be able to respond as fast as at lower frequency and we would, therefore, expect the dissipation factor to decrease.

There is a whole field of study, called dielectric spectroscopy, that studies the frequency dependence of both the dissipation factor and the dielectric constant as a way of analyzing the mechanical properties of the polymer chains. Monitoring the dissipation factor and its frequency dependence is sometimes a measure of the degree of cure of a polymer. The more highly cross-linked the polymer, the tighter the dipoles are held and the lower the dissipation factor.

Material	ε	tan(δ)	Relative Cost
FR-4	4.0–4.7	0.02	1
DriClad(IBM)	4.1	0.011	1.2
GETek	3.6–4.2	0.013	1.4
BT	4.1	0.013	1.5
Polyimide/glass	4.3	0.014	2.5
CyanateEster	3.8	0.009	3.5
NelcoN6000SI	3.36	0.003	3.5
RogersRF35	3.5	0.0018	5

Figure 9-10 Dissipation factor and dielectric constants of some common interconnect dielectrics.

As a rough rule of thumb, the more tightly the dipoles are held by the polymer, the lower the dielectric constant and the lower the dissipation factor. The polymers with lowest dielectric constants (i.e., Teflon, silicone rubber, and polyethylene) all have very low dissipation factors. Figure 9-10 lists some commonly used interconnect dielectrics and their dissipation factors and dielectric constants.

In most interconnect materials, the dissipation factor of the material is mostly constant with frequency. Often this small variation with frequency can be neglected and a constant value used to accurately predict the loss behavior. However, there may be variations in the dissipation factor from lot to lot, from board to board, and even across the same board, due to variations in the processing of the laminate material. If the material absorbs moisture from humidity, the higher density of absorbed water molecules will increase the dissipation factor. In Polyimide or Kapton flex films, humidity can more than double the dissipation factor.

TIP To completely describe the electrical properties of a dielectric material, two material terms are required. The dielectric constant of the material describes how the material increases the capacitance and decreases the speed of light in the material. The dissipation factor describes the number of dipoles and their motion and is a measure of how much the conductivity increases proportional to the frequency. Both terms may be weakly frequency dependent and will vary from lot to lot and board to board.

Since both terms affect electrical performance, to accurately predict performance, it is important to know how these material properties vary with frequency and how they vary from board to board. If there is uncertainty in the material properties, there will be uncertainty in the performance of the circuits. Some techniques to measure these material properties at high frequency are presented later in this chapter.

9.6 The Real Meaning of Dissipation Factor

Describing a term like dissipation factor as tan(δ) is a bit confusing. Why describe it as the tangent of an angle? What is it the angle of? In practice, it is irrelevant what the angle refers to and it's really okay to not worry about it. In that case, skip this section. But, if you are curious and want to understand the intrinsic behavior of materials, read on.

TIP To use the term tan(δ) and to describe lossy lines, neither the origin of the term nor the angle to which it refers is important. It is simply a material property that relates to the number of freely moving dipoles in a material and how far they can move with frequency.

To investigate the underlying mechanism for the origin of loss and to look at how to design materials to control their dissipation factor, it is important to look deeper at the real origin of dissipation factor.

A dielectric material has two important electrical material properties. One property, the relative dielectric constant, which we introduced in an earlier chapter, describes how dipoles re-align in an electric field to increase capacitance. This term describes how much the material will increase the capacitance between two electrodes and the speed of light in the material. However, it does not tell us anything about the losses in the material. A second, new material-property term, the dissipation factor, is needed to describe how the dipoles slosh back and forth and contribute to a resistance with current that is in phase with the applied-voltage sine wave.

However, both of these terms relate to the number of dipoles, how large they are and how they are able to move. When viewed in the frequency domain with applied sine-wave voltages, one term relates to the motion of the dipoles that are out of phase with the applied field and contributes to increasing capacitance, while the other term relates to the motion of the dipoles that are in phase with the applied voltage and contribute to the losses.

The current through a real capacitor, with an applied sine-wave voltage, can be described by two components. One component of the current is exactly out of

phase with the voltage and contributes to the current we think of passing through an ideal lossless capacitor. The other current component is exactly in phase with the applied-voltage wave and looks like the current passing through an ideal resistor, contributing to loss.

To describe these two currents, the out-of-phase and the in-phase components, a formalism has been established based on complex numbers. It is intrinsically a frequency-domain concept because it involves the use of sine-wave voltages and currents. To take advantage of the complex number formalism, we change the dielectric constant and make it a complex number. Using the complex number formalism, the applied voltage can be written as:

$$V = V_0 \exp(i\omega t) \tag{9-11}$$

The current through a capacitor is related to the capacitance by:

$$I = C\frac{dV}{dt} \tag{9-12}$$

Using the complex number formalism, the current through the capacitor, in the frequency domain, can be written as:

$$I = C\frac{dV}{dt} = i\omega CV \tag{9-13}$$

This relationship describes the current through an ideal, lossless capacitor as i out of phase with the applied voltage. The i describes the phase between the current and the voltage as 90 degrees out of phase.

If the capacitor, with an empty capacitance of C_0, is filled with a material that has a dielectric constant, ε_r, then the current through the ideal capacitor is:

$$I = C\frac{dV}{dt} = i\omega\varepsilon_r C_0 V \tag{9-14}$$

Now, here's the first trick. To describe these two material properties (i.e., the dielectric constant that influences currents out of phase and the dissipation factor that influences currents in phase with the voltage), we change the definition of the dielectric constant. If the dielectric constant is a single, real number, the only current generated is i, out of phase with the voltage. If we make the dielectric constant a complex number, the real part of it will still relate to the current i out of phase, but the imaginary term will convert some of the voltage into a current in phase with the voltage and relate to the losses.

Here is the second trick. If we just describe the new complex dielectric constant as a real and imaginary term, for example, a + ib, when it is multiplied by the factor of i from the voltage, the i from the b term will convert the i term from the voltage into –1. This would make the real part of the current negative, or 180 degrees out of phase from the current. To bring the real part of the current exactly in phase with the voltage, we define the complex dielectric constant with the negative of its imaginary term. The complex dielectric constant is defined in the form:

$$\varepsilon_r = \varepsilon_r' - i\varepsilon_r'' \tag{9-15}$$

where:

ε_r = the complex dielectric constant

ε_r' = the real part of the complex dielectric constant

ε_r'' = the imaginary part of the complex dielectric constant

We introduce the minus sign in the definition of the complex dielectric constant so that in this formalism, the real part of the current comes out as positive and exactly in phase with the voltage. The real part of the new, complex dielectric constant is actually what we have been calling simply the dielectric constant.

TIP We see now that what we have traditionally called the dielectric constant is actually the real part of the complex dielectric constant. The imaginary part of the complex dielectric constant is the term that creates the current in phase with the voltage and relates to the losses.

Using this definition, the current through an ideal, lossy capacitor is given by:

$$I = i\omega\varepsilon_r C_0 V = i\omega(\varepsilon_r' - i\varepsilon_r'')C_0 V = i\omega\varepsilon_r' C_0 V + \omega\varepsilon_r'' C_0 V \quad (9\text{-}16)$$

where:

I = the current though an ideal, lossy capacitor, in the frequency domain

ω = the angular frequency, = $2\pi \times f$

C_0 = the empty space capacitance of the capacitor

V = the applied sine-wave voltage, $V = V_0 \exp(i\omega t)$

ε_r = the complex dielectric constant

ε_r' = the real part of the complex dielectric constant

ε_r'' = the imaginary part of the complex dielectric constant

By turning the dielectric constant into a complex number, the relationship between the in-phase and out-of-phase currents is compact. Using complex notation, we can generalize the current through a real capacitor. What makes it a little confusing is that the imaginary component of the current, which is 90 degrees out of phase with the voltage and contributes to the capacitive current with which we have been familiar, is actually related to the real part of the complex dielectric constant. The real part of the current, which is in phase with the voltage, behaves like a resistor, and contributes to loss, is actually related to the imaginary part of the dielectric constant.

As a complex number, the dielectric constant has a real part and an imaginary part. We can describe this number as a vector in the complex plane, as shown in Figure 9-11. The angle of the vector with the real axis is called the loss angle, δ. As previously noted, the use of the Greek letter δ to label the loss angle is an unfortunate coincidence with the selection of the same Greek letter to label the skin depth. These two terms are completely unrelated, as the loss angle relates to the dielectric material and the skin depth relates to the conductor properties.

The tangent of the loss angle is the ratio of the imaginary to the real component of the dielectric constant:

$$\tan(\delta) = \frac{\varepsilon_r''}{\varepsilon_r'} \quad (9\text{-}17)$$

and

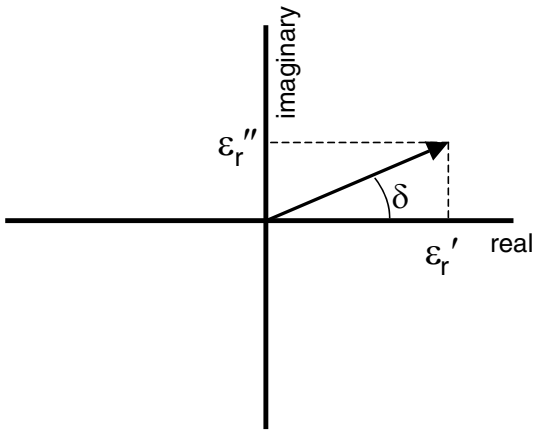


Figure 9-11 The complex dielectric constant plotted in the complex plane. The angle the dielectric-constant vector makes with the real axis is called the loss angle, δ.

$$\varepsilon_r'' = \varepsilon_r' \times \tan(\delta) \tag{9-18}$$

It is conventional that rather than using the imaginary part of the dielectric constant directly, the tangent of the loss angle, tan(δ), also referred to as Df, is used. In this way, the real part of the dielectric constant and tan(δ), and their possible frequency dependence, completely describe the important electrical properties of an insulating material. We also, by habit, omit the distinction that it is the real part of the dielectric constant and simply refer to it as the dielectric constant.

From the relationship above, we can relate the AC-leakage resistance of the transmission line to the imaginary part of the dielectric constant and the dissipation factor as:

$$R_{leakage} = \frac{V}{Real(I)} = \frac{V}{\omega\varepsilon_r''C_0V} = \frac{1}{\omega\varepsilon_r''C_0} = \frac{1}{\omega\varepsilon_r'\tan(\delta)C_0} = \frac{1}{\omega\tan(\delta)C} \tag{9-19}$$

In any geometrical configuration of conductors, the same geometrical features that affect the capacitance between the conductors also affect the resistance, but inversely. This is most easily seen for a parallel-plate configuration. The resistance and capacitance are given by:

$$C = \varepsilon_0 \varepsilon_r' \frac{A}{h} \tag{9-20}$$

$$\frac{A}{h} = \frac{C}{\varepsilon_0 \varepsilon_r'} \tag{9-21}$$

$$R = \frac{1}{\sigma}\frac{h}{A} = \frac{1}{\sigma}\frac{\varepsilon_0 \varepsilon_r'}{C} \tag{9-22}$$

Combining these two forms for the resistance between the conductors results in the connection between the bulk-AC conductivity of the material and the dissipation factor:

$$\sigma = \varepsilon_0 \varepsilon_r' \omega \tan(\delta) \tag{9-23}$$

where:

σ = the bulk-AC conductivity of the dielectric material

ε_0 = the permittivity of free space = 8.89×10^{-14} F/cm

ε_r' = the real part of the dielectric constant

ε_r'' = the imaginary part of the dielectric constant

$\tan(\delta)$ = the dissipation factor of the dielectric

δ = the loss angle of the dielectric

R = the AC-leakage resistance between the conductors

C = the capacitance between the conductors

h = the dielectric thickness between the conductors

A = the area of the conductors

ω = angular frequency = $2 \times \pi \times f$, with f = the sine-wave frequency

TIP Even though the dissipation factor itself is only weakly frequency dependent, once again, we see that the bulk-AC conductivity of a dielectric will increase linearly with frequency due to the ω term. Likewise, since the power dissipated by the leakage resistance is proportional to the bulk-AC conductivity, the power dissipation will also increase linearly with frequency. This is the fundamental origin of the main problem lossy lines create for signal integrity.

9.7 Modeling Lossy Transmission Lines

The two loss processes for attenuating signals in a transmission line are the series resistance through the signal- and return-path conductors and the shunt resistance through the lossy dielectric material. Both of these resistors have resistances that are frequency dependent.

It is important to note that an ideal resistor has a resistance that is constant with frequency. We have shown that in an ideal lossy transmission line, the two resistances used to describe the losses are more complicated than simple ideal resistors. The series resistance increases with the square root of frequency due to skin-depth effects. The shunt resistance decreases with frequency due to the dissipation factor of the material and the rotation of dipole molecules.

In a previous chapter, we introduced a new, ideal circuit element, the ideal distributed transmission line. It was described by a characteristic impedance and a time delay. This model distributes the properties of the transmission line throughout its length. An ideal lossy distributed transmission line model will add to this lossless model two loss processes: series resistance increasing with the square root of frequency and shunt resistance decreasing inversely with frequency. This is the basis of the new ideal lossy transmission line that is implemented in many simulators. The two factors that are specified, in addition to the characteristic impedance and time delay, are the dissipation factor and resistance per length, R_L, of the form:

$$R_L = R_{DC} + R_{AC}\sqrt{f} \tag{9-24}$$

where:

R_L = the resistance per length of the conductors

R_{DC} = the resistance per length at DC

R_{AC} = the coefficient of the resistance per length that is proportional to $f^{0.5}$

To gain insight into the behavior of ideal lossy lines, we can start with the approximation of a transmission line as an n-section LC circuit and add the loss terms and evaluate the behavior of the circuit model.

In a previous chapter, we showed that an ideal, distributed lossless transmission line can be approximated with an equivalent circuit model consisting of lumped-circuit-model sections with a shunt C and series L. This model is sometimes called the first order model of a transmission line, an n-section lumped-circuit model, or a lossless model of a transmission line. A piece of it is shown in Figure 9-12.

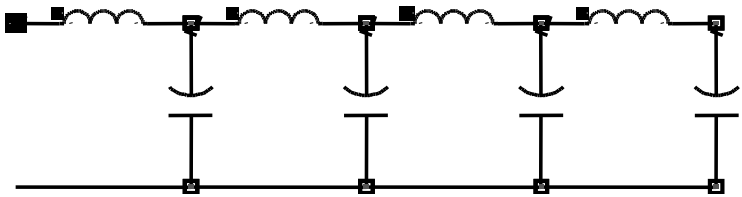


Figure 9-12 Four sections of an n-section LC model as an approximation of an ideal lossless, distributed transmission line.

This model is an approximation. However, it can be a very accurate approximation, to very high bandwidth, by using enough sections. We showed that the minimum number of sections required to achieve a bandwidth, BW, for a time delay, TD, is given by:

$$n = 10 \times BW \times TD \tag{9-25}$$

where:

n = the number of sections for an accurate LC model

BW = the bandwidth of the model, in GHz

TD = the time delay of the transmission line being approximated, in nsec

For example, if the bandwidth required for the model is 2 GHz, and the time delay of the line is 1 nsec, which is about 6 inches long physically, the minimum number of sections required for an accurate model is $n \sim 10 \times 2 \times 1 = 20$.

However, a significant limitation to this ideal lossless model is that it would still always be a lossless model. Using this first-order equivalent circuit model as a starting place, we can modify it to account for the losses. In each section, we can add the effect of the series resistance and the shunt resistance. A small section of the n-section lumped-circuit approximation for an ideal lossy transmission would have four terms that describe it:

C = the capacitance

L = the loop self-inductance

R_{series} = the series resistance of the conductors

R_{shunt} = the dielectric-loss shunt resistance

If we double the length of the transmission line, the total C doubles, the total L doubles, and the total R_{series} doubles. However, the total R_{shunt} is cut in half. If we double the length of the line, there is more area through which the AC-leakage current can flow, so the shunt resistance decreases.

For this reason, it is conventional to use the conductance of the dielectric-leakage resistance, rather than the resistance, to describe it. The conductance, denoted by the letter G, is defined as G = 1/R. Based on the resistance, the conductance is:

$$R_{leakage} = \frac{1}{\omega \tan(\delta) C} \tag{9-26}$$

$$G = \frac{1}{R_{leakage}} = \omega \tan(\delta) C \tag{9-27}$$

If the length of the transmission line doubles, the shunt resistance is cut in half but the conductance doubles. When we use G to describe the losses, we still model the losses as a resistor whose resistance decreases with frequency; we just use the G parameter to describe it. Using the conductance instead of the shunt resistance, the four terms that describe a lossy transmission line all scale with length. It is conventional to refer to their values per unit length. These four terms are referred to as the line parameters of a transmission line:

R_L = series resistance per length of the conductors
C_L = capacitance per length
L_L= series loop inductance per length
G_L= shunt conductance per length from the dielectric

We will use this ideal, second-order n-section lumped circuit model to approximate an ideal lossy transmission line, which in turn is an approximation to a real transmission line. An example of the equivalent n-section RLGC transmission-line model is shown in Figure 9-13.

The number of sections we use will depend on the length of the line and the bandwidth of the model. The minimum number of sections required is still roughly $10 \times BW \times TD$.

This is an equivalent circuit model. We can apply circuit theory to this circuit and predict the electrical properties. The mathematics is complicated, because

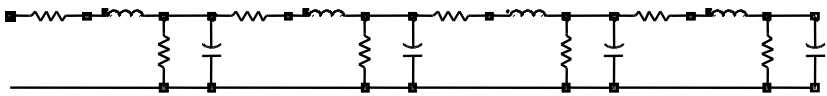


Figure 9-13 Four sections of an n-section RLGC model for an ideal lossy transmission line; an approximation of an ideal distributed lossy transmission line.

coupled, second-order differential equations are involved, but also because the equations are complex and the resistances have values that vary with frequency. The easiest domain in which to solve the equations is the frequency domain. We assume that signals are sine waves of voltage and, from the impedance, the sine waves of current can be calculated. The results only will be discussed.

In a lossless line, the resistance and conductance are equal to zero. This lossless circuit model would predict an interconnect that propagates a signal undistorted. The instantaneous impedance the signal sees at each step along the way is equal to the characteristic impedance of the line:

$$Z_0 = \sqrt{\frac{L_L}{C_L}} \tag{9-28}$$

The speed of a signal is given by:

$$v = \frac{1}{\sqrt{C_L \times L_L}} \tag{9-29}$$

where:

Z_0 = the characteristic impedance

v = the speed of the signal

C_L = the capacitance per length

L_L = the inductance per length

In this model, the ideal L, C, and Z_0 terms as well as the time delay are all constant with frequency. These are the only terms that define the ideal, lossless transmission line. A signal entering at one end of the line will exit the other end, with no change in amplitude. The only impact on a sine wave other than a reflection due to a possible impedance change would be a phase shift in transmission.

However, when the R and G terms are added to the model, the behavior of the ideal lossy transmission line is slightly different from the ideal lossless transmission line. When the differential equations are solved, the results are rather complicated. The solution, in the frequency domain, makes no assumption on how C_L, L_L, R_L, or G_L vary with frequency. At each frequency, they may change, or they may be constant.

The final solution ends up with three important features:

1. The characteristic impedance is frequency dependent and complex.

2. The velocity of a sine-wave signal is frequency dependent.

3. A new term is introduced that describes the attenuation of the sine-wave amplitude as it propagates down the line. This attenuation is also frequency dependent.

When the dust clears, the exact values for the characteristic impedance, velocity, and attenuation per length are given by:

$$Z_0 = \sqrt{\frac{R_L + i\omega L_L}{G_L + i\omega C_L}} \tag{9-30}$$

$$v = \frac{\omega}{\sqrt{\frac{1}{2}\left[\sqrt{(R_L^2 + \omega^2 L_L^2)(G_L^2 + \omega^2 C_L^2)} + \omega^2 L_L C_L - R_L G_L\right]}} \tag{9-31}$$

$$\alpha_n = \sqrt{\frac{1}{2}\left[\sqrt{(R_L^2 + \omega^2 L_L^2)(G_L^2 + \omega^2 C_L^2)} - \omega^2 L_L C_L + R_L G_L\right]} \tag{9-32}$$

where:

Z_0 = the characteristic impedance

v = the speed of a signal

α_n = the attenuation per length of the amplitude, in nepers/length

ω = the angular frequency of the sine wave, in radians/sec

R_L = series resistance per length of the conductors

C_L = capacitance per length

L_L= series loop inductance per length

G_L= shunt conductance per length from the dielectric

These are formidable, and though they can be implemented in a spreadsheet, are difficult to use to gain useful engineering insight. To simplify the algebra, one of the approximations that is commonly made is that the lines are lossy, but not too lossy. This is called the low-loss approximation. The approximation is that the series resistance, R_L, is $R_L << \omega L_L$ and the shunt conductance, G_L, is $G_L << \omega C_L$.

This approximation assumes the impedance associated with the series resistance of the conductors is small compared with the series impedance associated with the loop inductance. Likewise, the shunt current through the leakage resistance of the dielectric is small compared to the shunt current through the capacitance between the signal and return paths.

For 1-ounce copper traces, the series resistance above about 10 MHz will increase with the square root of frequency and ωL_L will increase linearly with frequency. At some frequency, this approximation will be good and it will get better at higher and higher frequencies.

The DC resistance per length of a 1-ounce copper trace is:

$$R_L = \frac{0.5}{w} \tag{9-33}$$

where:

R_L = the resistance per length, in Ohms/inch

w = line width, in mils

Above 10 MHz, the current will flow in a thinner cross section, and rather than the geometrical thickness of 34 microns for 1-ounce copper, the skin depth will determine the thickness of the current distribution. The skin depth for copper is:

$$\delta = 66\sqrt{\frac{1}{f}} \tag{9-34}$$

where:

δ = the skin depth, in microns

f = the sine-wave-frequency component, in MHz

The AC resistance per length of a 1-ounce copper trace, above about 10 MHz, is about:

$$R_L = \frac{0.5t}{w\delta} = \frac{0.5 \times 34}{w \times 66}\sqrt{f} = \frac{0.25}{w}\sqrt{\frac{\omega}{2\pi \times 10^6}} = \frac{1 \times 10^{-4}}{w}\sqrt{\omega} \tag{9-35}$$

where:

R_L = the resistance per length, in Ohms/inch

δ = the skin depth, in microns

t = the geometrical thickness, in microns

w = the line width, in mils

f = the sine-wave-frequency component, in MHz

ω = the sine-wave-frequency component, in radians/sec

The inductance per length is roughly 9 nH/inch for a 50-Ohm line. The low-loss regime happens when $\omega L_L >> R_L$or:

$$\omega \times 9 \times 10^{-9} \gg \frac{1 \times 10^{-4}}{w}\sqrt{\omega} \tag{9-36}$$

$$\omega \gg \left(\frac{1}{w}\right)^2\left(\frac{1 \times 10^{-4}}{9 \times 10^{-9}}\right)^2 = \frac{1 \times 10^8}{w^2} \tag{9-37}$$

$$f = \frac{\omega}{2\pi} \gg \frac{1 \times 10^8}{2\pi w^2} \sim \frac{2 \times 10^7}{w^2} \tag{9-38}$$

where:

R_L = the resistance per length, in Ohms/inch

ω = the sine-wave-frequency component, in radians/sec for the low-loss regime

f = the sine-wave-frequency component, in Hz for the low-loss regime

w = the line width, in mils

TIP This is a startling result. The conclusion is that for line widths wider than 3 mils, the low-loss regime is for sine-wave-frequency components above 2 MHz. In this regime, the impedance of the series resistance is much less than the reactance of the series inductance. For lines wider than 3 mils, the low-loss regime begins at even lower frequencies. The very lossy regime is actually in the low frequency, below the frequency where skin depth plays a role.

The conductance will roughly increase linearly with frequency and the capacitance will be roughly constant with frequency. The low-loss regime happens when $G_L << \omega C_L$. This is when $\tan(\delta) << 1$. For virtually all interconnect materials, the dissipation factor is less than 0.02, and the interconnect is always in the low-loss regime.

TIP The low-loss regime for circuit-board interconnects with 3-mil-wide traces and wider is for frequencies above 2 MHz, which is where most important frequency components are.

The conclusion is that the low-loss approximation is a very good approximation for all important frequency ranges of interest in high-speed digital applications.

9.8 Characteristic Impedance of a Lossy Transmission Line

In an ideal lossy transmission line, the characteristic impedance becomes frequency dependent and is complex. The characteristic impedance is given by:

$$Z_0 = \sqrt{\frac{R_L + \omega L_L}{G_L + \omega C_L}} \tag{9-39}$$

With a bit of algebra, after the dust clears, the real part and imaginary part of the characteristic impedance are given by:

$$\mathrm{Re}(Z_0) = \frac{1}{\sqrt{G_L^2 + \omega^2 C_L{}^2}} \sqrt{\frac{1}{2}\left[\sqrt{(R_L^2 + \omega^2 L_L{}^2)(G_L^2 + \omega^2 C_L{}^2)} + \omega^2 L_L C_L + R_L G_L\right]} \tag{9-40}$$

$$\mathrm{Imag}(Z_0) = \frac{1}{\sqrt{G_L^2 + \omega^2 C_L{}^2}} \sqrt{\frac{1}{2}\left[\sqrt{(R_L^2 + \omega^2 L_L{}^2)(G_L^2 + \omega^2 C_L{}^2)} - \omega^2 L_L C_L - R_L G_L\right]} \tag{9-41}$$

where:

$Re(Z_0)$ = the real part of the characteristic impedance

$Imag(Z_0)$ = the imaginary part of the characteristic impedance

R_L = series resistance per length of the conductors

C_L = capacitance per length

L_L= series loop inductance per length

G_L= shunt conductance per length from the dielectric

ω = the angular frequency

In the low-loss regime, the characteristic impedance reduces to:

$$Re(Z_0) = \sqrt{\frac{L_L}{C_L}} \tag{9-42}$$

$$Imag(Z_0) = 0 \tag{9-43}$$

The low-loss approximation for the characteristic impedance is exactly the same as for the lossless characteristic impedance. The terms that affect the characteristic impedance vary as R^2 and G^2 compared to ω^2L^2 or ω^2C^2. Our assumption of the low-loss regime is seen to introduce less than 1% errors, when we are at frequencies above 10 times the boundary of roughly 2 MHz for a 3-mil-wide line.

We can use the magnitude of the characteristic impedance as a rough measure of the impact from the losses. The magnitude is given by:

$$Mag(Z_0) = \sqrt{Re(Z_0)^2 + Imag(Z_0)^2} \tag{9-44}$$

Figure 9-14 plots the magnitude of the complex characteristic impedance of a 3-mil-wide, 50-Ohm microstrip in FR4 using the exact relationship above. This includes the conductor loss and the dielectric loss. We see that for frequencies above about 10 MHz, the complex characteristic impedance is very close to the lossless value. This transition frequency will move toward lower frequency for wider and less lossy lines.

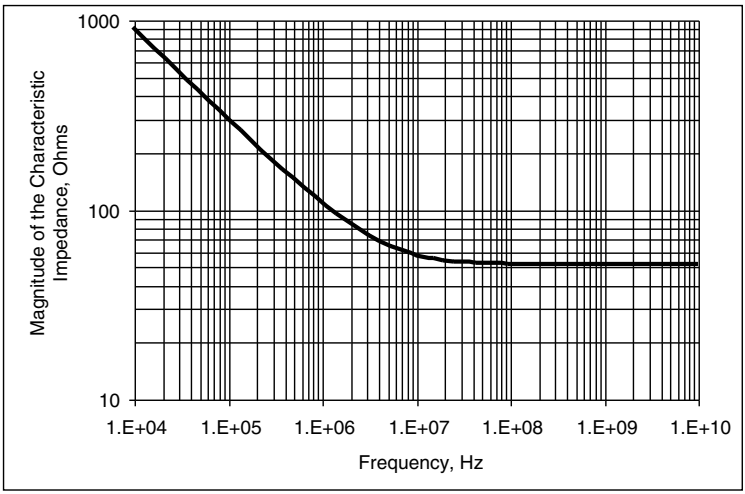


Figure 9-14 The magnitude of the complex characteristic impedance of a 50-Ohm lossy microstrip in FR4 shows that above about 10 MHz, the lossy characteristic impedance is very close to the lossless impedance. The low-loss regime is above 10 MHz.

TIP In the low-loss regime, there is no impact on the characteristic impedance from the losses.

As described earlier, there may be some frequency dependence to the inductance from skin-depth effects. Above about 100 MHz, the skin depth is much thinner than the geometrical thickness and the inductance is constant with frequency above this point. There may be some frequency dependence of the capacitance due to the real part of the dielectric constant changing with frequency. These terms may contribute to a slight frequency dependence of the characteristic impedance. In real interconnects, the impact from these effects is usually not noticeable.

9.9 Signal Velocity in a Lossy Transmission Line

A consequence of the solution to the circuit model of a lossy transmission line is that the velocity of a sine wave is complicated. The velocity is given by:

$$v = \frac{\omega}{\sqrt{\frac{1}{2}\left[\sqrt{(R_L^2 + \omega^2 L_L^2)(G_L^2 + \omega^2 C_L^2)} + \omega^2 L_L C_L - R_L G_L\right]}} \tag{9-45}$$

In the low-loss regime, where the impedance of the resistance is much less than the reactance of the inductance and the dissipation factor is << 0.1, the velocity can be approximated by:

$$v = \frac{1}{\sqrt{L_L C_L}} \tag{9-46}$$

This is exactly the same result as for a lossless line.

TIP The conclusion is that in the low-loss regime, the velocity of a signal is not affected by the losses.

Using the exact form for the velocity, we can evaluate just how constant the velocity is and at what point the speed will vary with frequency. This effect of a velocity that is frequency dependent is called *dispersion* and in this case the dispersion is caused by loss. Figure 9-15 shows the frequency dependence to the speed of a signal for a worst case of a 3-mil-wide line in an FR4 50-Ohm microstrip, including both the dielectric and conductor losses.

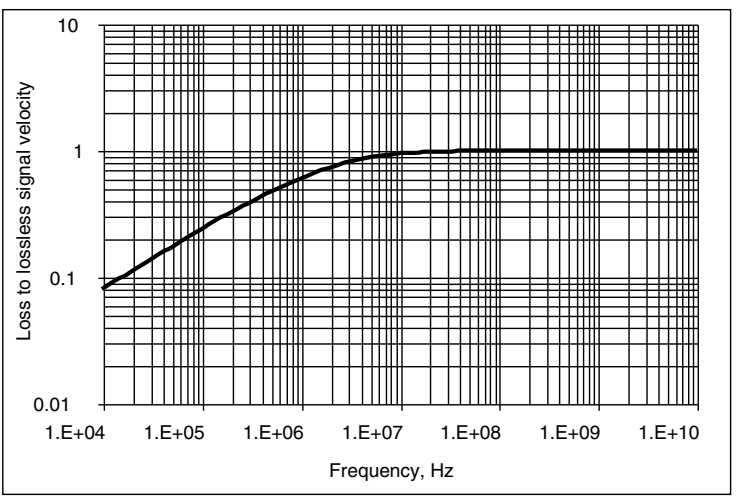


Figure 9-15 Dispersion due to losses for a 3-mil-wide, 50-Ohm line in FR4. Plotted is the ratio of the lossy velocity to the lossless velocity.

The impact from the losses is to slow down the lower frequencies more than the higher frequencies. At lower frequencies, the series resistive impedance dominates over the series reactive impedance from the loop inductance. As well, the line looks more lossy and the signal speed is reduced. When speed varies with frequency, we call this dispersion. It arises from two mechanisms: frequency-dependent dielectric constants and losses.

Dispersion will cause the higher-frequency components to travel faster than the lower-frequency components. In the time domain, the fast edge will arrive first, followed by a slowly rising tail, effectively increasing the rise time. However, if the losses are ever large enough to have a noticeable impact on rise-time degradation, the impact from the attenuation will usually be far larger than the impact from dispersion.

TIP For the worst-case line, 3 mils wide in FR4, the low-loss regime is above about 10 MHz. In this regime, the speed is independent of frequency and the dispersion from the losses is negligible.

9.10 Attenuation and the dB

The dominant effect on a signal from the losses in a line is a decrease in amplitude as the signal propagates down the length of the line. If a sine-wave-voltage signal with amplitude, V_{in}, is introduced in a transmission line, its amplitude will drop as it moves down the length. Figure 9-16 shows what the sine wave might look like at different positions, if we could freeze time and look at the sine wave as it exists on the line. This is for the case of a 1-GHz sine wave on a 40-inch-long, 50-Ohm microstrip in FR4 with a 10-mil-wide trace.

The amplitude drops off not linearly, but exponentially, with distance. This can be described with an exponent either to base e or to base 10. Using base e, the output signal is given by:

$$V(d) = V_{in}\exp(-A_n) = V_{in}\exp(-d \times \alpha_n) \tag{9-47}$$

where:

V(d) = the voltage on the line at position d

d = the position along the line, in inches

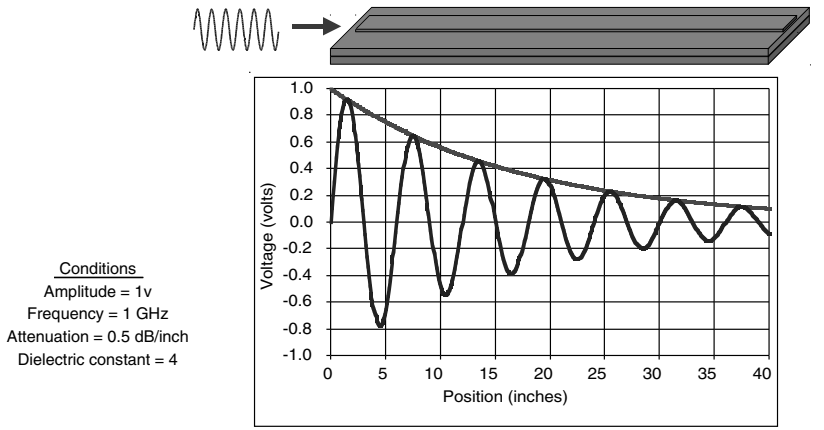


Figure 9-16 Amplitude of a 1-GHz sine-wave signal as it would appear on a 10-mil-wide, 50-Ohm transmission line in FR4.

V_{in} = the amplitude of the input voltage wave

A_n = the total attenuation, in nepers

α_n = the attenuation per length, in nepers/inch

When using the base e, the units of attenuation are dimensionless, but are still labeled as nepers, after John Napier, the Scotsman who is credited with the introduction of the base e exponent published in 1614. One confusing aspect of Napiers is the spelling. Napiers, napers, and nepers all refer to the same unit and are commonly used, alternative spellings of his name. Though dimensionless, we use the label to remind us it is the attenuation using base e.

For example, if the attenuation were 1 napier, the final amplitude would be exp(–1) = 37% of the input amplitude. If the attenuation were 2 napiers, the output amplitude would be exp(–2) = 13% of the input voltage.

Likewise, given the input and output amplitude, the attenuation can be found from:

$$A_n = -\ln\left(\frac{V(d)}{V_{in}}\right) \tag{9-48}$$

There is some ambiguity about the sign. In all passive interconnects, there will never be any gain. The output voltage is always smaller than the input volt-

age. An exponent of 0 will have exactly the same output amplitude as input amplitude. The only way to get a reduced amplitude is by having a negative exponent. Should the minus sign be placed explicitly in the exponent or be part of the attenuation? Both ways are conventionally done. The attenuation is sometimes referred to as –2 napiers or 2 napiers. Since it is always referred to as an attenuation, there is no ambiguity.

It is more common to describe attenuation using a base 10 than a base e. The output amplitude is of the form:

$$V(d) = V_{in}10^{-\frac{A_{dB}}{20}} = V_{in}10^{\left(-d \times \frac{\alpha_{dB}}{20}\right)} \tag{9-49}$$

where:

$V(d)$ = the voltage on the line at position d

d = the position along the line, in inches

V_{in} = the amplitude of the input voltage wave

A_{dB} = the total attenuation in dB

α_{dB} = the attenuation per length, in dB/inch

20 = the factor to convert dB into amplitude, described below

TIP The units used to describe attenuation are in decibels, or dB. These units are used throughout engineering and wherever they appear, they leave confusion in their wake. Understanding the origin of this unit will help remove the confusion.

The decibel was created over 100 years ago by Alexander Graham Bell. He started his career as a physician studying and treating children with hearing problems. To quantify the degree of hearing loss, he developed a standardized set of sound intensities and quantified individuals' ability to hear them. He found that the sensation of loudness depended not on the power intensity of the sound, but on the log of the power intensity. He developed a scale of loudness that started with the quietest whisper that could be heard as a unit of 0 and the sound that produced pain as 10.

All other sounds were distributed on this scale, in a log ratio of the actual measured power level. If the loudness increased from 1 to 2, for example, the loudness was perceived to have doubled. However, the actual power level in the

sound was measured as having increased by $10^2/10^1 = 10$. What Bell established is that the quality of perceived loudness change depends not on the power-level change, but on a unit that is proportional to the log of the power-level change.

The units of the Bell scale of loudness were called Bells, with 0 Bells as the quietest whisper that can be heard. The actual power density at the ear has since been quantified for each sound level. The quietest sound level that can be perceived at our peak sensitivity at about 2 kHz is the start of the scale, 0 Bell. It corresponds to a power density of 10^{-12} watts/m^2. For the loudest sound just at the pain threshold, 10 Bells, the power level is 10^{-2} watts/m^2.

As the Bell scale of perceived loudness gained wide acceptance, the last L was dropped and it become the Bel scale of perceived loudness. Over time, it was found that the range of the scale, 0 Bels to 10 Bels, was too small given the incredible range of perceived loudness. Instead of loudness measured in Bels, the scale was changed to deciBels, where the preface "deci" means 1/10. The perceived loudness scale now starts with 0 deciBels at the low end for the whisper and has 100 deciBels at the onset of pain. The deciBel is usually abbreviated as dB.

TIP Over the years, the deciBel scale has been adopted for other applications in addition to loudness, but in every case, the deciBel has retained its definition as the log of the ratio of two powers. The most important property of the deciBel scale is that it always refers to the log of the ratio of two powers.

In virtually all engineering applications, the log of the ratio of two powers, P_1 and P_0, is also measured in Bels: the number of Bels = $\log(P_1/P_0)$. Since 1 Bel = 10 deciBels, the ratio, in dB, is:

$$\text{ratio(dB)} = 10 \times \log\frac{P_1}{P_0} \tag{9-50}$$

For example, if the power increases by a factor of 1000, the increase, in Bels, is $\log(1000) = 3$ Bels. In dB, this is 10×3 Bels = 30 dB. A decrease in the power level, where the output power is only 1% the input power, is $\log(10^{-2}) = -2$ Bels, or 10×-2 Bels = –20 dB.

When the power level changes by any arbitrary factor, the change can be described in dB, but requires a calculator to calculate the log value. If the power doubles, the change in dB is $10 \times \log(2) = 10 \times 0.3 = 3$ dB. We often use the

expression "a 3 dB change," which refers to a doubling in the power level. If the change were to drop by 50%, the change in dB would be $10 \times \log(0.5) = -3$ dB.

The ratio of the actual power levels can be extracted from the ratio in dB by:

$$\text{ratio} = \frac{P_1}{P_0} = 10^{\frac{\text{ratio(dB)}}{10}} \tag{9-51}$$

The first step is to convert the dB into Bels. This is the exponent to base 10. For example, if the ratio in dB is 60, the power level ratio is $10^{60/10} = 10^6 = 1{,}000{,}000$. If the dB value is –3 dB, the power level ratio is $10^{-3/10} = 10^{-0.3} = 0.5$ or 50%.

There are two rules about the dB scale that are important to always keep in mind:

1. The dB scale *always* refers to the log of the ratio of two powers or energies.
2. When measured in dB, the exponent to base 10 of the ratio of two powers is just the dB/10.

The distinction about power is important whenever the ratio of two other quantities is measured. When the ratio, r, of two voltages is measured, for example, V_0 and V_1, the units are dimensionless: $r = \log(V_1/V_0)$. But we cannot use dB to measure the ratio, since the dB unit refers to the ratio of two powers or energies. A voltage is not an energy, it is an amplitude.

We can refer to the ratio of the powers associated with the two voltage levels, or $r_{dB} = 10 \times \log(P_1/P_0)$. How are the power levels related to the voltage levels? The energy in a voltage wave is proportional to the square of the voltage amplitude, or $P \sim V^2$.

The ratio of the powers, in dB, is:

$$r_{dB} = 10 \times \log\left(\frac{P_1}{P_0}\right) = 10 \times \log\left(\frac{V_1^2}{V_0^2}\right) = 10 \times 2\log\left(\frac{V_1}{V_0}\right) = 20\log\left(\frac{V_1}{V_0}\right) \tag{9-52}$$

TIP Whenever the ratio of two amplitudes is measured in units of dB, it is calculated by taking the log of the ratio of their associated powers. This is equivalent to multiplying the log of the ratio of the voltages by 20.

For example, a change in the voltage from 1 v to 10 v, measured in dB, is 20 × log(10/1) = 20 dB. The voltage increased by only a factor of 10. However, the underlying powers corresponding to the 1 v and 10 v increased by a factor of 100. This is reflected by the 20-dB change in the power level. dB is *always* a measure of the change in power. When the amplitude decreases to half its original value, or is reduced by 50%, the ratio of the final to the initial value, in dB, is 20 × log(0.5) = 20 × –0.3 = –6 dB. If the voltage is reduced by 50%, the power in the signal must have decreased by $(50\%)^2$ or 25%. If the ratio of two powers is 25% in dB, this is 10 × log(0.25) = 10 × –0.6 = –6 dB.

TIP When referring to an energy or power, the factor of 10 is used in calculating the dB value. When referring to an amplitude, a factor of 20 is used. An amplitude is a quantity such as voltage, a current, or an impedance.

From the ratio in dB, the actual ratio of the voltages can be calculated as:

$$\text{ratio} = \frac{V_1}{V_0} = 10^{\frac{\text{ratio}_{dB}}{20}} \tag{9-53}$$

For example, if the ratio in dB is 20 dB, the ratio of the amplitudes is $10^{20/20} = 10^1 = 10$. If the ratio in dB is –40 dB, the ratio of the voltages is $10^{-40/20} = 10^{-2} = 0.01$. If the ratio in dB is negative, this means the final value is always less than the initial value. Figure 9-17 lists a few examples of the ratio of the voltages, their associated powers, and the ratio in dB.

9.11 Attenuation in Lossy Lines

When a sine-wave signal propagates down a transmission line, the amplitude of the voltage decreases exponentially. The total attenuation, measured in dB, increases linearly with length. In FR4, a typical attenuation of a 1-GHz signal might be 0.1 dB/inch. In propagating 1 inch, the attenuation is 0.1 dB and the signal amplitude has dropped to $V_{out}/V_{in} = 10^{-0.1/20} = 99\%$. In propagating 10 inches, the attenuation is 0.1 dB/inch × 10 inches = 1 dB and the amplitude has dropped to $V_{out}/V_{in} = 10^{-1/20} = 89\%$.

The attenuation is a new term that describes a special property of lossy transmission lines. It is a direct result of the solution of the second-order, lossy RLCG

Voltage Ratio	Power Ratio	dB
100	10,000	40
10	100	20
2	4	6
1.4	2	3
1	1	0
0.7	0.5	-3
0.5	0.25	-6
0.1	0.01	-20
0.01	0.0001	-40

Figure 9-17 Ratio of the voltages, their corresponding powers, and their ratio in dB.

circuit model. The attenuation per length, usually denoted by alpha, α_n, in nepers/length, is given by:

$$\alpha_n = \sqrt{\frac{1}{2}[\sqrt{(R_L^2 + \omega^2 L_L^2)(G_L^2 + \omega^2 C_L^2)} - \omega^2 L_L C_L + R_L G_L]} \tag{9-54}$$

In the low-loss approximation, it is approximated by:

$$\alpha_n = \frac{1}{2}\left(\frac{R_L}{Z_0} + G_L Z_0\right) \tag{9-55}$$

There is a simple conversion from a ratio of two voltages in nepers to the same ratio in dB. If r_n is the ratio of the two voltages in nepers and r_{dB} is the ratio of the same voltages in dB, then, since they equal the same ratio of voltages:

$$10^{\frac{r_{dB}}{20}} = e^{r_n} \tag{9-56}$$

$$r_{dB} = r_n \times 20\log e = 8.68 \times r_n \tag{9-57}$$

Using this conversion, the attenuation per length of a transmission line, in dB/length, is:

$$\alpha_{dB} = 8.68\alpha_n = 8.68 \times \frac{1}{2}\left(\frac{R_L}{Z_0} + G_L Z_0\right) = 4.34\left(\frac{R_L}{Z_0} + G_L Z_0\right) \quad (9\text{-}58)$$

where:

α_n = the attenuation of the amplitude, in nepers/length

α_{dB} = the attenuation, in dB/length

R_L = series resistance per length of the conductors

C_L = capacitance per length

L_L= series loop inductance per length

G_L= shunt conductance per length from the dielectric

Z_0 = the characteristic impedance of the line, in Ohms

Surprisingly, though this is the attenuation in the frequency domain, there is no intrinsic frequency dependence to the attenuation.

TIP If the series resistance per length of the conductors were constant with frequency and the shunt dielectric conductance per length were constant with frequency, the attenuation of the transmission line would be constant with frequency. Every frequency would see the same amount of loss.

Every frequency would be treated exactly the same in propagating through the transmission line. Though the amplitude of the signal would decrease as it propagates through the transmission line, the shape of the signal's spectrum would be preserved and the rise time would be unchanged. The result would be the same rise time coming out of the line as going in.

However, as we saw earlier, this is not how real lossy transmission lines on typical laminate substrates behave. In the real world, to a very good approximation, the resistance per length will increase with the square root of frequency due to skin depth, and the shunt conductance per length will increase linearly with frequency due to the dissipation factor of the dielectric. This means the attenuation will increase with frequency. Higher frequency sine waves will be attenuated more

than lower frequency sine waves. This is the primary mechanism that will decrease the bandwidth of signals when propagating down a lossy line.

There are two parts to the attenuation per length. The first part relates the attenuation from the conductor. The attenuation from just the conductor loss is:

$$\alpha_{cond} = 4.34\left(\frac{R_L}{Z_0}\right) \tag{9-59}$$

The second part of the attenuation relates to the losses from just the dielectric materials:

$$\alpha_{diel} = 4.34(G_L Z_0) \tag{9-60}$$

The total attenuation is:

$$\alpha_{dB} = \alpha_{cond} + \alpha_{diel} \tag{9-61}$$

where:

α_{cond} = the attenuation per length from just the conductor loss, in dB/length

α_{diel} = the attenuation per length from just the dielectric loss, in dB/length

α_{dB} = the total attenuation, in dB/length

R_L = series resistance per length of the conductors

C_L = capacitance per length

L_L= series loop inductance per length

G_L= shunt conductance per length from the dielectric

Z_0 = the characteristic impedance of the line, in Ohms

In the skin-depth-limited regime, the resistance per length of a stripline, ignoring the slight resistance in the return path, was approximated above as:

$$R_L = \frac{0.5t}{w\delta} = \frac{0.5 \times 34}{w \times 66}\sqrt{f} \tag{9-62}$$

or for frequency in GHz:

$$R_L = \frac{0.5t}{w\delta} = \frac{8.14}{w}\sqrt{f} \tag{9-63}$$

where:

R_L = the resistance per length, in Ohms/inch of a stripline

δ = the skin depth, in microns

t = the geometrical thickness, in microns (for 1-ounce copper)

w = the line width, in mils

f = the sine-wave-frequency component, in GHz

Combining these results, the attenuation from just the conductor in a stripline is approximately:

$$\alpha_{cond} = 4.34\left(\frac{R_L}{Z_0}\right) = 4.34 \times \frac{1}{Z_0} \times \frac{8.14}{w}\sqrt{f} = \frac{36}{wZ_0}\sqrt{f} \tag{9-64}$$

The total attenuation from the conductor for the entire length of stripline transmission line is:

$$A_{cond} = Len \times \alpha_{cond} = Len\frac{36}{wZ_0}\sqrt{f} \tag{9-65}$$

where:

R_L = the resistance per length, in Ohms/inch of the stripline

w = the line width, in mils

f = the sine-wave-frequency component, in GHz

Z_0 = the characteristic impedance of the line, in Ohms

A_{cond} = the total attenuation from just the conductor loss, in dB

Len = the length of the transmission line, in inches

For example, at 1 GHz, a 50-Ohm stripline line with a line width of 10 mils, will have an attenuation per length from just the conductor of $\alpha_{cond} = 36/(10 \times 50) \times 1 = 0.07$ dB/inch. If the line is 36 inches long, typical of a backplane application, the total attenuation from one end to the other would be 0.07 dB/inch × 36 inches = 2.5 dB. The ratio of the output voltage to the input voltage is $V_{out}/V_{in} = 10^{-2.5/20} = 75\%$. This means that there will only be 75% of the amplitude left at the end of the line for 1-GHz frequency components as a result of *just* the conductor losses. Higher-frequency components will be attenuated even more. Of course, this is an approximation. A more accurate value can be obtained using a 2D field solver that allows calculation of the precise current distribution and how it changes with frequency.

Figure 9-18 compares this estimate for the attenuation from just the conductor loss to the calculated attenuation for the case of a 10-mil-wide, 50-Ohm microstrip trace using a 2D field solver. The approximation is a reasonable estimate.

It should be noted that these estimates assumed the surfaces of the copper trace were smooth. When the surface roughness is comparable to the skin depth, the series resistance of that surface can double. For typical surface roughness on the order of 10 microns, the series resistance of a surface will double at frequencies above about 100 MHz. Since most copper foils are rough on one side and smooth on the other, the series resistance will double for only one side of the conductor.

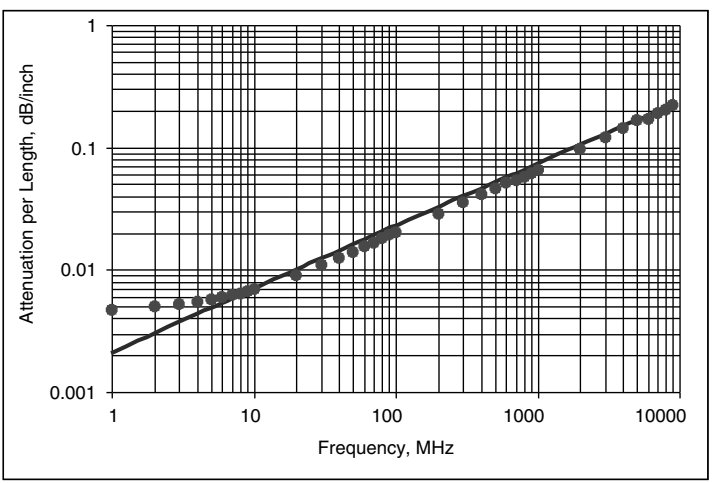


Figure 9-18 Calculated attenuation per length of a 10-mil-wide, 50-Ohm microstrip assuming only conductor loss and no dielectric loss, comparing the simple model above (line) and the simulation using Ansoft's SI2D field solver (circles).

This means the impact from surface roughness could increase the series resistance by 35% over the estimate above for smooth copper.

As shown previously, for all geometries, the conductance per length is related to the capacitance per length as:

$$G_L = \omega \tan(\delta) C_L \tag{9-66}$$

Likewise, for all geometries, the characteristic impedance is related to the capacitance as:

$$Z_0 = \frac{\sqrt{\varepsilon_r}}{cC_L} \tag{9-67}$$

From these two relationships, the attenuation per length from just the dielectric material can be rewritten as:

$$\alpha_{diel} = 4.34(G_L Z_0) = 4.34(\omega \tan(\delta) C_L)\left(\frac{\sqrt{\varepsilon_r}}{cC_L}\right) = \frac{4.34}{c}\omega \tan(\delta)\sqrt{\varepsilon_r} \tag{9-68}$$

where:

α_{diel} = the attenuation per length from just the dielectric loss, in dB/length

G_L = the conductance per length

ω = the angular frequency, in radians/sec

$\tan(\delta)$ = the dissipation factor

C_L = the capacitance per length

Z_0 = the characteristic impedance

ε_r = the real part of the dielectric constant

c = the speed of light in a vacuum

If we use units of inches/nsec for the speed of light, and GHz for the frequency, the attenuation per length from just the dielectric becomes:

$$\alpha_{diel} = 2.3 f \tan(\delta) \sqrt{\varepsilon_r} \tag{9-69}$$

where:

α_{diel} = the attenuation per inch from just the dielectric loss, in dB/inch

f = the sine-wave frequency, in GHz

$\tan(\delta)$ = the dissipation factor

ε_r = the real part of the dielectric constant

What is interesting is that the attenuation is independent of the geometry. If the line width is increased, for example, the capacitance will increase so the conductance will increase, but the characteristic impedance will decrease. The product stays the same.

TIP The attenuation due to the dielectric is only determined by the dissipation factor of the material. The attenuation due to the dielectric cannot be changed from the geometry; it is completely based on a material property.

This is not an approximation, but due to the fact that all the geometrical terms that affect the shunt conductance inversely affect the characteristic impedance, the product is always independent of geometry.

FR4 has a dissipation factor of roughly 0.02. At 1 GHz, the attenuation per length of a transmission line using FR4 would be about $2.3 \times 1 \times 0.02 \times 2 = 0.09$ dB/inch. This should be compared with the result above of 0.07 dB/inch for the attenuation per inch from the conductor for a 50-Ohm line and with a 10-mil width. At 1 GHz, the attenuation from the dielectric is slightly greater than the attenuation from the conductor. At even higher frequency, the attenuation from the dielectric will only get larger faster than the attenuation from the conductor. This means that if the dielectric loss dominates at 1 GHz, it will become more important at higher frequency and the conductor loss will become less important at higher frequency.

The attenuation from the dielectric will increase faster with frequency than the attenuation from the conductor. There is some frequency above which the attenuation will be dominated by the dielectric. Figure 9-19 shows the attenuation per length for a 50-Ohm line with an 8-mil-wide trace in FR4, comparing the attenuation from the conductor, from the dielectric, and for the total combined

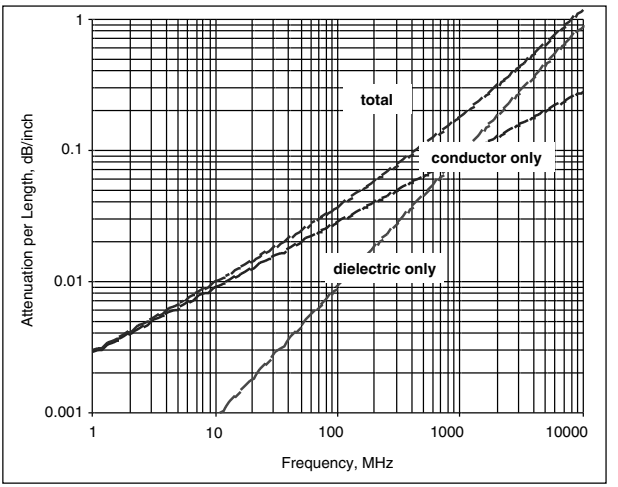


Figure 9-19 Attenuation per inch in a 50-Ohm microstrip with an 8-mil-wide trace, separating the attenuation from the conductor, the dielectric, and the total. In this geometry, with FR4, at frequencies above 1 GHz, the dielectric losses dominate the total losses.

attenuation per length. For 50-Ohm traces wider than 8 mils, the transition frequency where dielectric and conductor losses are equal is less than 1 GHz. Above 1 GHz, the dielectric loss dominates. If the line width is narrower than 8 mils, the transition frequency is above 1 GHz.

When referring to the losses in a transmission line, there are many terms used in the industry, and unfortunately used interchangeably, even though they all refer to different quantities.

The following are some of the terms used and their real definitions:

- Loss: This is the generic term referring to all aspects of lossy lines.
- Attenuation: This is the specific measure of the *total* attenuation of a line, which is a measure of the decrease in power of the transmitted signal (when measured in dB) or the decrease in amplitude (when described as a percentage transmitted signal). When measured in dB, the total attenuation of a signal will increase linearly with the length of the line. When measured in percent voltage at the output, it will decrease exponentially with increasing line length.
- Attenuation per length: This is the total attenuation of the power, measured in dB, normalized to the length of the line, which is constant as long as the line parameters of the transmission line are constant. The attenuation per length is intrinsic and does not depend on the length of the interconnect.

- Dissipation factor: This is the specific, intrinsic material property of all dielectrics, which is a measure of the number of dipoles and how far they can move in an AC field. This is the material property that contributes to dielectric loss and may be slightly frequency dependent.
- Loss angle: This is the angle, in the complex plane, between the complex dielectric constant vector and the real axis.
- tan(δ): The tangent of the loss angle, which is also the ratio of the imaginary part of the complex dielectric constant to the real part of the complex dielectric constant, is also known as the dissipation factor.
- Real part of the dielectric constant: The real part of the complex dielectric constant is the term associated with how a dielectric will increase the capacitance between two conductors, as well as how much it will slow down the speed of light in the material. It is an intrinsic material property.
- Imaginary part of the dielectric constant: The imaginary part of the complex dielectric constant is the term associated with how a dielectric will absorb energy from electric fields due to dipole motion. It is an intrinsic material property related to the number of dipoles and how they move.
- The dielectric constant: Normally associated with just the real part of the complex dielectric constant, the dielectric constant relates how a dielectric will increase the capacitance between two conductors.
- The complex dielectric constant: This is the fundamental intrinsic material property that describes how electric fields will interact with the material. The real part describes how the material will affect the capacitance; the imaginary part describes how the material will affect the shunt leakage resistance.

When referring to the loss in a transmission line, it is important to distinguish the term to which we are referring. The actual attenuation is frequency dependent. However, the dissipation factor of the material, or other properties of the material, are generally only slowly varying with frequency. Of course, the only way to know this is by measuring real materials.

9.12 Measured Properties of a Lossy Line in the Frequency Domain

The ideal model of a lossy transmission line introduced here has three properties:

1. A characteristic impedance that is constant with frequency
2. A velocity that is constant with frequency

3. An attenuation that has a term proportional to the square root of frequency and another term proportional to the frequency

The assumption here is that the dielectric constant and dissipation factor are constant with frequency. This does not have to be the case; it just happens that it is a pretty good approximation for most materials in most cases. In real materials that have a frequency-dependent material property, it usually varies so slowly with frequency that it can be considered constant over wide-frequency ranges. The only way to know how it varies is to measure it.

Unfortunately, at frequencies in the GHz range, where knowing the material properties is important, there is no instrument that is a dissipation-factor meter. We cannot put a sample of the material in a fixture and read out the final dissipation factor at different frequencies. Instead, a slightly more complicated method must be used to extract the intrinsic material properties from laminate samples.

The first step is to build a transmission line with the laminate, preferably a stripline, so there is uniform dielectric everywhere around the signal path. In order to probe the transmission line, there will be vias at both ends. Using microprobes, the behavior of sine-wave voltages can be measured with minimal impact from the probes.

A vector network analyzer (VNA) can be used to send in sine waves and measure how they are reflected and transmitted by the transmission line. The ratio of the reflected to the incident sine wave is called the return loss or S_{11}, and the ratio of the transmitted to the incident sine wave is called the insertion loss or S_{21}.

These two terms, at every frequency, completely describe how sine waves interact with the transmission line. At each frequency, S_{11} or S_{21}, called S or scattering parameters, describe the reflected or transmitted sine-wave amplitude and phase compared with the incident amplitude and phase. The further restriction in this definition is that reflected and transmitted sine waves are measured when a 50-Ohm source and load are connected to the ends of the transmission line.

If the characteristic impedance of the transmission line is different from 50 Ohms, there will be significant reflections, and depending on the length of the line, there will be periodic patterns in the S parameters, as the sine waves find resonances due to the length of the line and impedance discontinuities. However, we can account for all of these effects if we know the characteristic impedance of the line and the model for the vias or connectors on the ends.

The actual measured insertion loss in a 4-inch length of 50-Ohm transmission line is shown in Figure 9-20. The transmitted signal, S_{21}, in dB, drops off roughly as the frequency increases, as expected. In this example, the transmission

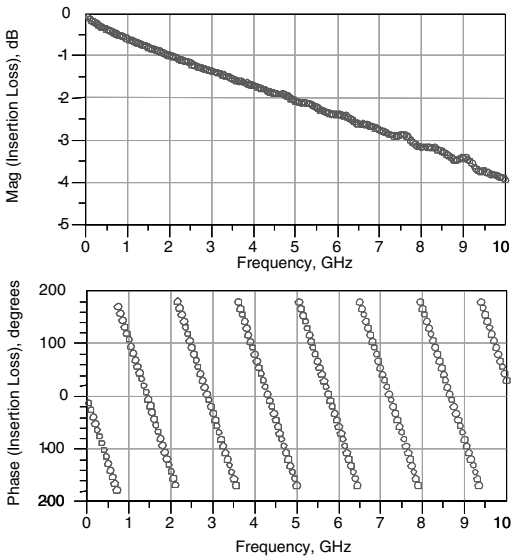


Figure 9-20 Measured insertion loss of a 4-inch-long, nearly 50-Ohm stripline made from FR4 measured with a GigaTest Labs Probe Station.

line is roughly 50 Ohms, and the vias do not contribute a very noticeable effect until above 8 GHz. The measured insertion loss is a rough approximation of the attenuation. When S_{21} is displayed in dB, the slope of the attenuation is seen to be pretty constant, as expected by the simple model.

For a perfectly matched transmission line with no vias, S_{21} is exactly a measure of the attenuation as a function of frequency. However, in practice, this is almost impossible to engineer, so we must interpret S_{21} based on a model for the transmission-line test structure, which includes the vias.

To account for the electrical effects of the vias at the ends, we can model them as a simple pi circuit with a C-L-C topology. The real transmission-line test structure can be modeled with an ideal circuit model as shown in Figure 9-21. This model is fully defined in terms of just eight parameter values:

C_{via1} = the capacitance of the first part of the via

L_{via} = the loop inductance associated with the via

C_{via2} = the capacitance of the second part of the via

Z_0 = the lossless characteristic impedance

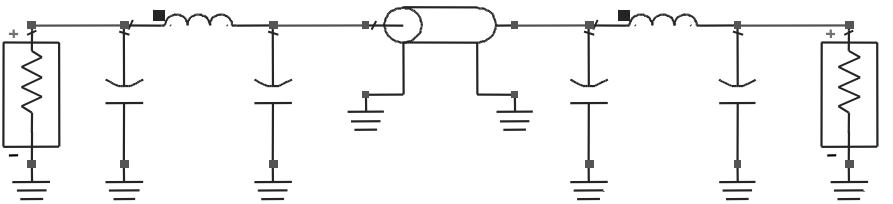


Figure 9-21 Circuit topology to model this 4-inch-long transmission line, including the VNA ports and vias at the two ends of the ideal lossy transmission line.

ε_r = the real part of the dielectric constant, assumed to be constant over frequency

len = the length of the transmission line, in this case, measured with a ruler as 4 inches

tan(δ) = the dissipation factor of the laminate, assumed to be constant over-frequency

$\alpha_{cond}/f^{0.5}$= the term related to conductor loss and normalized to the square-root of the frequency

If we knew the values for each parameter of this model, we could simulate its insertion loss. If we had the right topology and we knew the values of each parameter, we should get very good agreement between the measured insertion loss and the simulated insertion loss.

After optimization, the best set of parameter values for each of the eight terms in the model, for this particular measured transmission line, is:

C_{via1} = 0.025 pF

L_{via} = 0.211 nH

C_{via2} = 0.125 pF

Z_0 = 51.2 Ohms

ε_r = 4.05

len = 4 inches

tan(δ) = 0.015

$\alpha_{cond}/f^{0.5}$= 3 dB/m/sqr(f)

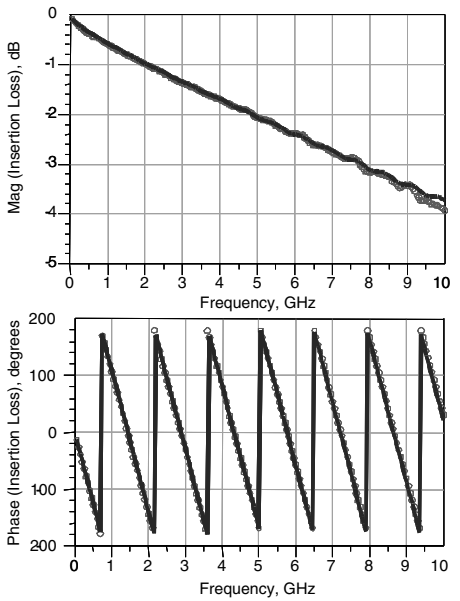


Figure 9-22 Comparing the measured and simulated insertion loss of a 4-inch-long, nearly 50-Ohm transmission line made from FR4.

In Figure 9-22, we show the comparison of the simulated insertion loss and the actual measured insertion loss using the ideal circuit model and these parameter values, which are all constant across the entire measurement bandwidth.

The agreement is excellent between the measured insertion loss and the predicted insertion loss based on the simple ideal circuit model. This gives us confidence we have a good topology and parameter values and can conclude that the dielectric constant is 4.05 and the dissipation factor is 0.015. The assumption they are both constant across the entire measurement bandwidth of 10 GHz fits the measured data.

In Figure 9-23, the measurement bandwidth is extended up to 20 GHz for this same sample. There is excellent agreement between the measured response and the predicted response, assuming a constant dielectric constant and dissipation factor, up to about 14 GHz. Above this frequency, it appears as though the actual dissipation factor of the laminate is increasing and the dielectric constant is decreasing very slightly. From this measurement, we can identify the limit to the assumption about constant material properties. For this FR4 sample, the assumption of constant material properties is good to about 14 GHz.

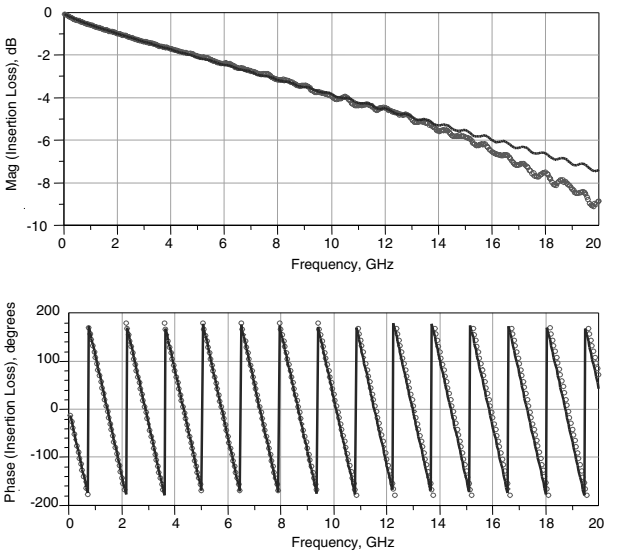


Figure 9-23 Measured and simulated insertion loss extended up to 20 GHz showing 14 GHz as the bandwidth of the model. Above this frequency, it appears the actual dissipation factor may be increasing with frequency.

TIP The fact that we get such excellent agreement between the measured insertion loss compared with the simple ideal lossy transmission-line model supports the use of this simple model to describe the high-frequency properties of real lossy transmission lines. The only caveat is that for different material systems, it is important to measure the specific material properties.

9.13 The Bandwidth of an Interconnect

If we start with the spectrum of an ideal square wave and preferentially attenuate the higher-frequency components more than the low-frequency components, the bandwidth of the transmitted signal—the highest sine-wave frequency that is significant—will decrease. The longer we let the wave propagate, the more the attenuation of the higher-frequency components and the lower the bandwidth.

The entire concept of bandwidth as the highest sine-wave-frequency component that is significant is inherently only a rough approximation. As we have stated previously, if a problem is so sensitive to bandwidth that knowing its value to within 20% is important, the bandwidth term should not be used. Rather, the entire spectrum of the signal and actual insertion or return loss behavior of the

interconnect across the whole frequency range should be used. However, the concept of bandwidth is very powerful in helping to feed our intuition and provide insight into the general behavior of interconnects.

There is a simple and fundamental connection between the bandwidth of an interconnect and the losses in the transmission line. The longer the line, the greater the high-frequency losses and the lower the bandwidth of the line. Being able to estimate the loss-limited bandwidth of an interconnect will allow us to establish some performance requirements on how much attenuation is too much and what material properties might be acceptable.

As we showed in an earlier chapter, the bandwidth of a signal is the highest frequency that has less than –3-dB amplitude compared to an ideal square-wave amplitude. At each distance, d, down the transmission line, we can calculate the frequency that has a 3-dB attenuation and is the bandwidth of the signal at that point. This frequency will be the intrinsic bandwidth of the transmission line, BW_{TL}.

Assuming we are in a frequency regime where dielectric loss dominates, so we can ignore the resistive loss, the total attenuation at a frequency, f, and having propagated a distance, d, is:

$$A_{dB} = \alpha_{diel} \times d = 2.3f \times \tan(\delta)\sqrt{\varepsilon_r} \times d \tag{9-70}$$

where:

A_{dB} = the total attenuation, in dB

α_{diel} = the attenuation per length from just the dielectric, in dB/inch

ε_r = the real part of the complex dielectric constant

d = the length of the transmission line, in inches

f = the frequency of the sine wave, in GHz

tan(δ) = the material's dissipation factor

The intrinsic bandwidth of the transmission line, BW_{TL}, corresponds to the frequency that has an attenuation of just 3 dB. By substituting BW_{TL} for the frequency, f, and the attenuation as 3 dB, the connection between the bandwidth and interconnect length can be found as:

$$BW_{TL} = \frac{3dB}{2.3 \times \tan(\delta) \times \sqrt{\varepsilon_r}} \times \frac{1}{d} = \frac{1.3}{\tan(\delta) \times \sqrt{\varepsilon_r}} \times \frac{1}{d} \quad (9\text{-}71)$$

where:

BW_{TL} = the intrinsic bandwidth of the interconnect, in GHz, for a length, d, in inches

ε_r = the real part of the complex dielectric constant

d = the length of the transmission line, in inches

$\tan(\delta)$ = the material's dissipation factor

This says, the longer the interconnect length, the lower the bandwidth and the lower the frequency at which the attenuation has increased to 3 dB. Likewise, the higher the value of the dissipation factor, the lower the bandwidth of the interconnect.

The rise time of an ideal square wave is 0, and the bandwidth of its spectrum is infinite. If we do something to the spectrum to decrease the bandwidth, the rise time will increase. The resulting rise time, RT, is:

$$RT = \frac{0.35}{BW} \quad (9\text{-}72)$$

where:

RT = the rise time, in nsec

BW = the bandwidth, in GHz

In a lossy interconnect, given its intrinsic bandwidth as limited by the dissipation factor of the material, we can calculate the resulting rise time of the waveform after propagating down the transmission line:

$$RT_{TL} = 0.35 \times \frac{\tan(\delta) \times \sqrt{\varepsilon_r}}{1.3} \times d = 0.27 \times \tan(\delta) \times \sqrt{\varepsilon_r} \times d \quad (9\text{-}73)$$

where:

RT_{TL} = the intrinsic rise time of the transmission line, in nsec

ε_r = the real part of the complex dielectric constant

d = the length of the transmission line, in inches

f = the frequency of the sine wave, in GHz

tan(δ) = the material's dissipation factor

For example, a transmission line using FR4, and having a dissipation factor of 0.02, would have an intrinsic interconnect rise time for a 1-inch length of line of about $0.27 \times 0.02 \times 2 \times 1 = 10$ psec. For a length of 10 inches, the intrinsic interconnect rise time would be about 100 psec. If a 1-psec rise-time signal were sent into such a transmission line, after traveling 10 inches, its rise time would have been increased to about 100 psec due to all the high-frequency components being absorbed by the dielectric and converted into heat.

TIP As a rough rule of thumb, the rise time of a signal propagating down an FR4 transmission line will increase its rise time by about 10 psec/inch of travel.

The actual rise time of the signal will get longer and longer as the signal propagates down the line. This intrinsic interconnect rise time is dominated by the length of the line and the dissipation factor of the laminate, and it is the shortest rise time the interconnect will support. Figure 9-24 shows the intrinsic interconnect

Material	ε	tan(δ)	Intrinsic Rise Time in psec/in
FR-4	4.0-4.7	0.02	10
DriClad (IBM)	4.1	0.011	5.4
GETek	3.6-4.2	0.013	7
BT	4.1	0.013	7
Polyimide/glass	4.3	0.014	8
CyanateEster	3.8	0.009	4.7
NelcoN6000SI	3.36	0.003	1.5
RogersRF35	3.5	0.0018	0.9

Figure 9-24 Intrinsic interconnect rise time for various laminate materials, assuming the bandwidth is limited only by dielectric loss.

rise time for a variety of laminate materials. It can range from 10 psec/inch in FR4 to less than 1 psec/inch in some Teflon-based laminates.

In these examples, we are assuming that the line width is wide enough that the chief limitation to attenuation is the dielectric. Of course, if the line is very narrow, especially for low-loss materials, the intrinsic rise time of the interconnect will be larger than the estimate based only on dielectric loss.

When the rise time that enters is not 1 psec but some longer value, RT_{in}, even comparable to the intrinsic rise time, the resulting output rise time, RT_{out}, is related to the intrinsic interconnect rise time, by:

$$RT_{out} = \sqrt{RT_{in}^2 + RT_{TL}^2} \tag{9-74}$$

where:

RT_{out} = the rise time coming out of the interconnect

RT_{in} = the rise time of the signal going into the interconnect

RT_{TL} = the intrinsic interconnect rise time

This is only a rough approximation, assuming the shape of the rising edge is Gaussian. As an example, Figure 9-25 shows a roughly 41-psec input-rise-time signal entering an 18-inch-long, FR4 transmission line. The intrinsic interconnect rise time is about RT_{TL} = 10 psec/inch × 18 inches = 180 psec. We would expect the output rise time to be about:

$$RT_{out} = \sqrt{41^2 + 180^2} = 185 \text{ psec} \tag{9-75}$$

In fact, what is measured is a rise time of about 150 psec, close to this estimate.

If the intrinsic interconnect rise time is much smaller than the input rise time, the output rise time is roughly the same and is unchanged. The relative change in the output rise time to the input rise time is:

$$\frac{RT_{out}}{RT_{in}} = \sqrt{1 + \left(\frac{RT_{TL}}{RT_{in}}\right)^2} \tag{9-76}$$

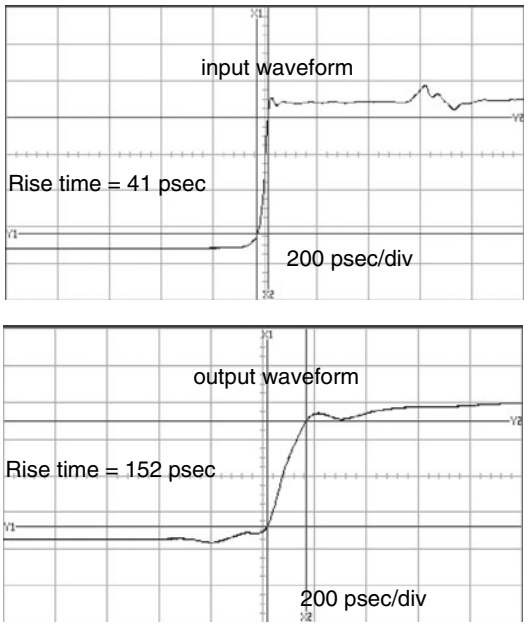


Figure 9-25 Measured rise-time degradation through an 18-inch length of FR4 trace with roughly a 50-Ohm characteristic impedance. Measured with an Agilent 54120 TDR.

To increase the output rise time by 25%, the intrinsic rise time has to be at least about 50% of the input rise time.

TIP This suggests that for a lossy transmission line to not appreciably degrade the rise time of the signal by more than about 25%, the intrinsic interconnect rise time must be less than 50% of the input rise time. If the initial rise time of the signal is 100 psec, the intrinsic interconnect rise time should be less than 50 psec. If it is longer, we will end up with an output rise time grossly increased.

In FR4, with an intrinsic-rise-time degradation of about 10 psec/inch, or 0.01 nsec/inch, there is a simple rule of thumb relating the rise time and the interconnect length at which the lossy effects will be important:

$$RT_{TL} > 0.5 \times RT_{signal} \tag{9-77}$$

$$0.01 \times d > 0.5 \times RT_{signal} \tag{9-78}$$

$$d > 50 \times RT_{signal} \tag{9-79}$$

where:

RT_{TL} = the intrinsic interconnect rise time, in nsec/inch

RT_{signal} = the rise time of the signal, in nsec

d = the interconnect length, in inches, where lossy effects are important

For example, if the rise time is 1 nsec, for transmission lines longer than 50 inches, the lossy effects will degrade the rise time and potentially cause ISI problems. If the lengths are shorter than 50 inches, the lossy effects in FR4 may not be a problem. However, if the rise time were 0.1 nsec, lossy effects may be a problem for lengths longer than only 5 inches.

This is why most motherboard applications with dimensions on the order of 12 inches and typical rise times of 1 nsec do not experience problems with lossy effects. However, for backplanes, where lengths are > 36 inches and rise times are less than 0.1 nsec, lossy effects can often dominate performance.

TIP This suggests a simple rule of thumb for estimating when to worry about lossy lines. In FR4, when the length of the line (in inches) is > 50 × the rise time (in nsec), the lossy effects may play a significant role.

Of course, this analysis is only a rough approximation. We have been assuming that we can actually use the 10–90 rise time to describe the output signal. In fact, the actual waveform gets distorted in a complex way since the higher frequency components are decreased gradually and the actual spectrum of the transmitted signal will change.

This simple rule of thumb describing the rise-time degradation through a lossy line is meant only as a way of estimating the point at which lossy-line properties will start to affect the signal quality. At this point, to accurately predict the actual waveforms and signal-quality effects, a lossy-line transient simulator should be used.

9.14 Time-Domain Behavior of Lossy Lines

If high-frequency components are attenuated more than low-frequency components, the rise time will increase as the signal propagates. Rise time is usually defined as the time for the edge to transition between 10% and 90% of the final values. This assumes the edge profile of the signal looks somewhat Gaussian, with the middle as the fastest slope region. For this waveform, the 10%–90% rise time makes sense and has value.

However, due to the nature of the attenuation in lossy lines, the rise time is distorted and the waveform is not a simple Gaussian edge. The initial part of the waveform is faster and there is a long tail to the rising edge. If we use just one number, like the 10–90 rise time, to describe the rise time, we will have a distorted sense of when the signal has reached a level associated with a trigger threshold. In the lossy regime, rise time has less meaning and is more of a rule of thumb indicator only.

Figure 9-26 is an example of the measured input waveform and output waveform through a 15-inch-long lossy transmission line in FR4 having a dissipation factor of about 0.01. The resulting rising-edge waveform is not very Gaussian.

In comparing the actual measured S-parameters of real lossy transmission lines, in the frequency domain, with the predictions of ideal lossy transmission-line models, it is clear that the simple, ideal models can work very well at frequencies at least above 10 GHz, providing we have the correct material properties.

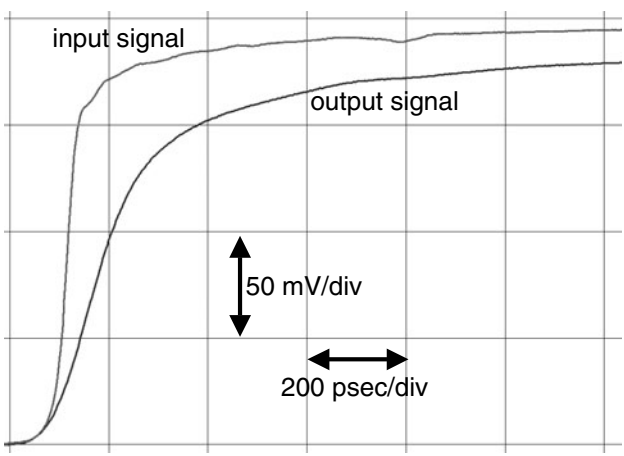


Figure 9-26 Measured rise-time degradation through a 15-inch length of FR4 trace with roughly a 50-Ohm characteristic impedance. Sample supplied by Doug Brooks and measured with an Agilent 86100, a GigaTest Labs Probe Station, and analyzed with TDA Systems IConnect.

This ideal lossy transmission-line model should be a good model to predict the time-domain behavior of real transmission lines as well. The basis of this ideal lossy transmission line model is that the series resistance is proportional to the square root of the frequency and the shunt conductance is proportional to the frequency. This is how most real lossy lines behave.

However, this is not how an ideal resistor behaves. An ideal resistor element is constant with frequency. If we use a time-domain simulator that simply has ideal resistor elements for both the series resistance and the shunt conductance, it will not be able to accurately simulate the effects of a lossy line. If the resistance is constant with frequency, the attenuation will be constant with frequency and there will be no rise-time degradation. The output rise time will look identical to the input rise time, just with slightly less amplitude.

TIP A simulator with only a model of a resistor element with resistance constant with frequency is worthless as a lossy-line simulator. It will miss the most important impact on performance.

Using a lossy-line simulator with the ideal lossy-line model having a frequency dependent resistance and conductance, the time-dependent waveform can be evaluated. An example of a transient simulation using a lossy-line simulator is shown in Figure 9-27.

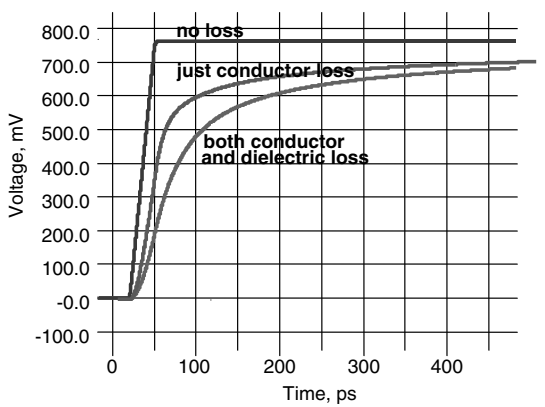


Figure 9-27 Simulated transmitted signals at the output of a 30-inch-long transmission line for a roughly 50-psec rise-time input signal with no loss, with conductor loss for an 8-mil-wide trace, and with combined conductor and dielectric loss for a dissipation factor of 0.02, showing the increasing rise-time degradation. Simulation performed with Mentor Graphics Hyperlynx.

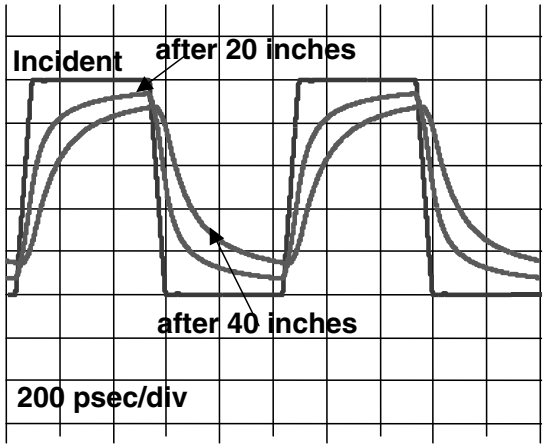


Figure 9-28 Simulated transmitted signals at the output of a 20-inch- and a 40-inch-long transmission line for a 1-Gbps signal, assuming 50-Ohm line, 8-mil-wide trace, and FR4 dielectric. These signals are compared to the received signal if there were no losses. Simulation performed with Mentor Graphics Hyperlynx.

If a similar interconnect were used with a 1-GHz clock, the resulting signal at the far end would look similar to that shown in Figure 9-28, comparing the lossless simulation with the lossy simulation for the case of 20 inches long and 40 inches long.

The most effective way of evaluating the impact of a lossy transmission line is by displaying the transmitted signal in an eye diagram. This shows the degree to which each bit pattern can be discerned for all combinations of bits. A pseudorandom bit pattern is synthesized, and the transmitted signal through the interconnect is simulated. Each bit is overlaid with the previous one, synchronized to the clock. If there is no ISI, the eye pattern will be perfectly open. In other words, each bit, no matter what the previous pattern was, would look the same and be identical to the previous bit. Its eye diagram would look like just one cycle.

The degree of ISI from losses and other effects, such as capacitive discontinuities of vias, will collapse the eye diagram. If the eye collapses more than the noise margin of the receiver, the bit error rate will increase and may cause faults.

Figure 9-29 shows the simulated eye diagram for the case of a 50-Ohm, 36-inch-long backplane trace in FR4 with no losses or discontinuities, and then successively turning on the resistive loss, the dielectric loss, and one 0.5-pF via at either end of the line. The line width in this example is 4 mils. The stimulating source has a bit period of 200 psec, corresponding to a bit rate of 5 Gbps.

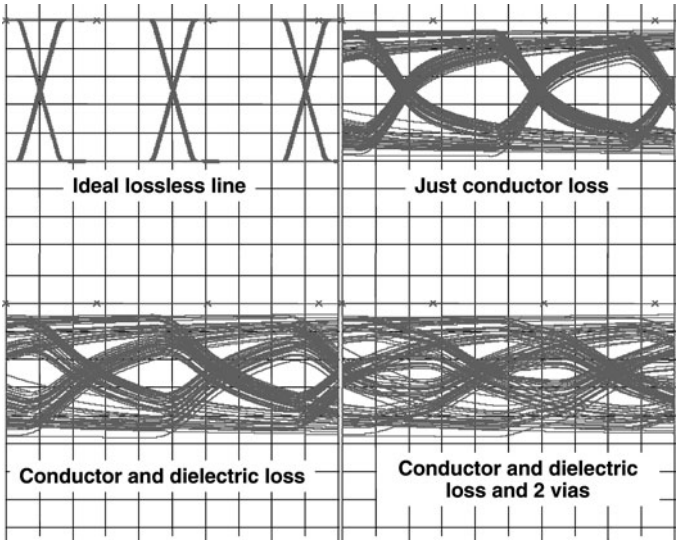


Figure 9-29 Simulated transmitted signals at the output of a 36-inch-long, 50-Ohm FR4 backplane trace, successively adding the effects of conductor loss, dielectric loss, and two vias. Simulation performed with Mentor Graphics Hyperlynx.

In the final simulation, which includes the losses and capacitive load from the vias, the eye is substantially closed and would not be usable at this bit rate. For acceptable performance, the transmission line would have to be improved or signal-processing techniques used to enhance the opening of the eye diagram.

9.15 Improving the Eye Diagram of a Transmission Line

There are three board-design factors that influence the quality of the eye diagram:

1. Discontinuities from the via stubs

2. Conductor loss

3. Dielectric loss

If the rise-time degradation is a problem, these are the only board-level features that will affect the performance.

The first step is to design the sensitive lines with vias having a minimum stub length by restricting layer transitions, using blind and buried vias or back-drilling long stubs. The next step is to reduce the size of capture pads and increase

the size of antipad clearance holes to match the via impedance closer to 50 Ohms. This will minimize the rise-time degradation.

> **TIP** In general, the biggest impact a via will have is from its stub. Reduce the stub length to less than 10 mils, and the via may look pretty transparent up to 10 GHz. Then worry about matching the via to 50 Ohms.

The line width of the signal trace will be the dominant term affecting the attenuation from the conductor losses, provided the dielectric thickness is changed to keep the impedance of the line at a fixed target. A wider line width will decrease the conductor loss. If the line width is increased, the dielectric thickness would have to increase as well. This is not always practical and thus sets a limit to how wide a line can be used.

Depending on the bandwidth of interest, there is little impact from making the line too wide, because the dielectric loss may dominate. Figure 9-30 shows the impact on the attenuation per length of a 50-Ohm line, with FR4 dielectric. If decreasing attenuation is important, the first goal is to use wide lines and avoid line widths less than 5 mils if at all possible. However, making the line width wider than 10 mils does not significantly decrease the attenuation because of the dielectric losses in the FR4.

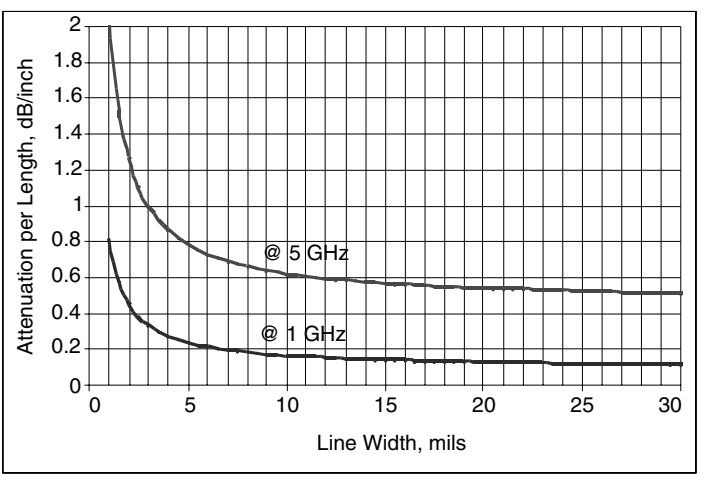


Figure 9-30 Total attenuation per length for a 50-Ohm line in FR4, as the line width is increased, assuming the dielectric thickness is also increased to keep the impedance constant.

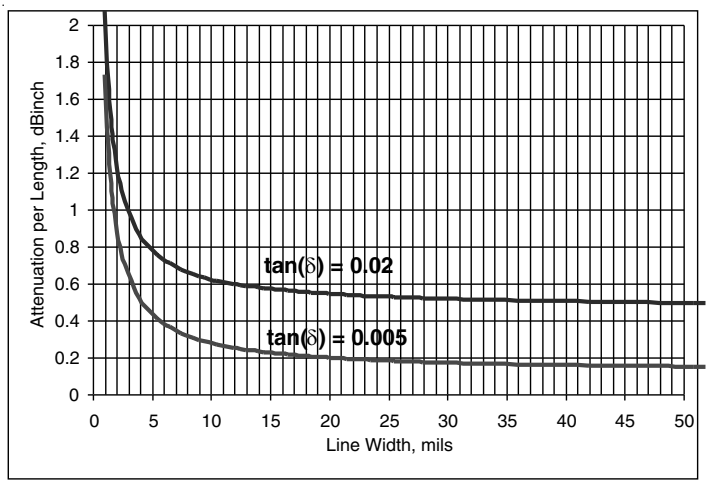


Figure 9-31 Total attenuation per length at 5 GHz for a 50-Ohm line as the line width is increased, for two different dissipation-factor materials.

TIP This suggests that with FR4-laminate lines, the optimum line width for minimal attenuation is around 5 mils to 10 mils wide.

If vias are optimized and the line width is kept to 10 mils, the only other term to adjust is the dissipation factor of the laminate. Figure 9-31 shows a similar attenuation curve for two different dissipation-factor materials, both at 5 GHz. A lower-dissipation factor will always contribute to a lower attenuation. Again, we see there is a point of diminishing returns in increasing the line width even for a low-loss laminate. Much beyond 20 mils, the attenuation is dominated by the laminate material. This is why an important element to predicting high-speed performance of interconnects is having accurate values for the material properties.

9.16 Pre-emphasis and Equalization

If the interconnect losses cause the eye to close, there is another approach to open the eye back up. If we send a sharp voltage step through the interconnect, it will be distorted by the time it exits the interconnect. If we can predict the degree of distortion, we can predistort the signal so that after propagating down the interconnect, the signal will approximate the sharp voltage step.

There are two ways of distorting the waveform. The fast-rising edge, which has the high-frequency components, will be attenuated more than the slow rising

edge. If we add a filter that cuts out the low frequencies and lets through the high frequencies, the product of the attenuation from the interconnect and the filter will be equal across the bandwidth and there will not be frequency dependent loss.

When we filter out the low-frequency components to match them to the attenuated high-frequency components, we call this *equalization*. If we add gain to the filter and increase the strength of the high frequency components, we call this *active equalization*.

Alternatively, we can add extra high-frequency components to the initial signal, so that by the time the edge reaches the far end, these higher frequency components have attenuated back in line with the low-frequency components. This is called *pre-emphasis*.

Using either technique requires that the distortions of the interconnect are predictable and reproducible. This can happen only if the material properties of the laminate are known. Equalization and pre-emphasis are powerful techniques to compensate for lossy interconnects and are used in all high-end, high-speed serial links.

9.17 The Bottom Line

1. The fundamental problem caused by lossy lines is rise-time degradation, which results in pattern-dependent noise, also called intersymbol interference, or ISI.
2. The frequency-dependent losses in circuit-board interconnects arise from the conductor losses and dielectric losses.
3. The rise time of a signal is increased when propagating down a lossy line because the higher frequencies are attenuated more than the low frequencies. This results in a decrease in the bandwidth of the propagating signal.
4. At about 1 GHz, for line widths of about 8 mils, the contributions from the two losses are comparable. At a higher frequency, the dielectric losses increase in a manner that is proportional to the frequency, while the conductor losses increase in a manner that is proportional to the square root of the frequency.
5. The characteristic impedance of the line and the speed of a signal are not affected by the losses above just a few MHz.
6. At frequencies above 1 GHz, the dielectric losses dominate and the dielectric's dissipation factor is the most important material term that describes the

lossy behavior of a material. Better materials will have a lower dissipation factor. FR4, with a dissipation factor of about 0.02, is the worst behaving material.

7. The lossy behavior of a transmission line is very accurately predicted with a lossy-line model that uses a resistance per length proportional to the square root of frequency and a shunt conductance that is proportional to frequency. This model accounts for ISI.

8. In addition to the material losses, any impedance discontinuities such as via stubs, can degrade the rise time and contribute to ISI. The chief impact from a via is not from the through part of the via, but from the via stub. Removing the via stub can make the via almost disappear.

9. As a rough rule of thumb, the dielectric losses in FR4 will degrade the rise time by about 10 psec per inch of travel. After 10 inches, the rise time can be increased to 100 psec.

CHAPTER 10

Cross Talk in Transmission Lines

Cross talk is one of the four families of signal-integrity problems. It is the transfer of an unwanted signal from one net to an adjacent net and will occur between every pair of nets. A net includes both the signal and the return path, and connects one or more nodes in a system. We typically call the net with the source of the noise the *active net* or the *aggressor net*. The net on which the noise is generated is called the *quiet net* or the *victim net*.

TIP Cross talk is an effect that happens between the signal and return paths of one net and the signal and return paths of a second net. The entire signal-return path loop is important, not just the signal path.

The noise margin is typically about 15% of the total signal-voltage swing, but varies among device families. Of this 15%, about 1/3 or 5% of the signal swing, is typically allocated to cross talk. If the signal swing were 3.3 v, the maximum allocated cross talk might be about 160 mV. This is a good starting place for the maximum allowable cross-talk noise. Unfortunately, the magnitude of the noise generated in typical traces on a board can often be larger than 5%. This is why it is important to be able to predict the magnitude of cross talk, identify the origin of excessive noise, and actively work to minimize the cross talk in the

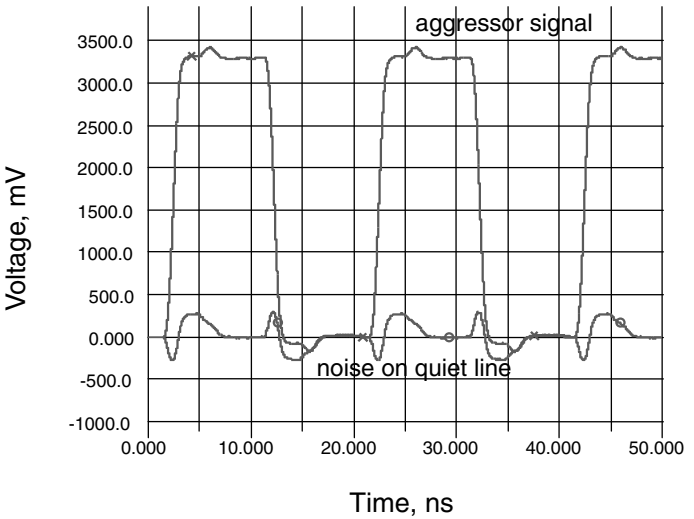


Figure 10-1 Simulated cross talk on a quiet line with aggressor lines on either side. Each line is a source-series terminated 50-Ohm microstrip in FR4 with a 10-mil line and space. Simulated with Mentor Graphics Hyperlynx.

design of packages, connectors, and board-level interconnects. Understanding the origin of the problem and how to design interconnects with reduced cross talk is increasingly important as rise time decreases.

Figure 10-1 shows the noise on the receiver of a quiet line when aggressor lines on either side have 3.3-v signals. In this example, the noise at the receiver is more than 300 mV.

10.1 Superposition

Superposition is an important principle in signal integrity and is critical when dealing with cross talk. Superposition is a property of all linear, passive systems, of which interconnects are a subset. It basically says multiple signals on the same net do not interact and are completely independent of each other. The amount of voltage that might couple onto a quiet net from an active net is completely independent of the voltage that might already be present on the quiet net.

Suppose the noise generated on a quiet line were 150 mV from a 3.3-v driver when the voltage level on the quiet line is 0 v. There would also be 150 mV of noise generated on the quiet line when the quiet line is driven directly by a driver to a level of 3.3 v. The total voltage appearing on the quiet line would be the direct sum of the signals that may be present and the coupled noise. If there are two

active nets coupling noise to the same quiet line, the amount of noise appearing on the quiet line would be the sum of the two noise sources. Of course they may have a different time dependency based on the voltage pattern on the two active lines.

Based on superposition, if we know the coupled noise when the quiet line has no additional signal on it, we can determine the total voltage on the quiet line by adding the coupled noise and any signal that might also be present.

Once the noise is on the quiet line, it is subject to the same behavior as the signal: It sees the same impedance and will suffer reflections and distortions from any impedance discontinuities that may be present in the quiet line.

If a quiet line has an active line on either side of it, and each active line couples an equal amount of noise to the quiet line, the maximum allowable noise between one pair of lines would be $1/2 \times 5\% = 2.5\%$. In a bused topology, it will be important to be able to calculate the worst case total number of adjacent traces that might couple to determine the worst case coupled noise. This will put a limit on how much noise might be allowable between just two traces.

10.2 Origin of Coupling: Capacitance and Inductance

When a signal propagates down a transmission line, there are electric-field lines between the signal and return paths and rings of magnetic-field lines around the signal and return path conductors. These fields are not confined to the immediate space between the signal and return paths. Rather, they spread out into the surrounding volume. We call these fields that spread out *fringe fields*.

TIP As a rough rule of thumb, the capacitance contributed by the fringe fields in a 50-Ohm microstrip in FR4 is about equal to the capacitance from the field lines that are directly beneath the signal line.

Of course, the fringe fields drop off very quickly as we move farther away from the conductors. Figure 10-2 shows the fringe fields between a signal path and a return path and how they might interact with a second net when it is far away and then when it is close.

If we are unfortunate enough to route another signal and its return path in a region where there are still large fringe fields from another net, the second trace may pick up noise from these fringe fields. The only way noise will be picked up in the quiet line is when the signal voltage and current in the active line change. This will cause current to flow through the changing electric and magnetic fields.

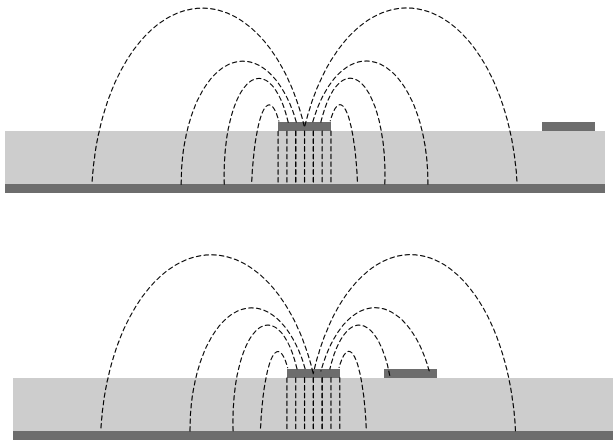


Figure 10-2 Fringe fields near a signal line. When a second trace is far away, there is little fringe-field coupling and little cross talk. When the second net is in the vicinity of the fringe fields, there can be excessive coupling and cross talk.

We can describe this coupling with circuit models by using capacitors and mutual inductors.

TIP Ultimately, it is the fringe fields that cause cross talk. The most important way to minimize cross talk is to space nets far enough apart so their fringe fields are reduced to acceptable levels.

Between every two nets in a system, there will always be some combination of capacitive coupling and inductive coupling arising from these fringe fields. We refer to the coupling capacitance and the coupling inductance as the mutual capacitance and the mutual inductance. Obviously, if we were to move the two adjacent signal- and return-path traces farther apart, the mutual-capacitance and mutual-inductance parameter values would decrease.

Being able to predict the cross talk based on the geometry is an important step in evaluating how well a design will meet the performance specification. This means being able to translate the geometry of the interconnects into the equivalent mutual capacitance and inductance, and relating how these two terms contribute to the coupled noise.

Though both mutual capacitance and mutual inductance play a role in cross talk, there are two regimes to consider. When the return path is a wide, uniform plane, as is the case for most coupled transmission lines in a circuit board, the capacitively coupled current and inductively coupled currents are of the same

order of magnitude, and both must be considered to accurately predict the amount of cross talk. This is the regime of cross talk in transmission lines on circuit boards as part of a bus and the noise will have a special signature.

When the return path is not a wide uniform plane, but is a single lead in a package or a single pin in a connector, there is still capacitive and inductive coupling, but in this case, the inductively coupled currents are much larger than the capacitively coupled currents. In this regime, the noise behavior is dominated by the inductively coupled currents. The noise on the quiet line is driven by a dI/dt in the active net, which usually happens at the rising and falling edges of the signal when the driver switches. This is why this type of noise is usually referred to as switching noise.

These two extremes are considered separately.

10.3 Cross Talk in Transmission Lines: NEXT and FEXT

The noise between two adjacent transmission lines can be measured in the configuration shown in Figure 10-3. A signal is injected into one end of the line, with the far end terminated to eliminate the reflection at the end of the line. The voltage noise is measured on the two ends of the adjacent quiet line. Connecting the ends of the quiet line to the input channels of the fast scope will effectively terminate the quiet line. Figure 10-4 shows the measured voltage noise in a quiet line adjacent to an active signal line that is driven by a fast rising edge. In this case, the two 50-Ohm microstrip transmission lines are about 4 inches long, with a spacing about equal to their line width. The ends of each line are terminated in 50 Ohms, so the reflections are negligible.

The measured noise voltage has a very different pattern on each end. To distinguish the two ends, we label the end nearest the source *the near end* and the end farthest from the source *the far end*. The ends are also defined in terms of the

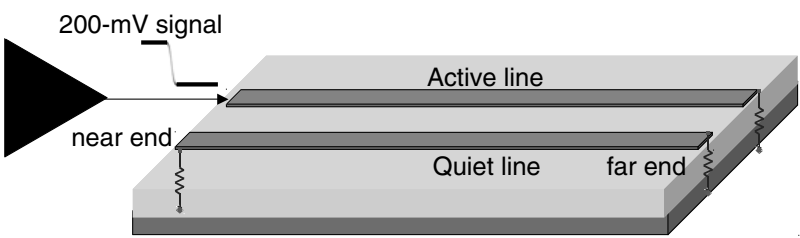


Figure 10-3 Configuration to measure the cross talk between an active and quiet net, looking on the near end and far end of the quiet line.

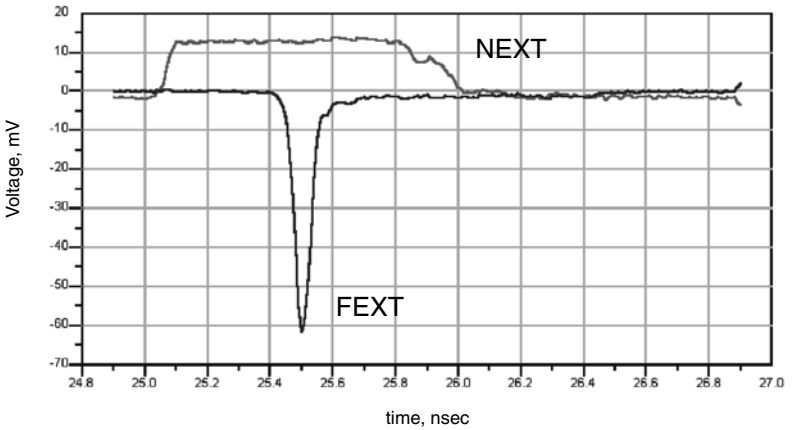


Figure 10-4 Measured noise on the quiet line when the active line is driven with a 200-mV, 50-psec rise-time signal. Measured with an Agilent DCA TDR and GigaTest Labs Probe Station.

direction the signal is traveling. The far end is in the *forward* direction to the signal propagation direction. The near end is in the *backward* direction to the signal propagation direction.

When the ends of the lines are terminated so multiple reflections do not play a role, the patterns of noise appearing at the near and far ends have a special shape. The near-end noise rises up quickly to a constant value. It stays up at this level for a time equal to twice the time delay of the coupling length and then drops down. We label the constant, saturated amount of near-end noise the *near-end cross talk* (or *NEXT*) coefficient. In the example shown above, the NEXT is about 13 mV. With an incident signal of 200 mV, the NEXT is about 6.5%.

The NEXT value is special in that it is defined as the near-end noise when the coupling length is long enough to reach the constant, flat value and in the special case of matched terminations. Changing the terminations at the end of the lines will not change the coupled noise into the quiet line. However, the backward traveling noise, when it hits the end of the line, may reflect if the termination is not matched. When the terminations are matched to the characteristic impedance of the line, the NEXT is a measure of the noise generated in the line. Once this is known, the impact on this voltage when it encounters a different termination can easily be estimated.

Obviously, the value of the NEXT will depend on the separation of the traces. Unfortunately, the only way of decreasing the NEXT is to move the traces farther apart.

The far end has a signature very different from the near end. There is no far-end noise until one time of flight after the signal enters the active line. Then it comes out very rapidly and lasts for a short time. The width of the pulse is the rise time of the signal. The peak voltage value is labeled as the *far-end cross talk* (or *FEXT*) coefficient. In the example above, the FEXT voltage is about 60 mV. This is with a signal of 200 mV, or a FEXT ratio of nearly 30%. This is a huge amount of noise.

If the terminations are not matched and reflections affect the magnitude of the noise appearing at the ends, we still refer to the far-end cross talk, but the magnitude is no longer labeled as FEXT. This coefficient is the special case when the terminations are matched.

TIP Three factors will decrease the FEXT: decreasing the coupling length, increasing the rise time, or moving the traces farther apart.

10.4 Describing Cross Talk

The most fundamental way of describing cross talk is in terms of the equivalent circuit model of the coupled lines. This model will allow simulations that will take into account the specific geometry and the terminations when predicting voltage waveforms. Two different models are generally used to model the coupling in transmission lines.

An ideal, distributed coupled transmission-line model for two lines describes a differential pair. The terms that describe the coupling are the odd- and even-mode impedances and the odd- and even-mode time delays. These four terms describe all the transmission-line and coupling effects. Many simulation engines, such as SPICE, especially those that have an integrated 2D field solver, use this type of model. The bandwidth of this model is as high as the bandwidth of an ideal lossless transmission line. This is the same model as a differential pair and is reviewed in detail in the next chapter.

An alternative, widely used model to describe coupling uses the n-section lumped-circuit-model approximation. In this model, each of the two transmission lines is described with an n-section lumped-circuit model and the coupling between them is described with mutual-capacitor and mutual-inductor elements. The equivalent circuit model of just one section is shown in Figure 10-5.

The actual capacitance and loop inductance between the signal and return paths and their mutual values are distributed uniformly down the length of the transmission lines. For uniform, coupled transmission lines, the per-length values

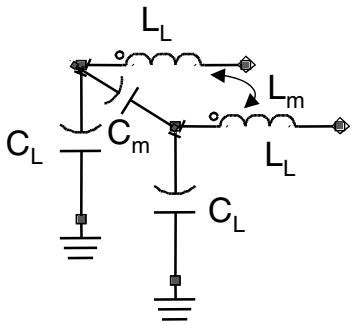


Figure 10-5 Equivalent circuit model of one section of an n-section coupled transmission-line model.

describe the transmission lines and the coupling. As shown below, these values can be displayed in a matrix, and this matrix formalism can be scaled and expanded to represent any number of coupled transmission lines. In some simulators, this matrix representation is the basis of describing the coupling, even though the actual simulation engine uses a true, distributed transmission line model.

We can approximate this distributed behavior by small, discrete lumped elements placed periodically down the length. The approximation gets better and better as we make the discrete lumped elements smaller. The number of sections needed, as shown in a previous chapter, depends on the required bandwidth and the time delay, with the minimum number being:

$$n > 10 \times BW \times TD \tag{10-1}$$

where:

n = the minimum number of lumped sections for an accurate model

BW = the required bandwidth of the model, in GHz

TD = the time delay of each transmission line, in nsec

Two coupled transmission lines can be described with two, independent, n-section lumped-circuit models. If the lines are symmetric, the L and C values in each segment will be the same for each line. To this uncoupled model, we need to add the coupling. In each section, the coupling capacitance can be modeled as a

capacitor between the signal paths. The coupling inductance can be modeled as a mutual inductor between each of the loop inductors in the n-section model.

Each single-ended transmission line is described by a capacitance per length, C_L, and a loop self-inductance per length, L_L. The coupling is described by a mutual capacitance per length, C_{ML}, and a loop mutual inductance per length, L_{ML}. For a pair of uniform transmission lines, the mutual capacitance and mutual inductance are distributed uniformly down the two lines.

> **TIP** Everything about the two coupled transmission lines can be described by these four line parameters. When there are more than two transmission lines, the model can be scaled directly, but looks more complicated. Between every pair of sections of the transmission lines, there is a mutual capacitor. Between every pair of signal- and return-loop sections, there is a mutual inductor.

Each of the mutual capacitors and loop mutual inductors scale with length, and we refer to their mutual capacitance and mutual inductance per length. To keep track of each of these additional mutual capacitors and inductors, we can take advantage of a simple formalism based on matrices.

10.5 The SPICE Capacitance Matrix

For a collection of multiple transmission lines, we can label each signal path with an index number. If there are five lines, for example, we would label each one from 1 to 5. The return path, by convention, we label as conductor 0. An example of the cross section of five conductors and a common return plane is shown in Figure 10-6. We first look at the capacitor elements. In a later section, we look at the inductor elements.

Every pair of conductors in the collection has a capacitance between them. For every signal line, there is a capacitor to the return path. Between every pair of signal lines, there is a coupling capacitor. To keep track of all the pairs, we can

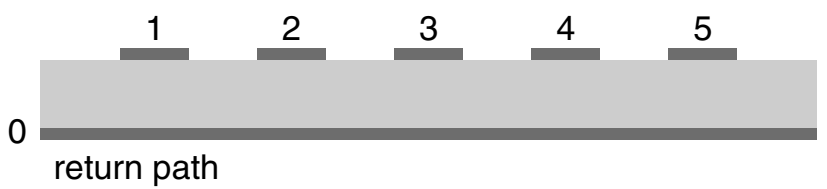


Figure 10-6 Five coupled transmission lines, in cross section, with each conductor labeled using the index convention.

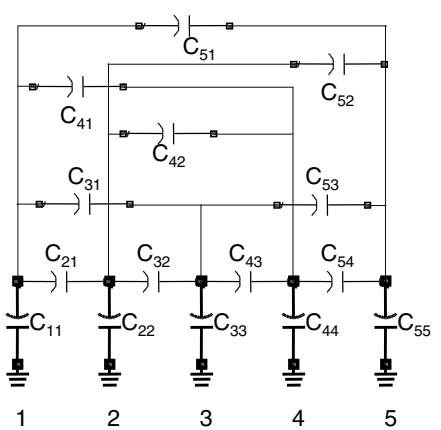
label the capacitors based on the index numbers. The capacitance between conductors 1 and 2 is labeled as C_{12}, and between capacitors 2 and 4 is labeled as C_{24}. The capacitance between the signal line and the return, we might label as C_{10} or C_{30}.

To take advantage of the powerful formalism of matrix notation, we rename the capacitor labels that describe the capacitance between the signal paths and the return paths. Instead of C_{10}, we reserve the diagonal element for the capacitance between the signal and its return path and label it as C_{11}. Likewise, the other capacitors between the signals and their returns become C_{22}, C_{33}, C_{44}, and C_{55}. In this way, we end up with a 5×5 matrix of capacitors that labels the capacitance between every pair of conductors. The equivalent circuit and corresponding matrix of parameter values are shown in Figure 10-7.

Of course, even though there is a matrix entry for C_{14} and C_{41}, it is the same capacitor value, and there is only one instance of this capacitor in the model.

TIP In the capacitor matrix, the diagonal elements are the capacitance between the signal and the return paths. The off-diagonal elements are the coupling or mutual capacitance. For uniform transmission lines, each matrix element is the capacitance per length, usually in units of pF/inch.

The matrix is a handy, convenient, compact way of keeping track of all the capacitor values. This matrix is often called the SPICE capacitance matrix to distinguish it from some of the other matrices. It is a place to store the parameter val-



$$\begin{matrix} C_{11} & C_{12} & C_{13} & C_{14} & C_{15} \\ C_{21} & C_{22} & C_{23} & C_{24} & C_{25} \\ C_{31} & C_{32} & C_{33} & C_{34} & C_{35} \\ C_{41} & C_{42} & C_{43} & C_{44} & C_{45} \\ C_{51} & C_{52} & C_{53} & C_{54} & C_{55} \end{matrix}$$

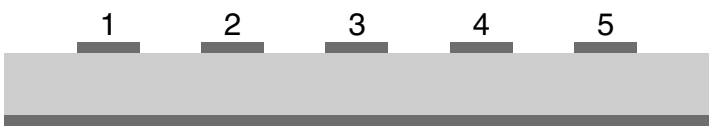
Figure 10-7 Equivalent capacitance model of five coupled transmission lines and the corresponding matrix of capacitance parameter values.

ues for the SPICE equivalent circuit model, shown above. Each matrix element represents the value of a capacitor that would be present in the complete circuit model for the coupled transmission lines.

Each element is the capacitance per length. To construct the actual transmission lines' approximate model, we would first identify how many sections are needed in the lumped-circuit model, from $n > 10 \times BW \times TD$. From the length of the transmission lines and the number of sections required, the length of each section can be calculated as Length per section = Len/n. The value of the capacitor for each section is the matrix element of the capacitance per length times the length of each section. For example, the coupling capacitance of each section would be $C_{21} \times Len/n$.

The actual values for each capacitor-matrix element can be found by either calculation or by measurement. There are few approximations that are very good. Rather, a few simple rules of thumb can be used and when a more accurate value of the coupling capacitance is required, a 2D field solver should be used. Many field solver tools are commercially available; they are easy to use, and generally very accurate. An example of the SPICE capacitance matrix for a collection of five microstrip conductors, as calculated with a 2D field solver, is shown in Figure 10-8.

It is sometimes difficult to get a good physical feel for the values of the capacitance-matrix elements and how quickly they drop off by just looking at the numbers. Instead, the matrix can be plotted in 3D. The vertical axis is the magnitude of the capacitance. These same matrix elements are shown in Figure 10-9. At a



2.812	0.151	0.016	0.008	0.005
0.151	2.682	0.149	0.016	0.008
0.016	0.149	2.675	0.149	0.017
0.008	0.016	0.149	2.684	0.151
0.005	0.008	0.017	0.151	2.813

Figure 10-8 Five coupled transmission lines, each of 5-mil line width and 5-mil space and the SPICE capacitance matrix, in pF/inch, calculated with the Ansoft SI2D field-solver tool.

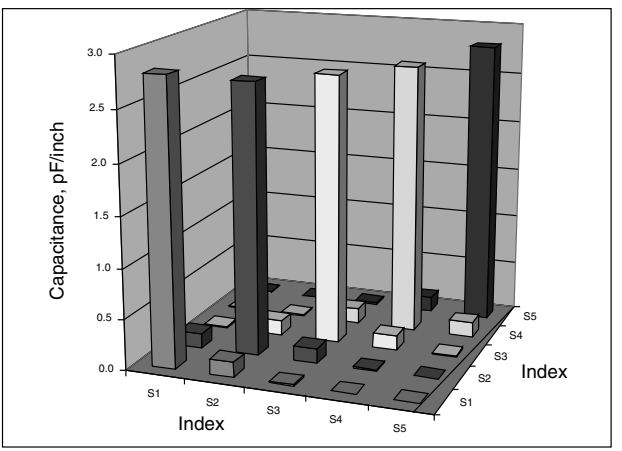


Figure 10-9 Plotting the SPICE capacitance-matrix elements showing how quickly off-diagonal elements drop off.

glance, it is apparent that all the diagonal elements have about the same values and the off-diagonal elements drop off very fast.

In this particular example, the conductors are 50-Ohm microstrips, with 5-mil line width and 5-mil space, placed as close together as possible. We see the coupling between conductors 1 and 3 is negligible compared to that between 1 and 2. The farther apart the traces, the more rapidly the off-diagonal elements drop off.

In the SPICE capacitance matrix, it is important to keep in mind that each element is the parameter value of a circuit element that appears in an equivalent circuit model. The value of each element is a direct measure of the amount of capacitive coupling between the two conductors. This will directly determine, for example, the capacitively coupled current that would flow between each pair of conductors for a given dV/dt. The larger the matrix element, the larger the capacitive coupling, and the more fringe fields between the conductors.

T I P In coupled transmission lines, the size of the off-diagonal element should always be compared with the diagonal element. In this geometry example of five coupled 50-Ohm lines, with a spacing equal to the line width (the tightest spacing manufacturable), the relative coupling between adjacent traces is about 5%. The coupling between one trace and the trace two traces away is down to less than 0.6%. These are good values to remember as a rough rule of thumb.

The physical configuration of the traces will affect the parameter values, but for a given configuration of traces, the circuit model itself will not change as the geometry of the traces changes. Obviously, if we move the traces farther apart, the parameter values will decrease.

If we change the line width of one line, to first order, it will affect the diagonal element of that line and the coupling between that line and the adjacent traces on either side. However, it may also affect, to second or third order, the coupling between the lines on either side of it. The only way to know for sure is to put in the numbers with a 2D field solver.

10.6 The Maxwell Capacitance Matrix and 2D Field Solvers

Unfortunately, there is more than one capacitance matrix and this creates confusion. Above, we introduced the SPICE capacitance matrix whose elements were the parameter values of the equivalent circuit model for the coupled lines. There is also a capacitance matrix that is the result of a field-solver calculation. This is referred to as the *Maxwell capacitance matrix*. Even though they are both called capacitance matrices, their definitions are different.

A field solver is basically a tool that solves one or more of Maxwell's Equations for a specific set of boundary conditions. A circuit topology for the collection of conductors is assumed and from the fields, all the parameter values are calculated. The equation that is solved to calculate the capacitances of an array of conductors is LaPlace's Equation. In its simplest differential form, it is:

$$\nabla^2 V = 0 \tag{10-2}$$

This differential equation is solved under the specific geometry boundary conditions of the conductors and dielectric materials. Solving this equation will allow the calculation of the electric fields at every point in space.

For example, suppose there is a collection of five conductors, as shown in Figure 10-10. Conductor 0 is defined as the ground reference and is always at 0-v potential. To calculate the capacitance between each conductor, there are six steps.

1. A 1-v potential is set on conductor k and the potential of every other conductor is set to 0 v.
2. Given this boundary condition, LaPlace's Equation is solved to find the potential everywhere in space.

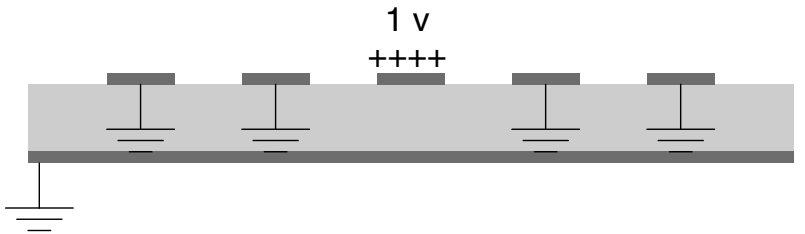


Figure 10-10 Conductor set up to calculate the capacitance matrix for this collection of transmission lines using a 2D field solver that solves LaPlace's Equation.

3. Once the potential is solved, the electric field is calculated at the surface of each conductor, from

$$E = -\nabla V \tag{10-3}$$

4. The total charge is calculated on each conductor by integrating the electric field on the surface of each conductor:

$$Q_j = \oint_j E \bullet da_j \tag{10-4}$$

5. From the charge on each conductor, the capacitance is calculated from the definition of the Maxwell Capacitance matrix:

$$C_{jk} = \frac{Q_j}{V_k} \tag{10-5}$$

6. This process is repeated with a 1-v potential sequentially placed on each of the conductors.

The definition of the Maxwell capacitance-matrix elements is different from the SPICE capacitance-matrix elements. The SPICE matrix elements are the parameter values for the corresponding equivalent circuit model. The value of each element is a direct measure of the amount of capacitively coupled current that would flow between each pair of conductors for a given dV/dt between them.

The Maxwell capacitance-matrix elements are really defined based on:

$$C_{jk} = \frac{Q_j}{V_k} \tag{10-6}$$

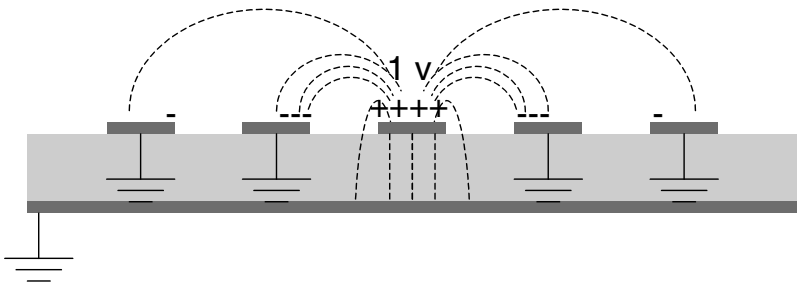


Figure 10-11 Charge distribution for the five conductors with conductor 3 set to 1 v and all others at ground potential.

Each capacitance-matrix element between two conductors is a measure of how much excess charge will be on one conductor when the other is at a 1-v potential and *all other conductors* are grounded. This is a very specialized condition and gives rise to much confusion.

Suppose a 1-v potential is placed on conductor 3 and all the other conductors are at 0-v potential. To do this will require placing some extra plus charge on conductor 3 to raise it to a 1-v potential with respect to ground. This plus charge will attract some negative charge to all the other nearby conductors. The negative charge will come out of the ground reservoir to which each other conductor is connected. How much charge is attracted to each of the other conductors is a measure of how much capacitive coupling there is to the conductor with the 1 v applied. The charge distribution is illustrated in Figure 10-11.

From the definition of the Maxwell capacitance matrix, the capacitance between conductor 3 and the reference ground, which is the diagonal element, is the ratio of the charge on conductor 3, Q_3, and the voltage on it, $V_3 = 1$ v. This is the charge on conductor 3 when all the other conductors are also connected to ground. This capacitance is often called the loaded capacitance of conductor 3:

$$C_{33} = \frac{Q_3}{V_3} = C_{loaded} \tag{10-7}$$

The loaded capacitance will always be larger than the SPICE diagonal capacitance.

When there is plus charge on conductor 3 to raise it up to the 1-v potential, the charge induced on all the other conductors is negative. Even though all the

other conductors are at ground potential, they will have some net negative charge due to the coupling to the conductor with the 1 v.

TIP The diagonal elements of the Maxwell capacitance matrix are the loaded capacitances of each conductor. It is not just the capacitance to the return path, the ground reference, it is the capacitance to the return path *and* to all the other conductors that are also tied to ground. This is not the same as the diagonal element of the SPICE capacitance matrix, which just includes the coupling of the conductor to the return path and not to any other signal path.

By the definition of the Maxwell capacitance matrix, the off-diagonal-matrix element between conductor 3 and 2 will be:

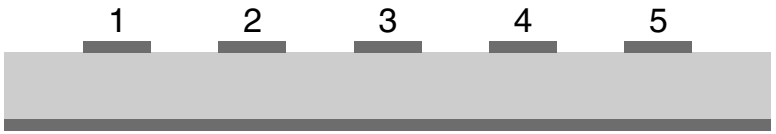
$$C_{23} = \frac{Q_2}{V_3} \tag{10-8}$$

TIP Since the charge induced on conductor 2 is negative and the induced charge on every other conductor is also negative, every off-diagonal capacitor-matrix element must be negative.

The negative sign really means that the induced charge on conductor 2 will be negative when a 1-v potential is placed on conductor 3.

An example of the Maxwell capacitance matrix for a collection of five microstrip traces is shown in Figure 10-12. At first glance, it is bizarre to see capacitance values that are negative. What could it possibly mean to have negative capacitances? Is this an inductance? In fact, they are negative because they are not SPICE capacitance values, they are Maxwell capacitance values, and the definition of the Maxwell capacitor-matrix elements is different from the SPICE elements.

The output from most commercially available field solvers is typically in the form of Maxwell capacitance values. This is usually because the software developer who wrote the code did not really understand the end user's applications and did not realize that what most signal-integrity engineers want to see are the SPICE capacitance-matrix elements. The Maxwell capacitance matrix is not wrong, it is just not the first choice of what an engineer wants to see.



2.992	-0.151	-0.016	-0.008	-0.005
-0.151	3.005	-0.149	-0.016	-0.008
-0.016	-0.149	3.006	-0.149	-0.017
-0.008	-0.016	-0.149	3.007	-0.151
-0.005	-0.008	-0.017	-0.151	2.994

Figure 10-12 Maxwell capacitance matrix, in pF/inch, for the collection of five closely spaced, 50-Ohm transmission lines calculated with Ansoft's SI2D field solver.

It is very easy to convert from one matrix to the other. The off-diagonal elements are very similar, with just the sign difference:

$$C_{ij}(\text{SPICE}) = -C_{ij}(\text{Maxwell}) \tag{10-9}$$

The off diagonals relate to the number of field lines that couple the two conductors and directly relate to the capacitively coupled current that might flow between the conductors for a given dV/dt between them.

However, the diagonals are a little more complicated. The diagonal elements of the Maxwell matrix are the loaded capacitance of each conductor. The diagonal elements of the SPICE matrix are the capacitance between just the diagonal conductor and the return path. The SPICE diagonal element is counting only the field lines coupling between the signal line to the return path. Based on this comparison, diagonal elements of the SPICE and Maxwell matrices can be converted by:

$$C_{jj}(\text{Maxwell}) = \sum_{i} C_{ij}(\text{SPICE}) \tag{10-10}$$

$$C_{jj}(\text{SPICE}) = \sum_{i} C_{ij}(\text{Maxwell}) \tag{10-11}$$

The easiest way to determine which matrix is reported by the field solver is to look for negative signs. If there are negative signs, it's usually not numerical accuracy, it is the Maxwell capacitance matrix.

In either matrix, the off-diagonal elements are a direct measure of the coupling between signal lines and the strength of the fringe fields that couple the conductors. The greater the spacing, the fewer the fringe-field lines between the traces and the lower the coupling. In both matrices, the physical presence of any conductor between two traces will affect the field lines between them and will be taken into account by the matrix-element values.

Each matrix element will depend on the presence of the other conductors. For example, for two conductors and their return path, the diagonal SPICE capacitance of one line, C_{11}, will depend on the position of the adjacent conductor. C_{11} is the capacitance of line 1 to the return path. If we bring the adjacent trace in proximity, it will begin to steal some of the fringe-field lines between line 1 and the return and decrease C_{11}. This is illustrated in Figure 10-13.

When the spacing is more than about two line widths or four dielectric thicknesses, the presence of the adjacent trace has very little impact on the diagonal element of the SPICE capacitance matrix.

Since the loaded capacitance of line 1 is a measure of all the fringe-field lines between the signal line and all the other conductors, it doesn't change much

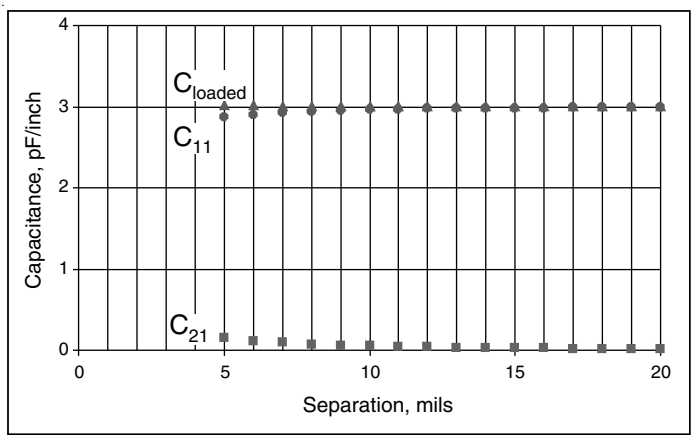


Figure 10-13 Variation of the diagonal and off-diagonal elements of the SPICE capacitance matrix and the loaded capacitance of conductor 1 as the spacing between the two 5-mil-wide, 50-Ohm lines increases. Simulated with Ansoft's SI2D.

as the adjacent trace is brought closer. Field lines not going from 1 to the return, stolen by trace 2, are accounted for by the new field lines between 1 and 2.

The off-diagonal elements will also depend on the geometry and the presence of other conductors. If the spacing is increased, the off-diagonal element will decrease. Also, if another conductor is added between two traces, this conductor will steal some of the field lines between the two conductors and decrease the off-diagonal SPICE capacitance element.

Figure 10-14 shows three geometry configurations and the resulting SPICE capacitance-matrix element that is calculated. In each case, the signal line is 5 mils wide and roughly 50 Ohms.

TIP When the spacing is also 5 mils wide, the coupling capacitance is 0.155 pF/in. This is about 5% of the on-diagonal element of 2.8 pF/inch. When the spacing is increased to 15 mils, so the spacing is three times the width, the capacitive coupling is 0.024 pF/inch, or 0.9% of the diagonal element. If another 5-mil-wide trace is added in this space, the coupling capacitance between the two outer conductors is reduced to 0.016 pF/inch, or 0.6% of the on-diagonal element.

By adding a conductor between the two signal lines the coupling capacitance between them is reduced. This is the basis of the use of guard traces, discussed in detail later in this chapter. Of course, we have only considered one type of coupling; there is also inductive coupling to consider.

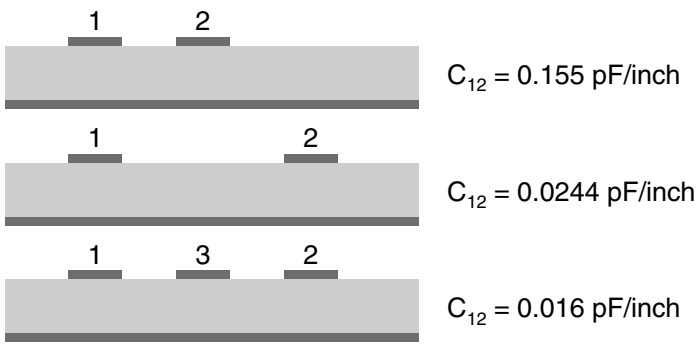


Figure 10-14 Three geometries and the corresponding capacitance-matrix elements between the two signal lines. The presence of the metal between the conductors decreases the capacitive coupling about 35%.

10.7 The Inductance Matrix

Just as a matrix is used to store all the capacitor values in a collection of signal- and return-path conductors, a matrix is used to store the values of the loop self-inductance and the loop mutual inductances associated with a collection of conductors. It is important to keep in mind that the inductance elements are loop inductances. As the signal propagates down a transmission line, the current loop travels down the signal path and immediately returns through the return path. This current loop is probing the loop inductance in the immediate vicinity of the edge of the signal transition. Of course, the loop self-inductance is related to the partial self-inductances of the signal and return paths, and their mutual inductance, by:

$$L_{loop} = L_{self\text{-}signal} + L_{self\text{-}return} - 2 \times L_{mutual} \tag{10-12}$$

where:

L_{loop} = the loop inductance per length of the transmission line

$L_{self\text{-}signal}$ = the partial self-inductance per length of the signal path

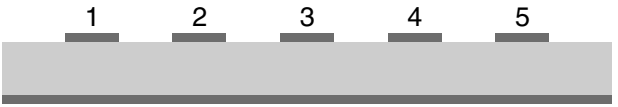
$L_{self\text{-}return}$ = the partial self-inductance per length of the return path

L_{mutual} = the partial mutual inductance per length between the signal and return paths

In the inductance matrix, the diagonal elements are the loop self-inductances of each signal and return path. The off-diagonal elements are the loop mutual inductance between every pair of signal and return paths. The units are in inductance per length, typically nH/inch.

For example, the collection of five microstrip traces above have an inductance matrix shown in Figure 10-15. When plotted in 3D, the loop-inductance matrix reveals the basic properties of inductance. The diagonal elements, the loop self-inductances of each conductor and its return path, are all basically the same. The off-diagonal elements, the loop mutual inductances, drop off very rapidly the farther apart the pair.

The combination of the capacitance and inductance matrices contain all the information about the coupling between a collection of transmission lines. From these values, all aspects of cross talk between two or more transmission lines can be calculated. A SPICE equivalent circuit model can be built that could be used to simulate the behavior of a collection of coupled traces.



7.517	1.012	0.292	0.137	0.088
1.012	7.464	0.996	0.287	0.138
0.292	0.996	7.456	0.995	0.296
0.137	0.287	0.995	7.454	1.008
0.088	0.138	0.296	1.008	7.496

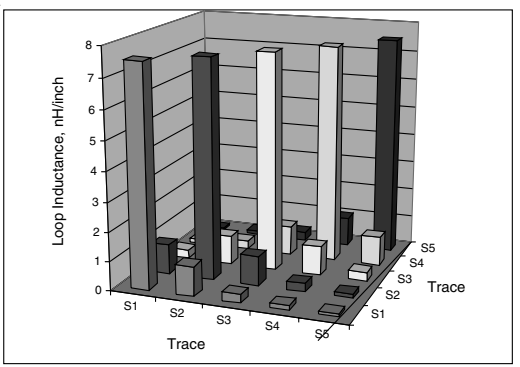


Figure 10-15 Five transmission lines, each 50 Ohms and with a 5-mil width and 5-mil spacing, their inductance matrix and the values of each matrix element as extracted using the Ansoft SI2D field solver.

TIP These two matrices contain all the fundamental information about the coupling between multiple transmission lines.

10.8 Cross Talk in Uniform Transmission Lines and Saturation Length

For the case of two coupled transmission lines, the C and L matrices are simple, two-by-two matrices. The off-diagonal elements describe the amount of mutual capacitance and mutual inductance. The easiest way to understand the generation of the noise in the quiet line and the particular near-end and far-end signature is to walk down the line and observe the noise coupling over at each step.

Consider two 50-Ohm microstrip transmission lines that have some coupling distributed down their length. In addition, we will terminate the ends of the lines in their characteristic impedance of 50 Ohms to eliminate any effects from reflections. This equivalent circuit model is illustrated in Figure 10-16.

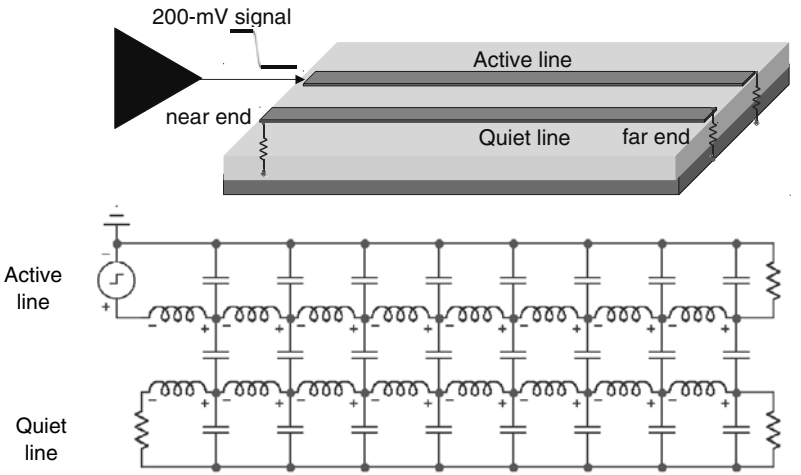


Figure 10-16 A pair of tightly coupled transmission lines and an equivalent circuit model using an n-section lumped-circuit approximation.

As a signal propagates down the active line, it will see the mutual capacitors and mutual inductors connecting it to the quiet line. The only way noise current will flow from the active to the quiet line is through these elements, and the only way current flows through a capacitor, or is induced in a mutual inductor, is if either the voltage or current changes.

TIP As the signal propagates down the active line, the only place there is coupled-noise current to the quiet line is in the specific region where the edge of the signal is, where there is a dV/dt or a dI/dt. Everywhere else along the line, the voltage and current are constant and there is no coupled-noise current.

This is illustrated in Figure 10-17. If the leading edge is approximated by a linear ramp with a rise time of RT, the noise is approximately proportional to V/RT and I/RT.

The edge of the signal acts as a current source moving down the line. At any instant, the total current that flows through the mutual capacitors is:

$$I_C = C_m \frac{dV}{dt} \tag{10-13}$$

where:

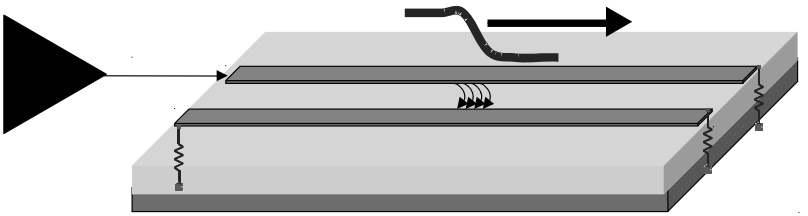


Figure 10-17 The only region in which coupled noise flows from the active line to the quiet line is at the signal wavefront where the voltage and current change.

I_C = the capacitively coupled-noise current from the active line to the quiet line

V = the signal voltage

C_m = the mutual capacitance that couples over the length of the signal rise time

The total capacitance that is coupling is the capacitance along the spatial extent of the rise time:

$$C_m = C_{mL} \times \Delta x = C_{mL} \times v \times RT \tag{10-14}$$

where:

C_m = the mutual capacitance that couples over the length of the signal rise time

C_{mL} = the mutual capacitance per length (C_{12})

Δx = the spatial extent of the leading edge as it propagates over the active line

v = the signal propagation speed

RT = the signal rise time

The total, instantaneous, capacitively coupled current injected into the quiet line is:

$$I_C = C_{mL} \times v \times RT \times \frac{V}{RT} = C_{mL} \times v \times V \tag{10-15}$$

where:

I_C = the capacitively coupled-noise current from the active line to the quiet line

C_{mL} = the mutual capacitance per length (C_{12})

v = the signal-propagation speed

RT = the signal rise time

V = the signal voltage

The capacitively coupled current from the active line is injected into the quiet line locally only where the edge of the signal is in the active line. Surprisingly, the coupled-noise current has a total magnitude that is independent of the rise time. The faster the rise time, the larger the dV/dt, so we would expect a larger amount of capacitively coupled current. But, the faster the rise time, the shorter the region of coupled line where there is a dV/dt and the less capacitance is available to do the coupling. The capacitively coupled current depends only on the mutual capacitance per length.

By the same analysis, the instantaneous voltage induced in the mutual inductor in the quiet line is:

$$V_L = L_m \frac{dI}{dt} = L_{mL} \times v \times RT \times \frac{I}{RT} = L_{mL} \times v \times I \qquad (10\text{-}16)$$

where:

V_L = the inductively coupled-noise voltage from the active line to the quiet line

I = the signal current in the active line

L_{mL} = the mutual inductance per length (L_{12})

v = the signal-propagation speed

RT = the signal rise time

Again, we see that the inductively coupled noise crosses to the quiet line only where the signal voltage is changing on the active line. Also, the amount of noise voltage generated in the quiet line does not depend on the rise time of the signal, only on the mutual inductance per length.

Four important properties emerge about the coupled noise to the quiet line:

1. The amount of instantaneous coupled voltage and current noise depends on the signal strength. The larger the signal voltage and current, the larger the amount of instantaneously coupled noise.
2. The amount of instantaneously coupled voltage and current noise depends on the amount of coupling per length as measured by the mutual capacitance and mutual inductance per length. If the coupling per length increases as the conductors are brought closer together, the instantaneously coupled noise will increase.
3. It appears that the higher the velocity, the higher the instantaneously coupled total current. This is due to the fact that the higher the speed, the longer the spatial extent of the rise time and the longer the region that we will see coupling at any one instant. If the velocity of the signal increases, the coupled length over which current flows will increase and the total coupling capacitance or inductance will increase.
4. Surprisingly, the rise time of the signal does not affect the total instantaneously coupled-noise current or voltage. While it is true that the shorter rise time will increase the coupled noise through a single mutual C or L element, with a shorter rise time, the spatial extent of the edge is shorter, and there is less total mutual C and mutual L that couple at any one instant.

This last property is based on an assumption that the length of the coupled region is longer than half the spatial extent of the rise time. This is the most confusing and subtle aspect of near-end noise.

Consider a pair of coupled lines with a TD of the coupling region very long compared to the rise time of the signal. As the signal starts out from the driver and enters the coupling region, the amount of coupled noise flowing between the aggressor and victim lines will begin to increase and appear as increasing near-end noise. The near-end noise will continue to increase as long as more rising edge enters the coupling region. The near-end noise increases for a time equal to the rise time. After this period of time, the near-end noise has reached its maximum value and has "saturated."

When the beginning of the leading edge of the signal finally leaves the coupled region, the coupled current flowing from the aggressor line to the victim line will begin to decrease. Of course, it will take one TD of the coupled region for the beginning of the rising edge to traverse the coupled region. Once the coupled current begins to decrease, at the far end of the line, it will take another TD for this

reduced current to make it back to the near end and be recorded as reduced near-end noise. The near-end noise will begin to decrease in a time equal to 2 × TD from the very beginning of the near-end noise starting.

As the coupling TD between the two transmission lines is decreased, there will reach a point where the near-end noise reaches its peak value, a rise time after it begins, just as it begins to decrease, 2 × TD after it begins. This condition is that the rise time = 2 × TD. This is the condition for the coupling length being just long enough to saturate the near-end noise.

When the rise time of the signal is 2 × TD, the coupled lines are saturated. This condition is when the TD is half the rise time, or when the coupling length is half the spatial extent of the rising edge. We give this length the special name of *saturation length*:

$$\text{Len}_{\text{sat}} = \frac{1}{2} \times \text{RT} \times \text{v} \sim \text{RT} \times 3 \frac{\text{inch}}{\text{nanosecond}} \tag{10-17}$$

where:

Len_{sat} = the saturation length for near-end cross talk, in inches

RT = the rise time of the signal in nsec

v = the speed of the signal down the active line in inches/nsec

If the rise time is 1 nsec, in a transmission line composed of FR4, with a velocity of roughly 6 inches/nsec, the saturation length is 1/2 nsec × 6 inches/nsec = 3 inches. If the rise time were 100 psec, the saturation length would be only 0.3 inch. For short rise times, the saturation length is usually shorter than a typical interconnect length, and near-end noise is independent of coupled length. The saturation length is illustrated in Figure 10-18.

Once the noise current transfers from the active line to the quiet line, it will propagate in the quiet line and give rise to the effects we see as near-end and far-end noise. Even though a constant current is transferring to the quiet line, the features of the propagation in the quiet line will shape this distributed-current source into very different patterns at the near and far ends. To understand the details of the origin of the near- and far-end signatures, we will first look at how the capacitively coupled currents behave at the two ends, then at the inductively coupled currents and add them up.

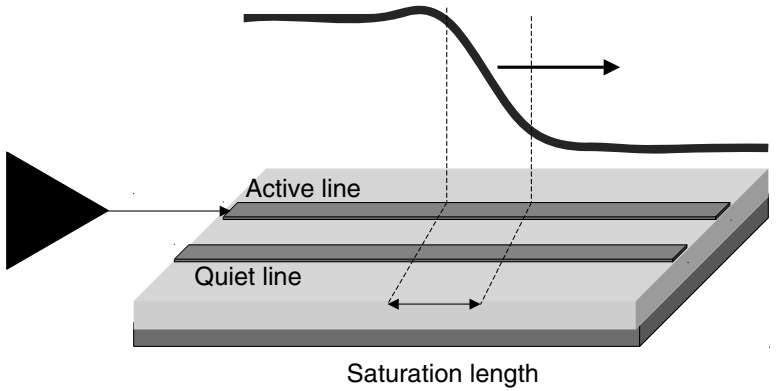


Figure 10-18 The saturation length is half the spatial extent of the leading edge. If the length of the coupled region is longer than the saturation length, the amount of near-end noise on the quiet line is independent of the rise time and independent of the coupling length.

10.9 Capacitively Coupled Currents

Figure 10-19 shows the redrawn equivalent circuit model with just the mutual-capacitance elements. In this example, we assume the coupled length is longer than the saturation length. The rising edge will act as a current source moving down the active line. Because current flows through the mutual capacitors only when there is a dV/dt, it is only where the rising edge is that there is capacitively coupled current flowing into the quiet line.

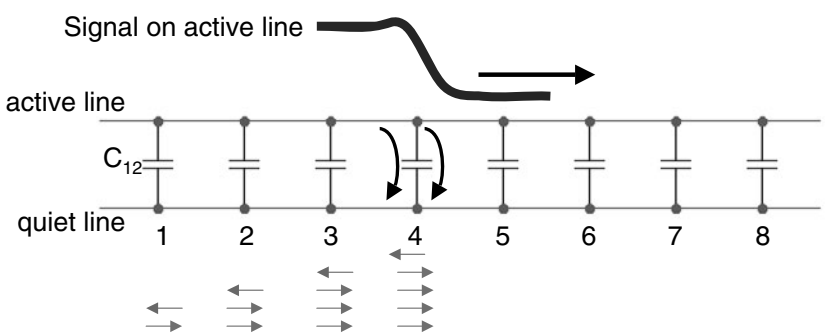


Figure 10-19 Equivalent circuit model of two coupled lines just showing the coupling capacitors, the coupled current, and the spatial extent of the signal edge.

Once this current appears in the quiet line, which way will it flow? The primary factor that will determine the direction of the current flow is the impedance the noise current sees. When the noise current looks up and down the quiet line, it sees exactly the same impedance in both directions: 50 Ohms. An equal amount of noise current will flow in both the forward and backward directions.

TIP The direction of the capacitively coupled-current loop in the quiet line is from the signal path to the return path. It is a positive voltage between the signal and return paths of the quiet line that will propagate in both directions.

As the signal initially emerges from the driver, there will be some capacitively coupled current to the quiet line. Half of this will travel backward to the near end. The other half will travel in the forward direction. The current, flowing through the terminating resistor on the near end of the quiet trace, will flow in the positive direction, from the signal path to the return path. It will start out at 0 v and as the rising edge emerges from the driver, it will rise up. As the signal edge moves down the line, the backward-flowing capacitively coupled-noise current will continue back to the near end, at a steady rate. It is as though the active signal is leaving a constant, steady amount of current in its wake.

After a time equal to the rise time, the current appearing at the near end will reach its peak value. After the beginning of the rising edge in the active line has left the coupled region and reached the far-end terminating resistor, the coupled-noise current will begin to decrease, taking a time equal to the rise time. There is still the backward-moving current in the quiet line that has yet to reach the near end. It will continue flowing back to the near end of the quiet line, taking additional time equal to the time delay, TD, of the coupled region.

The signature of the near-end, capacitively coupled current is a rise up to a constant value in a time equal to the signal's rise time and staying at a constant value lasting for a time equal to 2 × TD – 2 × rise time, and then falling to zero. This is illustrated in Figure 10-20.

The magnitude of the saturated, capacitively coupled current at the near end will be:

$$I_C = \frac{1}{2} \times \frac{1}{2} \times C_{mL} \times v \times V = \frac{1}{4} \times C_{mL} \times v \times V \qquad (10\text{-}18)$$

where:

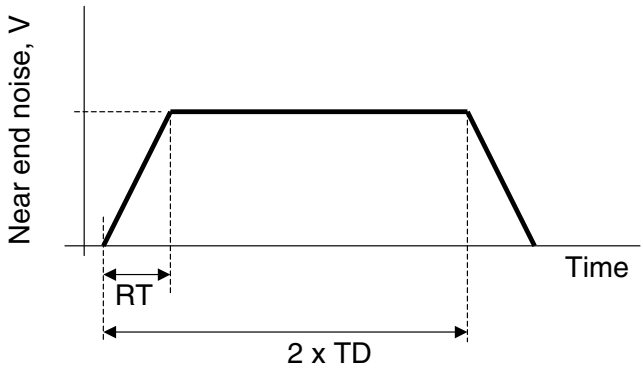


Figure 10-20 Typical signature of the capacitively coupled voltage at the near end of the quiet line, through the terminating resistor.

I_C = the capacitively coupled, saturated noise current at the near end of the quiet line

C_{mL} = the mutual capacitance per length (C_{12})

v = the signal-propagation speed

V = the signal voltage

1/2 factor = comes from half the current going to the near end and the other half to the far end

1/2 factor = accounts for the backward-flowing noise current spread out over $2 \times TD$

The second factor of 1/2 comes from the fact that the current source is moving in the forward direction, while the induced current is moving in the backward direction. In every short interval of time, a total amount of charge is transferred into the quiet line, which is moving in the backward direction, but over a spatial extent that is expanding in both directions. The total current, which is the charge that flows past a point per unit time, is spread over two units of time.

While half the capacitively coupled-noise current is flowing backward to the near end, the other half of the capacitively coupled-noise current is moving in the forward direction. The forward current in the quiet line is moving to the far end at exactly the same speed as the signal edge is moving to the far end in the active line. At each step along its path, half of it is added to the already present noise moving in the forward direction. It is as though the forward moving capacitively

current were growing like a snowball down a hill, building up more and more each step along its way.

At the far end, no current is present until the signal edge reaches the far end. Coincident with the signal hitting the far end, the forward-moving, capacitively coupled current reaches the far end. This current is flowing from the signal path to the return path. Through the terminating resistor across the quiet line, the voltage drop will be in the positive direction.

Since the capacitively coupled current to the quiet line scales with dV/dt, the actual noise profile in the quiet line, moving to the far end, will be the derivative of the signal edge. If the signal edge is a linear ramp, the capacitively coupled-noise current will be a short rectangular pulse, lasting for a time equal to the rise time. The capacitively induced noise signature at the far end of the quiet line is illustrated in Figure 10-21.

The total amount of current that couples over from the active to the quiet line will be concentrated in this narrow pulse. The magnitude of the current pulse, translated into a voltage by the terminating resistor, will be:

$$I_C = \frac{1}{2} \times C_{mL} \times Len \times \frac{V}{RT} \tag{10-19}$$

where:

I_C = the total capacitively coupled-noise current from the active line to the quiet line

1/2 factor = the fraction of capacitively coupled current moving to the far end

C_{mL} = the mutual capacitance per length (C_{12})

RT = the signal rise time

V = the signal voltage

TIP This says the magnitude of the capacitively coupled current at the far end scales directly with the mutual capacitance per length and with the coupled length of the pair of lines, and it scales inversely with the rise time. A shorter rise time will increase the far-end noise.

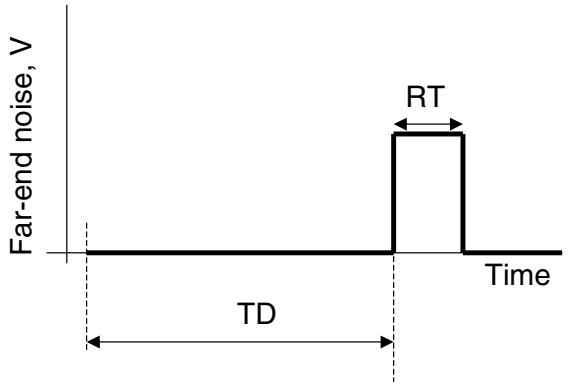


Figure 10-21 Typical signature of the capacitively coupled voltage at the far end of the quiet line, through the terminating resistor.

Unlike the backward propagating noise, the far-end propagating noise voltage will scale with the length of the coupled region and scale inversely with the rise time of the signal. The current direction for the forward propagating capacitively coupled current will be in the positive direction, from the signal line to the return path, therefore generating a positive voltage across the terminating resistor.

10.10 Inductively Coupled Currents

The inductively coupled currents behave in a similar way to the capacitively coupled currents. These currents are driven by a dI/dt in the active line through the mutual inductor, which creates a voltage in the quiet line. The noise voltage induced in the quiet line will see an impedance and will drive an associated current.

The changing current in the active line is moving from the signal to the return path, propagating down the line. If the direction of propagation is from left to right, the direction of the current loop is clockwise, as a signal-return path loop. This is also the direction of the increasing current, the dI/dt in the active line, a clockwise circulating loop. This changing current loop will ultimately induce a current loop in the quiet line. But in what direction will be the induced current loop? Will it be in the same direction as the signal current loop, the clockwise direction, or the opposite direction, the counterclockwise direction?

The direction of the induced current loop is based on the results of Maxwell's Equations. While it is tedious to go through the analysis to determine the direction of the induced current, it is easy to remember based on Lentz's law. This states that the direction of the induced current loop in the quiet line will be in the opposite direction of the inducing current loop in the active line.

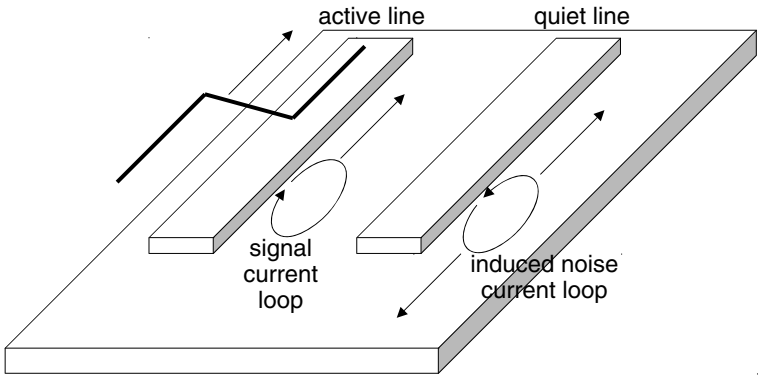


Figure 10-22 A dI/dt in the active line induces a voltage in the quiet line, which in turn creates a dI/dt in the quiet line. Half of the current loop will propagate in each direction in the quiet line.

The direction of the induced current loop in the quiet line will be circulating in the counterclockwise direction, located in the quiet line right where the signal edge is in the active line. This is illustrated in Figure 10-22.

Once this counterclockwise current loop is generated in the quiet line, which direction will it propagate? Looking up and down the line, it will see the same impedance, so it will propagate equal amounts of current in the two directions. This is a very subtle and confusing point. Half of the current in the induced current loop in the quiet line will propagate back to the near end. The other half of the current in the induced current loop in the quiet line will propagate in the forward direction.

TIP Moving in the backward direction, the counterclockwise current loop will be flowing from the signal path to the return path. This is the same direction the capacitively coupled currents flow. At the near end, the capacitively and inductively coupled-noise currents will add together.

TIP Moving in the forward direction, the counterclockwise current loop in the quiet line is flowing from the return path up to the signal path. The capacitively coupled and inductively coupled currents circulate in the opposite directions when they are moving in the forward direction. When the coupled currents reach the far-end terminating resistor on the quiet line, the net current through the terminating resistor will be the difference between the capacitively coupled current and the inductively coupled current.

The backward-moving inductively coupled-noise current will have exactly the same signature as the capacitively generated noise current. It will start out at zero and rise up as the signal emerges from the driver. After a time equal to the rise time, the backward-flowing current will reach a constant value and stay at that level. The signal edge will act as a current source for the inductively coupled currents and couple a constant amount of current as it propagates down the whole coupled length.

After the rising edge of the signal has just reached the terminating resistor at the far end of the active line, there will still be backward-flowing inductively coupled-noise current in the quiet line. It will take another TD for all this current to finally flow back to the near end of the quiet line. The current flows for the backward- and forward-moving noise currents are shown in Figure 10-23.

Moving in the forward direction, the inductively coupled noise will travel at the same speed as the signal edge in the active line. Each step along the way, there will be more and more inductively coupled noise current coupled over. The far-end noise will grow larger with coupled length. The shape of the inductively coupled current at the far end will be the derivative of the rise time, as it is directly proportional to the dI/dt of the signal.

The direction of the inductively coupled current at the far end is counterclockwise, from the return path up to the signal path. This is the opposite direction of the capacitively coupled current. At the far end, the capacitively coupled noise and inductively coupled noise will be in opposite directions. The net far-end noise will actually be the difference between the two.

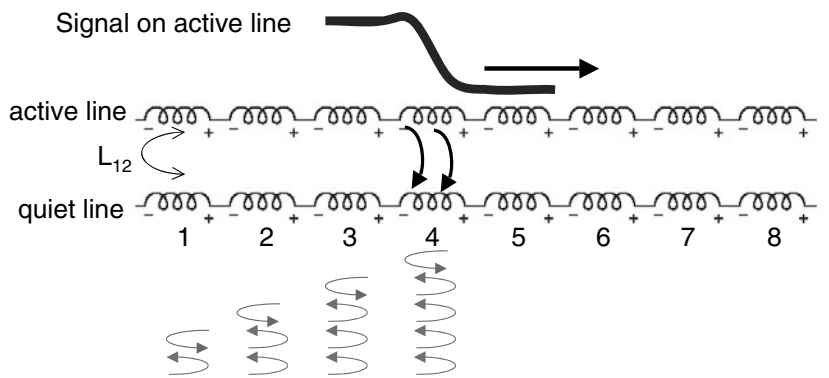


Figure 10-23 Induced-current loops propagating in the forward and backward direction as the signal in the active trace moves down the line.

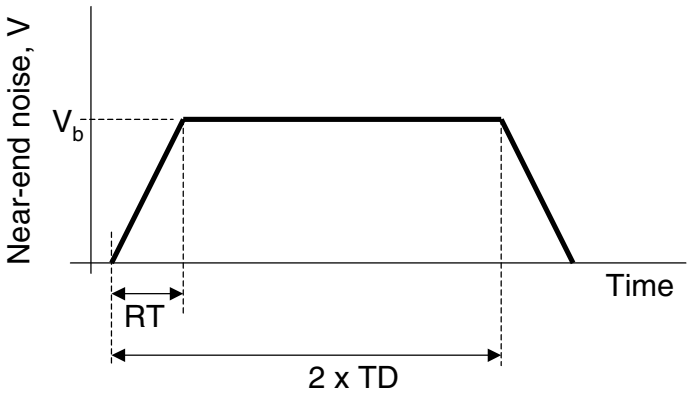


Figure 10-24 Near-end cross-talk voltage signature when the signal is a linear ramp.

10.11 Near-End Cross Talk

The near-end noise voltage is related to the net coupled current through the terminating resistor on the near end. The general signature of the waveform is displayed in Figure 10-24. There are four important features of the near-end noise:

1. If the coupling length is longer than the saturation length, the noise voltage will reach a constant value. The magnitude of this maximum voltage level is defined as the near-end cross-talk value (NEXT). It is usually reported as a ratio of the near-end noise voltage in the quiet line to the signal in the active line. If the voltage on the active line is V_a and the maximum backward voltage on the quiet line is V_b, the NEXT is NEXT = V_b/V_a. In addition, this ratio is also defined as the near-end cross-talk coefficient, $k_b = V_b/V_a$.

2. If the coupling length is shorter than the saturation length, the voltage will peak at a value less than the NEXT. The actual noise-voltage level will be the peak value, scaled by the actual coupling length to the saturation length. For example, if the saturation length is 6 inches, i.e., the signal has a rise time of 2 nsec in FR4, and the coupled length is 4 inches, the near-end noise is V_b / V_a = NEXT × 4 inches/6 inches = NEXT × 0.66. Figure 10-25 shows examples of the near-end noise for coupling lengths ranging from 20% of the saturation length to two times the saturation length.

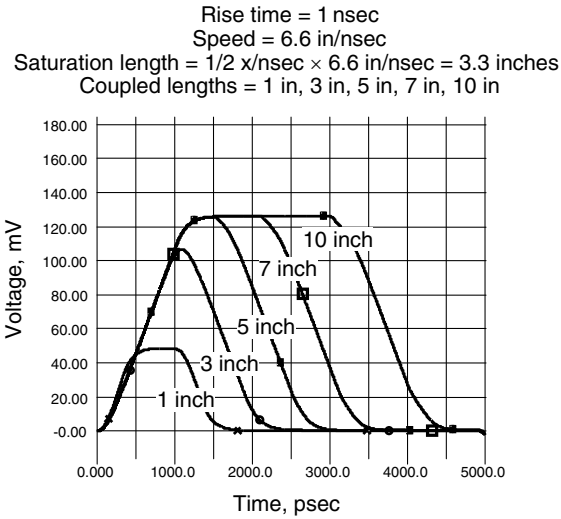


Figure 10-25 Near-end cross-talk voltage as the coupling length increases from 20% of the saturation length to two times the saturation length. Rise time is 1 nsec, speed is 6.6 inch/nsec, saturation length is 0.5 nsec × 6.6 in/nsec = 3.3 inches. Simulated with Mentor Graphics Hyperlynx.

3. The total time the near-end noise lasts is 2 × TD. If the time delay for the coupled region is 1 nsec, the near-end noise will last for 2 nsec.

4. The turn on for the near-end noise is the rise time of the signal.

The magnitude of the NEXT will depend on the mutual capacitance and the mutual inductance. It is given by:

$$\text{NEXT} = \frac{V_b}{V_a} = k_b = \frac{1}{4}\left(\frac{C_{mL}}{C_L} + \frac{L_{mL}}{L_L}\right) \tag{10-20}$$

where:

NEXT = the near-end cross-talk coefficient

V_b = the voltage noise on the quiet line in the backward direction

V_a = the voltage of the signal on the active line

k_b = the backward coefficient

C_{mL} = the mutual capacitance per length, in pF/inch (C_{12})

C_L = the capacitance per length of the signal trace, in pF/inch (C_{11})

L_{mL} = the mutual inductance per length, in nH/inch (L_{12})

L_L = the inductance per length of the signal trace, in nH/inch (L_{11})

As the two transmission lines are brought together, the mutual capacitance and mutual inductance will increase and the NEXT will increase.

The only practical way of calculating the matrix elements and the backward cross-talk coefficient is with a 2D field solver. Figure 10-26 shows the calculated near-end cross-talk coefficient, k_b, for two geometries, a microstrip pair and a stripline pair. In each case, each line is 50 Ohms and the line width is 5 mils. The spacing is varied from 4 mils up to 50 mils. It is apparent that when the spacing is greater than about 10 mils, the stripline geometry has lower-near-end cross talk.

As a rough rule of thumb, the maximum acceptable cross talk allocated in a noise budget is about 5% of the signal swing. If the quiet line is part of a bus, there could be as much as two times the near-end noise appearing on a quiet line. This is due to the sum of the noise from the two adjacent traces on either side and any that are farther away. To estimate a design rule for near-end noise, the spacing should be large enough so that the near-end noise between just two adjacent traces is less than 5% ÷ 2 ~ 2%.

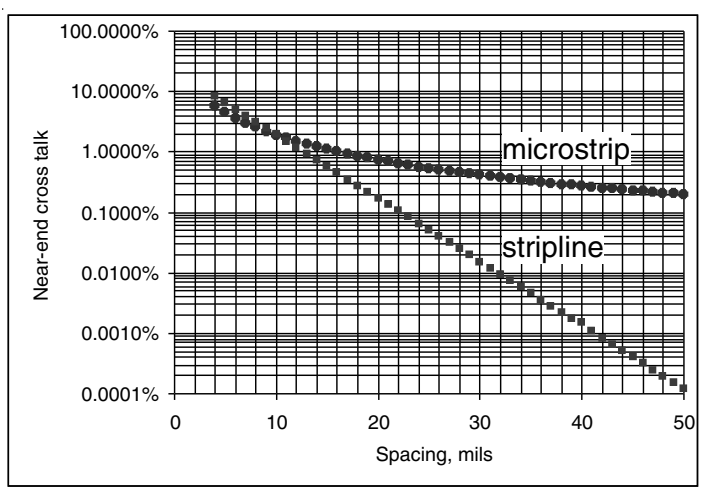


Figure 10-26 Calculated near-end cross-talk coefficient for microstrip and stripline traces, each 50 Ohms and 5 mils wide, in FR4, as the spacing is increased. Results from the Ansoft SI2D field solver.

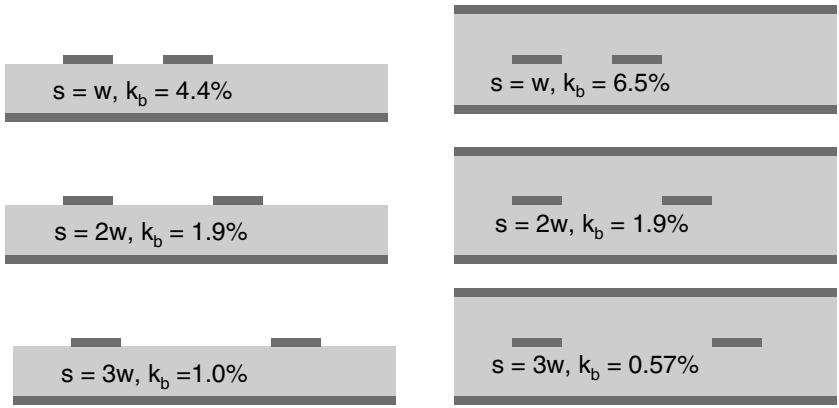


Figure 10-27 Near-end cross-talk coefficients for microstrip and stripline for a few specific spacings. These are handy rules of thumb to remember.

TIP In the case of 5-mil-wide lines in microstrip and stripline, the minimum spacing that creates less than 2% near-end noise is about 10 mils. This is a good rule of thumb for acceptable noise: The edge-to-edge spacing of signal traces should be at least two times the line width.

If the spacing between adjacent signal lines is greater than two times the line width, the maximum near-end noise will be less than 2%. In the worst-case coupling between one victim line and many aggressor lines on both sides, the maximum possible near-end noise coupled to the victim line will be less than 5%, within most typical noise budgets.

From this plot two other rules of thumb can be generated. These apply for the special case of 50-Ohm lines in FR4 with dielectric constant of 4. The near-end cross talk will scale with the ratio of the line width to the spacing. Of course, it is the dielectric thickness that is important, but a specific line width and a 50-Ohm characteristic impedance also defines a dielectric thickness.

Figure 10-27 summarizes the coupling in microstrip and stripline for spacings of 1 × w, 2 × w, and 3 × w. These are handy values to remember.

10.12 Far-End Cross Talk

The far-end noise voltage is related to the net coupled current through the terminating resistor on the far end. This, after all, is the voltage that is propagating

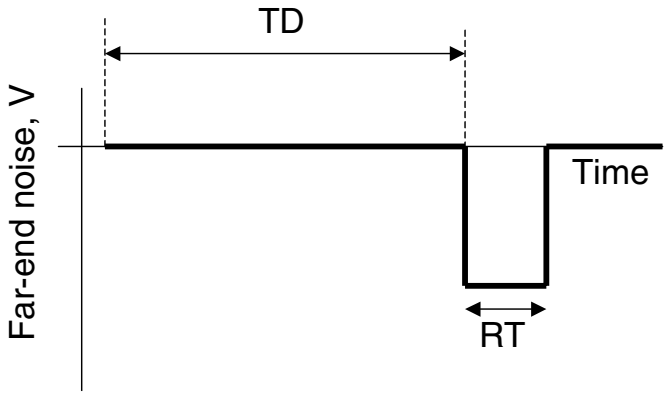


Figure 10-28 General signature of far-end cross-talk voltage noise when the signal is a linear ramp.

down the quiet line in the forward direction. The general signature of the waveform is displayed in Figure 10-28. There are four important features of the far-end noise:

1. No noise appears until a TD after the signal is launched. The noise has to travel down the end of the quiet line at the same speed as the signal.

2. The far-end noise appears as a pulse that is the derivative of the signal edge. The coupled current is generated by a dV/dt and a dI/dt, and this generated noise pulse travels down the quiet line in the forward direction, coincident with the active signal traveling down the aggressor line. The width of the pulse is the rise time of the signal. Figure 10-29 shows the far-end noise with different rise times. As the rise time decreases, the width of the far-end noise decreases and its peak value increases.

3. The peak value of the far-end noise scales with the coupling length. Increase the coupling length and the peak value increases.

4. The FEXT coefficient is a direct measure of the peak voltage of the far-end noise, V_f, usually expressed relative to the active signal voltage, V_a. FEXT = V_f/V_a. This noise value scales with the two extrinsic terms (coupling length and rise time) in addition to the intrinsic terms that are based on the cross section of the coupled lines. The FEXT is related to:

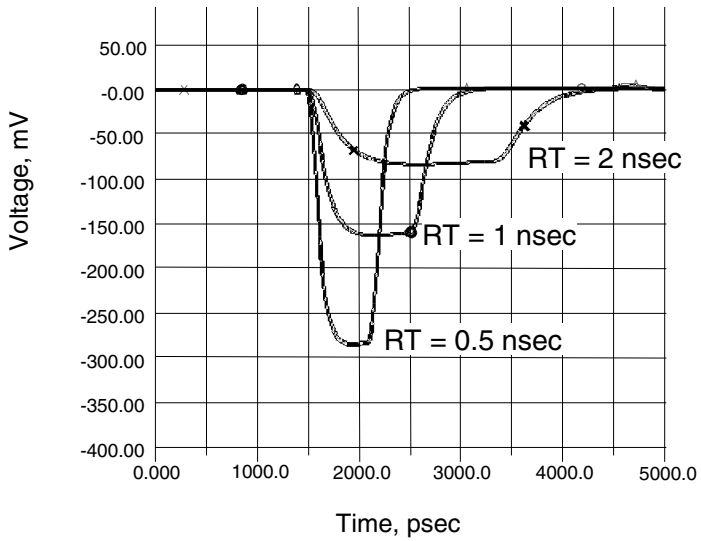


Figure 10-29 Far-end noise between two 50-Ohm microstrips in FR4 with 5-mil line and space, for the case of three different signal rise times, but the same 10-inch-long coupled length. Simulated with Mentor Graphics Hyperlynx.

$$\text{FEXT} = \frac{V_f}{V_a} = \frac{\text{Len}}{\text{RT}} \times k_f = \frac{\text{Len}}{\text{RT}} \times \frac{1}{2v} \times \left(\frac{C_{mL}}{C_L} - \frac{L_{mL}}{L_L}\right) \tag{10-21}$$

and

$$k_f = \frac{1}{2v} \times \left(\frac{C_{mL}}{C_L} - \frac{L_{mL}}{L_L}\right) \tag{10-22}$$

where:

FEXT = the far-end cross-talk coefficient

V_f = the voltage at the far end of the quiet line

V_a = the voltage on the signal line

Len = the length of the coupled region between the two lines

k_f = the far-end coupling coefficient that depends only on intrinsic terms

v = the speed of the signal on the line

C_{mL} = the mutual capacitance per length, in pF/inch (C_{12})

C_L = the capacitance per length of the signal trace, in pF/inch (C_{11})

L_{mL} = the mutual inductance per length, in nH/inch (L_{12})

L_L = the inductance per length of the signal trace, in nH/inch (L_{11})

The k_f term, or the far-end coupling coefficient, depends only on intrinsic qualities of the line: The relative capacitive and inductive coupling and the speed of the signal. It does not depend on the length of the coupling region nor on the rise time of the signal. What does this term mean? The inverse of k_f, $1/k_f$, has units of a speed, inches/nsec. What speed does it refer to?

As we show in the next chapter, $1/k_f$ is really related to the difference in speed between an odd-mode signal and an even-mode signal. Another way of looking at far-end noise is that it is really created when the odd mode has a different speed than the even mode. In a homogeneous distribution of dielectric material, the effective dielectric constant is independent of any voltage pattern and both the odd mode and even modes travel at the same speed. There is no far-end cross talk.

TIP If the dielectric material distribution surrounding all the conductors is uniform and homogeneous, as with two coupled, fully embedded microstrip traces or with two coupled striplines, the relative capacitive coupling is exactly the same as the relative inductive coupling. There will be no far-end cross talk in this configuration.

If there is any inhomogeneity in the distribution of dielectric materials, the fields will see a different effective dielectric constant depending on the specific voltage pattern between the signal lines and the return path, and there will be a difference in the relative capacitive and inductive coupling. This will result in far-end noise.

If all the space surrounding the conductors in a pair of coupled lines is filled with air, and there is no other dielectric nearby, the relative capacitive coupling and inductive coupling are both equal and the far-end coupling coefficient, k_f, is 0.

If all the space surrounding the conductors is filled with a material with dielectric constant ε_r, the relative inductive coupling will not change, as magnetic fields do not interact with dielectric materials at all.

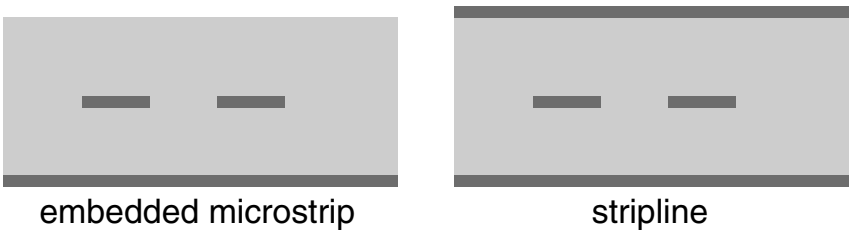


Figure 10-30 Two structures with homogeneous dielectric and no far-end cross talk: fully embedded microstrip and stripline.

The capacitive coupling will increase proportional to the dielectric constant. The capacitance to the return path will also increase proportional to the dielectric constant. However, the ratio will stay the same. There will be no far-end cross talk. An example of a fully embedded microstrip with no far-end cross talk is shown in Figure 10-30.

If the dielectric is removed above the embedded microstrip traces, the relative inductive coupling will not change at all, since inductance is completely independent of any dielectric materials. However the capacitance terms will be affected by the dielectric distribution. Figure 10-31 shows the change in the two capacitance terms, as the dielectric thickness above the traces is decreased. Though the capacitance to the return path decreases as the dielectric thickness above decreases, it does so only by a relatively small amount. The coupling capacitance decreases much more. The coupling capacitance C_{mL}, is strongly dependent on the dielectric constant of the material where the coupling fields are strongest—between the signal traces. As the dielectric on top is removed, it dramatically reduces the coupling capacitance.

In the fully embedded case, the relative coupling capacitance is as large as the relative coupling inductance. In the pure microstrip case with no dielectric above the traces, the relative coupling capacitance has actually decreased from the fully embedded case.

TIP Here is a situation where we are actually decreasing the coupling capacitance, yet the far-end noise increases.

What is often reported is not k_f, but $v \times k_f$ which is dimensionless:

$$v \times k_f = \frac{1}{2} \times \left(\frac{C_{mL}}{C_L} - \frac{L_{mL}}{L_L} \right) \tag{10-23}$$

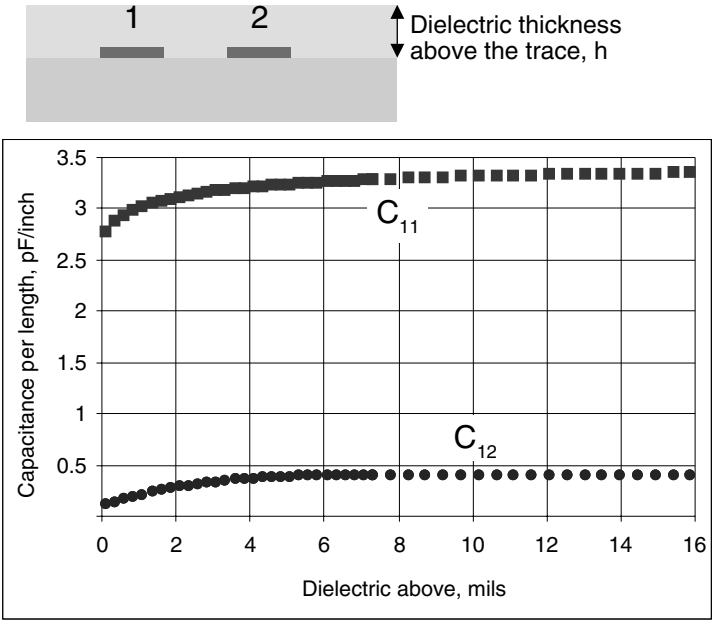


Figure 10-31 Calculated C_{11} and C_{12} terms as the dielectric thickness above the traces is increased. The diagonal element increases slightly, but the off-diagonal term increases much more. Results calculated with Ansoft's SI2D field solver.

Using this term, the FEXT can be written as:

$$FEXT = \frac{V_f}{V_a} = \frac{Len}{RT \times v} \times v \times k_f = \frac{TD}{RT} \times v \times k_f \qquad (10\text{-}24)$$

The term $v \times k_f$ is also an intrinsic term and depends only on the cross-sectional properties of the coupled lines. This is a measure of how much far-end noise there might be when the time delay of the coupled region is equal to the rise time, or when TD = RT. When $v \times k_f = 5\%$, there will be 5% far-end cross talk on the quiet line when TD = RT. If the coupled length doubles, the far-end noise will double to 10%.

Figure 10-32 shows how $v \times k_f$ varies with separation for the case of two 50-Ohm FR4 microstrip traces with line widths of 5 mils. From this curve, we can develop a simple rule of thumb to estimate the far-end noise. If the spacing is equal to the line width, $v \times k_f$ will be about 4%. For example, if the rise time is

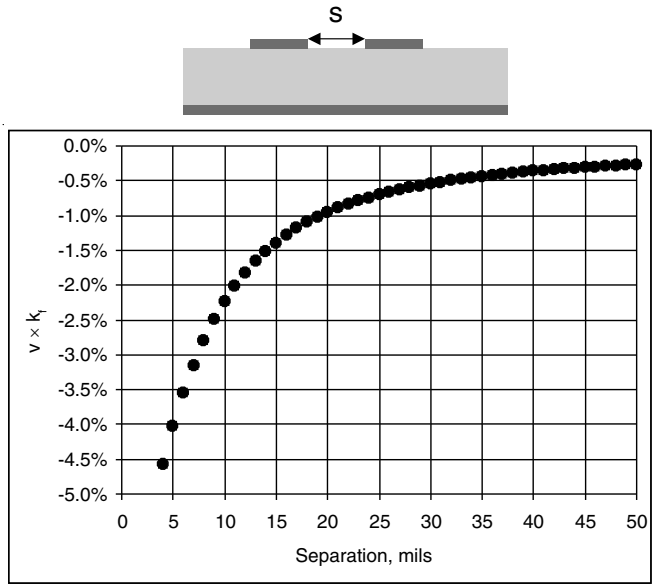


Figure 10-32 Calculated value of $v \times k_f$ for the case of two 50-Ohm coupled microstrip traces in FR4 with 5-mil-wide traces, as the spacing increases. Simulated with Ansoft's SI2D.

1 nsec, the coupled length is 6 inches, and the TD = 1 nsec, the far-end noise from just one adjacent aggressor line will be $v \times k_f \times TD/RT = 4\% \times 1 = 4\%$.

> **TIP** As a good rule of thumb, for two 50-Ohm microstrip lines with FR4, and with the tightest spacing manufacturable, a spacing equal to the line width, the far-end cross-talk noise will be $-4\% \times$ TD/RT.

If the coupled length increases, the far-end noise voltage will increase. If the rise time is decreased, the far-end noise will increase. If there are aggressors on either side of the signal line, there will be an equal amount of far-end noise from each active line. With a line width of 5 mils and spacing of 5 mils, the far-end noise on the quiet line would be 8% when TD = RT.

The length of surface traces on a board usually does not shrink much from one product generation to the next generation. The coupled time delay will also be roughly the same. However, with each product generation, rise times generally decrease. This is why far-end noise will become an increasing problem.

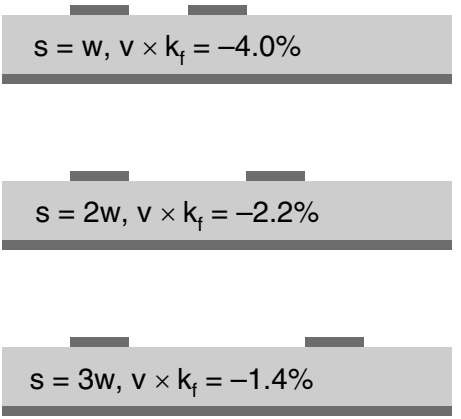


Figure 10-33 Simple rules of thumb for estimating the far-end cross talk for a pair of coupled 50-Ohm microstrips in FR4 for different spacings.

TIP As rise times shrink, far-end noise will increase. For a circuit board with the closest pitch, and 6-inch-long coupled traces, the far-end noise can easily exceed the noise budget for rise times of 1 nsec or less.

One important way of decreasing far-end noise is by increasing the spacing between adjacent signal paths. Figure 10-33 lists the value of $v \times k_f$ for three different spacings of two coupled, 50-Ohm microstrip transmission lines in FR4. Since the capacitance and inductance matrix elements scale with the ratio of line width to dielectric thickness, this table offers a handy way of estimating the amount of far-end cross talk for any line width, as long as each line is 50 Ohms.

10.13 Decreasing Far-End Cross Talk

There are four general guidelines to decrease far-end noise:

1. Increase the spacing between the signal traces. Increasing the spacing from $1 \times w$ to $3 \times w$ will decrease the far-end noise by 65%. Of course the interconnect density will also decrease and may make the board more expensive.
2. Decrease the coupling length. The amount of far-end noise will scale with the coupling length. While $v \times k_f$ can be as large as 4% for the tightest spacing (i.e., equal to the line width), if the coupled length can be kept very short, the

magnitude of the far-end noise can be kept small. For example, if the rise time is 0.5 nsec, and the coupling length, TD, is less than 0.1 nsec, the far-end noise can be kept below $4\% \times 0.1/0.5 = 0.8\%$. A TD of 0.1 nsec is a length of about 0.6 inch. A tightly coupled region under a BGA or connector field, for example, may be acceptable if it is kept short. The maximum coparallel run length is a term that can be set up in the constraint file of many layout tools.

3. Add dielectric material to the top of the surface traces. When surface traces are required and the coupling length cannot be decreased, it is possible to decrease the far-end noise by adding a dielectric coating on top of the traces. This could be with a thicker solder mask, for example. Figure 10-34 shows how $v \times k_f$ varies with the thickness of a top coat, assuming a dielectric constant the same as the FR4, or 4, and assuming the trace-to-trace separation is equal to the line width.

Adding dielectric above the trace will also increase the near-end noise and decrease the characteristic impedance of the traces. These have to be taken into account when adding a top coating.

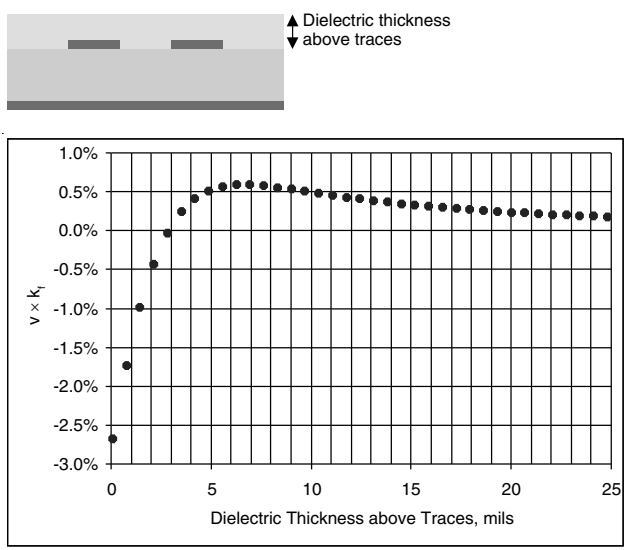


Figure 10-34 Variation of $v \times k_f$ as the coating thickness above the signal lines increases. The bus uses tightly coupled 50-Ohm microstrip traces with 5-mil-line and 5-mil spacing in FR4, assuming a dielectric coating with the same dielectric constant. Simulated with Ansoft's SI2D.

As the coating thickness is increased, the far-end noise initially decreases and actually passes through zero. Then it goes positive and finally drops back and approaches zero. In a fully embedded microstrip, the dielectric is homogeneous and there is no far-end noise. This is when the dielectric thickness is roughly five times the line width. This complex behavior is due to the precise shape of the fringe fields between the traces and also between the traces and the return plane as well as to how they penetrate into the dielectric material as the coating thickness is increased.

It is possible to find a value of the coating thickness so the far-end noise for surface traces is exactly zero. In this particular case, it is with a thickness equal to the dielectric thickness between the traces and the return path, or about 3 mils.

TIP In general, the optimum coating thickness will depend on all the geometry features and dielectric constants. Even a thin solder-mask coating will provide some benefit by decreasing the far-end noise a small amount.

4. Route the sensitive lines in stripline. Coupled lines in buried layers, as stripline cross sections, will have minimal far-end noise. If far end is a problem, the surest way of minimizing it is to route the sensitive lines in stripline.

   In practice, it is usually not possible to use a perfectly homogenous dielectric material even in a stripline. There will always be some variations in the dielectric constant due to combinations of core and prepreg materials. Usually, the prepreg is resin rich and has a lower dielectric constant than the core laminate. This will give rise to an inhomogenous dielectric distribution and some far-end noise.

Far-end noise can be the dominant source of noise in microstrip lines. These are the surface traces. When the spacing equals the line width, and the rise time is 1 nsec, the far-end noise will be 8% on a victim line in the middle of a bus with a coupling length longer than 6 inches. As the rise time decreases or coupling length increases, the far-end cross talk will increase. This is why far-end noise can often be the dominant problem in low-cost boards where many of the signal lines are routed as microstrip.

Whenever you use microstrip traces, a small flag should go off that there is the potential for too much far-end noise and an estimate should be performed of the expected far-end noise to have confidence it won't be a problem. If it is, either increase the routing pitch, decrease the coupling length, add some more solder mask, or route the long lines in stripline.

10.14 Simulating Cross Talk

In the special case of two uniform transmission lines, with perfect terminations on each end, the expected voltage noise on the near and far ends can be calculated given the cross-section geometry and material properties. A 2D field solver will allow the calculation of k_b and NEXT. The 2D field solver will also calculate k_f, and from the TD of the coupled length and the rise time, we can get the FEXT.

But what if the termination changes, and what if the coupled length is just part of a larger circuit? This requires a circuit simulator that includes a model for a coupled transmission line. If the mutual capacitance per length and mutual inductance per length are known, an n-section, coupled transmission-line model can be created to allow the simulation for any termination strategy, providing enough sections are used. The difficulty with an n-section lumped-circuit model is its complexity and computation time.

For example, if the time delay is 1 nsec, such as in a 6-inch-long interconnect, and the required model bandwidth is 1 GHz, a total of 10×1 GHz $\times$ 1 nsec = 10 lumped-circuit sections would be required in the coupled model. For longer lengths, this means even more sections are required.

There is a class of simulators that allow the creation of coupled transmission lines using an ideal, distributed, coupled transmission-line model. These tools usually have an integrated 2D field solver. The geometry of the cross section is input and the distributed coupled model is automatically created. The details of an ideal, distributed, coupled transmission line, often called an ideal differential pair, are reviewed in the next chapter.

TIP Coupled transmission-line circuits with arbitrary drivers, loads, and terminations can be simulated with a tool that has an integrated 2D field solver and allows automatic generation of a distributed, coupled transmission-line model from the cross-sectional information. These tools are incredibly powerful in predicting the performance of coupled noise in real systems.

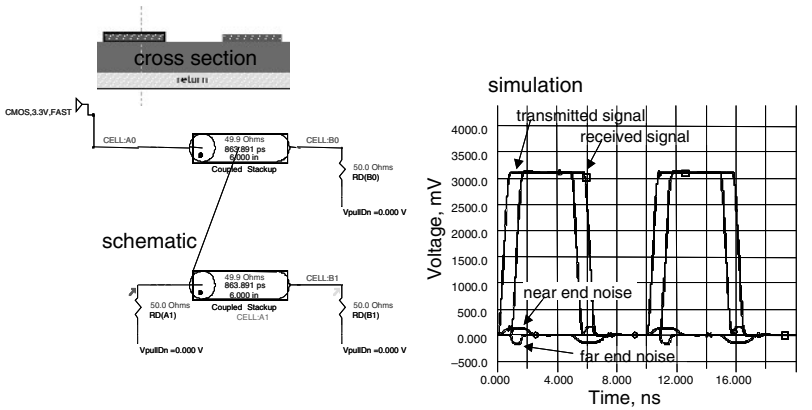


Figure 10-35 Simulating coupled transmission-line circuits with Mentor Graphics Hyperlynx, showing the cross section of a tightly coupled pair of microstrip lines, the resulting circuit with the coupled lines with drivers and terminating resistors, and the simulated voltages on the active and quiet lines.

Figure 10-35 is an example of the near- and far-end noise predicted in an ideal case, with a low-impedance driver on an active line and terminating resistors on all the ends. The geometry in this case is closely coupled, 50-Ohm microstrips, with 5-mil-wide traces and 5-mil-wide spacing.

When source-series termination is used, the results are slightly more complicated. Figure 10-36 shows the circuit for source-series termination. In this case, the noise at the far end of the quiet line is important as this is where the receiver is

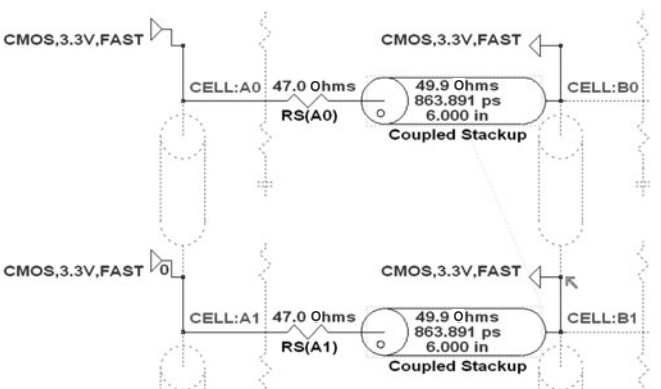


Figure 10-36 Circuit schematic for a source-series pair of coupled lines using Mentor Graphics Hyperlynx.

that would be sensitive to any noise. The actual noise appearing at the receiver end of the quiet line is subtle.

Initially, the noise at the receiver end of the quiet line is due to the far-end noise from the active signal. This is a large negative spike. However, as the signal on the active line reflects off the open end of the receiver on the active line, it travels back to the source of the active line. As it travels back to the source, the receiver on the quiet line is now the backward end for this reflected signal wave. There will be backward noise on the quiet line at the receiver. Even though the receiver on the quiet line is at the far end of the quiet line, it will see both far-end and near-end noise because of the reflections of the signal in the active line.

The near-end noise on the quiet-line receiver is about the same as what is expected based on the ideal NEXT. Figure 10-37 shows the comparison of the ideal case of NEXT and FEXT in the quiet line, with the noise at the receiver on the quiet line for this source-series terminated topology. This illustrates that the NEXT and FEXT terms are actually good figures of merit for the expected noise on the quiet line. Of course, the parasitics associated with the package models and the discrete termination components will complicate the received noise compared with the ideal NEXT and FEXT. A simulator tool will automatically take these details into account.

However, this is the case of having just one aggressor line coupling to the victim line. In a bus, there will be multiple traces coupling to a victim line. Each of the aggressor lines can add additional noise to the victim line. How many

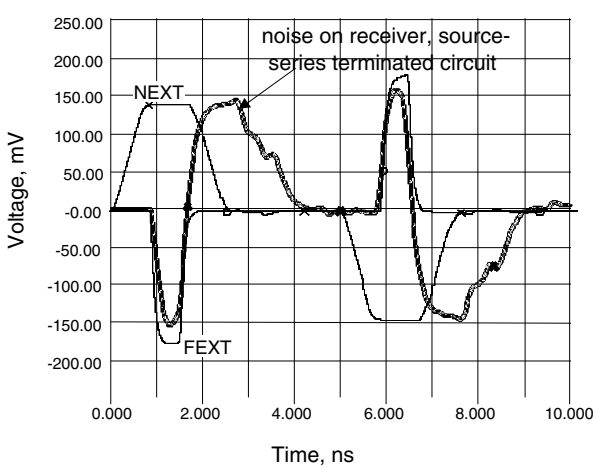


Figure 10-37 Simulated NEXT and FEXT in a far-end terminated circuit compared with the receiver noise on the quiet line of a similar pair of lines but with source-series termination. Simulated with Mentor Graphics Hyperlynx.

aggressors on either side should be included? The only way to know is to put in the numbers from a calculation.

In Figure 10-38, a bus circuit is set up, with one victim line and five aggressor lines on either side. Each line is configured with a source-series termination and receivers on the far ends. In this example, aggressive design rules are used for 50-Ohm striplines with 5-mil-wide traces and 5-mil spaces. The speed of a signal in this transmission line is about 6 inches/nsec. The signal has a rise time of about 1 nsec and a clock frequency of 100 MHz. The saturation length is 1/2 nsec × 6 inches/nsec = 3 inches. We use a length of 10 inches so the near-end noise will be saturated and thus a worst case.

There will be no far-end noise in this stripline. However, the signal reflecting off the far end of the aggressor lines will change direction and its backward noise will appear at the victim's far end. The signal in each aggressor is 3.3 volts.

Figure 10-39 shows the simulated noise at the receiver of the victim line. With just one of the adjacent lines switching, the noise on the quiet line is 195 mV.

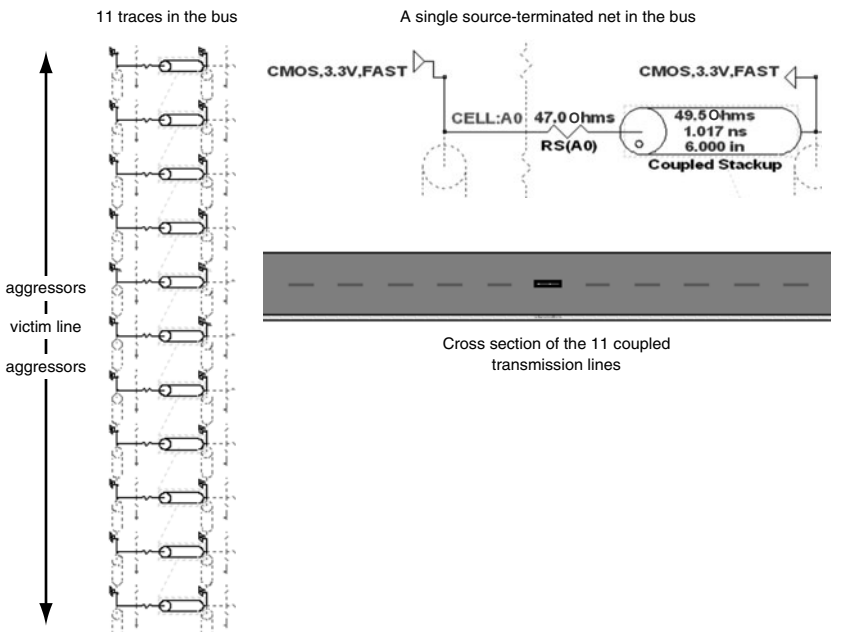


Figure 10-38 Circuit diagram used to simulate the noise on a victim line when it is surrounded by five aggressor lines on each side. Each line has source-series termination, 5-mil-wide line, 5-mil space, 10-inch length, as 50-Ohm striplines in FR4. Simulation set up by Mentor Graphics Hyperlynx.

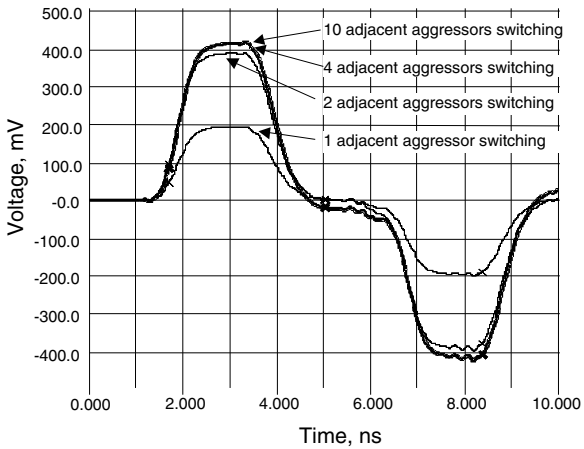


Figure 10-39 Simulated noise at the receiver of the victim line under the case of one aggressor, two aggressors, four aggressors, and ten aggressors switching. Most of the noise is taken into account by including just the nearest two aggressors. Simulated with Mentor Graphics Hyperlynx.

This is about 6%, probably too large all by itself for any reasonable noise budget. This geometry really is a worst case.

With the second aggressor on the other side of the victim line also switching, the noise is about twice, or 390 mV. This is 12% voltage noise from the adjacent two aggressor lines, one on either side of the victim line. Obviously, we get the same contribution of noise from each aggressor and its noise contribution to the victim line just adds.

In stripline, the coupled noise drops off rapidly as the trace separation increases. We would expect the noise from the next two distant aggressor traces to be very small compared with the immediately adjacent traces. With the four nearest aggressors switching, the amount of noise at the receiver of the victim line is 410 mV, or 12.4%. Finally, the worst case is when all the aggressor lines switch. The noise is still 410 mV, indistinguishable from the noise with just the adjacent four nearest aggressors switching.

We see that the absolute worst-case noise in the bus is about 2.1 times the basic NEXT noise level. If we take into account only the adjacent traces switching, we would still be including about 95% of the total amount of noise, given this worst-case geometry. By including the switching from the two adjacent aggressor lines on both sides, we are including 100% of the coupled noise.

In this example, very aggressive design rules were used. The amount of cross talk, 12%, is far above any reasonable noise budget. If, instead, the spacing equals

TIP Using just the noise from the adjacent traces on either side of a victim line in cross-talk analysis is usually sufficient for most system-level simulations. This will account for 95% of the cross talk in a closely coupled bus.

two times the line width, the amount of noise from just one adjacent trace switching would be 1.5%. The noise from two adjacent traces switching would be 3%. When the four nearest aggressor traces switch, the coupled noise at the receiver of the quiet line would still be only 3%. It is not affected to any greater degree if all 10 adjacent traces switch. The same simulation, using this practical design rule of 5-mil-line width and 10-mil spacing, is shown in Figure 10-40.

It is only necessary to look at the adjacent trace on either side of the victim line for a worst-case analysis.

TIP When the most aggressive design rules are used (i.e., a spacing equal to the line width), more than 95% of the noise on a victim line that is part of a bus is due to coupling from the two nearest aggressor traces, one on either side of the victim line. When conservative design rules are used and the spacing is twice the line width, virtually all of the noise from the bus comes from coupling from the most adjacent aggressor on either side of the victim line.

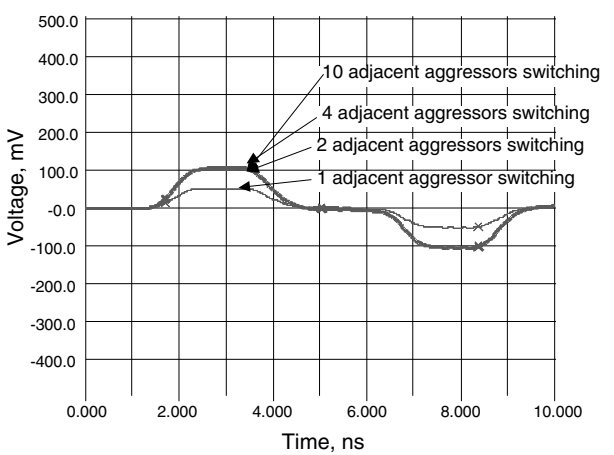


Figure 10-40 Noise on the receiver of a central trace in a bus composed of eleven, 50-Ohm striplines in FR4 with 5-mil-wide traces and 10-mil spacings. The number of active traces switching is changed from just one to two to four and then all 10. Simulated using Mentor Graphics Hyperlynx.

This suggests that when establishing design rules for adequate spacings for long, parallel signal buses, the worst-case noise we might expect to find is at most 2.1 times the basic NEXT values. If the noise budget allocates 5% the voltage swing for cross talk on any victim net, the actual amount of NEXT that we should allow between adjacent nets should be 5% ÷ 2.1, which is a NEXT value of about 2%. We can use this as the design goal in establishing a spec for the closest spacing allowed in either microstrip or stripline traces.

10.15 Guard Traces

One way to decrease cross talk is by spacing the traces farther apart. Keeping the spacing twice the line width will ensure the worst-case cross talk is less than 5%. In some cases, especially mixed-signal cases, keeping the cross talk significantly less than 5% is important. For example, a sensitive rf receiver may need as much as –100-dB isolation from any digital signals. A noise level of –100 dB is less than 0.001% of the active signal appearing on the sensitive quiet line.

It is often suggested that using guard traces will significantly decrease cross talk. This is the case, but only by the correct design and configuration. A much more effective way of isolating two traces is to keep them on separate layers with different return planes. This can offer the greatest isolation of any routing alternative. However, if it is essential to route an aggressor and victim line on the same signal layers and adjacent, then guard traces can offer an alternative way of controlling the cross talk, under specific conditions.

Guard traces are separate traces that are placed between the aggressor lines and the victim lines that we want to shield. Examples of the geometry of guard traces are shown in Figure 10-41. The guard trace should be as wide as will fit between the signal lines, consistent with the fabrication design rules for spacing.

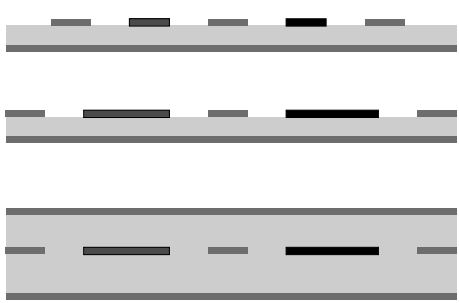


Figure 10-41 Examples of cross sections of traces that incorporate guard traces. The guard trace is hatched; all other traces are signal traces.

Guard traces can be used in both microstrip and stripline cross sections. The value of a guard trace in microstrip is not very great, as a simple example will illustrate.

If we stick to design rules that specify the smallest allowable spacing is equal to the line width, then before we can even add a guard trace, we must increase the spacing between two traces to three times the line width just to fit a trace between the aggressor and victim lines. We can evaluate the added benefit of a guard trace in microstrip structures by comparing the three cases:

1. Two closely spaced microstrip lines, 5-mil trace width, 5-mil spacing
2. Increasing the spacing to the minimum spacing (15 mils) that will allow a guard trace to fit
3. Increased spacing of 15 mils *and* with the guard trace in place

These three examples are illustrated in Figure 10-42.

In addition, we can evaluate different termination strategies for the guard trace. Should it be left open, terminated at the ends, or shorted at the ends?

In Figure 10-43, the peak noise on the receiver of the quiet line for all of these alternatives is compared. The noise seen on the receiver is the combination of the reflections of the signal on the active line and the near- and far-end noise on the quiet line and their reflections. Combined with these first-order effects are the second-order contributions from the parasitics of the packages and resistor elements.

The peak noise with the traces close together is 130 mV. This is about 4%. Simply by pulling the traces far enough apart to fit a guard trace, the noise is reduced to 39 mV, or 1.2%. This is almost a factor-of-four reduction and is almost always a low enough cross talk, except in very rare situations. When the guard

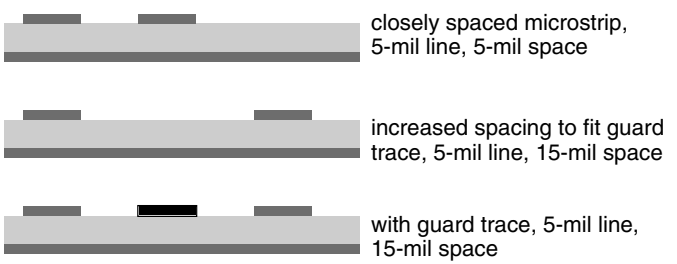


Figure 10-42 Three different microstrip structures evaluated for noise at the receiver of the victim line.

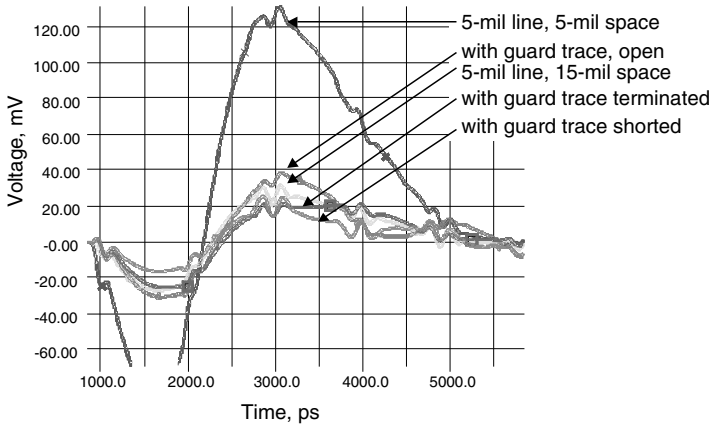


Figure 10-43 Three different geometries tested and with the guard trace left open on the ends, shorted on the ends, and terminated on the ends. Simulated with Mentor Graphics Hyperlynx.

trace is inserted, but left open, floating, we see the noise on the quiet line actually increases a slight amount.

However, if the guard trace is terminated with 50-Ohm resistors on both ends, the noise is reduced to about 25 mV, or 0.75%. When the guard trace is shorted at the ends, the noise on the quiet line is reduced to 22 mV, or about 0.66%.

TIP You will get the most benefit by increasing the spacing. The noise is reduced by a factor of four. By adding the guard trace and shorting it on both ends, the noise is reduced further by a factor of two.

TIP The guard trace affects the electric- and magnetic-field lines between the aggressor line and the victim line, and will always decrease both the capacitance and inductance matrix elements.

The reduction in magnitude of the off-diagonal matrix elements by a guard trace is purely related to the geometry. It is independent of how the guard trace is electrically connected to the return path. This would imply that a guard trace would always be beneficial. However, the guard trace will also act as another signal line. Noise will couple to the guard trace from the aggressor line. This noise on the guard trace can then couple over to the quiet line. The amount of noise generated on the guard trace, that can couple to the quiet line, will depend on how it is terminated.

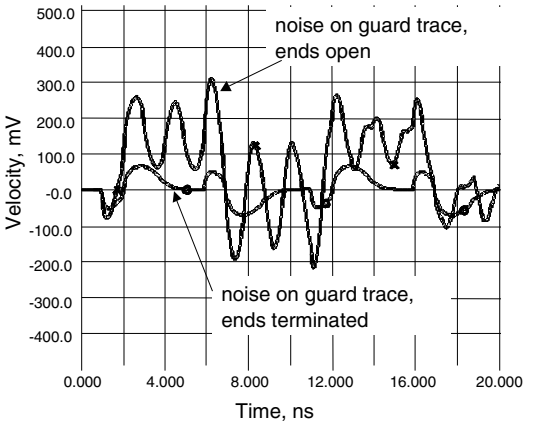


Figure 10-44 Noise at the end of a guard trace when it is open and when it is terminated. This noise will couple to the quiet line. Simulated with Mentor Graphics Hyperlynx.

If the guard trace is left open, the maximum amount of noise is generated on the guard trace. If it is terminated with 50 Ohms on each end, less noise appears. Figure 10-44 shows the noise at the far end of the guard trace under the same conditions as above, with 5-mil line width, 5-mil space, and 50-Ohm microstrips with a 3.3-v, 100-MHz signal on the active line using source-series termination. It is clear there is more noise generated on the guard trace when the guard trace is left open at the ends. This extra noise will couple more noise to the quiet line, which is why a guard trace with open ends can sometimes create more noise on the victim line than if it were not there.

The biggest benefit of a guard trace is when the guard trace is shorted on the ends. As the signal moves down the aggressor line, it will still couple noise to the guard trace. The backward-moving noise on the guard trace will hit the short at the near end and reflect, with a reflection coefficient of –1. This means that much of the near-end noise on the guard trace, moving in the backward direction, will be cancelled out with the coincident, negative, reflected near-end noise traveling in the forward direction on the guard trace.

TIP Shorting the end of the guard trace will eliminate any near-end noise that would appear along the guard trace.

There will still be the buildup of the forward-moving far-end noise on the guard trace. This noise will continue to grow until it hits the shorted far end of the

guard trace. At this point, it will see a reflection coefficient of –1 and reflect back. At the very far end of the guard trace, the net noise will be zero since there is, after all, a short. But, the reflected far-end noise will continue to travel back to the near end of the guard trace. If all we do is short the two ends of the guard trace, the far-end noise on the guard trace will continue to reflect between the two ends, acting as a potential noise source to the victim line that is supposed to be protected. If there were no losses in the guard trace, the far-end noise would continue to rattle around the guard trace, always acting as a low-level noise source to couple back onto the victim line.

TIP We can minimize the amount of far-end noise generated on the guard trace by adding more shorting vias, distributed down the length of the guard trace. These vias will have no impact on the noise directly coupled from the aggressor line to the victim line. They will only suppress the noise voltage generated on the guard trace.

The spacing between the shorting vias on the guard trace affects the amount of voltage noise generated on the guard trace in two ways. The far-end noise on the guard trace will only build up in the region between the vias. The closer the spacing, the lower the maximum far-end noise voltage that can build up on the guard trace. The more vias, the lower the far-end noise on the guard trace. This will mean lower voltage noise available to couple to the victim line.

Figure 10-45 shows the comparison of the coupled noise on the victim line, for two coupled, 10-inch-long microstrip traces with a guard trace between them, one simulation with one shorting via at each end of the guard trace and a second with 11 shorting vias, spaced at 1-inch intervals along the guard trace.

The initial noise on the victim line is the same, independent of the number of shorting vias. This noise is due to the direct coupled noise between the aggressor and victim lines. The amount of coupled noise is related to the reduction in the size of the matrix element due to the presence of the guard trace. By adding multiple vias, we limit the buildup of the noise on the guard trace and this eliminates the possibility of this additional noise coupling to the victim line.

The second role of multiple shorting vias is in generating a negative reflection of the far-end noise, which will cancel out the incident far-end noise. However, this will only be cancelled out where the incident and reflected far-end noise

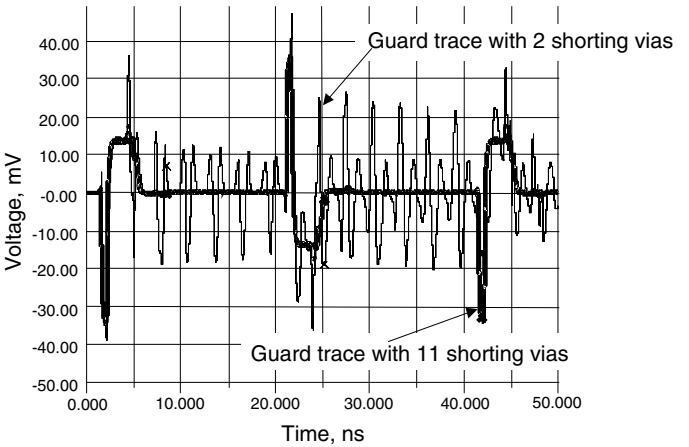


Figure 10-45 Comparing the noise on the victim line using a guard trace with only two shorting vias and one with 11 shorting vias. The signal is 3.3 v with a 0.7-nsec rise time and a low clock frequency of 25 MHz to show the effect of the noise in the guard trace with only shorting vias on its ends. The maximum noise on the victim line is 35 mV out of 3.3 v or 1%. Simulated with Mentor Graphics Hyperlynx.

overlaps. If the spacing between the shorting vias is longer than the width of the far-end noise, which is the rise time, this effect will not occur.

TIP As a rough rule of thumb, the shorting vias should be distributed along the guard trace so that there are at least three vias within the spatial extent of the signal rise time. This will guarantee overlap of far-end noise and its negative reflection, causing cancellation of noise voltage on the guard trace.

If the rise time is 1 nsec, the spatial extent is 1 nsec × 6 inches/nsec = 6 inches. The spacing between shorting vias should be 6 inches/3 = 2 inches. If the rise time were 0.7 nsec, as in the previous example, the spatial extent would be 6 inches × 0.7 nsec = 4.2 inches and the optimum spacing between shorting vias should be 4.2 inches/3 = 1.4 inches. In the previous example, vias were placed every 1 inch. Any closer would have no impact on the noise coupled to the victim line.

Of course, the shorter the signal rise time, the closer the spacing for the shorting vias required for optimum isolation. There will always be a compromise between the cost of the vias and the number that should be added. This should be an issue only when high isolation is required. In practice, there will be only a

small impact on the noise in the victim line as the number of shorting vias is increased.

However, when high isolation is important, the coupled noise to the quiet line will be much less in stripline, and stripline should always be used. In stripline structures, there will be much less far-end noise and much less need for shorting vias distributed along the length of the guard trace.

A guard trace can be very effective at limiting the isolation in a stripline structure. Figure 10-46 is the calculated near-end cross talk between an aggressor and victim, for a 50-Ohm stripline pair, with and without a guard trace between the aggressor and victim lines. In this example, the spacing between the two lines was increased and the guard trace was made as wide as would fit, consistent with the design rule of the spacing always greater than 5 mils. The guard trace, in the stripline configuration, offers a very significant amount of isolation over a configuration with no guard trace. Even isolations as large as –160 dB can be achieved using a wide guard trace in stripline. With separations of 30 mils between the active and aggressor lines, the guard trace can reduce the amount of isolation by almost three orders of magnitude.

A guard trace shields more than just the electric fields. There will also be induced currents in the guard trace from the proximity of the signal currents in the

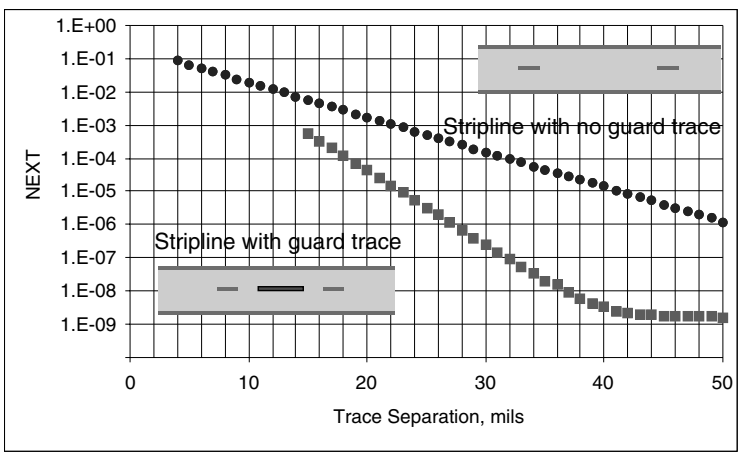


Figure 10-46 Near-end cross talk in stripline pairs, with and without a guard trace. The signal-line width is 5 mils and each line is 50 Ohms in FR4. The guard-trace width is increased as the signal-trace separation is increased to maintain a 5-mil gap. The flattening of the near-end noise at about 10^{-9} is due to the numerical noise floor of the simulator. Simulated with Ansoft's SI2D.

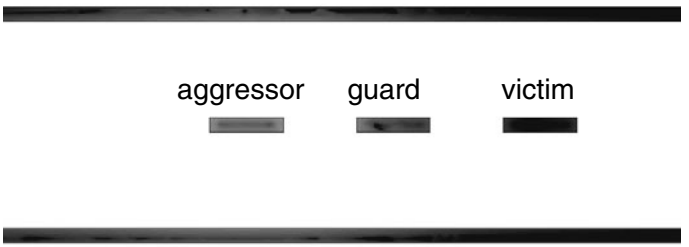


Figure 10-47 Current distribution in the conductors when a 100-MHz signal is driven in the aggressor line and the guard trace is shorted to the return path. Lighter color is higher current density. Gray scale is a log scale. Simulated with Ansoft's SI2D.

adjacent active lines. Figure 10-47 shows the current distribution in an active trace, a guard trace, and a quiet trace, with 5-mil width and spacing. Current is driven only in the active line, yet, there is induced current in the guard trace, comparable to the current density in the return planes.

This induced current will be in the opposite direction of the current in the active line. The magnetic-field lines from the induced current in the guard trace will further act to cancel the stray magnetic-field lines from the active line at the location of the quiet line. Both the electric-field shielding and magnetic-field shielding effects are fully accounted for by a 2D field solver when calculating the new capacitance- and inductance-matrix elements for a collection of conductors when one or more are used as a guard trace.

10.16 Cross Talk and Dielectric Constant

Near-end noise is related to the sum of the relative capacitive and inductive coupling: $C_{12}/C_{11} + L_{12}/L_{11}$. Of course, the inductive coupling is completely unaffected by the dielectric material around the conductors.

In a 50-Ohm stripline geometry with a uniform dielectric constant everywhere, if the dielectric constant of all surrounding materials were decreased, the capacitance between the signal and return paths, C_{11}, would decrease. However, the fringe-field capacitance between the two signal lines, C_{12}, would also decrease by the same amount. There would be no change to the cross talk in a stripline.

However, if the dielectric constant were decreased everywhere, the characteristic impedance would increase from the 50 Ohms. One way to get back to 50 Ohms would be to decrease the dielectric thickness. If the dielectric thickness were decreased to reach 50 Ohms, the cross talk would be decreased.

TIP In a very subtle way, a lowered dielectric constant will decrease cross talk, but only indirectly. The lower dielectric constant will allow a closer spacing between the signal and return paths for the same target impedance, which will cause the lower cross talk.

Figure 10-48 shows the characteristic impedance of a stripline trace as the plane-to-plane separation changes, for a line width of 5 mils for two different dielectric constants. If a design were to use a laminate with a dielectric constant of 4.5, such as FR4, and then switch to 3.5, such as with Polyimide, the plane-to-plane spacing to maintain 50 Ohms would have to decrease from 14 mils to 11.4 mils. If the spacing between the 5-mil traces were kept at 5 mils, the near-end cross talk would reduce from 7.5% to 5.2%. This is a reduction in near-end cross talk of 30%.

TIP Using a lower dielectric-constant material would either allow a lower cross-talk for the same routing pitch or tighter pitch routing for the same cross-talk spec. This could result in a smaller size board if it were limited by cross-talk design rules.

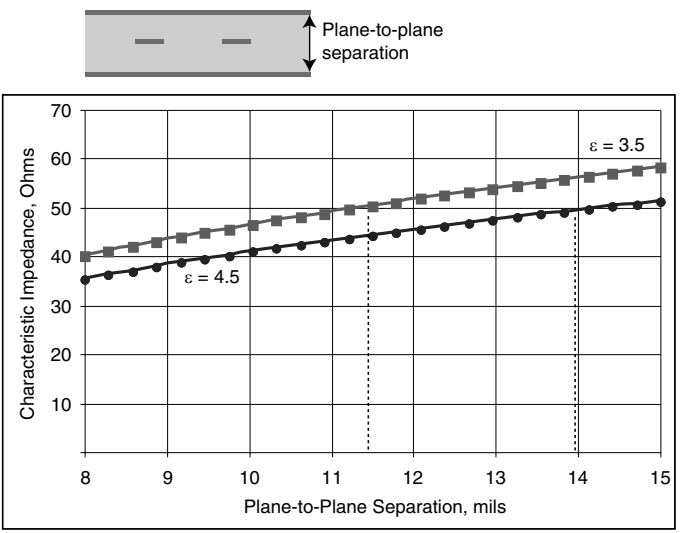


Figure 10-48 Variation in characteristic impedance for two dielectric materials in 5-mil-wide stripline traces as the total separation between the planes is changed. Lower dielectric constant allows closer spacing between the planes for the same impedance, which would result in lower cross talk. Simulated with Ansoft's SI2D.

10.17 Cross Talk and Timing

The time delay, TD, of a signal line depends on the length of the interconnect and the speed of the signal in the line. The speed of the signal depends on the dielectric constant of the surrounding material. In principle, cross talk from adjacent aggressor lines should not affect the time delay on a victim line. After all, how could the signal on adjacent traces affect the speed of the signal on the victim line?

TIP In stripline, this is true. The speed of a signal on a victim line is completely independent of the signals appearing on any nearby aggressor lines and there is no impact on timing from cross talk.

However, in microstrip traces, there is a subtle interaction between cross talk and the timing. This is due to the combination of the asymmetry of the dielectric materials and the different fringe electric fields between the signal lines depending on the data pattern on the aggressor lines.

Consider three closely spaced, 10-inch-long microstrip signal lines, each 5 mils wide with 5-mil space. The outer two lines will be aggressor lines and the center line will be a victim line. The time delay for the victim line when there is no signal on the active lines is about 1.6 nsec. This is shown in Figure 10-49.

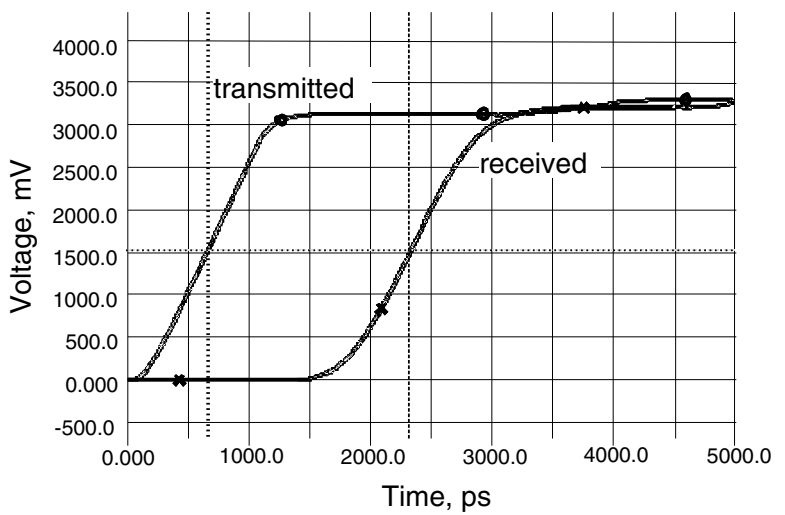


Figure 10-49 Time delay of 1.6 nsec between the signal leaving the victim driver and when it reaches the receiver on the victim line, both aggressors off. Simulated with Mentor Graphics Hyperlynx.

TIP The time delay of the signal on the victim line will depend on the voltage pattern on the aggressor lines. When the aggressors are switching opposite to the victim line, the time delay is decreased. When the aggressor lines are switching with the same signal as the victim line, the time delay of the victim line is increased.

The relationship of the voltage pattern on the aggressor lines to the time delay of the signal on the victim line is illustrated in Figure 10-50. The simulated results are shown in Figure 10-51. It is as though when the aggressors are switching opposite to the victim line, the speed of the signal is increased and the time delay is decreased. When the aggressors are switching with the same voltage pattern as the victim line, the speed of the victim signal is decreased and the time delay increased.

When the aggressor lines are off, the field lines from the signal voltage on the victim line see a combination of the bulk-material dielectric constant and the air above the line. This effective dielectric constant determines the speed of the signal.

When the aggressor lines are switching opposite from the signal on the victim line, there are large fields between the victim and aggressor traces, and many of these field lines are in air with a low dielectric constant. The effective dielectric constant the victim signal sees would have a larger fraction of air and be reduced compared to if the aggressors were not switching. A lower effective dielectric

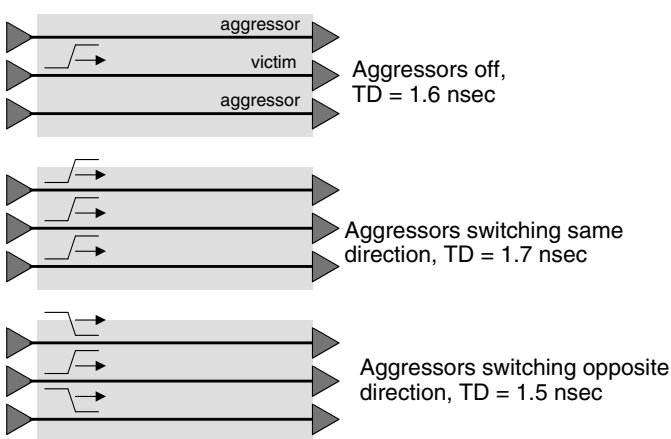


Figure 10-50 Three configurations for signals on the victim line and aggressor lines. The voltage pattern on the aggressors affects the time delay of the victim line.

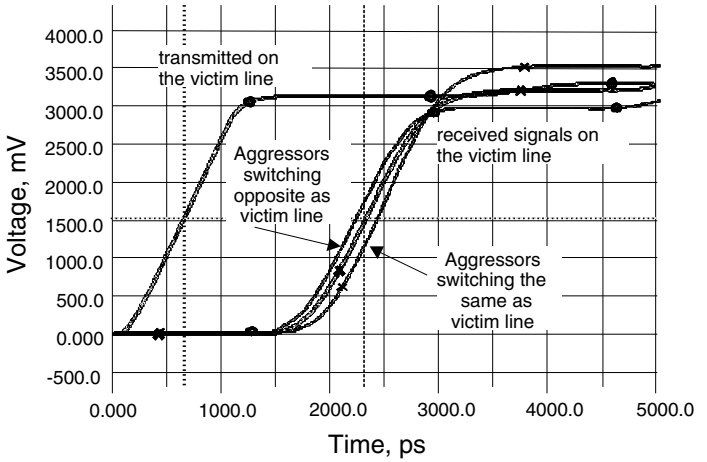


Figure 10-51 Received signal on victim line when aggressors switch in the same direction and opposite to the signal on the victim line. Simulated with Mentor Graphics Hyperlynx.

constant would result in a faster speed and a shorter time delay for the signal on the victim line. This is illustrated in Figure 10-52.

When the aggressor lines are switching in the same direction as the victim line, each trace is at the same potential and there are few field lines in the air; most of them are in the bulk material. This means the effective dielectric constant the victim line sees will be higher, dominated by the bulk-material value. This will increase the effective dielectric constant for the victim-line signal when the

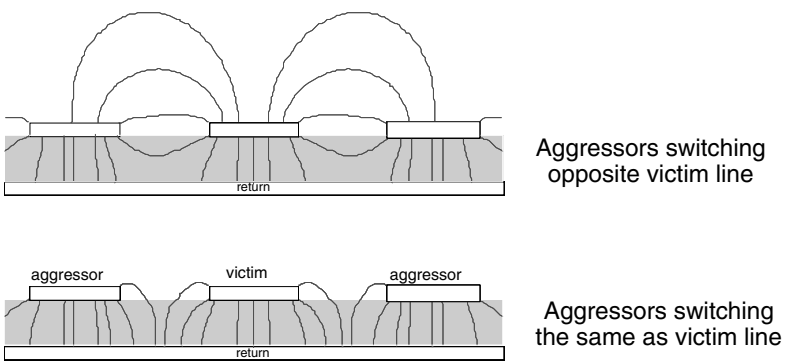


Figure 10-52 Electric-field distribution around victim and aggressor lines when they switch opposite and then when they switch in the same direction. Simulated with Mentor Graphics Hyperlynx.

aggressor lines switch in the same direction. This will decrease the signal speed and increase the time delay of the victim signal.

This impact of cross talk on time delay will only happen in tightly coupled lines where there is significant overlap of the fringe fields. If the lines are far enough apart where the cross-talk voltage is not a problem, the fringe fields will not overlap and the time delay of the victim line will be independent of how the other traces are switching. This effect is fully taken into account when simulating with a distributed, coupled, transmission-line model.

10.18 Switching Noise

So far, we have been discussing cross talk when the return path is a wide, continuous plane, as with most transmission lines on circuit boards. In this environment, there is near- and far-end noise, and all the coupling occurs between adjacent signal lines, and virtually no coupling from farther away traces.

When the return path is not a uniform plane, the inductive coupling will increase much more than the capacitive coupling and the noise will be dominated by the loop mutual inductance. This usually occurs in small localized regions of the interconnect such as in packages, connectors, or local regions of the board where the return path is interrupted by gaps.

When loop mutual inductance dominates, and it occurs in a small region, we can model the coupling by a single lumped mutual inductor. The noise generated in the quiet line by the mutual inductance only arises when there is a dI/dt in the active line, which is when the edge switches. For this reason, the noise created when mutual inductance dominates is sometimes called *switching noise*, or *dI/dt noise*, or *delta I noise*. Ground bounce, as discussed previously, is a form of switching noise for the special case when the loop mutual inductance is dominated by the total inductance of the common return lead. Whenever there are shared return paths, ground bounce will occur. As pointed out, there are three ways of reducing ground bounce:

1. Increase the number of return paths so the total dI/dt in each return path is reduced.
2. Increase the width and decrease the length of the return path so it has a minimum partial self-inductance.
3. Bring each signal path in proximity to its return path to increase their partial mutual inductance.

TIP However, even if there is no shared return path, there can still be cross talk between two or more signal/return-path loops when dominated by mutual inductance. Loop mutual inductance in connectors and packages can often be the limiting feature to their high-speed performance.

By using a simple model, we can estimate how much loop mutual inductance between two signal/return loops is too much. As a signal passes through one pair of pins in a connector (the active path), there is a sudden change in the current in the loop right at the wave front. This changing current causes a voltage noise to be induced in an adjacent, quiet loop, due to the mutual inductance between the two loops.

The voltage noise that is induced in the quiet loop is approximated by:

$$V_n = L_m \frac{dI_a}{dt} = L_m \frac{V_a}{RT \times Z_0} \tag{10-25}$$

where:

V_n = the voltage noise in the quiet loop

L_m = the loop mutual inductance between the active and quiet loops

I_a = the current change in the active loop

Z_0 = the typical impedance the signal sees in the active and quiet loops

V_a = the signal voltage in the active loop

RT = the rise time of the signal (i.e., how fast the current turns on)

The only term that is really influenced by the connector or package design is the loop mutual inductance between the loops. The impedance the signal sees, typically on the order of 50 Ohms, is part of the system specification, as are the rise time and signal voltage.

The magnitude of acceptable switching noise depends on the allocated noise in the noise budget. Depending on the negotiating skills of the engineer who is responsible for selecting the connector or IC package, the switching noise would probably have to be below about 5% to 10% of the signal swing.

TIP A signal-integrity engineer who has good negotiating skills and knows how hard it might be to find a connector or package with low enough mutual inductance might push for a higher allowable mutual inductance spec, therefore forcing another part of the system to require a tighter spec.

If the maximum acceptable switching noise is specified, this will define a maximum allowable loop mutual inductance between an active and quiet net. The maximum allowable loop mutual inductance is:

$$L_m = \frac{V_n}{V_a}(RT \times Z_0) \tag{10-26}$$

where:

L_m = the loop mutual inductance between the active and quiet loops

V_n = the voltage noise in the quiet loop

V_a = the signal voltage in the active loop

RT = the rise time of the signal, or how fast the current turns on

Z_0 = the typical impedance the signal sees in the active and quiet loops

As a starting place, we will use values of:

$V_n/V_a = 5\%$

$Z_0 = 50$ Ohms

RT = 1 nsec

The maximum allowable loop mutual inductance in this example would be 2.5 nH. If the rise time were to decrease, more switching noise would be created by the dI/dt and the loop mutual inductance would have to be decreased to keep the switching noise down.

TIP This suggests a simple rule of thumb. To keep switching noise to an acceptable level between a pair of signal and return paths, the loop mutual inductance between them should be less than $L_m < 2.5\ nH \times RT$, with RT in nsec.

If the rise time were 0.5 nsec, the maximum allowable loop mutual inductance would be 1.2 nH. As rise times decrease, the maximum allowable mutual inductance will also decrease. This will make connector and package design more difficult. There are three primary geometrical features that will decrease loop mutual inductance.

The most important term that influences loop mutual inductance is the length of the loops. Decrease the length of the loops and the mutual inductance will decrease. This is why the trend in packages and connectors is to make them as short as possible, such as with chip scale packages.

The second term that influences the loop mutual inductance is the spacing between the loops. Increase the spacing and the loop mutual inductance will decrease. There are practical limits to how far apart signal and return paths can be placed, though.

The third term that influences the loop mutual inductance is the proximity of the signal to return path of each loop. The loop mutual inductance is related to the loop self-inductance of either loop. Decreasing the loop self-inductance of either loop will decrease the loop mutual inductance between them. Bringing the signal closer to the signal path of one loop will decrease its impedance. In general, switching noise will be decreased with lower impedance signal paths. Of course, using too low a value will introduce a new set of problems related to impedance discontinuities.

This analysis has assumed that the switching noise is generated between just two adjacent signal paths. If there is significant coupling between two aggressor lines and the same quiet line, keeping the same total switching noise on the victim line would require half the mutual inductance between either pair. The more aggressor lines that couple to the same victim line, the lower the allowable loop mutual inductance.

If we know how much loop mutual inductance there is between signal pairs in a package or connector, we can use the approximation in equation 10-26 to immediately estimate their shortest usable rise time or highest usable clock frequency. For example, if the loop mutual inductance were 2.5 nH, the shortest rise time before the switching noise were more than 5% the signal swing would be 1 nsec, if there was coupling between just the two signal and return paths.

TIP This suggests a simple rule of thumb. The shortest usable rise time, in nsec, limited by switching noise is $RT > L_m/2.5$ nH. The highest usable clock frequency would be about $10 \times 1/RT = 250\ MHz/L_m$. This assumes the clock period is $10 \times RT$.

For example, if the loop mutual inductance between a pair of signal paths is 1 nH, the maximum operating clock frequency might be on the order of 250 MHz. Of course, if there were five aggressor pairs that could couple to a signal victim line, each with 1-nH loop mutual inductance, the highest operating clock frequency would be reduced to 250 MHz/5 = 50 MHz. This is why leaded packages, with typically 1 nH of loop mutual inductance between signal paths, have maximum clock operating frequencies in the 50-MHz range.

10.19 Summary of Reducing Cross Talk

Cross talk can never be completely eliminated; it can only be reduced. The general design features that will reduce cross talk include the following:

1. Increase the pitch of the signal traces.
2. Use planes for return paths.
3. Keep coupled lengths short.
4. Route on stripline layers.
5. Decrease the characteristic impedance of signal traces.
6. Use laminates with lower dielectric constants.
7. Do not share return pins in packages and connectors.
8. When high isolation between two signal lines is important, route them on different layers with different return planes
9. Guard traces have little value in microstrip traces. In stripline, use guard traces with shorting vias on the ends and throughout the length.

TIP Unfortunately, the features that reduce cross talk also invariably increase system costs. It is critically important to be able to accurately predict the expected cross talk so that the right trade-offs can be established to achieve the lowest cost at acceptable cross talk.

For cross talk in uniform transmission lines, the right tool is a 2D field solver with an integrated circuit simulator. For nonuniform transmission-line sections, the right tool is a 3D field solver, either static (if the section can be approximated by a single-lumped section) or full wave (if it is electrically long). With a predictive tool, it will be possible to evaluate how large a bang can be expected for the buck it will cost to implement the feature.

10.20 The Bottom Line

1. Cross talk is related to the capacitive and inductive coupling between two or more signal and return loops. It can often be large enough to cause problems.
2. The lowest cross talk between adjacent signal lines is when their return paths are a wide plane. In this environment, the capacitive coupling is comparable to the inductive coupling and both terms must be taken into account.
3. Since cross talk is primarily due to the coupling from fringe fields, the most important way of decreasing cross talk is to space the signal paths farther apart.
4. The noise signature will be different on the near end and far end of a quiet line adjacent to a signal path. The near-end noise is related to the sum of the capacitively and inductively coupled currents. The far-end noise is related to the difference between the capacitively and inductively coupled currents.
5. To keep the near-end noise below 5% for the worst-case coupling in a bus, the separation should be at least two times the line width for 50-Ohm lines.
6. Near-end noise will reach a maximum value when the coupling length equals half the spatial extent of the rise time.
7. Far-end noise will scale with the ratio of the time delay of the coupling length to the rise time. For a pair of microstrip traces with a spacing equal to their line width, there will be 4% far-end noise when the time delay of the coupling length equals the rise time.
8. In a tightly coupled bus, 95% of the coupled noise can be accounted for by considering just the nearest aggressor line on either side of the victim line.
9. There is no far-end cross talk in striplines.
10. For very high isolation, signal lines should be routed on different layers with different return planes. If they have to be on the same layer and adjacent, use stripline with guard traces. Isolation greater than –160 dB is possible. In this case, watch out for ground bounce cross talk when signals change reference layers through vias.
11. Mutual inductance will dominate the coupled noise in some packages and connectors. As rise times decrease, the maximum amount of allowable mutual inductance between signal- and return-path loops must decrease. This will make it harder to design high-performance components.

CHAPTER 11

Differential Pairs and Differential Impedance

A differential pair is simply a pair of transmission lines with some coupling between them. The value of using a pair of transmission lines is not so much to take advantage of the special properties of a differential pair as to take advantage of the special properties of differential signaling, which uses differential pairs. Differential signaling is the use of two output drivers to drive two independent transmission lines, one line carrying one bit and the other line carrying its complement. The signal that is measured is the difference between the two lines. This difference signal carries the information.

Differential signaling has a number of advantages over single-ended signaling, such as the following:

1. The total dI/dt from the output drivers is greatly reduced over signal-ended drivers, so there is less ground bounce, rail collapse and potentially less EMI.
2. The differential amplifier at the receiver can have higher gain than a single-ended amplifier.
3. The propagation of a differential signal over a tightly coupled differential pair is more robust to cross talk and discontinuities in the return path.

4. Propagating differential signals through a connector or package will be less susceptible to ground bounce and switching noise.
5. A low cost, twisted-pair cable can be used to transmit a differential signal over long distances.

The most important downside to differential signals comes from the potential to create EMI. If differential signals are not properly balanced or filtered, and there is any common signal component present, it is possible for real world differential signals that are driven on external twisted-pair cables to cause EMI problems.

The second downside is that to transmit a differential signal requires twice the number of signal lines as to transmit a single-ended signal. The third downside is that there are many new principles and a few key design guidelines to understand about differential pairs. Due to the complicated effects in differential pairs, there are many myths in the industry that have needlessly complicated and confused their design.

Ten years ago, less than 50% of the circuit boards fabricated had controlled-impedance interconnects. Now, more than 90% of them have controlled impedance interconnects. Today, less than 50% of the boards fabricated have differential pairs. In a few years, it may be over 90% with differential pairs.

11.1 Differential Signaling

Differential signals are widely used in the small computer scalable interface (SCSI) bus, in Ethernet, in many of the telecommunications optical-carrier (OC) protocols such as OC-48, OC-192, and OC-768, and for all high-speed serial protocols. One of the popular signaling schemes is low-voltage differential signals (LVDS).

When we look at a signal voltage, it is important to keep track of where we are measuring the voltage. When a driver drives a signal on a transmission line, there is a signal voltage on the line that is measured between the signal conductor and the return path. We usually refer to this as the single-ended transmission-line signal. When two drivers drive a differential pair, in addition to the two single-ended signals, there is also a voltage difference between the two signal lines. This is the difference voltage, or the differential signal. Figure 11-1 illustrates how these two different signals are measured in a differential pair.

In LVDS, two output pins are used to drive a single bit of information. Each signal has a voltage swing from 1.125 v to 1.375 v. Each of these signals drives separate transmission lines. The single-ended voltage between the signal and return paths of each line is shown in Figure 11-2.

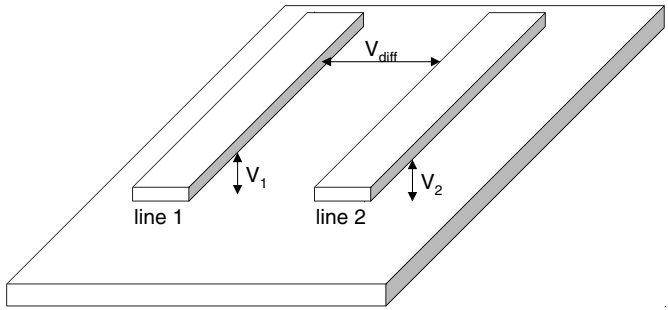


Figure 11-1 The single-ended signal is measured between the signal conductor and return conductor. The differential signal is measured between the two signal lines in a differential pair.

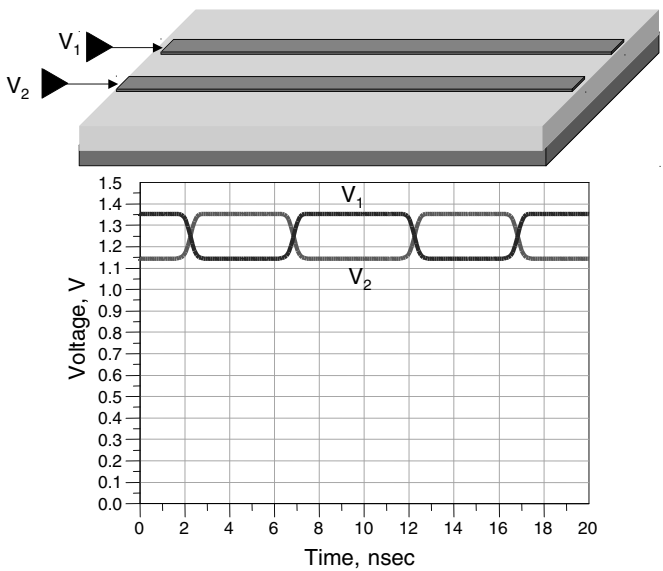


Figure 11-2 Voltage signaling scheme for LVDS signals, a typical differential signal.

At the receiver end, the voltage on line 1 is V_1 and the voltage on line 2 is V_2. A differential receiver measures the differential voltage between the two lines and recovers the differential signal:

$$V_{diff} = V_1 - V_2 \tag{11-1}$$

where:

V_{diff} = the differential signal

V_1 = the signal on line 1 with respect to its return path

V_2 = the signal on line 2 with respect to its return path

In addition to the intentional differential signal that carries information, there may be some common signal. The common signal is the average voltage that appears on both lines, defined as:

$$V_{comm} = \frac{1}{2}(V_1 + V_2) \tag{11-2}$$

where:

V_{comm} = the common signal

V_1 = the signal on line 1 with respect to its return path

V_2 = the signal on line 2 with respect to its return path

TIP These definitions of differential and common signals apply universally to all signals. When any arbitrary signal is generated on a pair of transmission lines, it can always be completely and uniquely described by a combination of a common- and a differential-signal component.

Given the common and differential signals, the single-ended voltages on each line with respect to the return plane can also be recovered from:

$$V_1 = V_{comm} + \frac{1}{2}V_{diff} \tag{11-3}$$

$$V_2 = V_{comm} - \frac{1}{2}V_{diff} \tag{11-4}$$

where:

V_{comm} = the common signal

V_{diff} = the differential signal

V_1 = the signal on line 1 with respect to its return path

V_2 = the signal on line 2 with respect to its return path

In LVDS signals, there are some differential-signal and some common-signal components. These signal components are displayed in Figure 11-3. The differential signal swings from –0.25 v to +0.25 v. As the differential-signal propagates down the transmission line, its voltage is a 0.5-v transition.

There is also a common signal. The average value is 1.25 v. This is more than a factor of 2 larger than the differential-signal component.

TIP Even though an LVDS signal is called a differential signal, it has a very large common-signal component, which is nominally constant.

When we call an LVDS signal a differential signal, we are really lying. It has a differential component, but it also has a very large common component. It is not a pure differential signal. In ideal conditions, the common signal is constant. The common signal normally will not have any information content and will not affect signal integrity or system performance. Unfortunately, as we will see, very small perturbations in the physical design of the interconnects can cause the common-

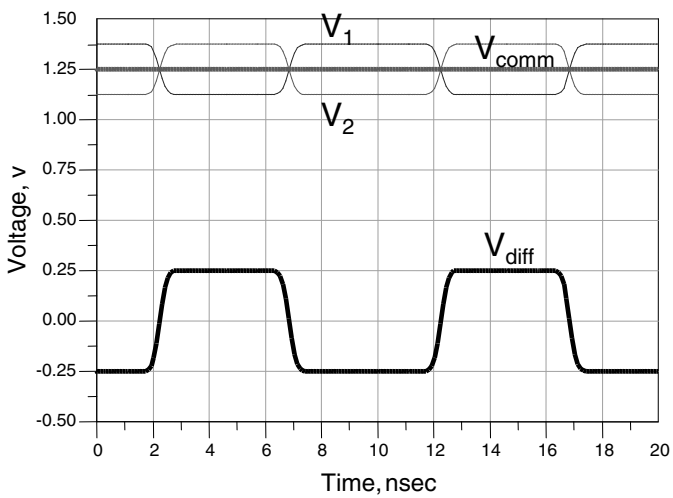


Figure 11-3 The common- and differential-signal components in LVDS signals. Note there is a very large common-signal component, which, in principle, is constant.

signal component to change, and a changing common-signal component has the potential of causing two very important problems:

1. If the value of the common signal gets too high, it may saturate the input amplifier of the differential receiver and prevent it from accurately reading the differential signal.
2. If any of the changing common signal makes it out on a twisted-pair cable, it has the potential of causing excessive EMI.

TIP The terms *differential* and *common* always and only refer to the properties of the signal, never to the properties of the differential pair of transmission lines themselves. Misuse of these terms is one of the major causes of confusion in the industry.

11.2 A Differential Pair

All it takes to make up a differential pair is two transmission lines. Each transmission line can be a simple, single-ended transmission line. Together, the two lines are called a differential pair. In principle, any two transmission lines can make up a differential pair.

Just as there are many cross sections for a single-ended transmission line, there are many cross sections for differential pairs of transmission lines. Figure 11-4 shows the cross sections of the most popular geometries.

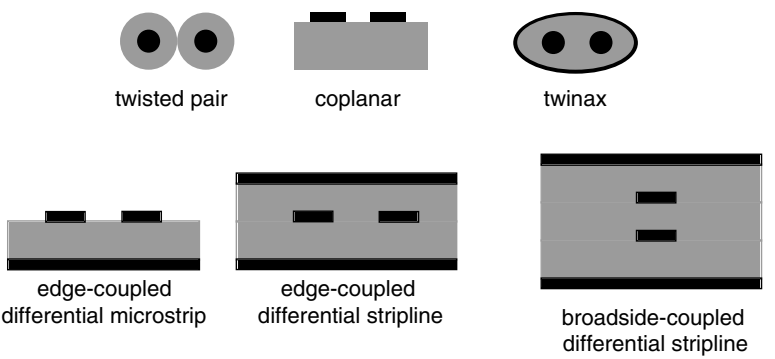


Figure 11-4 Some of the more popular cross sections for differential pair transmission.

Though, in principle, any two transmission lines can make up a differential pair, five features will optimize their performance for transporting high-bandwidth difference signals:

1. The most important property of a differential pair is that it be of uniform cross section down its length and provide a constant impedance for the difference signal. This will ensure minimal reflections and distortions for the differential signal.
2. The second most important property is that the time delay between each line is matched so that the edge of the difference signal is sharp and well defined. Any time delay difference between the two lines, or skew between them, will also cause a distortion of the differential signal and some of the differential signal to be converted into common signal.
3. Both transmission lines should look exactly the same. The line width and dielectric spacing of the two lines should be the same. This property is called symmetry. There should be no other asymmetries such as a test pad on one and not on the other, a neck down on one but not the other. Any asymmetries will convert differential signals into common signals.
4. Each line in a pair should have the same length. The total length of each line should be exactly the same. This will help maintain the same delay and minimize skew.
5. There does not have to be any coupling between the two lines, but most of the noise immunity benefits of differential pairs is lost if there is no coupling. Coupling between the two signal lines will allow differential pairs to be more robust to ground bounce noise picked up from other active nets than single-ended transmission lines. The greater the coupling, the more robust the differential signals will be to discontinuities and imperfections.

An example of a differential pair of transmission lines is shown in Figure 11-5, with a differential signal propagating. In this example, the signal on one line is a transition from 0 v to 1 v and the other signal a transition from 1 v to 0 v. As the signals propagate down the transmission lines, they have the voltage distribution shown in the figure.

Of course, even though we called this a differential signal, we see right away that we are also lying. It is not a pure differential signal, but contains a large com-

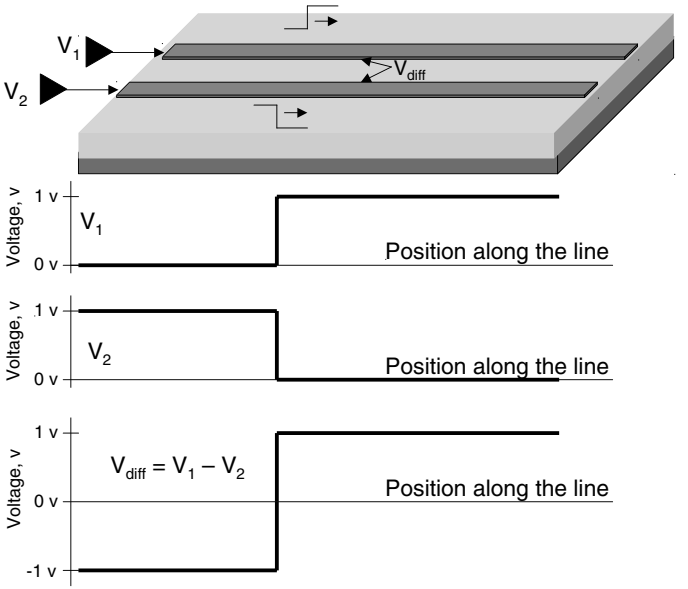


Figure 11-5 A differential pair with a signal propagating on each line, and the differential signal between the two lines.

mon signal component of 0.5 v. However, this common signal is constant and we will ignore it, paying attention only to the differential component.

Given these voltages on each transmission line, the difference voltage can be easily calculated. By definition, it is $V_1 - V_2$. The pure differential-signal component that is propagating down the differential pair is also shown in the figure. With a 0-v to 1-v signal transition on each single-ended transmission line, the differential-signal swing is a 2-v transition propagating down the interconnect. At the same time, there is a common-signal component, $1/2 \times (V_1 + V_2) = 0.5$ v, which is constant along the line.

TIP The most important electrical property of a differential pair is the impedance the differential signal sees, which we call the differential impedance.

11.3 Differential Impedance with No Coupling

The impedance the differential signal sees, the differential impedance, is the ratio of the voltage of the signal to the current of the signal. This definition is the basis

of calculating differential impedance. What makes it subtle is identifying the voltage of the signal and the current of the signal.

The simplest case to analyze is when there is no coupling between the two lines that make up the differential pair. We will look at this case, determine the differential impedance of the pair, and then turn on coupling and look at how the coupling changes the differential impedance.

To minimize the coupling, assume the two transmission lines are far enough apart, i.e., at least a spacing equal to twice the line width, so they do not appreciably interact with each other and so each line has a single-ended characteristic impedance, Z_0, of 50 Ohms. The current going into one signal trace and out the return is:

$$I_{one} = \frac{V_{one}}{Z_0} \tag{11-5}$$

where:

I_{one} = the current into one signal line and out its return

V_{one} = the voltage between the signal line and the adjacent return path of one line

Z_0 = the single-ended characteristic impedance of the signal line

For example, when a signal of 0 v to 1 v is imposed on one line and at the same time a 1-v to 0-v signal is imposed on the other line, there will also be a current loop into each line. Into the first line will be a current of I = 1 v/50 Ohms = 20 mA flowing from the signal conductor down to and out its adjacent return path. Into the second line will also be a current loop of 20 mA, but flowing from the return conductor up to and out its signal conductor.

The differential-signal transition propagating down the line is the differential signal between the two signal line conductors. This differential signal swing is twice the voltage on either line: $2 \times V_{one}$. In this case, it is a 2-volt transition that propagates down the signal line pair. At the same time, just looking at the signal line conductors, it looks like there is a current loop of 20 mA going into one signal conductor and coming out the other signal conductor.

By the definition of impedance, the impedance the differential signal sees is:

$$Z_{diff} = \frac{V_{diff}}{I_{one}} = \frac{2 \times V_{one}}{I_{one}} = 2 \times \frac{V_{one}}{I_{one}} = 2 \times Z_0 \quad (11\text{-}6)$$

where:

Z_{diff} = the impedance the differential signal sees, the differential impedance

V_{diff} = the difference or differential-signal transition

I_{one} = the current into one signal line and out its return

V_{one} = the voltage between the signal line and the adjacent return path of one line

Z_0 = the single-ended characteristic impedance of the signal line

The differential impedance is twice the characteristic impedance of either line. This is reasonable since the voltage across the two lines is twice the voltage across either one and its return path, but the current going into one signal and out the other is the same. If the single-ended impedance of either line is 50 Ohms, the differential impedance of the pair would be 2 × 50 Ohms = 100 Ohms.

If a differential signal propagates down a differential pair and reaches the end, where the receiver is, the impedance the differential signal sees at the end will be very high and the differential signal will reflect back to the source. These multiple reflections will cause signal-quality noise problems. Figure 11-6 is an

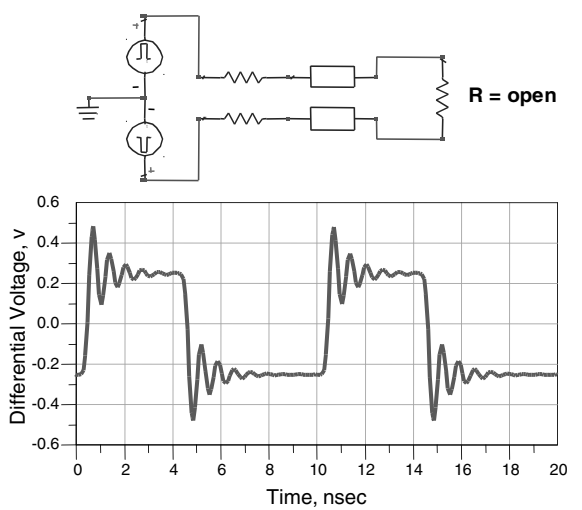


Figure 11-6 Differential circuit and the received signal at the far end of the differential pair when the interconnect is not terminated. There is no coupling in this differential pair. Simulated with Agilent ADS.

example of the simulated differential signal at the end of a differential pair. The ringing is due to the multiple bounces of the differential signal between the low impedance of the driver and the high impedance of the end of the line.

One way of managing the reflections is to add a terminating resistor across the ends of the two signal lines that matches the impedance the difference signal sees. The resistor should have a value of $R_{term} = Z_{diff} = 2 \times Z_0$. When the differential signal encounters the terminating resistor at the end of the line, it will see the same impedance as in the differential pair, and there will be no reflection. Figure 11-7 shows the simulated received differential signal with a 100-Ohm terminating resistor between the two signal lines.

Differential impedance can also be thought of as the equivalent impedance of the two single-ended lines in series. This is illustrated in Figure 11-8. Looking into the front of each line, each driver sees the characteristic impedance of the line, Z_0. The impedance between the two signal lines is the series combination of the impedances of each line to the return path. The equivalent impedance of the two signal lines, or the differential impedance, is the series combination:

$$Z_{diff} = Z_0 + Z_0 = 2 \times Z_0 \tag{11-7}$$

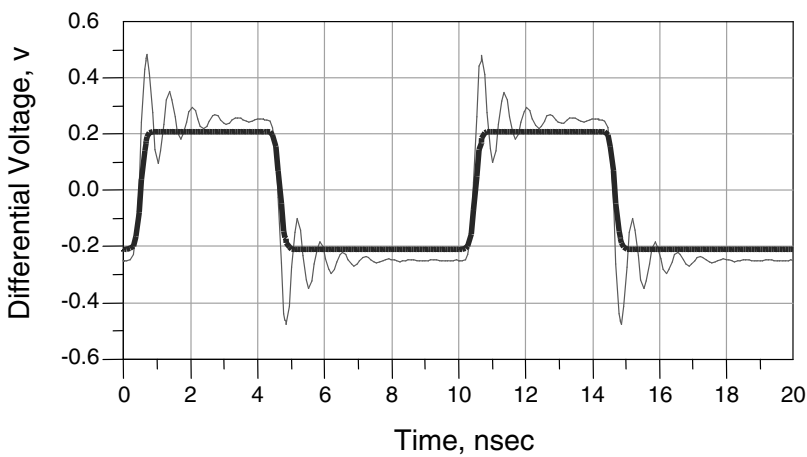


Figure 11-7 Received differential signal at the far end of the differential pair when the interconnect is terminated. There is no coupling in this differential pair. Simulated with Agilent ADS.

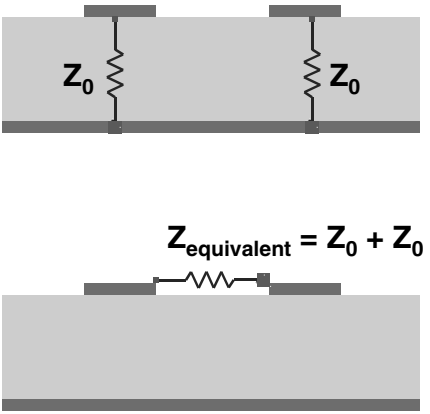


Figure 11-8 The impedance between each line and its return path in a differential pair and the equivalent impedance between the two signal lines.

where:

Z_{diff} = the equivalent impedance between the signal lines, or the differential impedance

Z_0 = the impedance of each line to the return path

If all we had to worry about was the differential impedance of uncoupled transmission lines, we would essentially be done. The differential impedance of a differential pair is always two times the single-ended impedance of either line, as seen by either driver. Two important factors complicate the real world. First is the impact from coupling between the two lines and second is the role of the common signal and its generation and control.

11.4 The Impact from Coupling

If we bring two stripline traces closer and closer together, their fringe electric and magnetic fields will overlap and the coupling between them will increase. Coupling is described by the mutual capacitance per length, C_{12}, and mutual inductance per length, L_{12}. (Unless otherwise noted, the capacitance-matrix elements always refer to the SPICE capacitance-matrix elements, not the Maxwell capacitance-matrix elements.)

As the traces are brought together, both C_{11} and C_{12} will change. C_{11} will decrease as some of the fringe fields between signal line 1 and its return path are

TIP It is important to note that the coupling, described by the capacitance- and inductance-matrix elements are completely independent of any applied voltages. They are purely related to the geometry and material properties of the collection of conductors.

intercepted by the adjacent trace, and C_{12} will increase. However, the loaded capacitance, $C_L = C_{11} + C_{12}$, will not change very much. Figure 11-9 shows the equivalent capacitance circuit of two stripline traces and how C_{11}, C_{12}, and C_L vary for the specific case of two 50-Ohm stripline traces in FR4 with 5-mil-wide traces.

Both L_{11} and L_{12} will change as the traces are brought together. L_{11} will decrease very slightly (less than 1% at the closest spacing) due to induced eddy currents in the adjacent trace and L_{12} will increase.This is shown in Figure 11-10.

As the two traces are brought together, the coupling increases. However, even in the tightest spacing, the spacing equal to the line width, the maximum relative coupling, C_{12}/C_L or L_{12}/L_{11}, is less than 15%. When the spacing is more than 15 mils, the relative coupling is reduced to 1%—a negligible amount. Figure 11-11 shows how this ratio, both the relative capacitive coupling and the relative inductive coupling, changes as the separation changes.

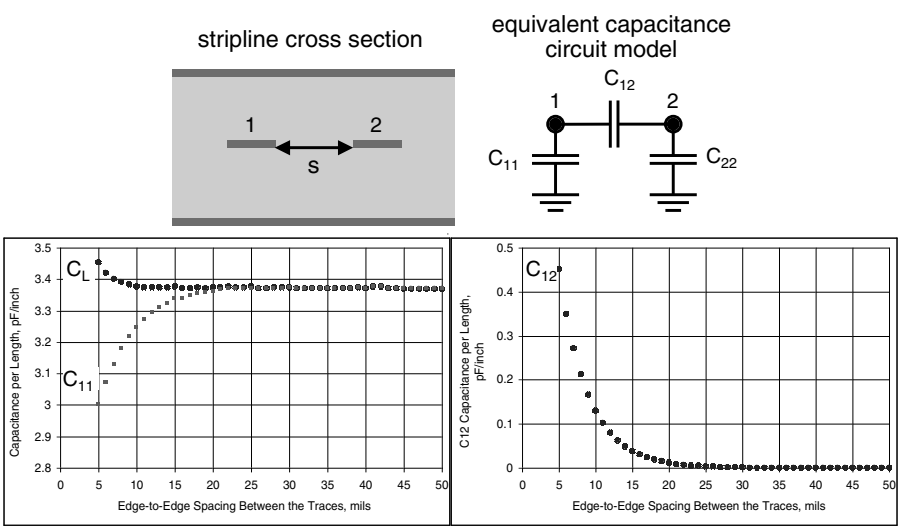


Figure 11-9 Variation in the loaded capacitance per length, C_L, and the SPICE diagonal capacitance per length, C_{11}. Also plotted is the variation in the coupling capacitance, C_{12}. Simulated with Ansoft's SI2D.

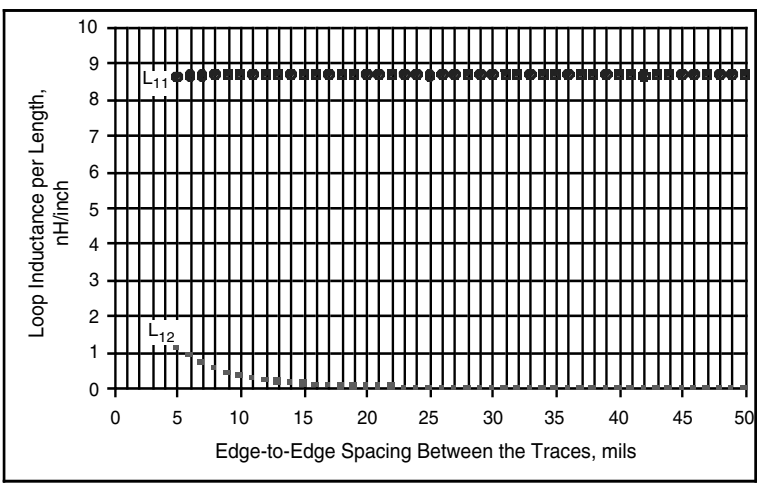


Figure 11-10 Variation in the loop self-inductance per length, L_{11}, and the loop mutual inductance per length, L_{12}. Simulated with Ansoft's SI2D.

When the lines are far apart, the characteristic impedance of line 1 is completely independent of the other line. The characteristic impedance varies inversely with C_{11}:

$$Z_0 \sim \frac{1}{C_{11}} \tag{11-8}$$

where:

Z_0 = the characteristic impedance of the line

C_{11} = the capacitance between the signal trace and the return path

When the lines are brought closer together, the presence of the other line will affect the impedance of line 1. This is called the *proximity effect*. If the second line is tied to the return path, i.e., a 0-v signal is applied to line 2, and only line 1 is being driven, the impedance of line 1 will depend on its loaded capacitance in the proximity of the other line. The characteristic impedance of the driven line will be related to the capacitance per length that is being driven:

$$Z_0 \sim \frac{1}{C_{11} + C_{12}} = \frac{1}{C_L} \tag{11-9}$$

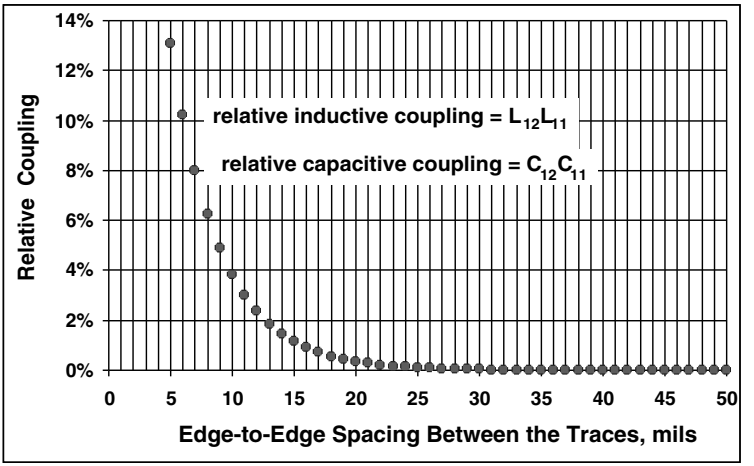


Figure 11-11 Relative mutual capacitance and mutual inductance as the spacing between two 50-Ohm, 5-mil stripline traces in FR4 changes. The relative coupling capacitance and coupling inductance are identical for a homogeneous dielectric structure such as stripline. Simulated with Ansoft's SI2D.

where:

Z_0 = the characteristic impedance of the line

C_{11} = the capacitance per length between the signal trace and the return path

C_{12} = the capacitance per length between the two adjacent signal traces

C_L = the loaded capacitance per length of one line

As the traces move closer together, the impedance of line 1 will decrease, but only very slightly, less than 1%. Figure 11-12 shows the change in the single-ended characteristic impedance of line 1 as the two traces are brought into proximity. The single-ended characteristic impedance is essentially unchanged if the second line is pegged low and the pair of traces brought closer together.

However, suppose the second trace is also being driven and the signal on line 2 is the opposite of the signal on line 1. As the signal on line 1 increases from 0 v to 1 v, the signal on line 2 is simultaneously dropped from 0 v to –1 v. As the driver on line 1 turns on, it will drive a current through the C_{11} capacitance due to the dV_{11}/dt between line 1 and the return path. In addition, there will also be a current from line 1 to line 2, due to the changing voltage between them, dV_{12}/dt. This voltage will be twice the voltage change between line 1 and its return, $V_{12} = 2 \times V_{11}$.

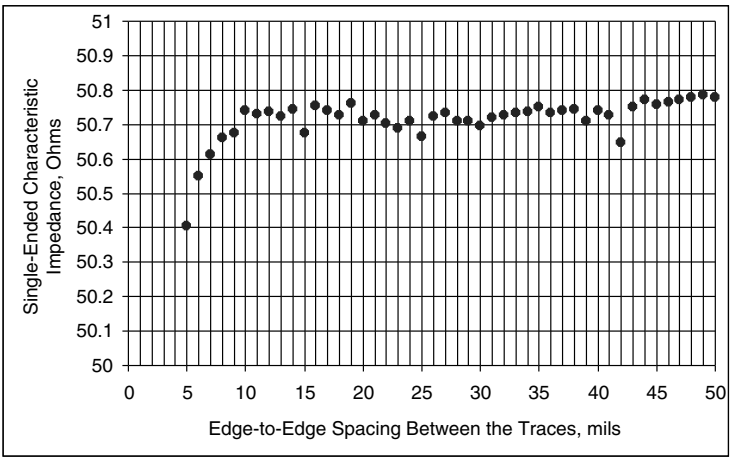


Figure 11-12 Single-ended characteristic impedance of one line when the second is shorted to the return path and the separation changes, for 5-mil-wide, 50-Ohm striplines in FR4. The small fluctuations, on the order of 0.05 Ohm, are due to the numerical noise of the simulation tool. Simulated with Ansoft's SI2D.

The current into one signal line will be related to:

$$I_{one} = v \times RT \times \left(C_{11}\frac{dV_{11}}{dt} + C_{12}\frac{dV_{12}}{dt} \right) \sim C_{11}V_{one} + 2C_{12}V_{one} = V_{one}(C_L + C_{12}) \quad (11\text{-}10)$$

where:

I_{one} = the current going into one line

v = the speed of the signal moving down the signal line

C_{11} = the capacitance between the signal line and its return path per length

V_{11} = the voltage between the signal line and the return path

C_{12} = the capacitance between each signal line per length

V_{12} = the voltage between each signal line

V_{one} = the voltage change between the signal line and return path of one line

RT = the rise time of the transition

As the two traces move together and they are both being driven by signals transitioning in opposite directions, the current from the driver into line 1, and out

the return path, will increase in order to drive the higher capacitance of the single-ended line.

> **TIP** If the current increases for the same applied voltage, the input impedance the driver sees will decrease. The characteristic impedance of the single-ended line will drop if the second line is driven with the opposite signal.

Suppose the second line is driven with exactly the same signal as the first line. The driver to line 1 will see less capacitance to drive, since there will be no voltage between the two signal lines and the driver will see only the C_{11} capacitance. When the second line is driven with the same voltage as line 1, the current into signal line 1 will be:

$$I_{one} = v \times RT \times \left(C_{11} \frac{dV_{11}}{dt} \right) \sim C_{11} V_{one} = V_{one}(C_L - C_{12}) \quad (11\text{-}11)$$

where:

I_{one} = the current going into one line

v = the speed of the signal moving down the signal line

C_{11} = the capacitance between the signal line and its return path per length

V_{11} = the voltage between the signal line and the return path

C_{12} = the capacitance between each signal line per length

V_{one} = the voltage change between the signal line and return path of one line

RT = the rise time of the transition

We see that the single-ended characteristic impedance of one line, when in the proximity of a second line, is not a unique value. It depends on how the other line is being driven. If the second line is pegged low, the impedance will be close to the uncoupled impedance. If the second line is switching opposite the first line, the impedance of the first line will be lower. If the second line is switching exactly the same as the first line, the impedance of the first line will be higher. Figure 11-13 shows how the characteristic impedance of the first line varies as the separation changes for these three cases.

This is a critically important observation. When we only dealt with single-ended signals, a transmission line only had one impedance that described it. But,

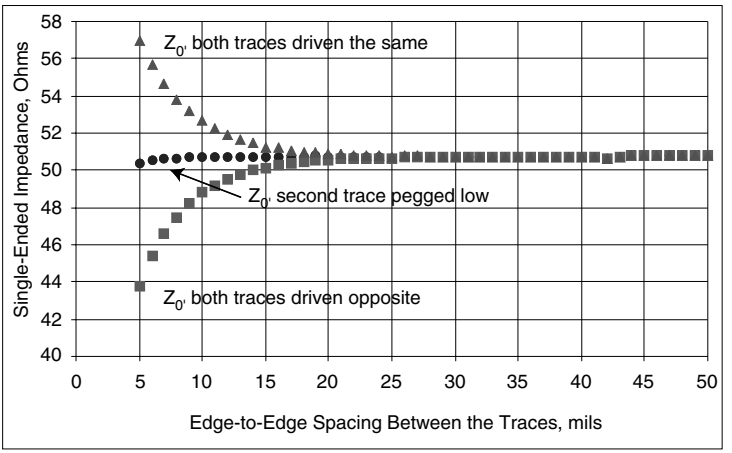


Figure 11-13 The characteristic impedance of line 1 when the second line is pegged low, switching opposite and switching the same, as the separation between the traces changes, for 5-mil-wide, 50-Ohm striplines in FR4. Simulated with Ansoft's SI2D.

when it is part of a pair and coupling plays a role, it now has three different impedances that describe it. We need to identify new labels to describe which of the three different impedances we talk about when referring to the impedance of a transmission line when it is part of a pair. This is why we will introduce the terms odd and even mode impedance in a later section, to provide a clear and unambiguous language for describing the properties of differential pairs.

TIP When trace separations are closer than about three line widths, the presence of the adjacent trace will affect the characteristic impedance of the first trace. Its proximity as well as the manner in which it is being driven must be considered.

When a differential signal travels down a differential pair, the impedance the differential signal sees will be the series combination of the impedance of each line to its respective return path. The differential signal drives opposite signals on each of the two lines. As we saw above, the impedance of each line, when the pair is driven by a differential signal, will be reduced due to the coupling between the two lines. The differential impedance, with coupling, will still be twice the characteristic impedance of either line. Figure 11-14 shows the differential impedance as the spacing between the lines decreases. At the closest spacing that can be realistically manufactured (i.e., a spacing equal to the line width), the differential imped-

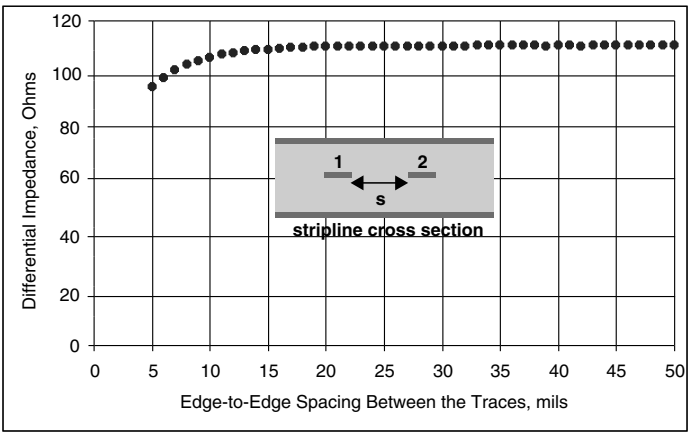


Figure 11-14 The differential impedance of a pair of 5-mil-wide, 50-Ohm stripline traces in FR4, as the spacing between them decreases. Simulated with Ansoft's SI2D.

ance of a pair of coupled stripline traces is reduced only about 12% from when the striplines are three line widths apart and uncoupled.

11.5 Calculating Differential Impedance

To describe this at most 12% effect, additional formalism must be introduced. The complexity arises when quantifying the decrease in differential impedance as the traces are brought closer together and coupling begins to play a role. There are five different approaches to this analysis:

1. Use the direct results from an approximation.

2. Use the direct results from a field solver.

3. Use an analysis based on modes.

4. Use an analysis based on the capacitance and inductance matrix.

5. Use an analysis based on the characteristic impedance matrix.

There is only one reasonably close approximation useful for calculating the differential impedance of either an edge-coupled microstrip or an edge-coupled stripline, originally offered in a National Semiconductor application note (AN-905) by James Mears. These approximations are based on empirical fitting of measured data.

For edge-coupled microstrip using FR4 material, the differential impedance is approximately:

$$Z_{diff} = 2 \times Z_0\left[1 - 0.48\exp\left(-0.96\frac{s}{h}\right)\right] \tag{11-12}$$

where:

Z_{diff} = the differential impedance in Ohms

Z_0 = the uncoupled single-ended characteristic impedance for the line geometry

s = the edge-to-edge separation between the traces in mils

h = the dielectric thickness between the signal trace and the return plane

For an edge-coupled stripline with FR4, the differential impedance is:

$$Z_{diff} = 2 \times Z_0\left(1 - 0.37\exp\left(-2.9\frac{s}{b}\right)\right) \tag{11-13}$$

where:

Z_{diff} = the differential impedance, in Ohms

Z_0 = the uncoupled single-ended characteristic impedance for the line geometry

s = the edge-to-edge separation between the traces, in mils

b = the total dielectric thickness between the planes

We can evaluate how accurate these approximations are by comparing them to the predicted differential impedance calculated with an accurate field solver. In Figure 11-15, we compare these approximations to three different cross sections for edge-coupled microstrip and edge-coupled stripline. In each case, we use the field-solver-calculated single-ended characteristic impedance in the approximation, and the approximation predicts the slight perturbation on the differential impedance from the coupling. The accuracy varies from 1%–10%, provided we have an accurate starting value for the characteristic impedance.

A more accurate tool, which will also predict the single-ended characteristic impedance of either line, is a 2D field solver. The 2D field solver requires input regarding the cross-sectional geometry and material properties. It offers output including, among other quantities, the differential impedance of the pair of lines.

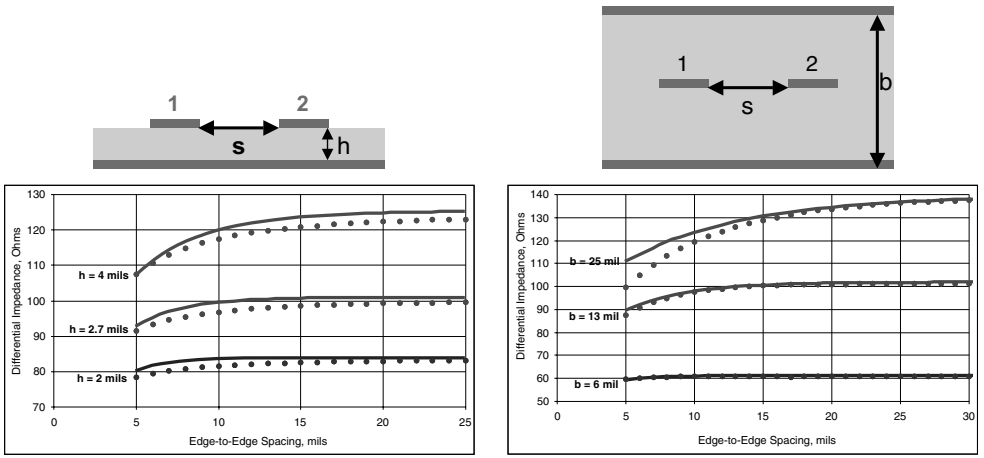


Figure 11-15 Comparing the accuracy of the differential impedance approximation with the results from a 2D field solver. In each case, the line width is 5 mils and the material is FR4. The lines are the approximations and the points are the simulation using Ansoft's SI2D.

The advantage of a field solver is that some tools can be accurate to better than 1% across a wide range of geometrical conditions. They account for the first-order effects, such as line width, dielectric thickness, and spacing, as well as second-order effects such as trace thickness, trace shape, and inhomogeneous dielectric distributions.

TIP When accuracy is important, such as in signing off on a drawing for the fabrication of a circuit board, the only tool that should be used is a verified 2D field solver. Approximations should never be used to sign off on a design.

However, limited to using only a field solver to calculate differential impedance, we will lack the terms needed to describe the behavior of common signals, terminations, and cross talk. In the following sections, we will introduce the concepts of odd and even modes and how they relate to differential and common impedance. From this foundation, we will evaluate termination strategies for differential and common signals.

We will end up introducing the description of two coupled transmission lines in terms of the capacitance and inductance matrix and the characteristic impedance matrix to calculate the odd and even modes and the differential and common impedance. This is the most fundamental description and can be generalized to *n*

different coupled lines. Ultimately, this is what is really going on inside most field solver and simulation tools.

Even without invoking more complex descriptions at this point, we have the terminology to be able to use the results from a field solver to design for a target differential impedance and evaluate another important quality of differential pairs—the current distribution.

11.6 The Return-Current Distribution in a Differential Pair

When the spacing between two edge-coupled microstrip transmission lines is greater than three times their line width, the coupling between the signal lines will be small. In this situation, as we might expect, when we drive them with a differential signal, there is some current into each signal line and an equal amount in the opposite direction, in the return plane. An example of the current distribution for a 100-MHz differential signal into microstrip conductors that are 1.4 mils thick, or for 1-ounce copper, is illustrated in Figure 11-16.

The direction of the current into line 1 is into the paper. The return current in the return plane below line 1 is out of the paper. Likewise, at the same time, the current into line 2 is out of the paper and the current into its return path is into the paper. In the return plane, the return-current distributions are localized under the signal lines. When driven by a differential signal, there is virtually no overlap between the two return-current distributions in the plane.

If we pay attention only to the current in the signal traces, it looks like the same current goes into one trace and comes out the other. We might conclude that the return current of the differential signal in one trace is carried by the second trace. While it is perfectly true that the same amount of current going into one signal trace comes out of the other signal trace, it is not the complete story.

Because of the wide spacing between the differential pairs, there is no overlap of the return currents in the planes when driven with a differential signal.

Figure 11-16 The current distribution, at 100 MHz, in a pair of coupled 50-Ohm microstrip transmission lines, 5 mils wide and separated by 15 mils. Brighter colors mean higher current density. The current density scale in the plane is 10 times more sensitive than the traces to show the current distribution more clearly. Simulated with Ansoft's SI2D.

While it is true the net current into the plane is zero, there is still a well-defined, localized current distribution in the plane under each signal path. Anything that will distort or change this current distribution will change the differential impedance of the pair of traces.

After all, the presence of the plane defines the single-ended impedance of either trace. Increase the plane spacing, and the single-ended impedance of the trace increases, which will change the differential impedance.

Even in the extreme case for edge-coupled microstrip, bringing the signal traces as close together as is practical with a spacing equal to the line width, the degree of overlap of the currents in the return plane is very slight. The comparison of the current distributions is shown in Figure 11-17.

TIP When the coupling between the signal line and the return plane is much larger than the coupling between adjacent signal traces in a differential pair, there are separate and distinct return currents in the planes and very little overlap in the return-current distributions. These current distributions will strongly affect the differential impedance of the pair. Disturb the current distribution and the differential impedance will be affected.

For any pair of single-ended transmission lines sharing a common return conductor, if the return conductor is moved far enough away, a differential signal's return-current distribution in the return conductor will completely overlap and cancel each other out, and the presence of the return conductor will play no role in

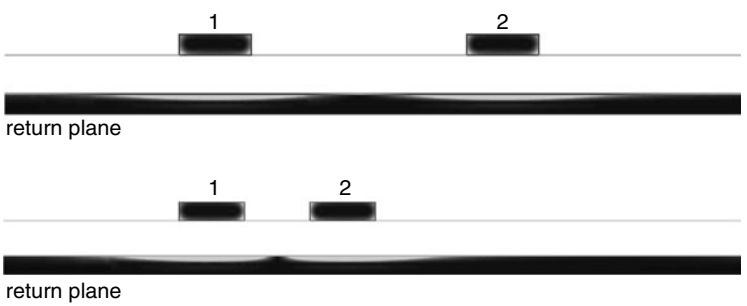


Figure 11-17 The current distribution, at 100 MHz, in a pair of coupled 50-Ohm microstrip transmission lines, 5-mils-wide and separated by 5 mils compared with a separation of 15 mils. Brighter color means higher current density. The current density scale in the planes is 10 times more sensitive than the traces to show the current distribution more clearly. Simulated with Ansoft's SI2D.

establishing the differential impedance. In this specific situation, it is absolutely true that the return current of one line would be carried by the other line.

There are three such cases that should be noted:

1. Edge-coupled microstrip with return plane far away

2. Twisted-pair cable

3. Broadside-coupled stripline with return planes far away

In edge-coupled microstrip, the coupling between the two signal traces is largest when the spacing between the signal lines is as close as can be fabricated, typically equal to the line width. For the case of near-50-Ohm lines and closest spacing, as shown in Figure 11-17, there is significant return-current distribution in the return plane, and the presence of the plane affects the differential impedance. If the plane is moved farther away, the single-ended impedance of either line will increase and the differential impedance will increase. However, as the plane is moved farther away, the return currents of the differential signal, in the plane, will increasingly overlap.

There is a point where the return currents overlap so much there is no current in the plane and the plane will not influence the differential impedance. This is illustrated in Figure 11-18. The single-ended impedance continues to get larger, but the differential impedance reaches a highest value of about 140 Ohms and then

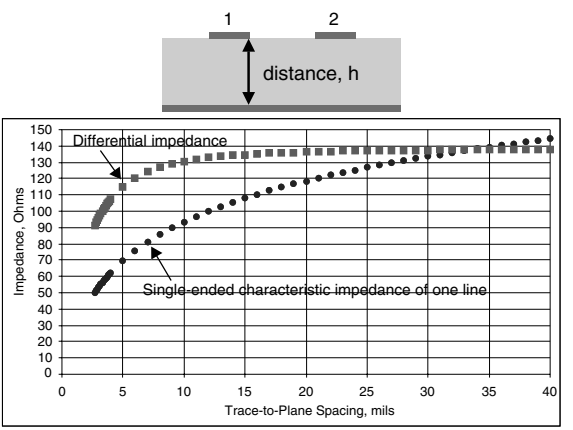


Figure 11-18 Single-ended and differential impedance of a pair of edge-coupled microstrips with 5-mil width and 5-mil spacing, as the distance to the return plane increases. Simulated with Ansoft's SI2D.

stops increasing. This is when the return currents completely overlap, at a trace-to-plane distance of about 15 mils.

> **TIP** As a rough rule of thumb, when the distance to the return plane is about equal to or larger than the total edge-to-edge span of the two signal conductors, the current distributions in the return path overlap and the presence of the return path plays no role in the differential impedance of the pair. In this case, it is true that for a differential signal, the return current of one signal line really is carried by the other line.

After all, in this geometry of edge-coupled microstrip with the plane far away, isn't this really like a coplanar transmission line with a single-ended signal? The signal in each case is the voltage between the two signal lines. The single-ended signal is the same as a differential signal, so the impedance the single-ended signal sees will be the same impedance as a differential signal sees. The single-ended characteristic impedance of a coplanar line with thick dielectric on one side is the same as the differential impedance of an edge-coupled pair with the return plane very far away.

In a shielded twisted pair, the return path of each signal line is the shield. The spacing between the twisted wires is determined by the thickness of the insulation. In some cables, the center-to-center pitch of the wires is 25 mils and the wire diameter is 16 mils, or 26-gauge wire. We can use a 2D field solver to calculate the differential impedance as the spacing to the shield increases.

When one wire of a twisted pair is driven single ended, with the shield as the return conductor, the signal current flows down the wire and the return current flows roughly symmetrically in the outer shield. Likewise, when the second wire is driven single ended, its return current will have about the same current distribution, but in the opposite direction. When driven by a differential signal and both wires are nearly at the center of the shield, their return-current distributions flow in opposite directions and overlap. There will be no residual-current distribution in the shield. In this case, the shield plays no role in influencing the differential impedance of the wires and can be eliminated.

When the shield is very close to the wires, the off-axis location of the two wires causes their current distributions to be slightly different in the shield, and the differential impedance will depend slightly on the location of the shield. When the shield is far enough away for the return current to be mostly symmetric, the two return currents overlap and the shield location has no effect on the differential

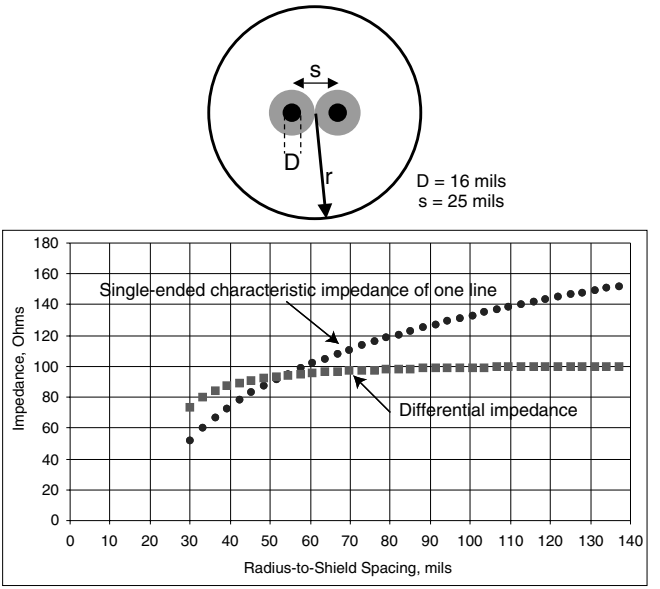


Figure 11-19 Variation in the single-ended impedance of one wire and the shield in a twisted pair, and the differential impedance of the two wires, as the radius of the shield expands. When the distance is more than about 2 times the center-to-center spacing of the twisted pair, the return currents in the shield overlap and cancel out and the presence of the shield plays no role in the differential impedance. Simulated with Ansoft's SI2D.

impedance. Figure 11-19 shows how the differential impedance varies as the shield radius increases. When the shield radius is about equal to twice the pitch of the wires, the currents mostly overlap and the differential impedance is independent of the shield position.

An unshielded differential pair has exactly the same differential impedance as a shielded differential pair with a large radius shield. As far as the differential impedance is concerned, the shield plays no role at all. As we will see, the shield plays a very important role in providing a return path for the common current, which will reduce radiated emissions.

The same effect happens in broadside-coupled stripline. When the two reference planes are close together, there is significant, separated return current in the two planes when the transmission lines are driven by a differential signal, and the presence of the planes affects the differential impedance. As the spacing between the planes increases so the return current from each line has roughly the same current distribution in each plane, the currents cancel out in each plane and the impact from the planes is negligible.

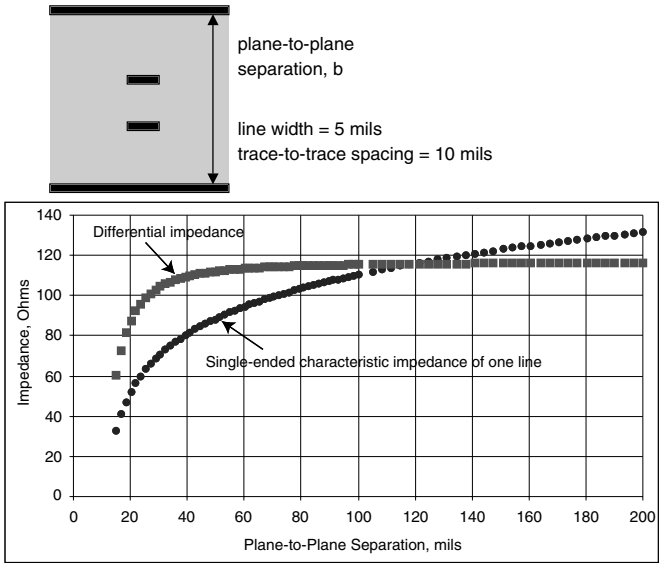


Figure 11-20 Variation in the single-ended impedance of one trace and the planes in a broadside-coupled stripline, and the differential impedance of the two traces, as the plane-to-plane spacing increases. Simulated with Ansoft's SI2D.

Figure 11-20 shows the differential impedance as the spacing between the planes is increased. In this example, the line width is 5 mils and the trace-to-trace spacing is 10 mils. This is a typical stack-up for a 100-Ohm differential-impedance pair, when the plane-to-plane spacing is about 25 mils. When the distance between a trace and the nearest plane is greater than about twice the separation of the traces, or 20 mils in this case, and the plane-to-plane spacing is more than 50 mils, the differential impedance is independent of the position of the planes.

These three examples illustrate a very important principle with differential pairs. When the coupling between any one trace and the return plane is large compared with the coupling between the two signal lines, there is significant return current in the plane, and the presence of the plane is very important in determining the differential impedance of the pair.

When the coupling between the two traces is much larger than the coupling between a trace and the return plane, there is a lot of overlap of the return currents and they mostly cancel each other out in the plane. In this case, the plane plays no role and will not influence the differential signal. It can be removed without affecting the differential impedance. In this case, it is absolutely true that the second line will carry the return current of the first line.

TIP As a rough rule of thumb, for the coupling between the two traces to be larger than the coupling between a trace and the return plane, the distance to the nearest plane must be about twice as large as the span of the signal lines.

In most board-level interconnects, the coupling between a signal trace and plane is much greater than the coupling between the two signal traces, so the return current in the plane is very important. In board-level interconnects, it generally is not true that the return current of one trace is carried in the other trace.

However, if the return path is removed, as in a gap, the coupling between the traces now dominates and in this region of discontinuity, it may be true that the return current of one line is carried by the other. In the case of a discontinuity in the return path, the change in differential impedance of the pair can be minimized by using tighter coupling between the pair in the region of the return-path discontinuity. This is discussed later in this chapter.

In connectors, the trace-to-trace coupling is usually greater than the trace-to-return-pin coupling, so it is generally true that the return current of one pin is carried by the other. The only way to know for sure is to put in the numbers using a field-solver tool.

11.7 Odd and Even Modes

Any voltage can be applied to the front end of a differential pair. If we were to launch the voltage pattern of a 0-v to 1-v signal in line 1 and a 0-v constant signal in line 2, we would find that as we move down the lines with the signal, the actual signal on the lines will change. There will be far-end cross talk between line 1 and line 2. Noise will be generated on line 2 and as the noise builds, the signal on line 1 will decrease.

Figure 11-21 shows the evolution of the voltages on the two lines as the signals propagate. The voltage pattern launched into the differential pair changes as it propagates down the line. In general, any arbitrary voltage pattern we launch into a pair of transmission lines will change as it propagates down the line.

However, in the case of an edge-coupled microstrip differential pair, there are two special voltage patterns we can launch into the pair that will propagate down the line undistorted. The first pattern is when exactly the same signal is applied to either line; for example, the voltage transitions from 0 v to 1 v in each line.

In this case, there is no dV/dt between the signal lines, so there is no capacitively coupled current. The inductively coupled current into one line is identical to

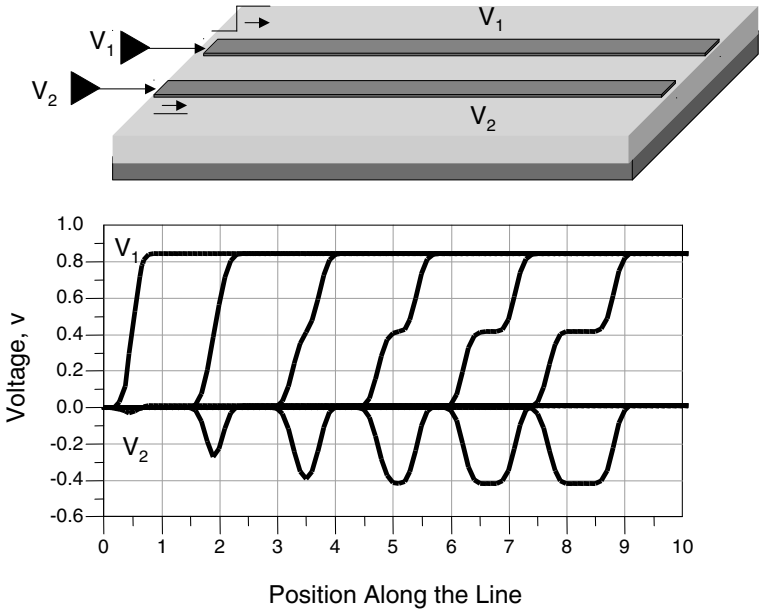


Figure 11-21 Voltage pattern on the two lines in an edge coupled microstrip, when one line is driven by a 0-v to 1-v transition and the other line is pegged low. Simulated with Agilent's ADS.

the inductively coupled current into the other line since the dI/dt is the same in each line. Whatever one line does to the other, the same action is returned. The result is that as this special voltage pattern propagates down the transmission lines, the voltage pattern on either line will remain exactly the same.

The second special voltage pattern that will propagate unchanged down the differential pair is when the opposite-transitioning signals are applied to each line; for example, one of the signals transitions from 0 v to 1 v and the other goes from 0 v to –1 v.

The signal in the first line will generate far-end noise in the second line that has a negative-going pulse. This will decrease the voltage on the first line as the signal propagates. However, at the same time, the negative-going signal in the second line will generate a positive far-end noise pulse in the first line. The magnitude of the positive noise generated in line 1 is exactly the same as the magnitude of the drop in the signal in line 1 from the loss due to its noise to line 2. The voltage pattern on the differential pair will propagate down the lines undistorted. Figure 11-22 shows the voltage pattern on the two signal lines as the signal propagates down the pair of lines when these two patterns are applied.

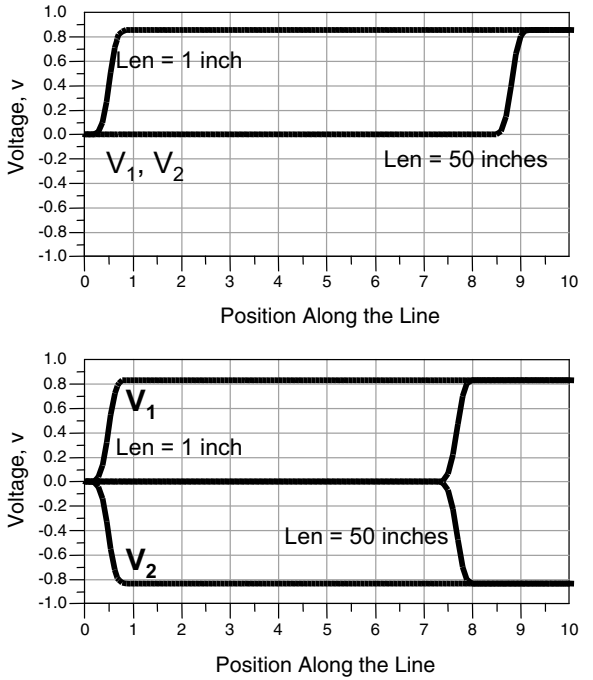


Figure 11-22 Voltage pattern on the two lines in an edge-coupled microstrip, when the pair is driven in the even mode and the odd mode. The voltage pattern is the same going into the line as coming out 50 inches later. Simulated with Agilent's ADS.

These two special voltage patterns that propagate undistorted down a differential pair correspond to two special states at which the pair of lines can be excited or stimulated. We call these special states *the modes of the pair.*

When a differential pair is excited into one of these two modes, the signal will have the special property of propagating down the line undistorted. To distinguish these two states, we call the state where the same voltage drives each line *the even mode* and the state where the opposite-going voltages drive each line *the odd mode*.

T I P Modes are special states of excitation for a pair of lines. A signal that excites one of these special states will propagate undistorted down a differential pair of transmission lines. For a two-signal-line differential pair, there are only two different special states or modes. For a three-signal conductor array of coupled lines, there are three modes. For a collection of four-signal conductors and their common return path, there are four special voltage states for which a voltage pattern will propagate undistorted.

Modes are intrinsic properties of a differential pair. Of course, any voltage pattern can be imposed on a pair of lines. When the voltage pattern imposed matches one of these special states, the voltage signal propagating down the line has special properties. The modes of a differential pair are used to define the special voltage patterns.

When the two conductors in a differential pair are geometrically symmetric with the same line widths and dielectric spacings, the voltage patterns that excite the even and odd modes correspond to the same voltages launched in both lines and the opposite-going voltages launched into both lines. If the conductors are not symmetric (e.g., have different line widths or dielectric thickness), the odd- and even-mode voltage patterns are not so simple. The only way to determine the specific odd- and even-mode voltage patterns is by using a 2D field solver. Figure 11-23 shows the field patterns of the odd- and even-mode states for a pair of symmetric lines.

It is important to keep separate, on one hand, the definition of modes as special, unique states into which the pair of lines can be excited based on the geometry of the pair, and, on the other hand, the applied voltages that can be any arbitrary values. Any voltage pattern can be applied to a differential pair—just connect a function generator to each signal and return path.

In the special case of a symmetric, edge-coupled microstrip differential pair, the odd-mode state can be excited by the voltage pattern corresponding to a pure differential signal. Likewise, the even-mode state can be excited by applying a pure common signal.

TIP A differential signal will drive a signal in the odd mode, and a common signal will drive a signal in the even mode for a symmetric, edge coupled microstrip differential pair. Odd and even modes refer to special intrinsic states of the differential pair. *Differential* and *common* refer to the specific signals that are applied to the differential pair. It's important to note that 90% of the confusion about differential impedance arises from misusing these terms.

Introducing the terms *odd mode* and *even mode* allows us to label the special properties of a symmetric, differential pair. For example, as we saw above, the impedance a signal sees on one line depends on the proximity of the adjacent line and the voltage pattern on the other line. Now we have a way of labeling the different cases. We refer to the impedance of one line, when the pair is driven in the odd-mode state as the odd-mode impedance of the line. The impedance of one line when the pair is driven in the even-mode state is the even-mode impedance of that line.

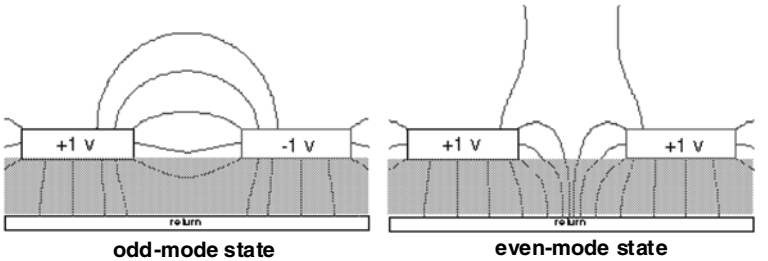


Figure 11-23 Field distribution for odd- and even-mode states of a symmetric microstrip, calculated with Mentor Graphics Hyperlynx.

The odd mode is often incorrectly labeled as the differential mode. If we equate the odd mode with the differential mode, then we might easily confuse the differential-mode impedance with the odd-mode impedance. If these were the same mode, why would there be any difference between the odd-mode impedance and the differential-mode impedance?

In fact, there is no such thing as the differential mode, so there is no such thing as the differential-mode impedance. Figure 11-24 emphasizes that we should remove the term *differential mode* from our vocabulary and then we will never be confused between odd-mode impedance and differential impedance. They are completely different quantities. There are odd-mode impedances, differential signals, and differential impedances.

TIP The odd-mode impedance is the impedance of one line when the pair is driven in the odd-mode state. The differential impedance is the impedance the differential signal sees as it propagates down the differential pair.

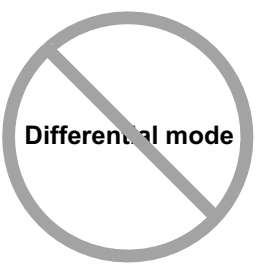


Figure 11-24 There is no such thing as differential mode. Forget the word and you will never confuse differential impedance with odd-mode impedance.

11.8 Differential Impedance and Odd-Mode Impedance

As we saw previously, the differential impedance a differential signal sees is the series combination of the impedance of each line to the return path. When there is no coupling, the differential impedance is just twice the characteristic impedance of either line. When the lines are close enough together for coupling to be important, the characteristic impedance of each line changes.

When a differential signal is applied to a differential pair, we now see that this excites the odd-mode state of the pair. By definition, the characteristic impedance of one line when the pair is driven in the odd mode is called the odd-mode characteristic impedance. As illustrated in Figure 11-25, the differential impedance is twice the odd-mode impedance. The differential impedance is thus:

$$Z_{diff} = 2 \times Z_{odd} \tag{11-14}$$

where:

Z_{diff} = the differential impedance

Z_{odd} = the characteristic impedance of one line when the pair is driven in the odd mode

TIP The way to calculate or measure the differential impedance is by first calculating or measuring the odd-mode impedance of the line and multiplying it by 2.

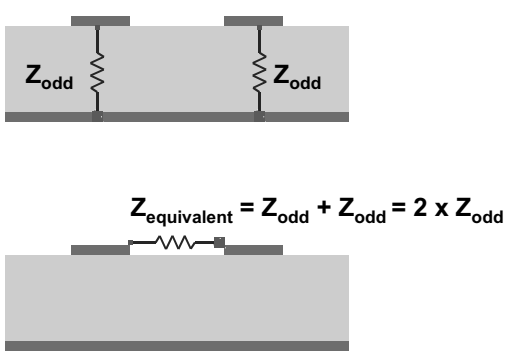


Figure 11-25 The impedance between each line and its return path is the odd-mode impedance when the pair is excited with a differential signal. The differential impedance is the equivalent impedance between the two signal lines.

The odd-mode impedance is directly related to the differential impedance, but they are not the same. The differential impedance is the impedance the differential signal sees. The odd-mode impedance is the impedance of one line when the pair is driven in the odd mode.

11.9 Common Impedance and Even-Mode Impedance

Just as we can describe the impedance a differential signal sees propagating down a line, we can also describe the impedance a common signal will see propagating down a differential pair. The common signal is the average voltage between the signal lines. A pure common signal is the absence of a differential signal, which is the same voltage applied to each line.

A common signal will drive a differential pair in its even-mode state. As the common signal propagates down the line, the characteristic impedance of each line will be, by definition, the even-mode characteristic impedance. The common signal will see the two lines in parallel, illustrated in Figure 11-26. The impedance the common signal sees will be the parallel combination of the impedances of each line. The parallel combination of the two even-mode impedances is:

$$Z_{comm} = Z_{equiv} = \frac{Z_{even} \times Z_{even}}{Z_{even} + Z_{even}} = \frac{1}{2} Z_{even} \tag{11-15}$$

where:

Z_{comm}= the common impedance

Z_{even} = the characteristic impedance of one line when the pair is driven in the even mode

The impedance the common signal sees is, in general, a low impedance. This is because the common signal is the same voltage as applied between each signal line and the return path, but the current going into the pair of lines and coming out the return path is twice the current going into either line. If a signal sees the same voltage, but twice the current, the impedance will look half as large.

TIP For two, uncoupled 50-Ohm transmission lines that make up a differential pair, the odd- and even-mode impedance will be the same—50 Ohms. The differential impedance will be 2 × 50 Ohms = 100 Ohms, while the common impedance will be 1/2 × 50 Ohms = 25 Ohms.

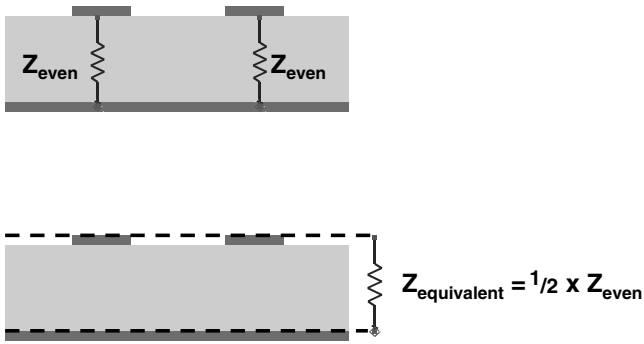


Figure 11-26 The impedance between each line and its return path is the even-mode impedance when the pair is excited with a common signal. The common impedance is the equivalent impedance between the two signal lines and the return plane.

As we turn on coupling, the odd-mode impedance of either line will decrease and the even-mode impedance of either line will increase. This means the differential impedance will decrease and the common impedance will increase. The most accurate way of calculating the differential impedance or common impedance is by using a field solver to first calculate the odd-mode impedance and even-mode impedance.

Figure 11-27 shows the complete set of impedances calculated with a field solver for the case of an edge-coupled microstrip in FR4 with 5-mil-wide traces and an uncoupled single-end characteristic impedance of 50 Ohms. As the separation between the traces gets smaller, coupling increases and the odd-mode impedance decreases, causing the differential impedance to decrease. The even-mode impedance increases causing the common impedance to increase. As this example illustrates, even with the tightest coupling manufacturable, the differential impedance and common impedance are only slightly affected by the coupling. The tightest coupling decreases the differential impedance by only 10%.

In many microstrip traces on boards, a solder-mask layer is applied to the top surface. This will affect the single-ended impedance as well as the odd-mode impedance. Figure 11-28 shows the impedance variation for a tightly coupled microstrip differential pair as the thickness of solder mask increases. Because of the stronger field lines between the traces in the odd mode, the presence of the solder mask will affect the odd-mode impedance more than it affects the other impedances.

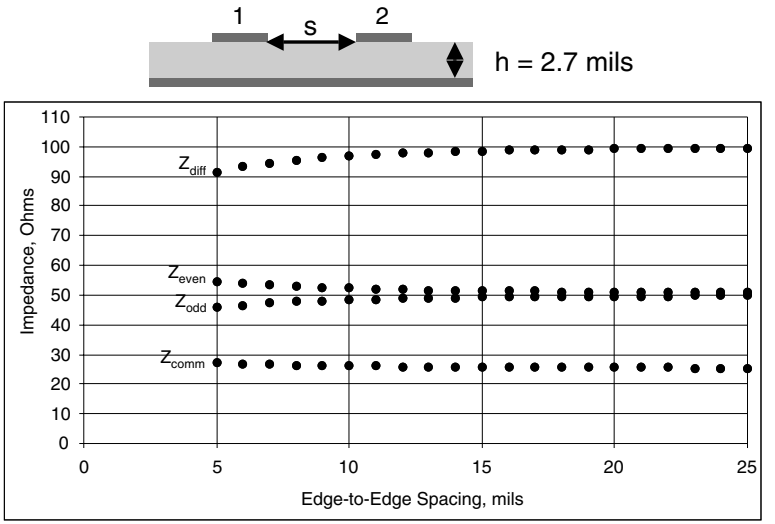


Figure 11-27 All the impedances associated with a pair of edge-coupled microstrip traces, with 5-mil-wide traces in FR4 and nominally 50 Ohms, as the separation increases. Simulated with Ansoft's SI2D.

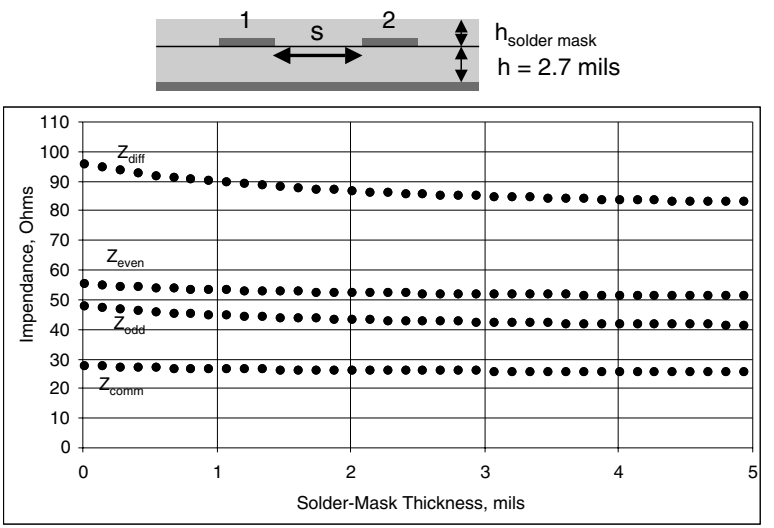


Figure 11-28 Effect on all the impedances as a solder-mask thickness applied to the top of the surface traces increases, for the case of tightest spacing of 5-mil-wide and 5-mil-spaced traces in FR4. Simulated with Ansoft's SI2D.

This is why it is so important to take into account the presence of solder mask when designing the differential impedance of a surface trace. Additionally, this effect can cause the fabricated differential impedance to be off by as much as 10%.

11.10 Differential and Common Signals and Odd- and Even-Mode Voltage Components

The terms *differential* and *common* refer to the signals imposed on a line. The part of an arbitrary signal that is the differential component is the difference between the two voltages on each signal line. The part of an arbitrary signal that is the common component is the average of the two signal lines.

For a symmetric differential pair, a differential signal travels in the odd mode of the pair and a common signal travels in the even mode of the pair. We can also use the terms *odd* and *even* to describe an arbitrary signal. The voltage component of an applied signal that is propagating in the even mode, V_{even}, is the common component of the signal. The component propagating in the odd mode, V_{odd}, is the differential component of the signal. These are given by:

$$V_{odd} = V_{diff} = V_1 - V_2 \tag{11-16}$$

$$V_{even} = V_{comm} = \frac{1}{2} \times (V_1 + V_2) \tag{11-17}$$

Likewise, any arbitrary signal propagating down a differential pair can be described as a combination of a component of the signal propagating in the even mode and a component of the signal propagating in the odd mode:

$$V_1 = V_{even} + \frac{1}{2} V_{odd} \tag{11-18}$$

$$V_2 = V_{even} - \frac{1}{2} V_{odd} \tag{11-19}$$

where:

V_{even} = the voltage component propagating in the even mode

V_{odd} = the voltage component propagating in the odd mode

V_1 = the signal on line 1, with respect to the common return path

V_2 = the signal on line 2, with respect to the common return path

For example, if a single-ended signal is imposed on one line of 0 v to 1 v and the other line is pegged low at 0 v, the voltage component that propagates in the even mode will be $V_{even} = 0.5 \times (1\text{ v} + 0\text{ v}) = 0.5$ v. The voltage component that propagates in the odd mode will be $V_{odd} = 1\text{ v} - 0\text{ v} = 1$ v. Simultaneously on the differential pair, a 0.5-v signal will be propagating down the line in the even mode, seeing the even-mode characteristic impedance of each line, and a 1-v component will be propagating down the line in the odd mode, seeing the odd-mode characteristic impedance of each line. The description of this signal in terms of odd- and even-mode components is illustrated in Figure 11-29.

Any imposed voltage can be described as a combination of an even-mode voltage component and an odd-mode component. The voltage components that travel in the odd mode are completely independent of the voltage components traveling in the even mode. They propagate independently and do not interact. Each signal component will see a different impedance for each signal line to its return path and each signal component may also travel at a different velocity.

The examples above used an edge-coupled microstrip to illustrate the different modes. When the dielectric material completely surrounding the conductors is everywhere uniform and homogeneous, there is no longer a unique voltage pattern that propagates down the differential pair in each mode. Every voltage pattern imposed on the differential pair will propagate undistorted. After all, when the

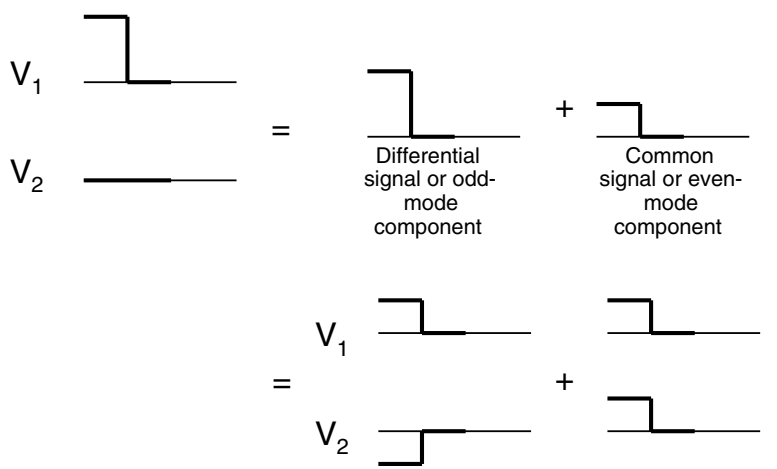


Figure 11-29 Three equivalent descriptions of the same signals on a differential pair: as the voltage on each line, as the differential and common signals, and as the components propagating in the odd mode and the even mode.

dielectric is homogeneous, as in a stripline geometry, there is no far-end cross talk. Any signal launched into the front end of the pair of lines will propagate undistorted down the line. However, by convention, we still use the voltage patterns above to define the odd and even modes for any symmetric differential pair.

11.11 Velocity of Each Mode and Far-End Cross Talk

The description of the signal in terms of its components propagating in each of the two modes is especially important in edge-coupled microstrip because signals in each mode travel at different speeds.

The velocity of a signal propagating down a transmission line is determined by the effective dielectric constant of the material the fields see. The higher the effective dielectric constant, the slower the speed, and the longer the time delay of a signal propagating in that mode. In the case of a stripline, the dielectric material is uniform all around the conductors and the fields always see an effective dielectric constant equal to the bulk value, independent of the voltage pattern. The odd- and even-mode velocities in a stripline are the same.

However, in a microstrip, the electric fields see a mixture of dielectric constants, part in the bulk material and part in the air. The precise pattern of the field distribution and how it overlaps the dielectric material will influence the value of the resulting effective dielectric constant and the actual speed of the signal. In the odd mode, more of the field lines are in air; in the even mode, more of the field lines are in the bulk material. For this reason, the odd-mode signals will have a slightly lower effective dielectric constant and will travel at a faster speed than do the even-mode signals.

Figure 11-30 shows the field patterns of the odd and even modes for a symmetric microstrip and stripline differential pair. In a stripline, the fields see just the bulk dielectric constant for each mode. There is no difference in speed between the modes for any homogeneous dielectric interconnect.

In an edge-coupled microstrip, a differential signal will drive the odd mode so it will travel faster than a common signal, which drives the even mode. Figure 11-31 shows the different speeds for these two signals. As the spacing between the traces increases, the degree of coupling decreases and the field distribution between the odd mode and the even mode becomes identical. If there is no difference in the field distributions, each mode will have the same effective dielectric constant and the same speed.

In this example, at closest spacing, the odd-mode speed is 7.4 inches/nsec, while the even-mode speed is 6.8 inches/nsec. If a signal were launched into the

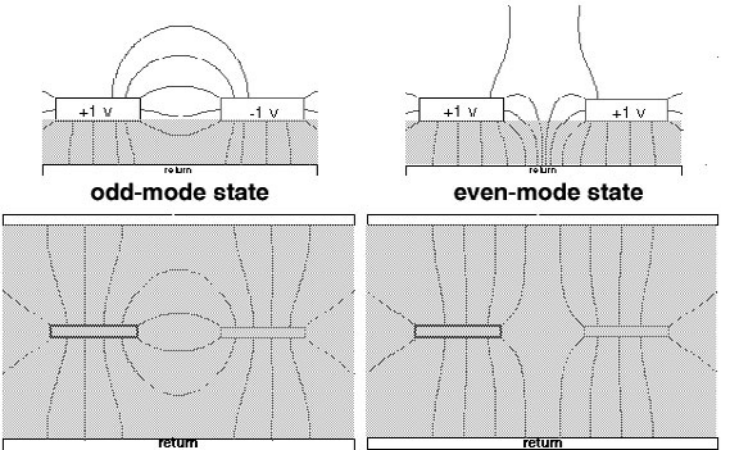


Figure 11-30 Electric-field distribution compared to the dielectric distribution for odd and even modes in microstrip and stripline. SImulated with Mentor Graphics Hyperlynx.

line with only a differential component, this differential signal would propagate down the line undistorted at a speed of 7.4 inches/nsec. If a pure common signal were launched into the line, it would propagate down the line undistorted at a speed of 6.8 inches/nsec.

For an interconnect that is 10 inches long, the time delay for a signal propagating in the odd mode would be TD_{odd} = 10 inches/7.4 inches/nsec = 1.35 nsec, while the time delay for an even-mode signal would be TD_{even} = 10 inches/6.8 inches/nsec = 1.47 nsec.

TIP The difference of only 120 psec may not seem like much, but this small difference in delay between signals propagating in the even and odd modes is the effect that gives rise to far-end cross talk in single-ended, coupled, transmission lines.

If instead of driving the differential pair with a pure differential or common signal, we drive it with a signal that has both components, each component will propagate down the line independently and at different speeds. Though they start out coincident, as they move down the line, the faster component, typically the differential signal, will move ahead. The wavefronts of the two signals—the differential and common components—will separate. At any point along the differential pair, the actual voltage on each line will be the sum of the

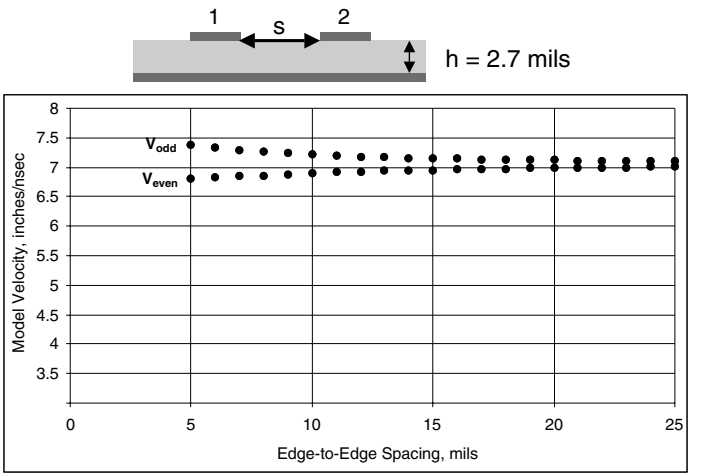


Figure 11-31 Speed of the even and odd modes for edge-coupled microstrip traces in FR4 with 5-mil-wide lines and roughly 50 Ohms.

differential and common components. As the edges spread out, the voltage patterns on both lines will change.

Suppose we launch a voltage signal into a differential pair that has a 0-v to 1-v transition on one line, but we keep the voltage at 0 v on the other line. This is the same as sending a single-ended signal down line 1 and pegging line 2 low. Line 1 becomes the aggressor line and line 2 becomes the victim line.

We can describe the voltage patterns on the two lines in terms of equal amounts of a signal propagating as a differential signal in the odd mode and as a common signal in the even mode. After all, a common signal of 0.5 v on line 1 and 0.5 v on line 2 and a differential signal of 0.5 v on line 1 and –0.5 v on line 2 will result in the same signal that is launched on the pair. This is shown in Figure 11-32.

In a homogeneous dielectric system, such as stripline, the odd- and even-mode signals will propagate at the same speed. The two modal signals will reach the far end of the line at the same time. When they are added up again, they will recompose, with no changes, the original waveform that was launched. In such an environment, there is no far-end cross talk.

In a microstrip differential pair, the differential component will travel faster than the common component. As these two independent voltage components propagate down the differential pair, their leading-edge wavefronts will separate. The leading edge of the differential component will reach the end of line 2 before the common-signal component reaches the end of line 2.

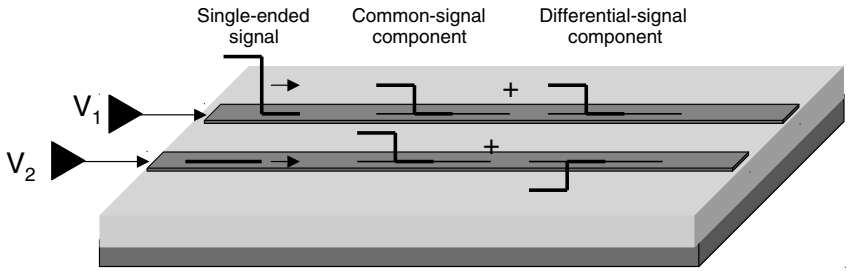


Figure 11-32 Describing a signal on an aggressor line and a quiet line as a simultaneous common- and differential-signal component on a differential pair.

The received signal at the far end of line 2 will be the recombination of the –0.5-v differential component, and the delayed 0.5-v common-signal component. These combine to give a net, though transient, voltage at the far end of line 2. We call this transient voltage the far-end noise.

TIP Far-end noise in a pair of coupled transmission lines can be considered as either due to the capacitively coupled current minus the inductively coupled current or as the sum of the shifted differential component and the common component. These two views are perfectly equivalent.

If the leading edge is a linear ramp, we can estimate the expected far-end noise due to the time delay between the two components. The setup is shown in Figure 11-33, showing the voltage on line 2 as a combination of a common and differential component. The net voltage on line 2 is just the sum of these two components. The differential-signal and common-signal magnitudes will be exactly 1/2 the voltage on line 1, $1/2V_1$. The time delay between the arrival of the differential component (traveling in the odd mode) and the common component (traveling in the even mode) will be:

$$\Delta T = \frac{Len}{v_{even}} - \frac{Len}{v_{odd}} \tag{11-20}$$

The initial part of the transient signal is the leading edge of the rise time. It reaches a peak value, the far-end voltage, related to the fraction of the rise time the time delay represents:

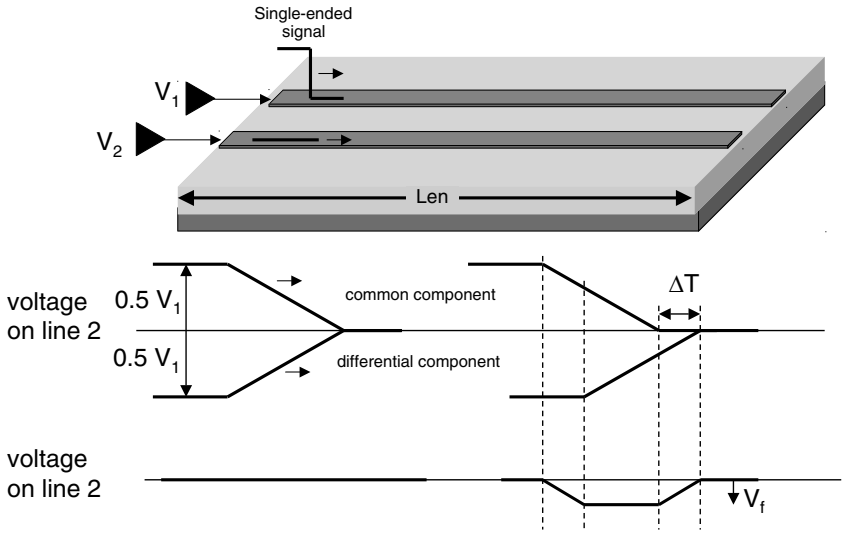


Figure 11-33 The signal on line 2 has a differential-signal component and common-signal component. The differential-signal component gets to the end of line 2 before the common component, causing a transient net signal on line 2.

$$V_f = -\frac{1}{2}V_1 \times \frac{\Delta T}{RT} = -\frac{1}{2}V_1 \frac{Len}{RT}\left(\frac{1}{v_{even}} - \frac{1}{v_{odd}}\right) = \frac{1}{2}V_1 \frac{Len}{RT}\left(\frac{1}{v_{odd}} - \frac{1}{v_{even}}\right) \quad (11\text{-}21)$$

where:

V_f = the far-end voltage peak on line 2, the victim line

V_1 = the voltage on line 1, the aggressor line

Len = the length of the coupled region

ΔT = the time delay between the arrival of the differential signal and the common signal

RT = the signal rise time

v_{even} = the velocity of a signal propagating in the even mode

v_{odd} = the velocity of a signal propagating in the odd mode

We can interpret far-end noise as the difference in speed between the odd and even modes. If the differential pair has homogeneous dielectric and there is no

difference in the speeds of the two modes, there will be no far-end noise. If there is air above the traces and the odd mode has a lower effective dielectric constant than the even mode, the odd mode will have a higher speed than the even mode has. The differential-signal component will arrive at the end of line 2 before the common component arrives. Since the differential-signal component on line 2 is negative, the transient voltage on line 2 will be negative.

As long as the time delay between the arrival of the differential and common signals is less than the rise time, the far-end noise will increase with the coupling length. However, if the time delay is greater than the rise time, the far-end noise will saturate at the differential-signal magnitude of 0.5 V_1.

The saturation length of far-end noise is when $V_f = 0.5\ V_1$, which can be calculated from:

$$Len_{sat} = -\frac{RT}{\frac{1}{v_{odd}} - \frac{1}{v_{even}}} \tag{11-22}$$

where:

Len_{sat} = the coupling length where the far-end noise will saturate

RT = the rise time

v_{even} = the velocity of a signal propagating in the even mode

v_{odd} = the velocity of a signal propagating in the odd mode

For example, in the case of the most tightly coupled microstrip, and a rise time of 1 nsec, the saturation length is:

$$Len_{sat} = -\frac{1\,ns}{\frac{1}{7.4\frac{in}{ns}} - \frac{1}{6.8\frac{in}{ns}}} = -\frac{1\,ns}{0.135\frac{ns}{in} - 0.147\frac{ns}{in}} = 83\ in \tag{11-23}$$

The smaller the difference between the odd- and even-mode velocities, the greater the saturation length. Of course, long before the far-end noise saturates, the magnitude of the far-end noise can become large enough to exceed any reasonable noise margin.

11.12 Ideal Coupled Transmission-Line Model or an Ideal Differential Pair

A pair of coupled transmission lines can be viewed as either two single-ended transmission lines with some coupling that gives rise to cross talk on the two lines, or as a differential pair with an odd- and even-mode characteristic impedance and an odd- and even-mode velocity. Both views are equivalent.

In the previous chapter, we saw that the near-end (or backward) noise, V_b, and the far-end (or forward) noise, V_f, in a pair of coupled transmission lines was related to:

$$V_b = V_a k_b \tag{11-24}$$

$$V_f = V_a \frac{Len}{RT} k_f \tag{11-25}$$

where:

V_b = backward noise

V_f = far-end noise

V_a = voltage on the active line

k_b = the backward cross-talk coefficient

k_f = the forward cross-talk coefficient

Len = the length of the coupled region

RT = the rise time of the signal

Using the view of a differential pair, the cross-talk coefficients are:

$$k_b = \frac{1}{2} \frac{Z_{even} - Z_{odd}}{Z_{even} + Z_{odd}} \tag{11-26}$$

$$k_f = \frac{1}{2} \left(\frac{1}{v_{odd}} - \frac{1}{v_{even}} \right) \tag{11-27}$$

The near-end noise is a direct measure of the difference in the characteristic impedance of the odd and even modes. The farther apart the traces, the smaller the difference between the odd- and even-mode impedances and the less the coupling.

When the traces are very far apart, there is no interaction between the traces and the characteristic impedance of one line is independent of the signal on the other line. The odd- and even-mode impedances are the same and the near-end cross-talk coefficient is zero.

> **TIP** This connection allows us to model a pair of coupled lines as a differential pair. An ideal distributed differential pair is a new ideal circuit model that can be added to our collection of ideal circuit elements. It can model the behavior of a differential pair or a pair of independent, coupled transmission lines.

Just as an ideal, single-ended transmission line is defined in terms of a characteristic impedance and time delay, an ideal differential pair is defined with just four parameters:

1. An odd-mode characteristic impedance
2. An even-mode characteristic impedance
3. An odd-mode time delay
4. An even-mode time delay

These terms fully take into account the effects of the coupling, which would give rise to near-end and far-end cross talk. This is the basis of most ideal circuit models used in circuit or behavior simulators.

If the transmission line is a stripline, the odd- and even-mode velocities and time delays are the same, so only three parameters are needed to describe a pair of coupled lines.

Usually, the parameter values for these four terms describing a differential pair are obtained with a 2D field solver.

11.13 Measuring Even- and Odd-Mode Impedance

A time domain reflectometer (TDR) is used to measure the single-ended characteristic impedance of a single-ended transmission line. The TDR will launch a voltage step into the transmission line and measure the reflected voltage. The magnitude of the reflected voltage will depend on the change in the instantaneous impedance the signal encounters in moving from the 50 Ohms of the TDR and its interconnect cables to the front of the transmission line. For a uniform line, the

instantaneous impedance the signal encounters will be the characteristic impedance of the line. The reflected voltage is related to:

$$\rho = \frac{V_{reflected}}{V_{incident}} = \frac{Z_0 - 50\Omega}{Z_0 + 50\Omega} \quad (11\text{-}28)$$

where:

ρ = the reflection coefficient

$V_{reflected}$ = the reflected voltage measured by the TDR

$V_{incident}$ = the voltage launched by the TDR into the line

Z_0 = the characteristic impedance of the line

50 Ohms = the output impedance of the TDR and cabling system

By measuring the reflected voltage and knowing the incident voltage, the characteristic impedance of the line can be calculated:

$$Z_0 = 50\Omega\frac{1 + \rho}{1 - \rho} \quad (11\text{-}29)$$

This is how we can measure the characteristic impedance of any single-ended transmission. In order to measure the odd-mode impedance or even-mode impedance of a line that is part of a differential pair, we must measure the characteristic impedance of one line while we drive the pair either into the odd mode or into the even mode.

To drive the pair into the odd mode, we need to apply a differential signal to the pair. Then the characteristic impedance we measure of either line is its odd-mode impedance. This means that if the incident signal is 0 v to +200 mV between the line and its return path, into the line under test, we need to also apply another signal to the second line of 0-v to –200-mV between the signal and its return path. Likewise, measuring the even-mode characteristic impedance of the line requires applying a 0-v to +200-mV signal to the second line.

This requires the use of a special TDR that has two active heads. This instrument is called a differential TDR or DTDR. Figure 11-34 shows the measured TDR voltages with the two outputs connected to opens to illustrate how the two

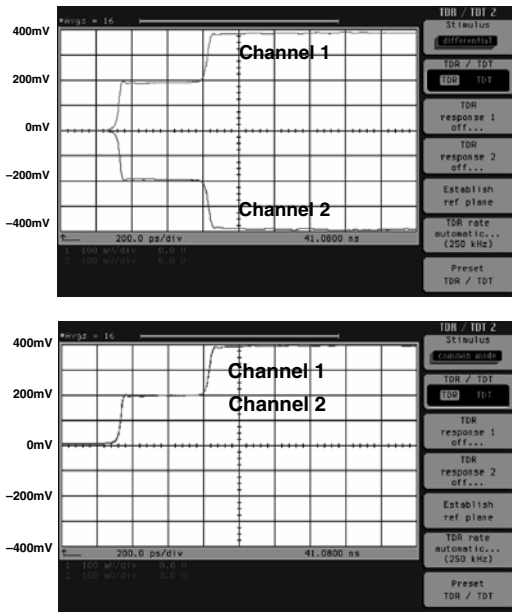


Figure 11-34 Measured voltages from the two channels of a differential TDR when driving a differential signal (top) and a common signal (bottom) into the differential pair under test. Measured with Agilent 86100 DCA with DTDR plug-in.

channels are driven when set up for differential-signal outputs and common-signal outputs.

In a DTDR, the reflected voltages from both channels can be measured, so the odd- or even-mode impedance of both lines in the pair can be measured. Figure 11-35 is an example of the measured odd- and even-mode characteristic impedance of one line in a pair of nearly 50-Ohm transmission lines when there is tight coupling. In this example, the odd-mode impedance is measured as 39 Ohms and the even-mode impedance of the same line is measured as 50 Ohms.

11.14 Terminating Differential and Common Signals

When a differential signal reaches the open end of a differential pair, it will see a high impedance and will be reflected. If the reflections at the ends of the differential pair are not managed, they may exceed the noise margin and cause excessive noise. A commonly used method of reducing the reflections is to have the differential signal see a resistive impedance at the end of the pair that matches the differential impedance. If the differential impedance of the line is designed for 100

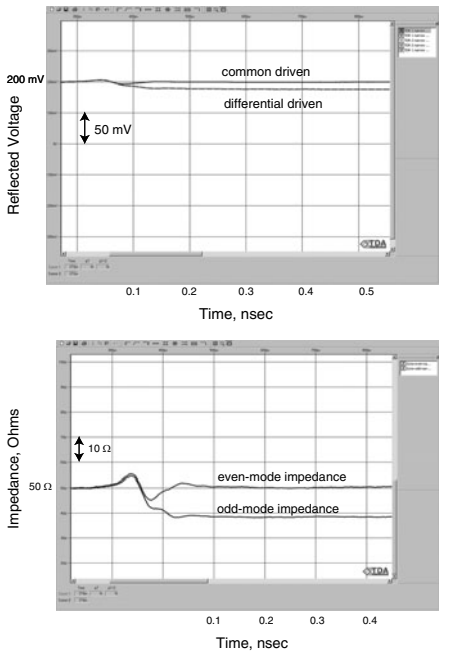


Figure 11-35 Measured voltage (top) from one channel of a differential TDR when the output is a differential signal and a common signal. These are interpreted as the impedance of the even mode (common driven) and odd mode (differential driven). The odd-mode impedance is 39 Ohms and the even-mode characteristic impedance is 50 Ohms. Measured with an Agilent 86100 DCA with DTDR plug-in and TDA Systems IConnect software, and a GigaTest Labs Probe Station.

Ohms, for example, the resistor at the far end should be 100 Ohms, illustrated in Figure 11-36. The resistor would be placed across the two signal lines so the differential signal would see its impedance. The differential signal will be terminated with this single resistor. But what about the common signal?

TIP While it is true that the common-signal component even in LVDS signal levels is very high, this voltage is nominally constant, even when the drivers are switching, and may not cause a problem.

Any transient common signal moving down the differential pair will see a high impedance at the end of the pair and reflect back to the source. Even if there were a 100-Ohm resistor across the signal lines, the common signal, having the same voltage between the two signal lines, would not see it. Depending on the

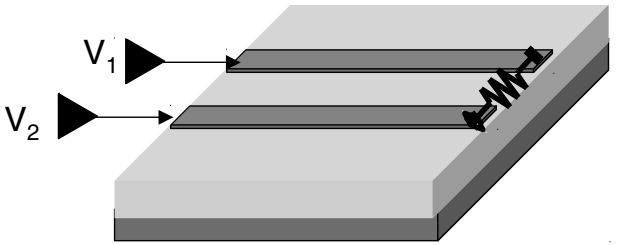


Figure 11-36 Terminating the differential signal at the far end of a differential pair with a single resistor that has a resistance equal to the differential impedance of the pair.

impedance of the driver, any common signal created will rattle back and forth and show the effect of ringing.

Is it important to terminate the common signal? If any devices in the circuit are sensitive to the common signal, then managing the signal quality of the common signal is important. If there is an asymmetry in the differential pair that converts some differential signal into common signal, if the common signal rattles around, and sees the asymmetry again, some of it may convert back into differential signal and contribute to differential noise.

TIP Terminating the common signal will not eliminate the common signal, it will simply prevent the common signal from rattling around. If a common signal is contributing to EMI, it's true that terminating the common signal will help reduce EMI slightly, but it will still be important to create a design that eliminates the source of the common signal.

One way of terminating the common signal is by connecting a resistor between the end of each signal line and the return path. The parallel combination of the two resistors should equal the impedance the common signal sees. If the common impedance were 25 Ohms, each resistor would be 50 Ohms, and the parallel impedance would be 25 Ohms. This is shown in Figure 11-37.

Unfortunately, if this termination scheme is used, while the common signal will be terminated, the differential signal will not be terminated, except for the special case of no coupling between the lines. The equivalent resistance a differential signal would see in this termination scheme is the series combination of the two resistors, $4 \times Z_{comm}$. It is only when $Z_{even} = Z_{odd}$ that this resistance is the dif-

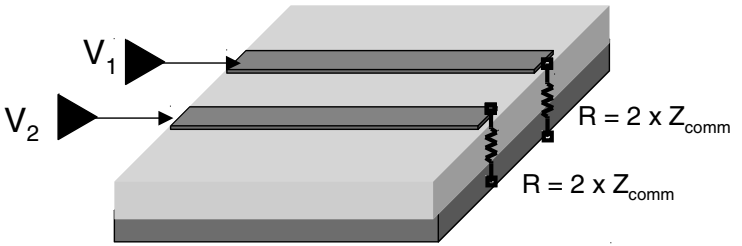


Figure 11-37 Terminating the common signal at the far end of a differential pair with two resistors, each with resistance equal to twice the common impedance of the pair.

ferential impedance. As coupling increases, the common impedance will increase and the differential impedance will decrease.

If it is important to terminate the common signal, we must use a method that will also simultaneously terminate the differential signal. This can be implemented with two different topologies, each using three resistors. The "pi" and "tee" topologies are shown in Figure 11-38.

In the pi topology, the values of the resistors can be calculated based on designing the equivalent resistance the common signal sees to equal the common impedance and the equivalent resistance the differential signal sees to equal the differential impedance. The equivalent resistance the common signal sees is the parallel combination of the two R_2 resistors:

$$R_{equiv} = \frac{1}{2}R_2 = Z_{comm} = \frac{1}{2}Z_{even} \tag{11-30}$$

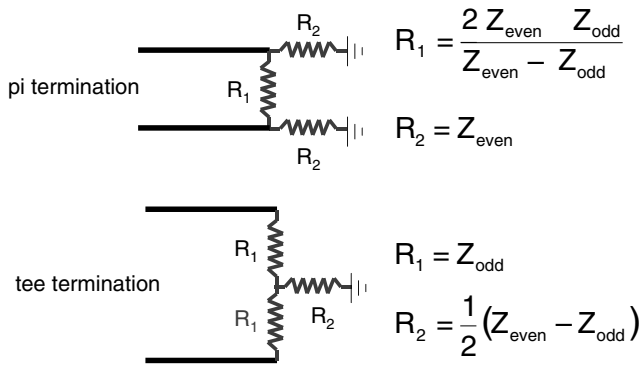


Figure 11-38 Pi- and tee-termination topologies for a differential pair that can terminate both the common and differential signals simultaneously.

where:

R_{equiv} = the equivalent resistance the common signal would see

R_2 = the resistance of each R_2 resistor

Z_{comm} = the impedance the common signal sees on the differential pair

Z_{even} = the even-mode characteristic impedance of the differential pair

This requirement defines $R_2 = Z_{even}$.

The equivalent resistance the differential signal sees is the parallel combination of the R_1 resistor with the series combination of the two R_2 resistors:

$$R_{equiv} = \frac{R_1 \times 2R_2}{R_1 + 2R_2} = Z_{diff} = 2 \times Z_{odd} \tag{11-31}$$

where:

R_{equiv} = the equivalent resistance the differential signal would see

R_1 = the resistance of each R_1 resistor

R_2 = the resistance of each R_2 resistor

Z_{diff} = the impedance the differential signal sees on the differential pair

Z_{odd} = the odd-mode characteristic impedance of the differential pair

Since $R_2 = Z_{even}$, we can calculate the value of R_1 as:

$$R_1 = \frac{2Z_{even}Z_{odd}}{Z_{even} - Z_{odd}} \tag{11-32}$$

When there is little coupling, and $Z_{even} \sim Z_{odd} \sim Z_0$, then $R_2 = Z_0$ and R_1 = open. After all, with little coupling this pi-termination scheme reduces to a simple far-end termination of a resistor equal to the characteristic impedance of either line. As the coupling increases, the resistor to the return path will increase to match the even-mode characteristic impedance. A high-value shunt resistor will short the two signal lines so the impedance the differential signal sees will drop as the coupling increases and will match the lower differential impedance.

For a typical tightly coupled differential pair, the odd-mode impedance might be 50 Ohms and the even-mode impedance might be 55 Ohms. In this case, in a pi termination, the resistors would be 1 KOhm between the signal lines and 55 Ohms from each line to the return path. This combination will simultaneously terminate the 100-Ohm differential impedance and the 27.5-Ohm common impedance.

In a tee topology, the differential signal will see an equivalent resistance of just the series combination of the R_1 resistors:

$$R_{equiv} = Z_{diff} = 2R_1 = 2Z_{odd} \tag{11-33}$$

where:

R_{equiv} = the equivalent resistance the differential signal would see

R_1 = the resistance of each R_1 resistor

Z_{diff} = the impedance the differential signal sees on the differential pair

Z_{odd} = the odd-mode characteristic impedance of the differential pair

This defines $R_1 = Z_{odd}$.

The equivalent resistance the common signal sees is the series combination of the two R_1 resistors that are in parallel with the R_2 resistor:

$$R_{equiv} = Z_{comm} = \frac{1}{2}R_1 + R_2 = \frac{1}{2}Z_{even} \tag{11-34}$$

The value of R_2 can be extracted as:

$$R_2 = \frac{1}{2}(Z_{even} - Z_{odd}) \tag{11-35}$$

In a tee termination, when there is little coupling, so $Z_{even} \sim Z_{odd} \sim Z_0$, the termination is simply the two R_1 resistors in series between the signal lines, each equal to the odd-mode characteristic impedance. In addition, there is a center tap connection to the return path that is a short. The tee termination reduces to the pi termination in the case of no coupling. As coupling increases, the differential

impedance decreases and the R_1 values decrease to match. The common impedance increases and the R_2 term increases to compensate.

If the odd-mode impedance were 50 Ohms and the even-mode impedance 55 Ohms, the tee termination would use two series resistors between the signal lines of 50 Ohms each, and it would use a 2.5-Ohm resistor between the center of the resistors and the return path.

The most important consideration when implementing a pi- or tee-termination strategy is the potential DC load on the drivers. In both cases, there is a resistance between either signal line to the return path on the order of the even-mode impedance. The lower the even-mode impedance, the more current flow from the driver. Typical differential drivers may not be able to handle low-DC resistance to the low-voltage side, so it may not be practical to terminate the common signal. Rather, care should be used to minimize the creation of common signals right from the beginning and simply to terminate the differential signals.

An alternative configuration to terminate both the differential and common signals is to use a tee topology with a DC-blocking capacitor. This circuit topology is shown in Figure 11-39. In this topology, the resistor values are the same as with a tee topology, and the value of the capacitor is chosen so that the time constant the common signal will see is long compared to the lowest frequency component in the signal. This will ensure a low impedance for the capacitor compared with the impedance for the resistors, at the lowest frequency component of the signal. As a first-order estimate, the capacitor value is initially chosen from:

$$RC = 100 \times RT \tag{11-36}$$

$$C = \frac{100 \times RT}{Z_{comm}} \tag{11-37}$$

where:

R = the equivalent resistance the common signal sees

C = the capacitance of the blocking capacitor

RT = the rise time of the signal

Z_{comm} = the impedance the common signal sees on the differential pair

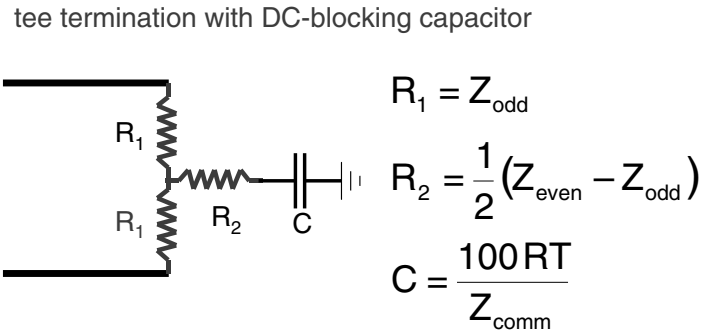


Figure 11-39 Tee termination with DC-blocking capacitor to terminate common signals but minimize the DC-current drain.

For example, if the common impedance is about 25 Ohms, and the rise time is 1 nsec, the blocking capacitor should be about 100 nsec/25 Ohms = 4 nF. Of course, whenever an RC termination is used, a simulation should be performed to verify the optimum value of the capacitor.

11.15 Conversion of Differential to Common Signals

All the information content in differential signaling is carried by the differential signal. First, it is important to maintain the signal quality of the differential signal. This is done by using the following guidelines:

1. Use controlled differential impedance lines.
2. Minimize any discontinuities in the differential pair.
3. Terminate the differential signals at the far end.

There is an additional source of distortion of the differential signal, due to asymmetries in the lines and skew between the drivers.

TIP Distortion due to asymmetries and skews are completely independent of the degree of coupling between the lines and they can occur in a differential pair with no coupling or the tightest coupling, both microstrip and stripline.

A skew between the transitioning of the two differential drivers will distort the differential signal. Figure 11-40 shows the edge of the differential signal as the

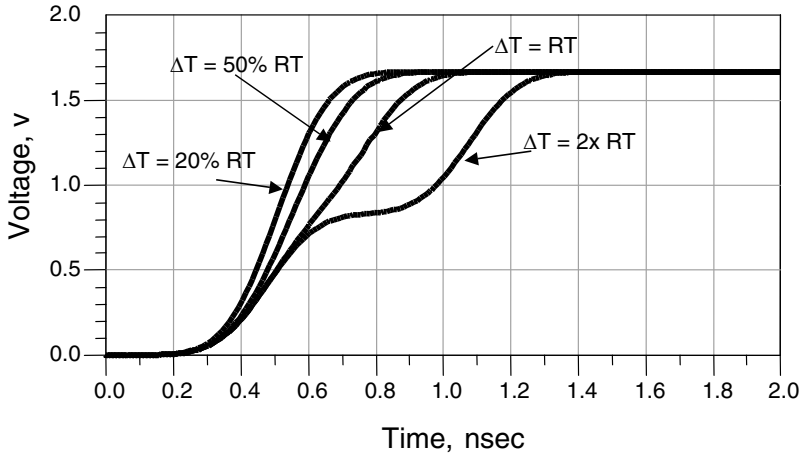


Figure 11-40 Received differential signal with driver skew varying from 20% of the rise time to 200% of the rise time. Simulated with Agilent's ADS.

skew between the drivers is increased from 20% of the rise time to two times the rise time. As a general rule, the skew should be kept to less than 10% of the rise time to avoid adding a time delay and distorting the edge of the differential signal. Any skew will directly affect timing margins.

Skew can also be created by having a length delay between the two lines in the pair. The same rule of thumb applies. To keep the skew to less than 10% of the rise time requires the lengths to be matched to within 10% of the spatial extent of the leading edge. This condition requires keeping the total lengths matched to:

$$\Delta L = 0.1 \times RT \times v \tag{11-38}$$

where:

ΔL = the maximum length skew between paths to keep skew less than 10% of the rise time

RT = the rise time of the signal

v = the speed of the differential signal

If the speed of a signal is roughly 6 inches/nsec and the rise time is 1 nsec, then the spatial extent of the leading edge is 6 inches. The lengths in a differential

pair would have to be matched to within less than 0.1 × 6 inches/nsec × 1 nsec ~ 0.6 inch. This is relatively easy to do.

If the rise time were 100 psec, the lengths would have to be matched to better than 60 mils. As the skew budget allocated to line-length differences decreases, it becomes more and more important to match the line lengths.

Other sources of asymmetry will potentially distort the differential signal. In general, if something affects one line and not the other, the differential signal will be distorted. For example, if one of the traces sees a test pad so there is a capacitive load and the other trace does not, the differential signal will be distorted. Figure 11-41 shows the simulated differential signal at the end of a differential pair with a capacitive load on one line.

There is another impact from skew and distortions. Any asymmetry can convert some of the differential signal into common signal. In general, if the drivers and receivers are not sensitive to common signals, the amount of common signal created may not be important. After all, differential receivers typically have good common-mode rejection ratio, or CMRR. However, if the common signal were to get out of the box on twisted pair, for example, it will dramatically contribute to EMI. It is critical to minimize any common signal that may have an opportunity, either intentionally or unintentionally, to escape outside the product enclosure through an aperture or on any cables.

TIP Any asymmetry will convert differential signal into common signal. This includes cross talk, driver skew, length skew, and asymmetrical loading. An important motivation to keep skews to a minimum is to minimize the conversion of differential signal into common signal.

A small skew that may not affect the differential-signal quality at all can have considerable impact on the common signal. Figure 11-42 shows the voltages on the signal lines when there is a skew of just 20% of the rise time, and the received differential and common signal when only the differential signal is terminated. The differential-signal component is terminated by the resistor at the far end, but the common-signal component sees an open at the far end and reflects back to the low impedance of the source. This creates the ringing.

Even if both the common and differential signals are terminated, there will still be common signal generated by any asymmetry. Figure 11-43 shows the voltages at the far end with a tee-termination network, where both differential and

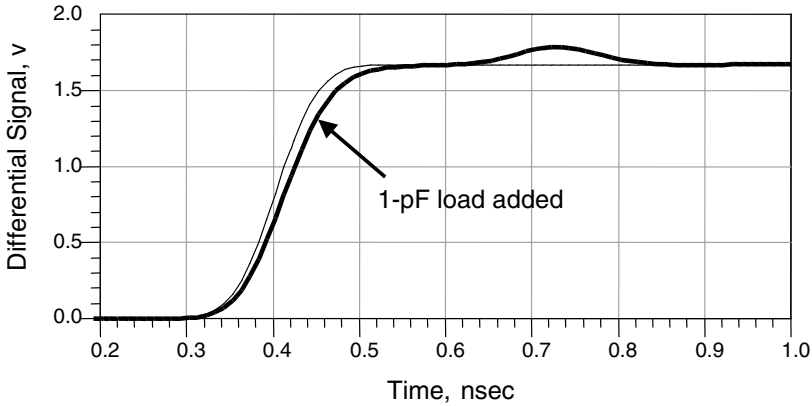


Figure 11-41 Received differential signal with and without a 1-pF capacitive load on one line of a differential pair. The rise time is 100 psec. Simulated with Agilent's ADS.

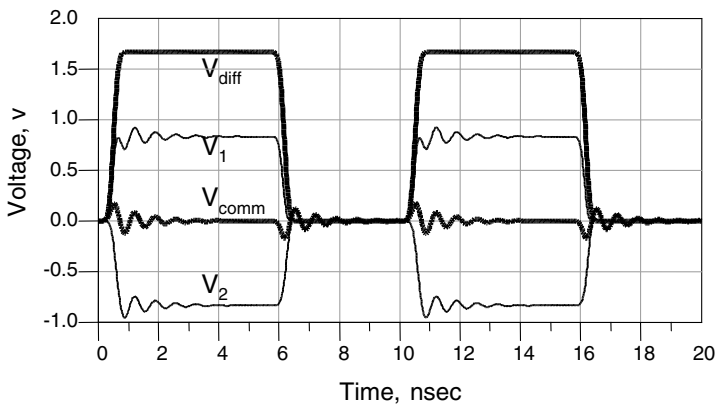


Figure 11-42 Signals at the far end of a differential pair with a differential terminating resistor. The driver skew is only 20% of the rise time of the signal. Note, the good quality of the differential signal, even with the voltage on each signal line distorted. Simulated with Agilent's ADS.

common signals are simultaneously terminated. The ringing is gone, but there is still a common signal, which can contribute to EMI if it were to get out of the box.

As the skew between the two signal lines increases, the magnitude of the common signal will increase, even with the common signal terminated. Figure 11-44 shows the common signal for skews of 20%, 50%, 100%, and 200% of the rise time.

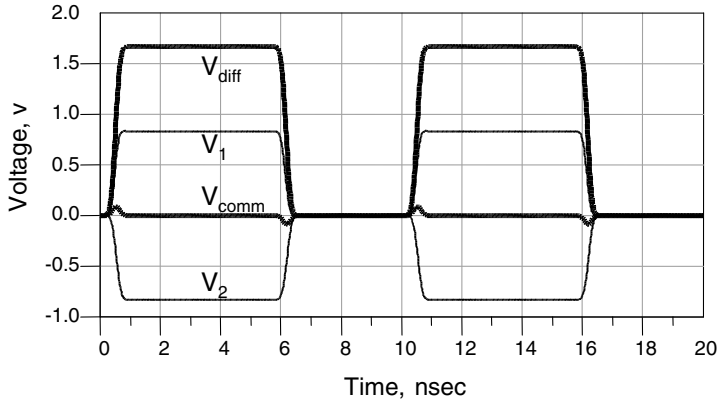


Figure 11-43 Received signals at the far end with a 20% skew and both differential- and common-signal termination. Simulated with Agilent's ADS.

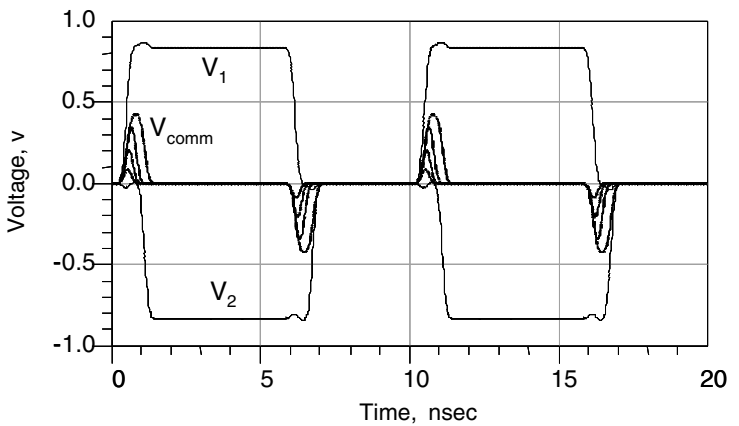


Figure 11-44 Common signal generated by skew even with the common signal terminated. Skew of 20%, 50%, 100%, 200% of the rise time. Simulated with Agilent's ADS.

TIP A very small driver skew can create significant common signal. This is why it is important to try to minimize all asymmetries.

To minimize the creation of common signals, care should be used to try to keep the paths as symmetric as possible. Imperfections should be modeled to anticipate the amount of common signal created. This will allow some prediction of the severity of potential EMI problems.

11.16 EMI and Common Signals

If an unshielded twisted-pair cable, such as category-5 cable, is connected to a differential pair on a circuit board, both differential and common signal may be transmitted onto the cable. The differential signal is the intended signal that carries information content. A twisted-pair cable is a very poor radiator of electromagnetic energy from the differential signals. However, any common current on the cable will radiate and contribute to EMI.

If there is common current on the twisted pair, where is the return of the common signal? On the circuit board where the differential signal is created, the common signal will return on the return plane of the board. When the signal switches from the circuit board to the twisted pair, there is no direct connection to the return plane of the circuit boards.

This poses no problem for the differential signal. It is possible to design the differential impedance of the circuit-board interconnects to match the differential impedance of the twisted pair. The differential signal may not see any impedance discontinuity in transitioning from the board to the twisted pair.

TIP In an unshielded twisted pair, the return path for the common signal is literally the ground or floor, or any other conductor that is nearby. The coupling to the nearest conducting surface is usually less than the coupling between adjacent signal paths, so the common impedance is usually high—on the order of a few hundred Ohms.

There may be a large impedance mismatch between what the common signal sees in the board and in the twisted pair. The connection between the two common return paths is through whatever path the current will find. At high frequency, the return path is typically dominated by stray capacitances between the circuit-board ground to the box chassis to the floor. A capacitance of only 1 pF will have an impedance at 1 GHz of about 160 Ohms.

As the common signal propagates down the twisted pair, the return current is continuously coupled to the nearest conductor by the stray capacitance between the twisted pair and the conductor. This is illustrated in Figure 11-45.

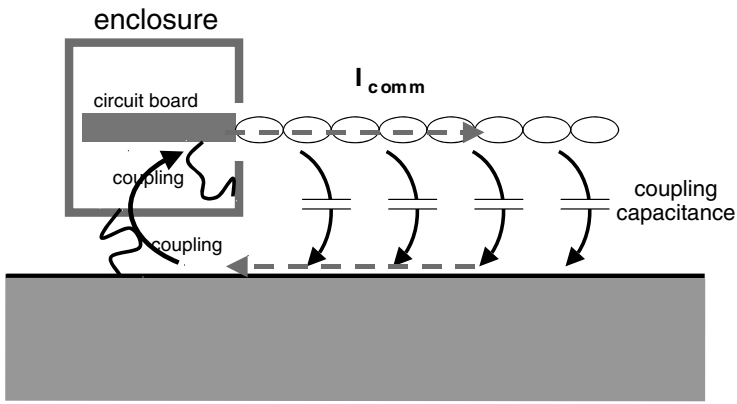


Figure 11-45 Schematic of the common current path when a twisted-pair cable is connected to a circuit board. The coupling paths are typically capacitive for high-frequency common-signal components.

The amount of common current on the twisted pair will depend on the common-signal voltage launched into the cable and on the impedance the common signal sees in the cable:

$$I_{comm} = \frac{V_{comm}}{Z_{comm}} \tag{11-39}$$

This common current will radiate. The amount of common current that will fail a certification test depends on the specific test spec, length of cable, and frequency. As a rough rule of thumb, it only takes 3 microAmps of common current, at 100 MHz, in a 1-meter-long external cable, to fail an FCC Class B certification test. This is a tiny amount of common current.

If the radiated field strength exceeds the allowed limits of the EMI certification regulation, the product will not pass certification and this may delay the ship date of the product. In most countries, it is illegal to sell a product for revenue unless it has passed the local certification regulations. In the United States the Federal Communications Commission (FCC) specifies the acceptable radiated field levels in two categories of products. Class A products are those that will be used in an industrial or manufacturing environment. Class B products are those that will be used in a home or office environment. Class B requirements limit acceptable radiated emissions to lower levels than Class A requirements.

Freq (MHz)	μ V/m
30–88	100
88–216	150
216–960	200
> 960	500

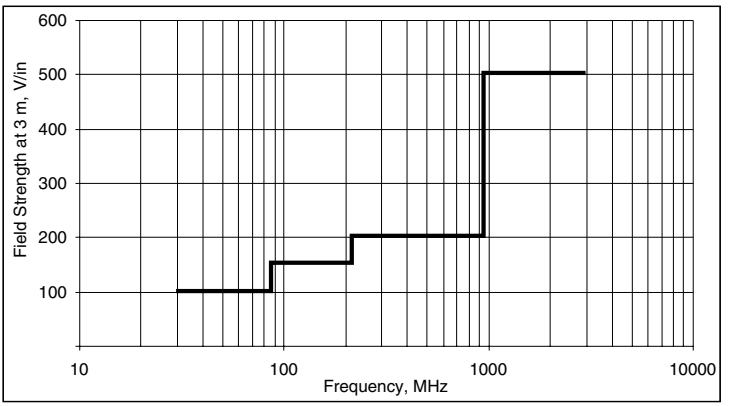


Figure 11-46 FCC Class B certification limits for the maximum allowable far-field electric-field strengths at a distance of 3 m.

Electric-field strength is measured in units of volts/m, at a specified frequency. For Class B certification, the maximum radiated field strengths are measured 3 m from the product. The maximum electric-field strengths vary within different frequency ranges. They are listed and plotted in Figure 11-46. As a reference point, the maximum acceptable field strength at 3 m, for 100 MHz, is 150 microV/m.

We can estimate the radiated electric-field strength from a common signal in a twisted pair by approximating the twisted pair as an electric monopole antenna. The far field exists at a distance about 1/6 of the wave length of the radiation. The FCC test condition of 3 m is in the far field for frequencies larger than about 16 MHz. The far-field electric-field strength from an electric monopole antenna is:

$$E = 4\pi 10^{-7} \times f \times I_{comm} \times \frac{Len}{R} \tag{11-40}$$

where:

I_{comm} = the common current on the twisted pair, in Amps

V_{comm} = the common signal that is launched into the twisted pair

Z_{comm} = the impedance the common signal sees

E = the field strength at a distance R from the wire, in volts/m

f = the sine-wave frequency of the common current component, in Hz

Len = the length of the twisted wire that is radiating, in m

R = the distance from the wire to the point where the field is measured, in m

For example, if the common signal is 100 mV and the impedance the common signal sees in the twisted pair is 200 Ohms, the common current may be as large as 0.1 v/200 Ohms = 0.5 mA. If the length of the twisted pair is 1 meter and the distance at which we measure the field strength is 3 meters, corresponding to the FCC Class B test, the radiated field strength at 100 MHz is:

$$E = 4\pi 10^{-7} \times 10^{8} \times 5 \times 10^{-4} \times \frac{1}{3} = 20000 \frac{\text{microV}}{\text{m}} \tag{11-41}$$

The radiated field strength from a common current that might easily be generated in an unshielded twisted pair is more than a factor of 100 larger than the FCC Class B certification limit.

TIP Even small amounts of common currents getting onto an unshielded twisted-pair cable can cause a product to fail an EMI certification test. Common currents larger than about 3 micro-Amps will fail FCC Class B Certification.

There are generally three techniques to reduce the radiated emissions from common currents in twisted-pair cables:

1. Minimize the conversion of differential signal into common signal by minimizing any asymmetries in the differential pairs and skews in the drivers. This minimizes the problem at the source.
2. Use a shielded twisted pair so the return path of the common current in the twisted pair is carried in the shield. Using a shielded cable may even increase the amount of common current since the common impedance may be lowered with the return path brought closer to the signal path. If the shield is connected to the chassis, the return current of the common signal may have an opportunity to flow on the inside of the shield. The common signal would then be flowing in a coax geometry—going out on the cable on the twisted-

pair center and back on the inside of the shield. In this geometry, there would be no external magnetic or electric field, and the common current will not radiate. This requires using a low-inductance connection between the shield and the chassis so that the common-return current can maintain its coaxial distribution.

3. Increase the impedance of the common-current path by adding common chokes. Common signal chokes are of two forms. Virtually all cables used by peripherals have cylinders of ferrite material on the outside of the cable. The placement of a ferrite around a cable is shown in Figure 11-47. The high permeability of the ferrite will increase the inductance and the impedance of any net currents going through the ferrite.

If a differential signal flows through the ferrite, there will be no net magnetic field through the ferrite. The currents in each line of the differential pair are equal and in opposite directions, so their external magnetic and electric fields mostly cancel each other. Only common-signal currents going through the ferrite, with return currents on the outside, would have magnetic-field line loops that would go through the ferrite material and see a higher impedance. The ferrite on the outside of the cable increases the impedance the common signal sees, reducing the common current, which decreases the radiated emissions. This type of ferrite choke can be used on the outside of any cable, twisted pair or shielded cable.

The second type of common-signal choke is primarily useful for twisted-pair cable. The goal is to dramatically increase the impedance a common signal would

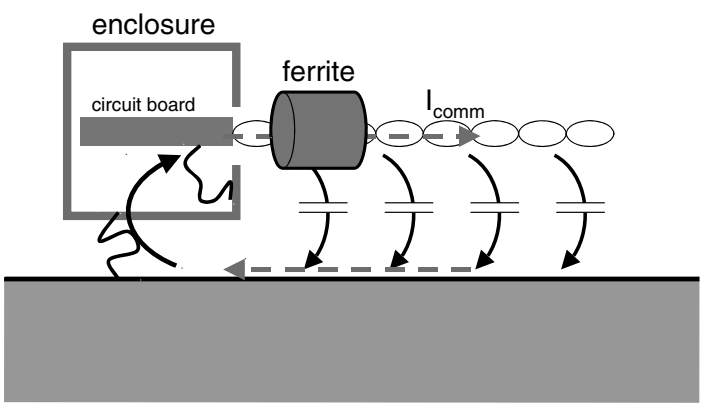


Figure 11-47 Placement of ferrite around the differential currents, but enclosing only the common-signal path, not the common return path.

see, but not affect the impedance a differential signal would see. To radically increase the impedance of the common signal, a twisted-pair wire is wound into a coil, sometimes with a ferrite core.

Any common current flowing down the twisted pair will see a much higher inductance from the coiling and the high-permeability ferrite core. However, the differential signal on the twisted pair will have very few magnetic-field lines outside the twisted pair and the differential current will not see the coil or the ferrite. The differential signal will be nearly unaffected by the coiling.

A twisted-pair coil can be built into a connector. Many RJ-45 connectors, used with Ethernet cables, for example, have built-in common-signal chokes.

TIP The higher impedance seen by the common signal can reduce the common current that would get onto the twisted-pair cable by more than 99%, or –40 dB. They are essential components to minimize radiated emissions from any common currents created by asymmetries in the circuit board.

11.17 Cross Talk in Differential Pairs

If we bring a single-ended transmission line close to a differential pair, some of the signal voltage will couple from the active, single-ended line to both lines of the differential pair. This is illustrated in Figure 11-48. The noise that couples to each line of the differential pair will be the same polarity, just of different magnitudes.

The closer line of the pair will end up with more noise than the farther line. The more tightly coupled the differential pair, the more equal the noise generated on the two lines, and the less the differential noise.

TIP In general, the more identical we can arrange the noise that is picked up on each line of the pair, the less the differential noise. This generally means a configuration of tightly coupled lines, with a large space between the pair and the aggressor line.

In Figure 11-49, we show the differential noise at the receiver generated on the differential pair constructed in stripline, with the far end differentially terminated and the near end having the low impedance of a typical driver.

In this configuration, the aggressor line is spaced a distance equal to the line width away from the nearest victim trace. Two different coupling levels are evaluated. Tight coupling has a spacing between lines in the differential pair equal to the line width, while weak coupling has the spacing twice the line width. In the

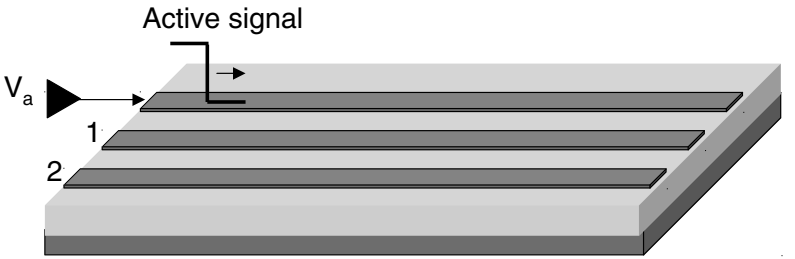


Figure 11-48 Cross talk from a single-ended signal to a differential pair.

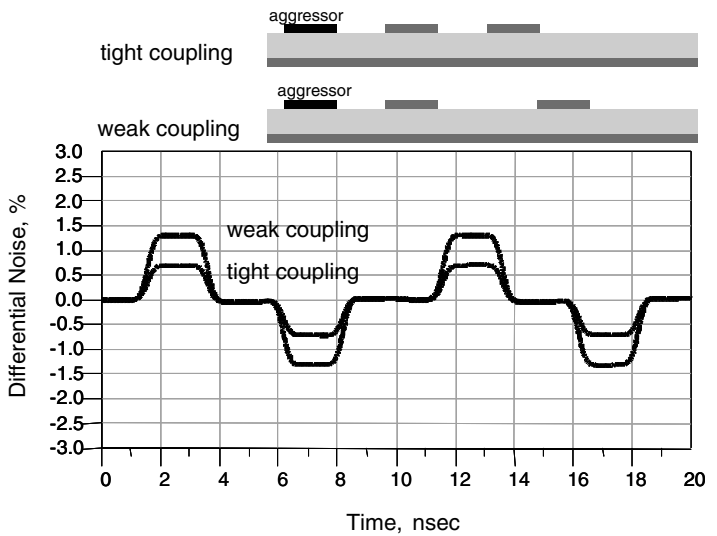


Figure 11-49 Differential noise in a differential pair from an adjacent single-ended aggressor line.

pair, the line that is farther away will have slightly less noise coupled to it than the line close to the aggressor line. The differential noise is the difference between the noise levels on each line. In this example, there is about 1.3% differential noise on the weakly coupled, victim differential pair while there is about half this value on the tightly coupled victim differential pair. Tighter coupling can decrease the differential noise by about 50%.

Though the differential noise from the active line to the differential line may be only 1.3% in the worst case, this may sometimes cause problems. If the aggressor line is a 3.3-v signal level, the differential noise on the differential pair may be

as large as 40 mV. If there is another active line on the other side of the victim lines, switching in the opposite direction of the first aggressor line, the differential noise from the two active lines will add to the victim line. This can contribute as much as 80 mV of noise. This may approach the allocated noise budget for some low-voltage differential signals.

The common noise on the victim differential pair is the average voltage on each line. Figure 11-50 shows the common noise for the two levels of coupling. The common noise will not be strongly affected as the coupling between the differential pair changes. With tight coupling (i.e., a spacing equal to the line width), the common noise is about 2.1%. With weak coupling (i.e., a spacing in the differential pair equal to twice the line width), the common noise is reduced to about 1.5%.

Tighter coupling will decrease the differential noise, but increase the common noise. Cross talk is one of the typical ways that common currents are created on a differential pair. Even if the differential pair is designed for perfect symmetry, cross talk can cause the creation of common voltages on a differential pair. This is why it is always important to provide common-signal chokes for any externally connected twisted-pair cables.

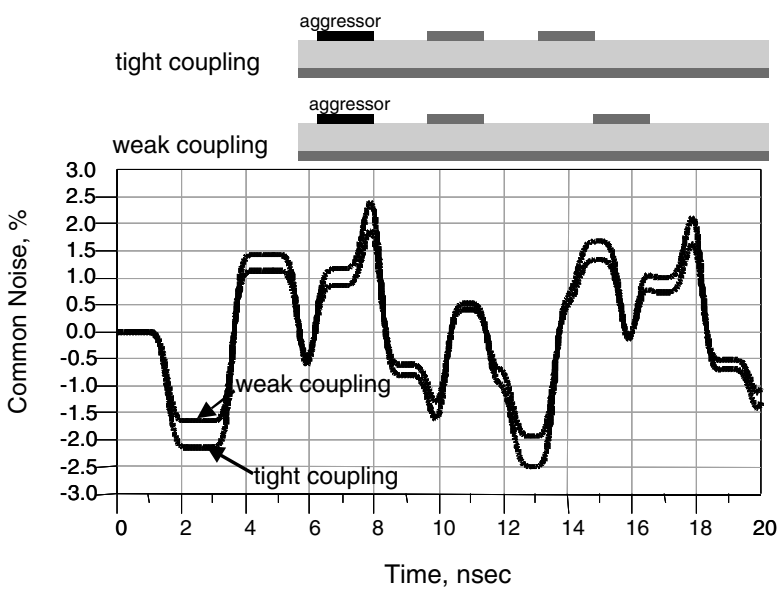


Figure 11-50 Common noise in a differential pair from an adjacent single-ended aggressor line. The ringing and distortion are due to the common signal not being terminated.

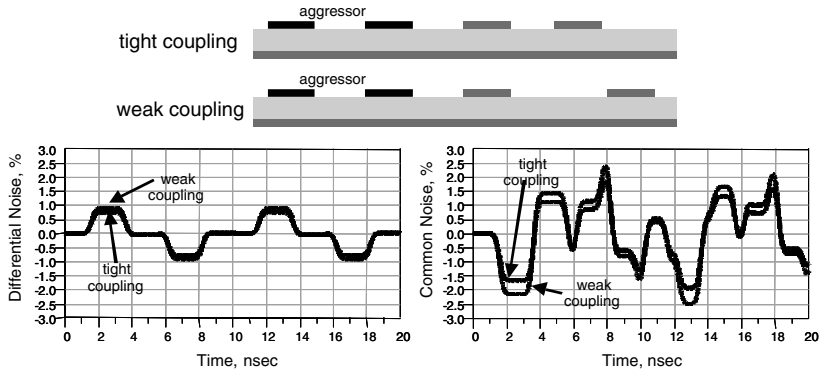


Figure 11-51 Differential (left) and common (right) noise on a differential pair when the aggressor line is also a differential pair for the case of tight and weak coupling within the victim pair.

TIP This suggests the general rule that to minimize the differential noise on a differential pair from a single-ended aggressor line, the coupling in the differential pair should be as tight as possible. Of course, the spacing between the aggressor and the victim pair should also be as large as possible to minimize the coupled noise.

The differential noise coupling between two differential pairs is slightly less than the differential noise coupling from a single-ended line. In Figure 11-51, we show the differential and common noise between two differential pairs. In this case, the edge-to-edge spacing between the two pairs is equal to the line width. The difference between tight and weak coupling is not very large. As a rough estimate, the differential noise can be less than 1% and the common noise can be less than 2% of the differential signal on the aggressor pair.

11.18 Crossing a Gap in the Return Path

Gaps in the return path are often used to isolate one region of a board from another. They also occur when a power plane is used as the reference layer and when a split power layer is used. Sometimes, a gap in the return path occurs unintentionally, as when the clearance holes in a return plane are overetched and overlap. In this case, any signal lines that pass through the via field will see a gap in the return path.

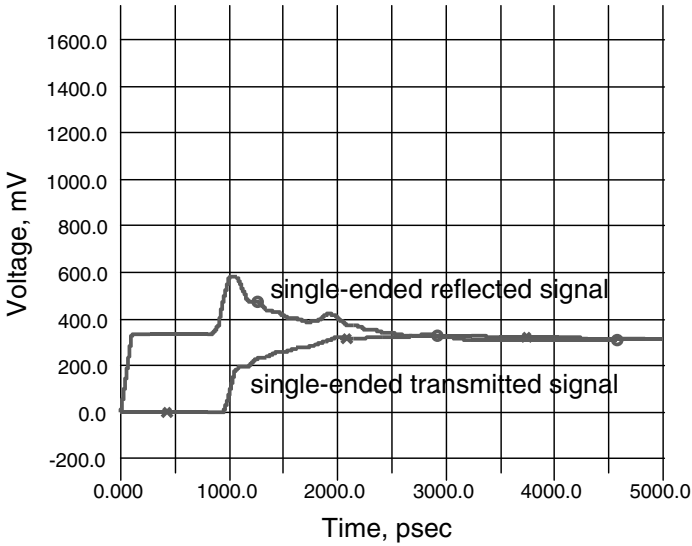


Figure 11-52 Reflected and transmitted 100-psec signal through a 50-Ohm transmission line with a 1-inch-wide gap in the return path in the middle. Simulated with Mentor Graphics Hyperlynx.

A single-ended signal will see a potentially disastrous discontinuity if the gap it encounters is wide. This will look like a large inductive discontinuity. Figure 11-52 shows the simulated reflected and transmitted single-ended signal passing over a 1-inch gap in the return path of an otherwise uniform 50-Ohm line with terminations at each end. The 100-psec initial rise time is dramatically increased due to the series inductive discontinuity.

While it may be possible in some situations to use a low-inductance capacitor to span the gap and provide a lower impedance path for the return, it is difficult to get good high-frequency performance.

An alternative approach to transport a signal across a gap in the return path with acceptable performance is using a differential pair. Figure 11-53 illustrates a typical case where the signal starts in one region of a board as an edge-coupled microstrip differential pair, in a second region where the return plane is far away, and in a third region where the interconnect is again an edge-coupled microstrip.

In regions 1 and 3, the differential pair has a differential impedance of about 90 Ohms. In the middle region, the return path is removed. With a 5-mil line and space and a 2.7-mil dielectric beneath the traces, but no other conductor, the differential impedance is about 160 Ohms. If the return plane is at least as far away as the span of the signal conductors, the differential impedance will be independent

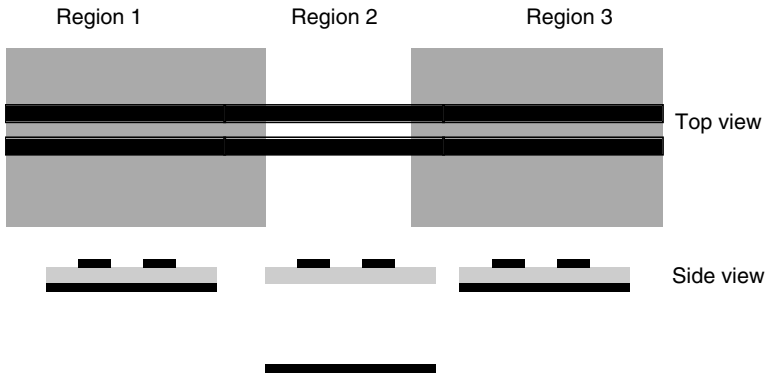


Figure 11-53 Region 2 of a circuit board is where there is a gap in the return path. This can be modeled as a region where the return path is very far away.

of the position of the return plane and it is as though it is not there. This trick allows us to build a circuit model for this discontinuity and to explore the impact of the discontinuity on the differential-signal quality. Figure 11-54 shows the simulated reflected and transmitted differential signal passing through the same gap as above. The rise time is preserved in the transmitted differential signal.

TIP Using a tightly coupled differential pair is one way of transmitting high-bandwidth signals in regions with a poorly defined return plane.

While the 160-Ohm differential impedance in the region of the gap is higher than the 90-Ohm differential impedance of the rest of the interconnect, it is a uniform transmission line. This will cause a degradation of signal quality.

11.19 To Tightly Couple or Not to Tightly Couple

A differential pair can transport a differential signal just as well if the coupling between the pairs is tight or loose. Of course, in fabricating for a target differential impedance, the cross section and stack-up must take into account the coupling. With an accurate 2D field solver tool, the stack-up for any target differential impedance and any level of coupling is equally easy to design.

The most significant advantage of using loosely coupled lines is the opportunity to use a wider line width. If we start with a differential pair that is loosely coupled and bring the traces closer together, keeping everything else the same, the

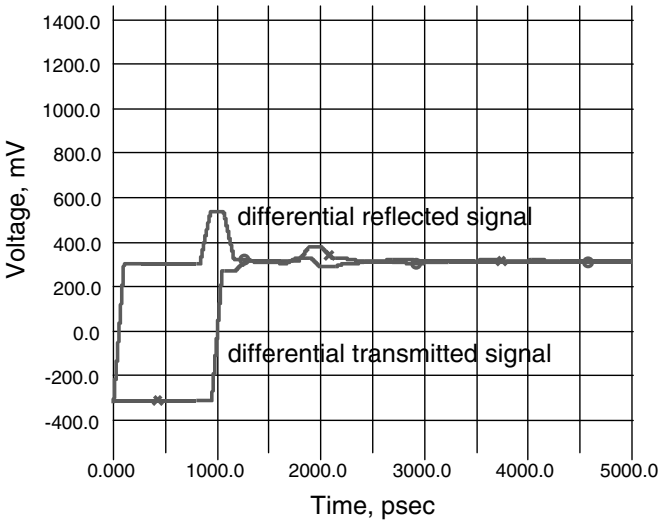


Figure 11-54 Differential signal with a 100-psec rise time transmitted and reflected across a 1-inch-wide gap in the return plane. The differential signal is minimally distorted across the gap. Simulated with Mentor Graphics Hyperlynx.

differential impedance will decrease. To achieve the target impedance, we have to make the lines narrower. This means higher resistance lines.

In comparing the line width of a loosely coupled and tightly coupled 100 Ohm differential pair, the line width has to decrease about 30% to maintain the 100 Ohms. This means a loosely coupled pair will have about 30% lower series resistive loss than a similar impedance tightly coupled pair. When loss is important, this can be an important reason to use loosely coupled pairs.

Another advantage in using loose coupling between the two lines in a differential pair is that the common-voltage noise on a victim line will be less than if the coupling were tight. However, if the common voltage is a problem because it causes EMI, the lowest cost and most effective way of solving this problem is by using a common-signal choke. The decrease in common-voltage noise from weak coupled lines compared with tightly coupled lines is never more than 20%.

Finally, another advantage of loosely coupled lines is that the differential impedance depends only on the single-ended impedance of either line, not on the spacing of the lines. This relaxes the routing constraints and may make it easier for the layout of a board. The two lines in a differential pair can take separate, independent paths around obstacles as long as they have a matched length. If their separation increases in localized areas, this will not affect their differential impedance.

There are also a number of considerations that offer an advantage in cost and performance if we use the tightest coupling between the lines in a differential pair that can be practically fabricated:

1. The interconnect density will be highest, so the board functional density will be highest and the board cost will be lowest.
2. The differential noise to a victim pair will be less.
3. The differential impedance discontinuity for a return-path imperfection will be reduced.

TIP When loss is important, loosely coupled differential pairs should be used. When interconnect density and noise immunity are important, tightly coupled differential pairs should be used. In general, with no overriding constraint, loose coupling with a spacing equal to twice the line width offers a reasonable compromise in providing the lowest loss at the highest interconnect density.

11.20 Calculating Odd and Even Modes from Capacitance- and Inductance-Matrix Elements

The first-order model of a single-ended transmission line is an n-section lumped-circuit model. This is described in terms of a capacitance per length and a loop inductance per length. The characteristic impedance and time delay of a single-ended transmission line are given by:

$$Z_0 = \sqrt{\frac{L_L}{C_L}} \tag{11-42}$$

$$TD = \sqrt{L_L C_L} \tag{11-43}$$

where:

Z_0 = the single-ended characteristic impedance of the line

L_L = the loop inductance per length of the line

C_L = the capacitance per length of the line

TD = the time delay of the line

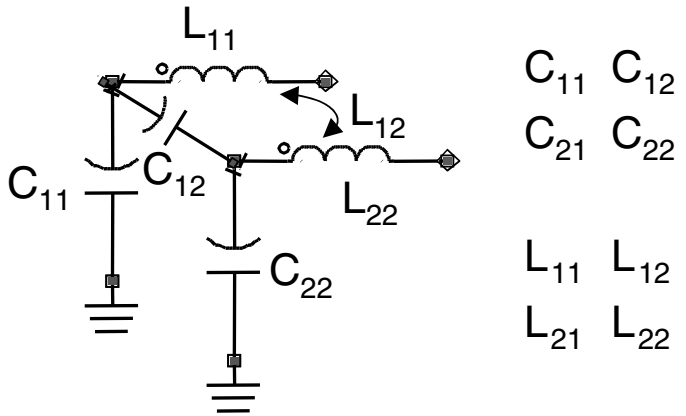


Figure 11-55 One section of an n-section lumped-circuit model for a pair of coupled transmission lines, with the SPICE capacitance matrix and loop-inductance matrix that defines all the model elements.

We can extend this model to include the coupling between two lines. Figure 11-55 shows the equivalent circuit model of a section of the two coupled transmission lines. The capacitor elements are defined in terms of the SPICE capacitor matrix and the inductance elements are defined in terms of the loop-inductance matrix. Of course, the values of the C and L matrices are obtained directly from the stack-up geometry of the differential pair by using a 2D field solver tool. We can use this model to determine the even- and odd-mode characteristic impedance values based on the values of the matrix elements.

When the pair is driven in the odd-mode state, the impedance of one line is the odd-mode characteristic impedance. The equivalent capacitance of one line is:

$$C_{odd} = C_{11} + 2C_{12} = C_{load} + C_{12} \tag{11-44}$$

where:

C_{odd} = the capacitance per length from the signal to the return path of one line when the pair is driven in the odd mode

C_{11} = the diagonal element of the SPICE capacitance matrix

C_{12} = the off-diagonal element of the SPICE capacitance matrix

C_{load} = the loaded capacitance of the signal line = $C_{11} + C_{12}$

In the odd mode, current is going into the signal line of trace 1 and back out its return path. At the same time, current is coming out of line 2 and into its return path. In this configuration, around transmission line 1 will be mutual magnetic-field lines from the signal in transmission line 2; these mutual magnetic-field lines are going opposite to the self-field lines of line 1. The equivalent loop inductance of line 1 will be reduced due to the current in line 2. The equivalent loop inductance of line 1 when the pair is driven in the odd mode is:

$$L_{odd} = L_{11} - L_{12} \tag{11-45}$$

where:

L_{odd} = the loop inductance per length from the signal to the return path of one line when the pair is driven in the odd mode

L_{11} = the diagonal element of the loop-inductance matrix

L_{12} = the off-diagonal element of the loop-inductance matrix

Looking into line 1 of the pair, we see a higher capacitance and lower loop inductance as the coupling between the lines increases. From these two terms, the odd-mode characteristic impedance and time delay can be calculated as:

$$Z_{odd} = \sqrt{\frac{L_{odd}}{C_{odd}}} = \sqrt{\frac{L_{11} - L_{12}}{C_{load} + C_{12}}} \tag{11-46}$$

$$TD_{odd} = \sqrt{L_{odd}C_{odd}} = \sqrt{(L_{11} - L_{12})(C_{load} + C_{12})} \tag{11-47}$$

When the pair is driven in the even mode, the capacitance between the signal and the return is reduced due to the shielding from the adjacent trace driven at the same voltage as trace 1. The equivalent capacitance per length is:

$$C_{even} = C_{11} = C_{load} - C_{12} \tag{11-48}$$

where:

C_{even} = the capacitance per length from the signal to the return path of one line when the pair is driven in the even mode

C_{11} = the diagonal element of the SPICE capacitance matrix

C_{12} = the off-diagonal element of the SPICE capacitance matrix

C_{load} = the loaded capacitance of the signal line = $C_{11} + C_{12}$

When driven in the even mode, current goes into signal line 1 and back out its return path. At the same time, current also goes into signal line 2 and back out its return path. The mutual-field lines from line 2 are in the same direction as the self-field lines in line 1. The equivalent loop inductance of line 1 when driven in the even mode is:

$$L_{even} = L_{11} + L_{12} \tag{11-49}$$

where:

L_{even} = the loop inductance per length from the signal to the return path of one line when the pair is driven in the even mode

L_{11} = the diagonal element of the loop-inductance matrix

L_{12} = the off-diagonal element of the loop-inductance matrix

From the capacitance per length and loop inductance per length looking into the front of line 1 when the pair is driven in the even mode, we can calculate the even-mode characteristic impedance and time delay:

$$Z_{even} = \sqrt{\frac{L_{even}}{C_{even}}} = \sqrt{\frac{L_{11} + L_{12}}{C_{load} - C_{12}}} \tag{11-50}$$

$$TD_{even} = \sqrt{L_{even}C_{even}} = \sqrt{(L_{11} + L_{12})(C_{load} - C_{12})} \tag{11-51}$$

These relationships based on the capacitance- and inductance-matrix elements are used by all field solvers to calculate the odd- and even-mode characteristic impedance and the time delays for any transmission, with any degree of coupling and any stack-up configuration. In this sense, the capacitance- and inductance-

matrix elements completely define the electrical properties of a pair of coupled transmission lines. This is a more fundamental description that shows the underlying effect of how the off-diagonal terms, C_{12} and L_{12}, that describe coupling affect the characteristic impedance and time delay for each mode. As coupling increases, the off-diagonal terms increase, the odd-mode impedance decreases, and the even-mode impedance increases.

11.21 The Characteristic Impedance Matrix

There is an alternative description of two or more coupled transmission lines that uses an impedance matrix. This is just as fundamental as the use of the capacitance and inductance matrices; it's just different. In this description, an impedance matrix is defined that relates the current and voltage into either line of a differential pair. Though the following analysis illustrates the impedance matrix for two coupled lines, it can be generalized to *n* coupled lines. It applies no matter what the stack-up geometry, symmetry, or materials distribution. These features will affect the parameter values, but not the general approach.

Two transmission lines are shown in Figure 11-56. Any arbitrary signal into either line can be described by the voltage between each line and the current into each signal line and out its return path. If there were no coupling between the lines then the voltage on each line would be independent of the other line. In this case, the voltages on each line would be given by:

$$V_1 = Z_1 I_1 \tag{11-52}$$

$$V_2 = Z_2 I_2 \tag{11-53}$$

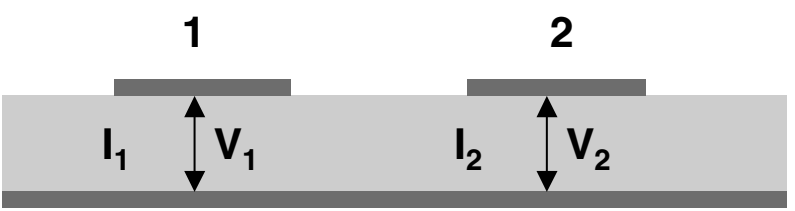


Figure 11-56 The voltage on each line and the current into each signal line and back out the return are labeled as shown.

However, if there is some coupling, this cross talk will cause the voltage on one line to be influenced by the current on the other line. We can describe this coupling with an impedance matrix. Each element of the matrix defines how the current into one line and out its return path influences the voltage in the other line. With the impedance matrix, the voltages on line 1 and 2 become:

$$V_1 = Z_{11}I_1 + Z_{12}I_2 \tag{11-54}$$

$$V_2 = Z_{21}I_1 + Z_{22}I_2 \tag{11-55}$$

The diagonal elements of the impedance matrix are the impedances of each line, when there is no current into the other line. Of course, when the coupling is turned off, the diagonal elements reduce to what we have called the characteristic impedance of the line.

The off-diagonal elements of the impedance matrix describe the amount of coupling but in a nonintuitive way. These are matrix elements. They are not the actual impedance between line 1 and line 2. Rather, they are the amount of voltage generated on line 1 per Amp of current in line 2. In this sense, they are really *mutual* impedances:

$$Z_{12} = \frac{V_1}{I_2} \tag{11-56}$$

$$Z_{21} = \frac{V_2}{I_1} \tag{11-57}$$

When there is little coupling, no voltage is created on one line for any current in the other line and the off-diagonal element between the lines is nearly zero. The smaller the off-diagonal term compared to the diagonal term, the smaller the coupling.

From this description, we can identify the odd- and even-mode impedances in terms of the characteristic impedance matrix. When a pure differential signal is applied to the two lines, the odd mode is driven. The odd mode is defined as when the current into one line is equal and opposite to the current into the other line, or $I_1 = -I_2$. Based on this definition, the voltages are:

$$V_1 = Z_{11}I_1 - Z_{12}I_1 = I_1(Z_{11} - Z_{12}) \tag{11-58}$$

From this, we can calculate the odd-mode impedance of line 1:

$$Z_{odd1} = \frac{V_1}{I_1} = Z_{11} - Z_{12} \tag{11-59}$$

Likewise, the even mode is when the currents into each line are the same, $I_1 = I_2$. The voltage on line 1 in the even mode is:

$$V_1 = Z_{11}I_1 + Z_{12}I_1 = I_1(Z_{11} + Z_{12}) \tag{11-60}$$

The even-mode impedance of line 1 is

$$Z_{even1} = \frac{V_1}{I_1} = Z_{11} + Z_{12} \tag{11-61}$$

In the same way, the odd- and even-mode impedance of the other line can be found. In this definition, the odd-mode impedance of one line is the difference between the diagonal and off-diagonal impedance-matrix elements. The larger the coupling, the larger the off-diagonal element and the smaller the differential impedance. The even-mode impedance of one line is the sum of the diagonal and off-diagonal impedance-matrix elements. As coupling increases, the off-diagonal element increases and the even-mode impedance gets larger.

Most 2D field solvers report the odd- and even-mode impedance, the capacitance and inductance matrices, and the impedance matrix.

From a signal's perspective, the only things that are important are the differential and common impedances, which can be described in three equivalent ways by:

1. Odd- and even-mode impedances

2. Capacitance- and inductance-matrix elements

3. Impedance matrix

These are all separate and independent ways of describing the electrical environment the common and differential signals see.

11.22 The Bottom Line

1. A differential pair is any two transmission lines.
2. Differential signaling has many signal-integrity advantages over single-ended signals, such as contributing to less rail collapse, less EMI, better noise immunity, and less sensitivity to attenuation.
3. Any signal on a differential pair can be described by a differential-signal component and a common-signal component. Each component will see a different impedance as it propagates down the pair.
4. The differential impedance is the impedance the differential signal sees.
5. A mode is a special state in which a differential pair operates. A voltage pattern that excites a mode will propagate down the line undistorted.
6. A differential pair can be fully and completely described by an odd-mode impedance, an even-mode impedance, and a time delay for the odd mode and for the even mode.
7. The odd-mode impedance is the impedance of one line when the pair is driven in the odd mode.
8. Forget the words *differential mode*. There is only *odd mode*, differential signal, and differential impedance.
9. Coupling between the lines in a pair will decrease the differential impedance.
10. The only reliable way of calculating the differential or common impedance is with an accurate 2D field solver.
11. Tighter coupling will decrease the differential cross talk picked up on a differential pair and will minimize the discontinuity the differential signal sees when crossing a gap in the return plane.
12. One of the most frequent sources of EMI is common signals getting onto an external twisted-pair cable. The way to reduce this is to minimize any asymmetries between the two lines in a differential pair and add a common-signal choke to the external cable.

13. All the fundamental information about the behavior of a differential pair is contained in the differential and common impedance. These can be more fundamentally described in terms of the even and odd modes, in terms of the capacitance and inductance matrices, or in terms of the characteristic impedance matrix.

CHAPTER 12

S-Parameters for Signal Integrity Applications

A new revolution hit the signal integrity field when signal bandwidths exceeded the 1 GHz mark. The rf world of radar and communications has been dealing with these frequencies and higher for more than 50 years. It is no wonder that many of the engineers entering the GHz regime of signal integrity came from the rf world and brought with them many of the analysis techniques common to the rf and microwave world. One of these techniques is the use of S-parameters.

TIP Though S-parameters are a technique with origins in the frequency domain, some of the principles and formalism can also be applied to the time domain. They have become the new universal standard to describe the behavior of any interconnect.

12.1 S-Parameters, the New Universal Metric

In the world of signal integrity, S-parameters have been relabeled a *behavioral model* as they can be used as a way of describing the general behavior of any linear, passive interconnect, which includes all interconnects with the exception of some ferrites.

In general, a signal is incident to the interconnect as the stimulus and the behavior of the interconnect generates a response signal. Buried in the stimulus-response of the waveforms is the behavioral model of the interconnect.

The electrical behavior of every sort of interconnect, such as that shown in Figure 12-1 can be described by S-parameters. These include:

- Resistors
- Capacitors
- Circuit board traces
- Circuit board planes
- Backplanes
- Connectors
- Packages
- Sockets
- Cables

That's why this technique has become and will continue to be such a powerful formalism and in such wide use.

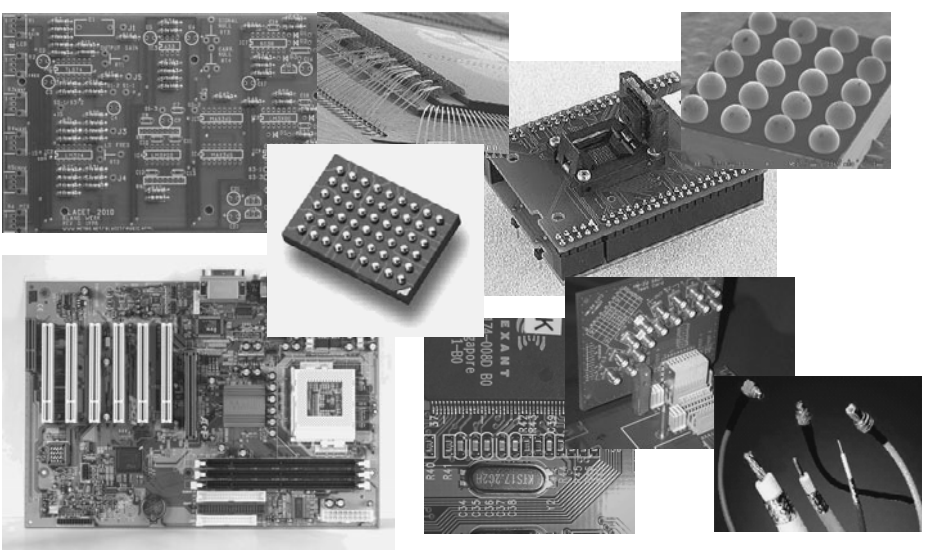


Figure 12-1 The formalism of S-Parameters applies to all passive, linear interconnects such as the ones shown here.

12.2 What Are S-Parameters?

Fundamentally, a behavioral model describes how an interconnect interacts with a precision, incident waveform. When describing in the frequency domain, the precision waveform is, of course, a sine wave. However, when describing the behavior in the time domain, the precision waveform can be a step edge or even an impulse waveform. As long as the waveform is well characterized, it can be used to create a behavioral model of the interconnect or device under test, DUT.

In the frequency domain, where sine waves interact with the DUT, the behavioral model is described by the S-parameters. In the time domain, we use the labeling scheme of S-parameters, but interpret the results differently.

> **TIP** Fundamentally, the S-parameters describe how precision waveforms, like sine waves, scatter from the ends of the interconnect. The term S-parameters is short for scattering parameters.

When a waveform is incident on an interconnect, it can scatter back from the interconnect, or it can scatter into another connection of the interconnect. This is illustrated in Figure 12-2.

For historical reasons, we use the term *scatter* when referring to how the waveform interacts. An incident signal can "scatter" off the front of the DUT, back into the source, or "scatter" into another connection. The "S" in S-parameters stands for *scattering*.

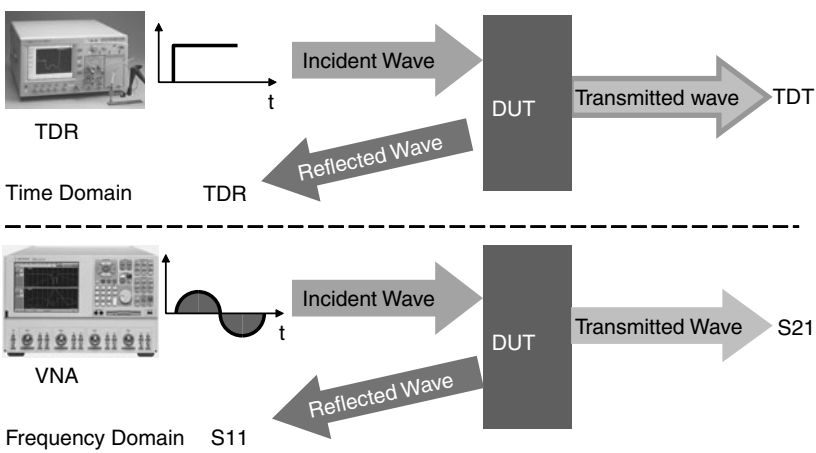


Figure 12-2 S-parameters are a formalism to describe how precision waveforms scatter from an interconnect or device under test (DUT).

We also call the wave that scatters back to the source the reflected wave and the wave that scatters through the device the transmitted wave. When the scattered waveforms are measured in the time domain, the incident waveform is typically a step edge, and we refer to the reflected wave as the time domain reflection (TDR) response. The instrument used to measure the TDR response is called a *time domain reflectometer (TDR)*. The transmitted wave is the *time domain transmitted (TDT)* wave.

In the frequency domain, the instrument used to measure the reflected and transmitted response of the sine waves is a *vector network analyzer (VNA)*. *Vector* refers to the fact that both the magnitude and phase of the sine wave are being measured. A *scalar network analyzer* just measures the amplitude of the sine wave, not its phase.

The frequency domain reflected and transmitted terms are referred to as specific S-parameters, such as S11 and S21 or the return and insertion loss.

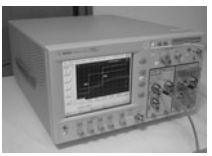
This formalism of describing the way precision waveforms interact with the interconnect can be applied to measurements or the output from simulations, as illustrated in Figure 12-3. All electromagnetic simulations, whether conducted in the time or the frequency domain, also use the S-parameter formalism.

TIP S-parameters, a formalism developed in the rf world of narrow band carrier waves, has become the de facto standard format to describe the wide bandwidth, high frequency behavior of interconnects in signal integrity applications.

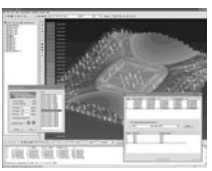
Regardless of where the S-parameter values come from, they represent a description of the way electrical signals behave when they interact with the interconnect. From this behavioral model, it is possible to predict the way any arbitrary signal might interact with the interconnect, and from this behavior predict output waveforms, such as an eye diagram. This process of using the behavior model to predict a system response is called emulation or simulation.

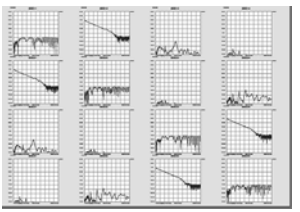
There is a wealth of information buried in the S-parameters, which describe some of the characteristics of an interconnect, such as its impedance profile, the amount of cross talk, and the attenuation of a differential signal.

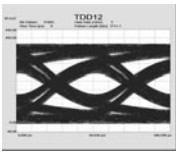
With the right software tool, the behavioral measurements of an interconnect can be used to fit a circuit topology based model of an interconnect, such as a connector, a via, or an entire backplane. With an accurate circuit model that matches physical features with performance, it is possible to "hack" into the model and identify what physical features contribute to the limitations of the interconnect


TDR

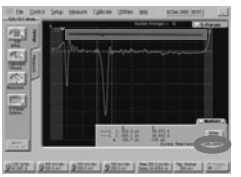

VNA


EM Sim


Single ended time domain
Single ended frequency domain
Differential time domain
Differential frequency domain



Emulation


Characterization

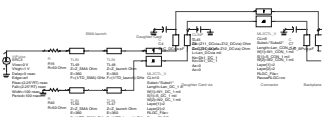

Hacking

Figure 12-3 Regardless of where the behavioral model comes from, it can be used to emulate system performance, characterize the interconnect, or hack into the performance limitations of the interconnect.

and suggest how to improve the design. This process is popularly called *hacking interconnects*.

12.3 Basic S-Parameter Formalism

S-parameters describe the way an interconnect affects an incident signal. We call the ends where signals enter or exit a device under test (DUT), a port. These are connections for both the signal and return paths of the DUT. The easiest way of thinking about a port is as a small coaxial connection to the DUT.

Unless otherwise stated, the impedance the signal sees inside the connection leading up to the DUT is 50 Ohms. In principle, the port impedance can be made any value.

> **TIP** S-parameters are confusing enough without also arbitrarily changing the port impedances. Unless there is a compelling reason, the port impedances should be kept at 50 Ohms.

Each S-parameter is the ratio of a sine wave scattered from the DUT at a specific port, to the sine wave incident to the DUT at a specific port.

For all linear, passive elements, the frequency of the scattered wave will be exactly the same as the incident wave. The only two qualities of the sine wave that can change are the amplitude and phase of the scattered wave.

To keep track of which port the sine wave enters and exits, we label the ports with consecutive index numbers and use these index numbers in each S-parameter.

Each S-parameter is the ratio of the output sine wave to the input sine wave:

$$S = \frac{\text{output sine wave}}{\text{input sine wave}} \tag{12-1}$$

The ratio of two sine waves is really two numbers. It is a magnitude that is the ratio of the amplitudes of the output to the input sine waves, and it is the phase difference between the output and the input sine waves. The magnitude of an S-parameter is the ratio of the amplitudes:

$$\text{mag}(S) = \frac{\text{amplitude(output sine wave)}}{\text{amplitude(input sine wave)}} \tag{12-2}$$

While the magnitude of each S-parameter is just a number from 0 to 1, it is often described in dB. As discussed in an earlier chapter, the dB ALWAYS refers to the ratio of two powers. Since an S-parameter is the ratio of two voltage amplitudes, the dB value must relate to the ratio of the powers within the voltage amplitudes. This is why when translating between the dB value and the magnitude value, a factor of 20 is used:

$$S_{dB} = 20 \times \log(S_{mag}) \tag{12-3}$$

where:

S_{dB} = the value of the magnitude in dB

S_{mag} = the value of the magnitude as a number

The phase of the S-parameter is the phase difference between the output wave minus the input wave:

$$\text{phase}(S) = \text{phase}(\text{output sine wave}) - \text{phase}(\text{input sine wave}) \quad (12\text{-}4)$$

As we will see, the order of the waveforms in the definition of the phase of the S-parameter will be important when determining the phase of the reflected or transmitted S-parameter and will contribute to a negative advancing phase.

TIP While the port assignment of a DUT can be arbitrary, and there is no industry standardized convention, there should be. If we are using the indexes to label each S-parameter, then changing the port assignment index labels will change the meaning of specific S-parameters.

For multiple, coupled transmission lines, such as a collection of differential channels, there is one port assignment scheme that is convenient and scalable and should be adopted as the industry standard. It is illustrated in Figure 12-4.

The ports are assigned as an odd port on the left end of a transmission line and the next higher number on the other end of the line. This way, as the number of coupled lines in the array increases, additional index numbers can be added following this rule. This is a very convenient port assignment approach. S-parameters are confusing enough without creating more confusion about mix-ups in the port assignments.

TIP You should always try to adopt a port assignment for multiple transmission lines that has port 1 connected to port 2 and port 3 adjacent to port 1, feeding into port 4. This is scalable to n additional transmission lines.

In order to distinguish to which combination of ports each S-parameter refers, two indices are used. The first index is the output port, while the second index is the input port.

For example, the S-parameter for the sine wave going into port 1 and coming out port 2 would be S21. This is exactly the opposite of what you would expect. It

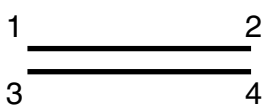


Figure 12-4 Recommended port assignment labeling scheme for multiple transmission line interconnects. The pattern continues as additional lines are added.

would be logical to use the first index as the going in port and the second number as the coming out port. However, part of the mathematical formalism for S-parameters requires this reverse-from-the-logical convention. This is related to the matrix math that is the real power behind S-parameters. By using the reverse notation, the S-parameter matrix can transform an array of stimulus voltage vectors, represented by the vector a_j, into an array of response voltage vectors, b_k:

$$b_k = S_{kj} \times a_j \tag{12-5}$$

Using this formalism, the transformation of a sine wave into and out of each port can be defined as a different S-parameter, with a different pair of index values. The definition of each S-parameter element is:

$$S_{kj} = \frac{\text{sine wave out port k}}{\text{sine wave into port j}} \tag{12-6}$$

This basic definition applies no matter what the internal structure of the DUT may be, as illustrated in Figure 12-5. A signal going into port 1 and coming out port 1 would be labeled as S11. A signal going into port 1 and coming out port 2 would be labeled as S21. Likewise, a signal going into port 2 and coming out port 1 would be labeled as S12.

12.4 S-Parameter Matrix Elements

With only one port on a DUT, there is only one S-parameter, which would be identified as S11. This one element may have many data points associated with it for many different frequency values. At any single frequency, S11 will be complex, so it will really be two numbers. It could be described by a magnitude and phase, or by a real and imaginary component. This single S-parameter value, at one frequency, could be plotted on either a polar plot or Cartesian plot.

In addition, S11 may have different values at different frequencies. To describe the frequency behavior of S11, the magnitude and phase could be plotted at each frequency. An example of the measured S11 from a short length transmission line, open at the far end, is shown in Figure 12-6.

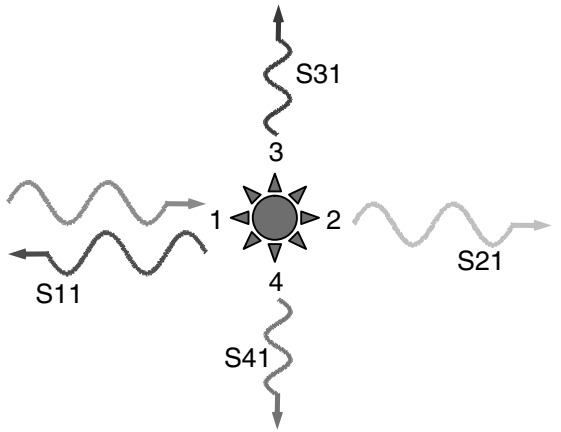


Figure 12-5 Index labeling definition for each S-parameter.

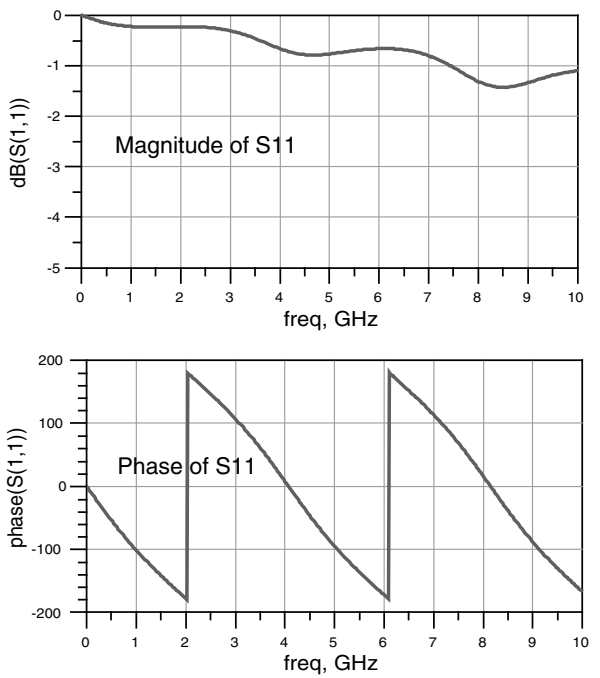


Figure 12-6 Measured S11 as a magnitude (top) and phase (bottom) for a transmission line, open at the far end. Connection was made to both the signal and return path at port 1. Measured with an Agilent N5230 VNA and displayed with Agilent's ADS.

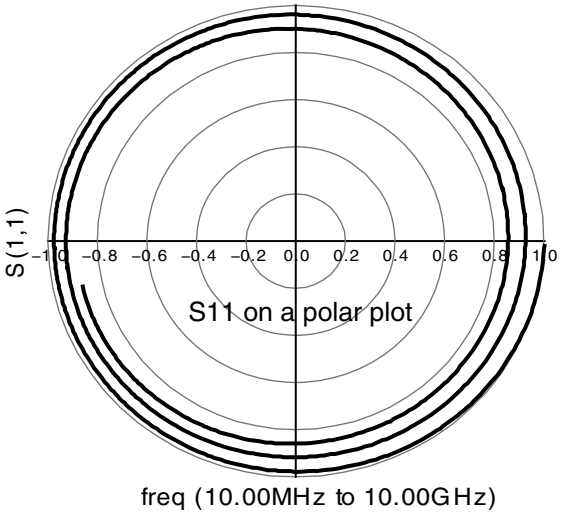


Figure 12-7 The same measured S11 data as above, but replotted in a polar plot The radius position is the magnitude and the angular position is the phase of S11.

Alternatively, the S11 value at each frequency can be plotted in a polar plot. The radial position of each point is the magnitude of the S-parameter and the angle from the real axis is the phase of S11. An example of the same measured S11, plotted in a polar plot is shown in Figure 12-7.

The information is exactly the same in each case, it just looks different, depending on how it is displayed. When displayed in a polar plot, it is difficult to determine the frequency values of each point unless a marker is used.

A two-port device would have four possible S-parameters. From port 1, there could be a signal coming out at port 1 and coming out at port 2. The same could happen with a signal going into port 2. The S-parameters associated with the two-port device can be grouped into a simple matrix:

$$\begin{matrix} S11 & S12 \\ S21 & S22 \end{matrix} \tag{12-7}$$

In general, if the interconnect is not physically symmetrical, S11 will not equal S22. However, for all linear, passive devices, S21 = S12, always. There are

only three unique terms in the four element S-parameter matrix. An example of the measured two-port S-parameters of a stripline transmission line is shown in Figure 12-8.

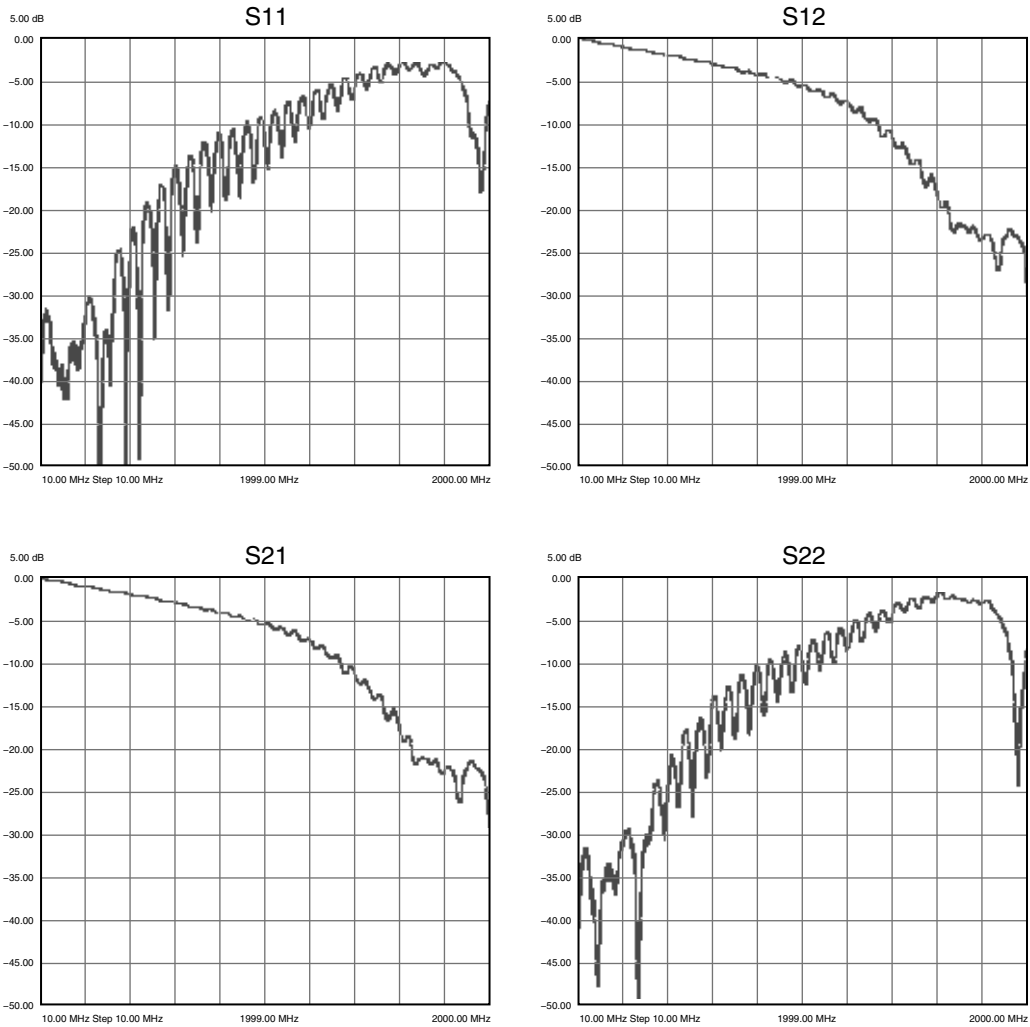


Figure 12-8 Measured two-port S-parameters of a simple 5-inch-long nearly 50 Ohm stripline transmission line, from 10 MHz to 20 GHz. Shown here are only the magnitude values. There are also phase values for each matrix element and at each frequency point, not shown. Measured with an Agilent N5230 VNA and displayed with Agilent's PLTS.

This example illustrates part of the power behind the S-parameter formalism. So simple a term as the S-parameter matrix represents a tremendous amount of data. Each of the three unique elements of the 2 × 2 matrix has a magnitude and phase at each frequency value of the measurement, from 10 MHz to 20 GHz at 10 MHz intervals. This is a total of 2,000 × 2 × 3 = 12,000 specific, unique data points, all neatly and conveniently catalog by the S-parameter matrix elements.

This formalism can be extended to include an unlimited number of elements. With 12 different ports, for example, there would be 12 × 12 = 144 different S-parameter elements. However, not all of them are unique. For an arbitrary interconnect, the diagonal elements are unique, and the lower half of the off-diagonal elements is unique. This is a total of 78 unique terms.

In general, the number of unique S-parameter elements is given by:

$$N_{unique} = \frac{n(n+1)}{2} \tag{12-8}$$

where:

N_{unique} = the number of unique S-parameter elements

n = the number of ports

In this example of 12 ports, there are (12 × 13)/2 = 78 unique elements. And, each element has two different sets of data, a magnitude and a phase. This is a total of 156 different plots. If there are 1,000 frequency values per plot, this is a total of 156,000 unique data points.

TIP With this huge amount of data, having a simple formalism to just keep track of the data is critically important.

There are really two aspects of S-parameters. First and most important is the analytical information contained in the S-parameter matrix elements. But there is also information that can be read from the patterns the S-parameter values make when plotted either in polar or Cartesian coordinates. A skilled eye can pick out important properties of an interconnect just from the pattern of the traces.

These S-parameter matrix elements and the data each contains is really the precise behavior of the interconnect. Everything you ever wanted to know about the behavior of an interconnect is contained in its S-parameter matrix elements.

Each S-parameter matrix element tells a different story about the 12-port device. The analytical information they contain can be immediately accessed with a variety of simulation tools.

12.5 Simulating Return and Insertion Loss

In a two-port device, there are three unique S-parameters: S11, S22, and S21. Each of these matrix elements contains complex numbers that may vary with frequency.

For historical reasons, the magnitude of the reflected S-parameter, S11, is called the return loss, and the magnitude of the transmitted S-parameter, S21, is called the insertion loss.

Insertion loss is a measure of what is lost from the signal when the interconnect is "inserted" between the two ports of a network analyzer. The *return loss* is what is returned to the incident port and lost to the transmitted signal. These terms arose in the early days of network analysis when just the magnitude of signals was recorded, and not the phase. The terms are still in common use today and are often a more convenient way of saying "the magnitude of S11" or "the magnitude of S21."

When the interconnect is symmetrical from one end to the other, the return losses, S11 and S22, are equal. In an asymmetric two-port interconnect, S11 and S22 will be different.

In general, calculating by hand the return and insertion loss from an interconnect line is complicated. It depends on the impedance profile and time delays of each transmission line segment that makes up the interconnect and the frequency of the sine waves.

In the frequency domain, any response to the sine waves is at steady state. The sine wave is on for a long time and the total reflected or transmitted response is observed. This is an important distinction between the frequency domain and the time domain.

In the time domain, the instantaneous voltage reflected or transmitted by the transmission line is observed. The reflected response, the TDR response, can map out the spatial impedance profile of the interconnect, by looking at when a reflection occurred and where the exciting edge must have been when the reflection happened. Of course, after the first reflection, an accurate interpretation of the impedance profile can only be obtained by post processing the reflected signal, but a good first order estimate can always be read directly from the TDR response.

In the frequency domain, the spatial information is intermixed throughout the frequency domain data and not directly displayed. Consider an interconnect with many impedance discontinuities down its length, as shown in Figure 12-9.

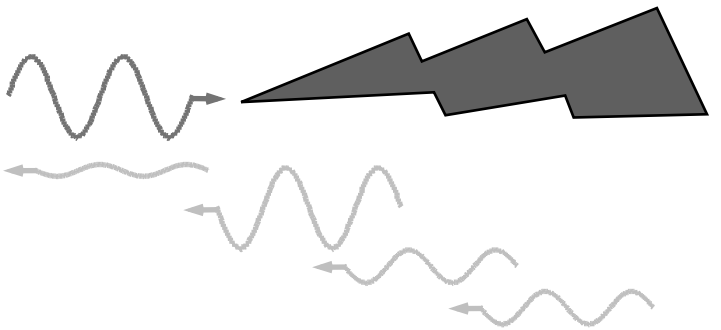


Figure 12-9 Each arbitrary impedance discontinuity in the irregularly shaped interconnect causes a reflected sine wave back to the incident port, each with a different magnitude and phase.

As the incident sine wave encounters each discontinuity, there will be a reflection and some of the sine wave will head back to the port. Some of the reflected waves will bounce multiple times between the discontinuities until they either are absorbed or eventually make it out to one of the ports where they are recorded.

What is seen as the reflected signal from port 1, for example, is the combination of all the reflected sine waves from all the possible discontinuities. For an interconnect 1 meter long, the typical time for a reflection down and back is about 6 nsec. In the 1 msec, a typical time for a VNA to perform one frequency measurement, a sine wave could complete more than 100,000 bounces, far more than would ever occur.

TIP For any single frequency, the reflected signal or transmitted signal is a steady state value. It represents all the possible combinations of reflections at all the different impedance interfaces. This is very different than the behavior in the time domain.

If the frequency incident to the port is fixed during the 1 msec measurement or simulation time, the frequency of every wave reflecting back from every discontinuity will be exactly the same frequency. However, the amplitude and phase of each resulting wave exiting the interconnect at each port, will be different.

Coming out of each port will be a large number of sine waves, all with the same frequency, but with an arbitrary combination of amplitudes and phases, as illustrated in Figure 12-10. Surprisingly, when we add together an arbitrary number of sine waves, each with the same frequency, but with arbitrary amplitude and phase, we get another sine wave.

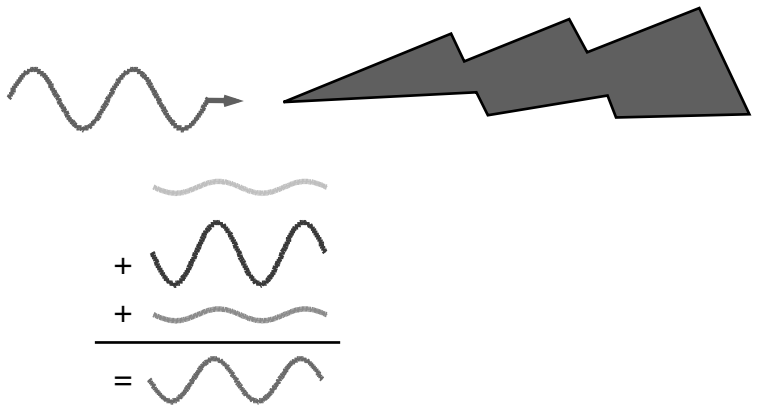


Figure 12-10 The sum of a large number of sine waves, all with the same frequency, but different amplitudes and phase, is another sine wave.

When a sine wave is incident into one of the ports, the resulting signal that comes back out one of the ports is also a sine wave, of exactly the same frequency, just with a different amplitude and phase. This is what is captured in the S-parameters. Unfortunately, other than a few special cases, the behavior of the S-parameters is a complicated function of the impedance profile and sine wave frequency. It is not possible to take the measured magnitude and phase of any S-parameter and back out each individual sine wave element that created it.

In general, other than a few simple cases, it is not possible to calculate the S-parameters of an interconnect by hand with pencil and paper. A simulator must be used. Though there are a number of sophisticated, commercially available simulators that can simulate the S-parameters of arbitrary structures, any SPICE simulator can calculate the return and insertion loss of any arbitrary structure with a simple circuit, as illustrated in Figure 12-11. An example of an interconnect with a few discontinuities is also shown.

While any arbitrary interconnect can be simulated, there are a few important patterns in the S-parameters that are indicative of specific features in the interconnect. Learning to recognize some of these patterns will enable the trained observer to immediately translate S-parameters into useful information about the interconnect.

One important pattern to recognize is that of a *transparent interconnect*. The pattern of the return and insertion loss can immediately indicate the "quality" of the interconnect, as "good" or "bad." In this context, we assume that "good" means the interconnect is transparent to signals, and "bad" means it is not transparent to signals.

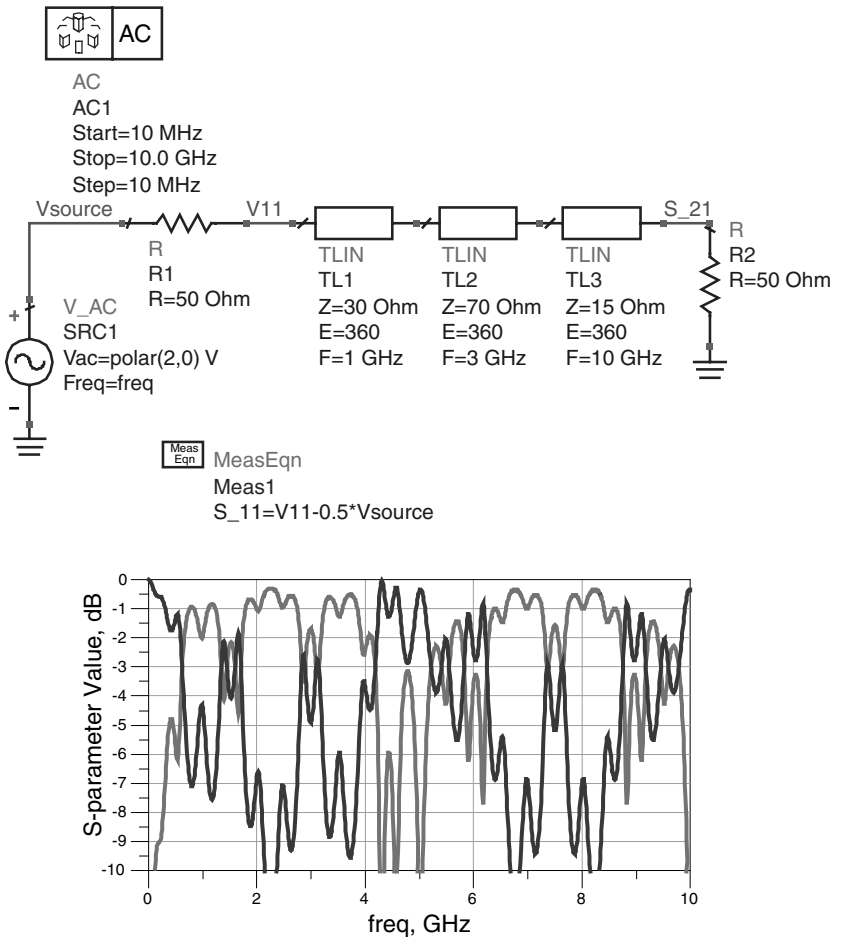


Figure 12-11 SPICE circuit to calculate the S11 and S21 of any interconnect circuit between the ports, circuit as set up in Agilent's ADS version of SPICE.

12.6 A Transparent Interconnect

There are three important features of a transparent interconnect:

- The instantaneous impedance down the length matches the impedance of environment in which it is embedded.
- The losses through the interconnect are low and most of the signal is transmitted.
- There is negligible coupling to adjacent traces.

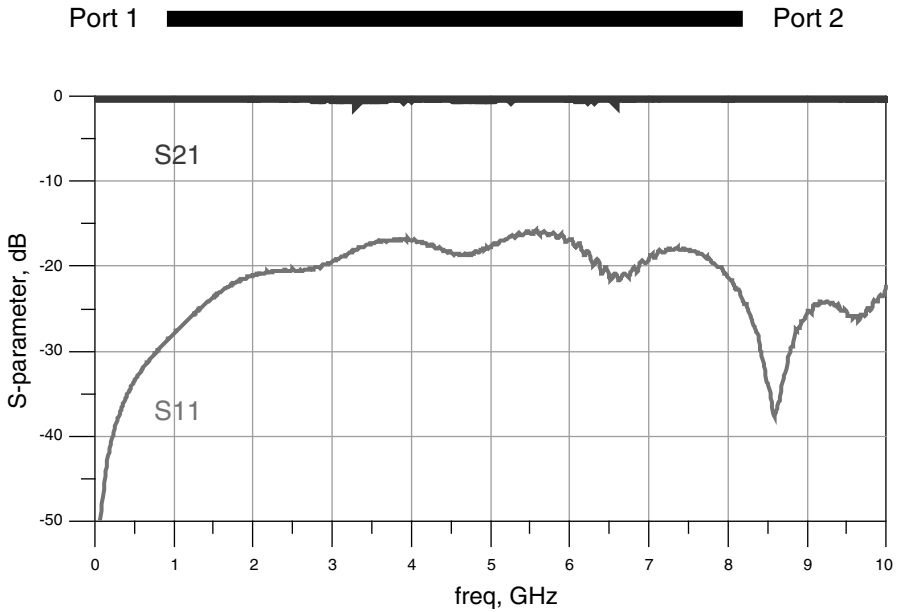


Figure 12-12 Measured return and insertion loss of a nearly transparent interconnect. Measured with an Agilent N5230 VNA and displayed with Agilent's ADS.

These three features are clearly displayed at a glance in the reflected and transmitted signals, which correspond to the S11 and S21 terms, when the ports are attached to opposite ends of the interconnect, labeled as port 1 and port 2.

Examples of the port configuration and measured return and insertion loss for a nearly transparent interconnect are shown in Figure 12-12.

When the impedance throughout the interconnect closely matches the port impedances, very little incident signal reflects and the reflected term, S11, is small. When displayed in dB, a smaller return loss is a larger, more negative dB value. The 50-Ohm impedance of port 2 effectively terminates the interconnect.

Of course, in practice, it is almost impossible to achieve a perfect match to 50 Ohms over a large bandwidth. Typically, the measured return loss of an interconnect gets worse at higher frequency, as seen in the above example as a smaller negative dB value at higher frequency.

TIP When displayed as a dB value, a transparent interconnect will have a large negative dB return loss.

The worse the interconnect and the larger the impedance mismatch to the port impedances, the closer the return loss will be to 0 dB, corresponding to 100% reflection.

The insertion loss is a measure of the signal that transmits through the device and out of port 2. The larger the impedance mismatch, the less transmitted signal. However, when there is close to a good match, the insertion loss is very nearly 0 dB and is insensitive to impedance variations.

There is a specific connection between the return loss and the insertion loss. Always keep in mind that the S-parameters are ratios of voltages. There is no law of conservation of voltage, but there is a law of conservation of energy.

If the interconnect is low loss, and there is no coupling to adjacent traces, and there is no radiated emissions, then the energy into the interconnect must be the sum of the reflected energy and the transmitted energy.

The energy in a sine wave is proportional to the square of the amplitude. This condition of energy in is equal to the energy reflected plus the energy transmitted, is described by:

$$1 = S11^2 + S21^2 \tag{12-9}$$

Given the return loss, the insertion loss is:

$$S21 = \sqrt{1 - S1^2} \tag{12-10}$$

For example, if the impedance somewhere on the interconnect is 60 Ohms in an otherwise 50-Ohm environment, the worst case return loss would be:

$$S11 = \frac{(60 - 50)}{(60 + 50)} = \frac{10}{110} = 0.091 = -21\,\text{dB} \tag{12-11}$$

The impact on insertion loss is:

$$S21 = \sqrt{1 - 0.091^2} = \sqrt{0.992} = 0.996 = -0.04\ \text{dB} \tag{12-12}$$

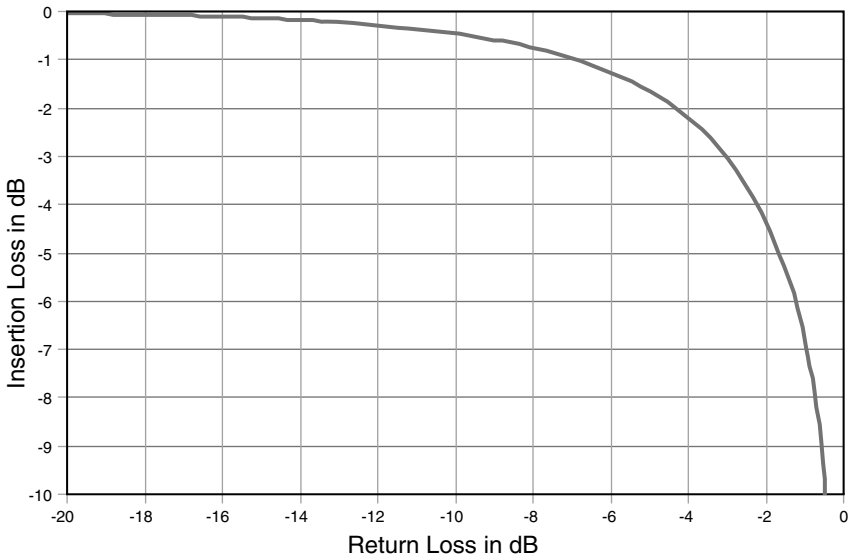


Figure 12-13 Impact on the insertion loss for different return losses. The return loss must be as high as -10 dB for there to be even 0.5 dB impact on the insertion loss, if there are no losses in the interconnect.

Even though the return loss is as high as -20 dB, the impact on insertion loss is very small and indistinguishable from 0 dB. Figure 12-13 illustrates this for a wide range of return loss values.

TIP Only for a return loss higher than -10 dB will there be a noticeable impact on the insertion loss.

12.7 Changing the Port Impedance

The industry standard port impedance is 50 Ohms. However, in principle, it can be made any value. As the port impedance changes, the behavior of the displayed return and insertion loss changes. To first order, moving the port impedance farther from the interconnect's characteristic impedance will increase the return loss. Other than this simple pattern, the specific values of the return loss and insertion loss are complicated functions of the port impedance.

Given the S-parameters at one port impedance, the S-parameters with any other port impedance can be calculated using matrix math. It can also be calculated directly in a SPICE circuit simulation.

TIP To describe the S-parameters of an interconnect it is not necessary to match the port impedance to the impedance of the device. Unless there is a compelling reason otherwise, 50 Ohms should always be used.

Regardless of the port impedance, an analytical analysis of each S-parameter element can be equally well performed. The only practical reason to switch the port impedance from 50 Ohms is to be able to qualitatively evaluate, from the front screen, the quality of the device in a non-50-Ohm environment.

If the device, like a connector or cable, is designed for a non-50-Ohm application, like the 75 Ohms environment of cable TV applications, its return loss will look "bad" when displayed with 50-Ohm port impedances.

The reflections from the impedance mismatches at the ends will cause ripples in the return loss. The excessive return loss will show ripples in the insertion loss. To the trained eye, the behavior will look complicated and be difficult to interpret, other than the impedance is way off from 50 Ohms.

But, if the application environment was 75 Ohms, the interconnect might be acceptable. By changing the port impedance to 75 Ohms, the application impedance, the behavior of the device in this application environment can be visually evaluated right from the front screen.

An example of the return and insertion loss of a nominally 75-Ohm transmission line with 50-Ohm connectors attached is shown in Figure 12-14 for the case of 50-Ohm and 75-Ohm port impedances.

In a 75-Ohm environment, the 75-Ohm cable looks nearly transparent at low frequency, below about 1 GHz. Above 1 GHz, the connectors dominate the behavior, and the interconnect is not very transparent, no matter what the port impedance. The behavioral model contained in the measured S-parameter response is exactly the same, regardless of the port impedances. It is merely redisplayed for different port impedances.

S-parameters are confusing enough without also leaving the port impedance ambiguous. Of course, when the S-parameters are stored in the industry standard, touchstone file format, the port impedances of each port to which the data is referenced is specifically called out at the top of the file. From one touchstone file, the S-parameters for any port impedance can easily be calculated.

12.8 The Phase of S21 for a Uniform 50-Ohm Transmission Line

The simplest interconnect to evaluate is when the impedance of the line is 50 Ohms, matched to the impedance of the ports. In this case, there are no reflections,

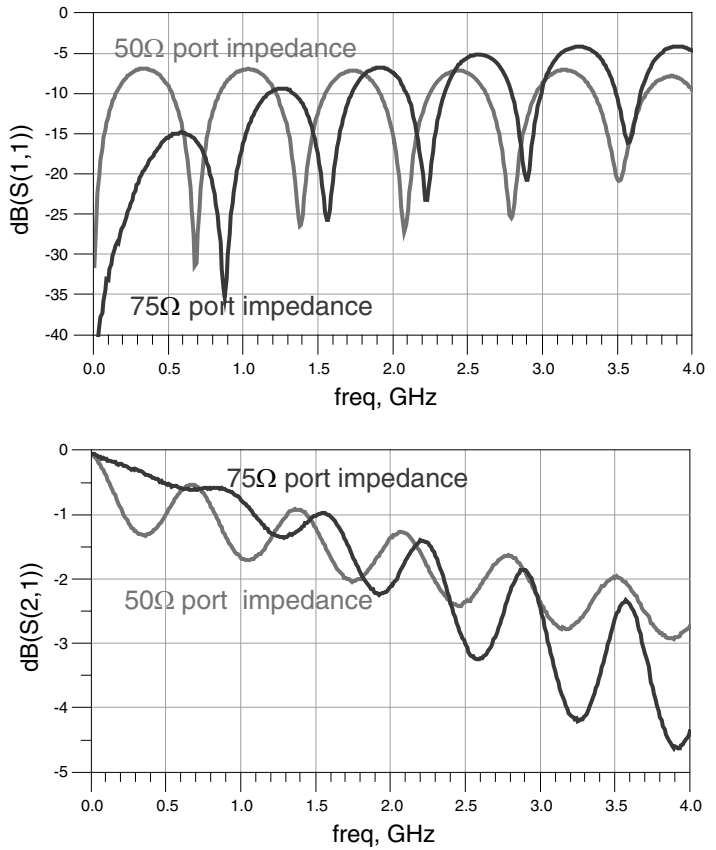


Figure 12-14 Measured return and insertion loss for a 75-Ohm transmission line with 50-Ohm connectors using a port impedance of 50 Ohms and 75 Ohms. Measured with an Agilent N5230 VNA and displayed with Agilent's ADS.

and the magnitude of S11 is 0. In dB, this is a large, negative dB, usually limited by the noise floor of the instrument or simulator, on the order of -100 dB.

All of the sine wave would be transmitted so the magnitude of S21 would be 1, which is 0 dB at each frequency. The phase of S21 would vary depending on the time delay of the transmission line and the frequency. The behavior of the phase is the most subtle aspect of S-parameters.

The definition of the S-parameters is that each matrix element is the ratio of the sine wave that comes out of a port to the sine wave that goes into a port.

For S21, the ratio of the sine wave coming out of port 2 to the sine wave going into port 1, we get two terms:

$$\text{mag(S21)} = \frac{\text{sine wave amplitude out of port 2}}{\text{sine wave amplitude into port 1}} \tag{12-13}$$

$$\text{phase(S21)} = \text{phase(sine wave out of port 2)} - \text{phase(sine wave into port 1)} \tag{12-14}$$

When we send a sine wave into port 1, it doesn't come out of port 2 until a time delay, TD, later. If the phase is 0 degrees when the sine wave went into port 1, it will also have a phase of 0 degrees when it comes out of port 2. It has merely been transmitted from one end of the line to the other.

However, when we compare the phase of the sine wave coming out of port 2 to the phase of the sine wave going into port 1, it is the phases that are present at the same instant of time. This is illustrated in Figure 12-15.

When we see the 0 degree phase sine wave coming out of port 2 and then immediately look at the phase of the wave going into port 1, we are looking not at the 0 degree phase when the sine wave entered the transmission line, TD nsec ago, but at the current phase of the sine wave entering port 1.

In the time the 0 degree wave front has been propagating through the transmission line, the phase of the sine wave entering port 1 has advanced. The phase of the sine wave entering port 1 is now f × TD.

When we calculate the phase of S21, as the difference between the phase coming out of port 2 minus the phase going into port 1, the phase coming out of

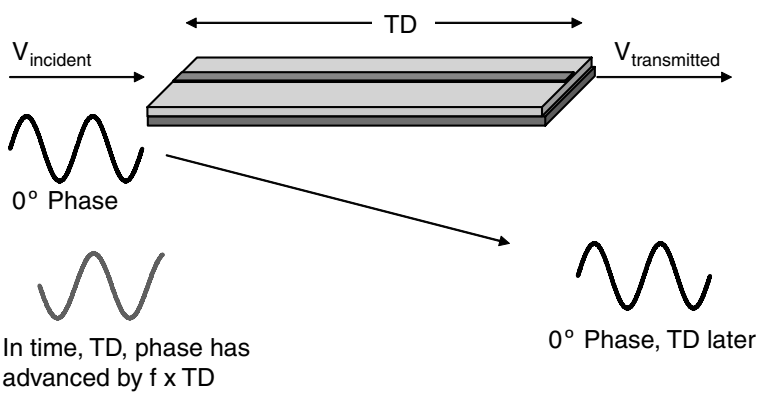


Figure 12-15 Phase of S21 is negative due to incident phase advancing while the sine wave is transmitting through the transmission line.

port 2 may be 0 degrees, but the phase going into port 1, NOW, has advanced to f × TD. This means the phase of S21 is:

$$\text{phase}(S21) = 0° - f \times TD \tag{12-15}$$

where:

f = the frequency of the sine wave going into port 1

TD = the time delay of the transmission line

TIP The phase of S21 will start out negative and increase more and more negative with frequency. This is the most bizarre and confusing aspect of S-parameters, the phase of S21 through a transmission line is increasingly negative.

We see this behavior is based on two features. First, the definition of the phase of S21 is that it is the phase of the sine wave coming out of port 2 minus the phase of the sine wave going into port 1. Second, it is the difference in phase of the two waves, at the same instant in time. The phase of S21 will always be negative and increase more negative as frequency increases. An example of the measured phase of S21 for a stripline transmission line is shown in Figure 12-16.

At low frequency, the phase of S21 starts out very nearly 0 degrees. After all, if the TD is small compared to a cycle, then the phase does not advance very far during the transit time of the wave through the interconnect. As frequency increases, the number of cycles the sine wave advances during the transit time increases. Since the phase of S21 is the negative of the incident phase, as the incident phase advances, the phase of S21 gets more and more negative.

We usually count phase from -180 degrees to +180 degrees. When the phase advances to -180, it is reset to +180 and continues to count down. This gives rise to the typical sawtooth pattern for the phase of S21.

12.9 The Magnitude of S21 for a Uniform Transmission Line

The magnitude of the insertion loss is a measure of all the processes that prevent the energy being transmitted through the interconnect. The total energy flowing into an interconnect must equal the total energy coming out. There are five ways energy can come out:

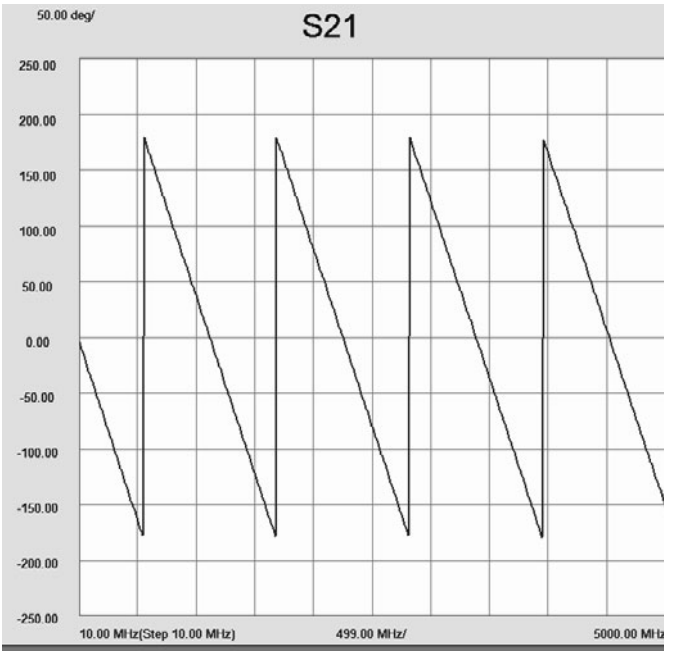


Figure 12-16 Measured phase of S21 for a uniform, 50-Ohm stripline, about five inches long, measured from 10 MHz to 5 GHz. Measured with an Agilent N5230 VNA and displayed with Agilent's PLTS.

- Radiated emissions
- Losses in the interconnect turning into heat
- Energy coupled into adjacent traces, whether measured or not
- Reflected energy back to the source
- Transmitted energy into port 2 and measured as part of S21

In most applications, the impact on S21 from radiated losses is negligible. Though radiated emissions are very important in causing FCC certification test failures, the amount of energy typically radiated is so small a fraction of the signal as to be difficult to detect in S21.

The losses in the interconnect that turn into heat are accounted for by the conductor loss and dielectric loss. As shown in a previous chapter, both of these effects increase monotonically with frequency. When all other mechanisms that affect S21 are eliminated, when measured in dB, the insertion loss is a direct measure of the attenuation in the line and will increase as a more negative dB value

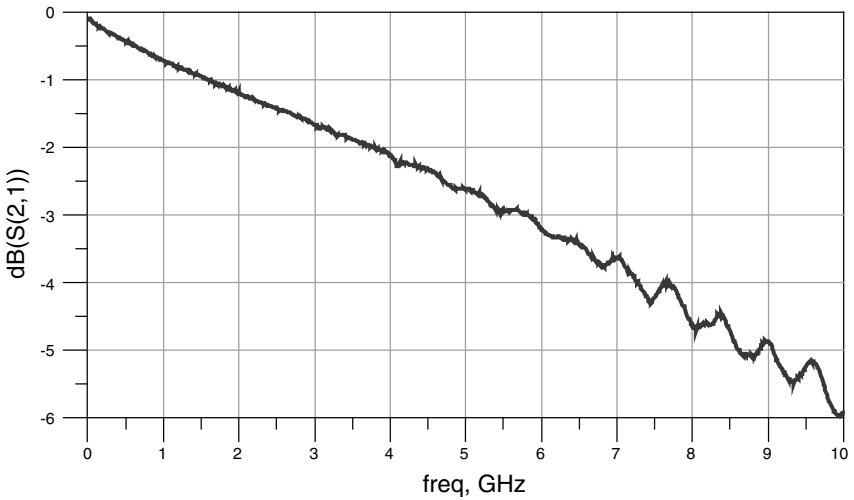


Figure 12-17 Measured insertion loss of 5-inch-long stripline. Measured with an Agilent N5230 VNA and displayed with Agilent ADS.

with increasing frequency. Figure 12-17 shows the measured insertion loss of a stripline transmission line about 5 inches long.

There is often some ambiguity of the sign of the attenuation. Usually, when describing the attenuation, the sign chosen is positive because we get lazy and look for the simplest description. The fact that it is explicitly stated as an attenuation means the transmitted amplitude is getting smaller. This is why attenuation is often plotted as an increasing positive number with frequency.

However, when the attenuation is described by the insertion loss and measured in dB, it is critically important to be absolutely consistent and always make the insertion loss the correct, negative sign.

TIP Insertion loss is a direct measure of the attenuation from both the conductor and dielectric losses. Given the conductor and dielectric properties of the transmission line, as described in a previous chapter, the insertion loss can be easily calculated.

The insertion loss, due to just the losses, is:

$$S21 = -(A_{diel} + A_{cond}) \tag{12-16}$$

where:

S21 is the insertion loss in dB

A_{diel} = the attenuation from the dielectric loss, in dB

A_{cond} = the attenuation from the conductor loss, in dB

When the there are no impedance discontinuities in the transmission lines, and conductor loss is small compared to the dielectric loss, the insertion loss is a direct measure of the dissipation factor:

$$S21 = -2.3 \times f \times Df \times \sqrt{Dk} \times Len \tag{12-17}$$

where:

S21 = the insertion loss in dB

f = the frequency in GHz

Df = the dissipation factor

Dk = the dielectric constant

Len = the interconnect length in inches

A measurement of the insertion loss, when scaled appropriately, is a direct measure of the dissipation factor of the laminate material:

$$Df = \frac{-S21 \text{ [in dB]}}{2.3 \times \sqrt{Dk} \times Len \times f} \tag{12-18}$$

where:

S21 = the insertion loss in dB

f = the frequency in GHz

Df = the dissipation factor

Dk = the dielectric constant

Len = the interconnect length in inches

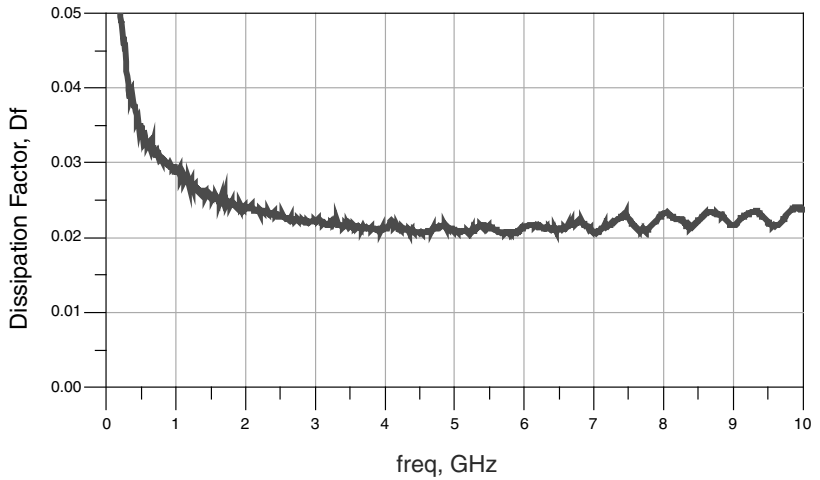


Figure 12-18 Measured FR4 stripline insertion loss re-plotted as dissipation factor, showing the value of about 0.022. The higher values at low frequency are due to conductor losses not taken into account. Measured with an Agilent N5230 VNA and displayed with Agilent ADS.

When the dielectric constant is close to 4, typical for most FR4 type materials, the insertion loss is approximated as –5 × Df dB/inch/GHz. Figure 12-18 shows the same measured insertion loss for the stripline case above, but scaled by 5 and by the length and frequency to plot the dissipation factor directly.

At low frequency, conductor loss can be a significant contributor to loss and is not taken into account in this simple analysis of insertion loss. When the approximation above is used, the extracted dissipation must be artificially higher to compensate for the conductor loss. At higher frequency, the impedance discontinuities, typically from connectors or vias, cause ripples in the insertion loss.

While the slope of the insertion loss per length is a rough indication of the dissipation factor of an interconnect, it is only a rough indication. In practice, the discontinuities from impedance discontinuities of both the impedance of the line and from the connectors, as well as contributions to the attenuation from the conductor losses will complicate the interpretation of the insertion loss. In general, for accurate measurements of the dissipation factor, these features must be taken into account.

However, the behavior of the insertion loss is a good indictor of dissipation factor of the materials. Figure 12-19 is an example of the measured insertion loss for two nearly 50-Ohm transmission lines, each about 36 inches long. One is made from a Teflon substrate, while the other is made from FR5, a very lossy dielectric.

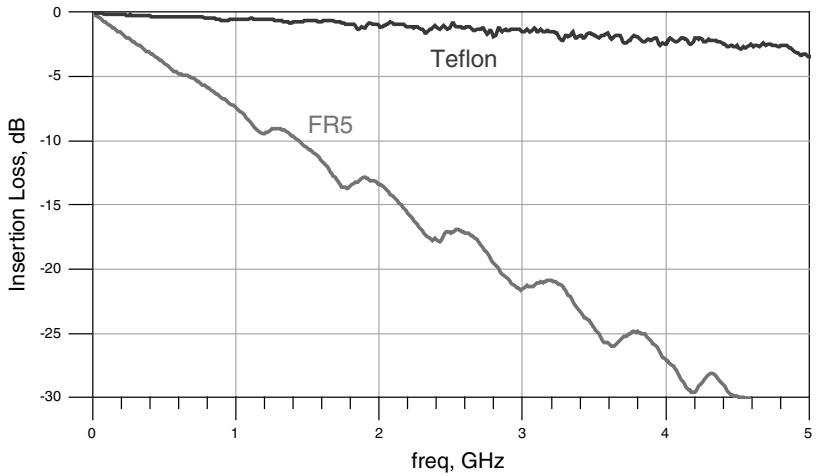


Figure 12-19 Measured FR5 and Teflon transmission lines showing the impact on the insertion losses from the dissipation factors of the laminates. Each line is a 50-Ohm microstrip and 36 inches long. Measured with an Agilent N5230 VNA and displayed with Agilent ADS.

The much larger slope of the FR5 substrate is an indication of this much higher dissipation factor.

When plotted as a function of frequency, the insertion loss shows a monotonic drop in magnitude. At the same time, the phase is advancing at a constant rate in the negative direction. When S21 is plotted in a polar plot, the behavior is a spiral, as shown in Figure 12-20.

The lowest frequency has the largest magnitude of S21 and a phase close to zero. As frequency increases, S21 rotates around in the clockwise direction, with smaller and smaller amplitude. It spirals into the center.

The spiral pattern of the insertion loss of a uniform transmission line, or the return loss when looking at the 1 port measurement of a transmission line open at the end, is a pattern indicative of losses in the line that increase with frequency, typical of conductor and dielectric losses.

12.10 Coupling to Other Transmission Lines

Even if the cross talk noise into adjacent lines is not measured, it will still happen and its impact will be the magnitude reduction of S11 and S21. The simplest case is a uniform microstrip line with an adjacent microstrip closely coupled to it.

The insertion loss of an isolated microstrip will show the steady drop off due to the dielectric and conductor loss. As an adjacent microstrip is brought closer,

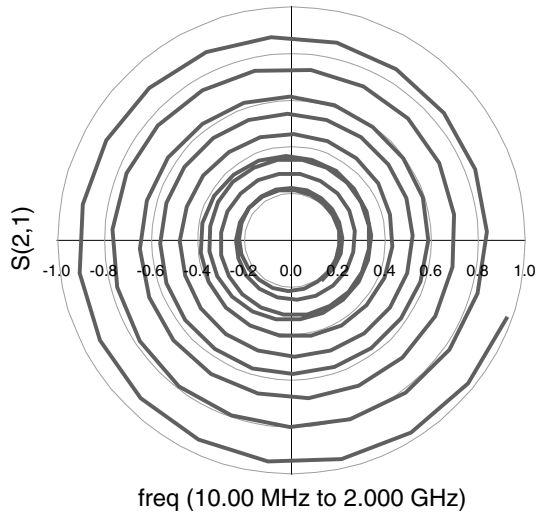


Figure 12-20 Measured insertion loss of the 36-inch-long FR5 microstrip from above, but plotted in a polar plot. The frequency range is 10 MHz to 2 GHz, every 10 MHz. Measured with an Agilent N5230 VNA and displayed with Agilent ADS.

some of the signal in the driven microstrip will couple into the adjacent one, giving rise to near- and far-end cross talk. In microstrip, the far-end noise can be much higher than the near-end noise.

If the ends of the adjacent microstrip are also terminated at 50 Ohms, then any cross talk will effectively be absorbed by the terminations and not reflect in the victim microstrip line.

As frequency increases, the far-end cross talk will increase and more signal will be coupled out of the driven line, reducing S21. In addition to the drop in S21 from attenuation, there will also be a drop in S21 from cross talk. As the two traces are brought closer together and coupling increases, more energy flows from the active line to the quiet line and less signal makes it out as S21. Figure 12-21 shows the S21 will decrease with frequency and with coupling as the spacing between the lines decreases.

TIP Just looking at the S21 response, it is difficult to separate out how much of S21 is due to attenuation and how much is due to coupling to other interconnects, unless the coupled signal into the other interconnects is also measured.

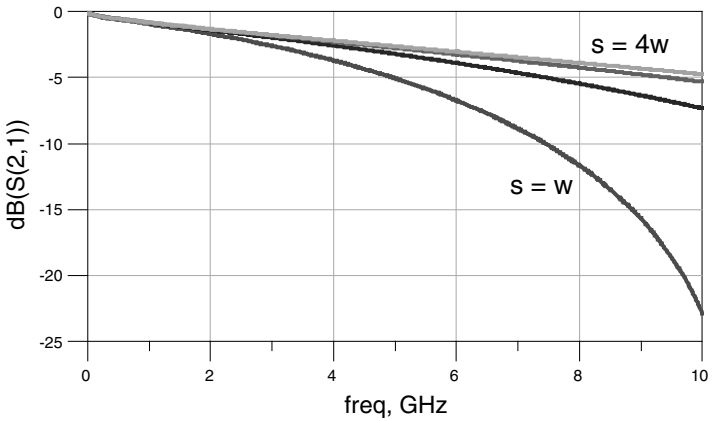


Figure 12-21 The insertion loss of one line in a microstrip pair. As the coupling increases, energy is coupled out of the line and S21 decreases more than just due to attenuation. Simulated with Agilent's ADS.

Using the standard port labeling convention, two adjacent microstrip lines can be described by four ports. The insertion loss of one line is S21, while the insertion loss of the other line is S43. The near-end noise is described by S31 and the far-end noise by S41. As the coupling increases and S21 drops, we can see the corresponding increase in the far-end noise, S41, as shown in Figure 12-22.

In this example, the reduction in magnitude of S21 from coupling increased with frequency, but was relatively slowly varying. When the coupling is to an

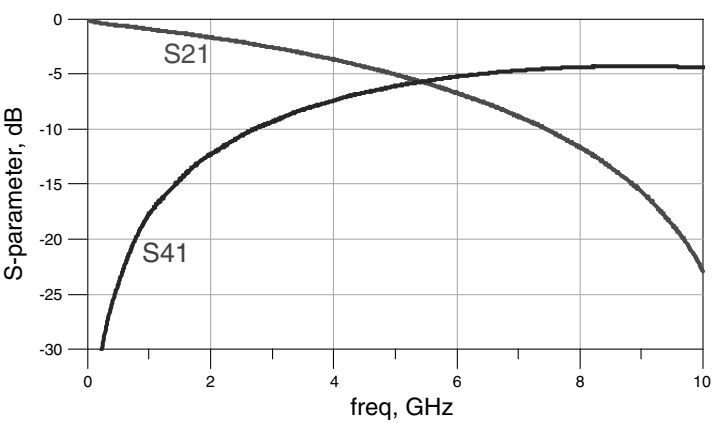


Figure 12-22 Insertion loss, S21, of one line in a microstrip pair and the far-end cross talk, S41, showing the insertion loss decreasing as the far-end noise increases. Simulated with Agilent ADS.

interconnect that is not terminated but is floating, the Q of the isolated floating line can be very high. If it were excited, the noise injected would rattle around between the open ends, attenuating slightly with each bounce. It could rattle around for as many as 100 bounces before eventually dying out from its own losses.

The impact from coupling to high Q resonators is very narrow band absorption in S21 or S11. Figure 12-23 shows the S21 of the same microstrip as above, but now the adjacent victim line is floating, open at each end. It has very narrow frequency resonances, which suck out energy in the active line in a very narrow frequency range.

Coupling to high Q resonators is also apparent when measuring the return loss in one-port configurations. In principle, when one port is connected to a transmission line with its other end open, S11 should have a magnitude of 1, or 0 dB. In practice, as we have seen, all interconnects have some loss, so S11 is always smaller than 0 dB and continues to get smaller as frequency increases.

If there is coupling between the transmission line being measured and adjacent transmission lines that are unterminated, they will act as high Q resonators and show very narrow absorption lines.

TIP Narrow, sharp dips in the return or insertion loss are almost always indications of coupling to high Q resonant structures. When the resonating structure has a complex geometry, the multiple resonant frequency modes may be difficult to calculate without a full wave field solver.

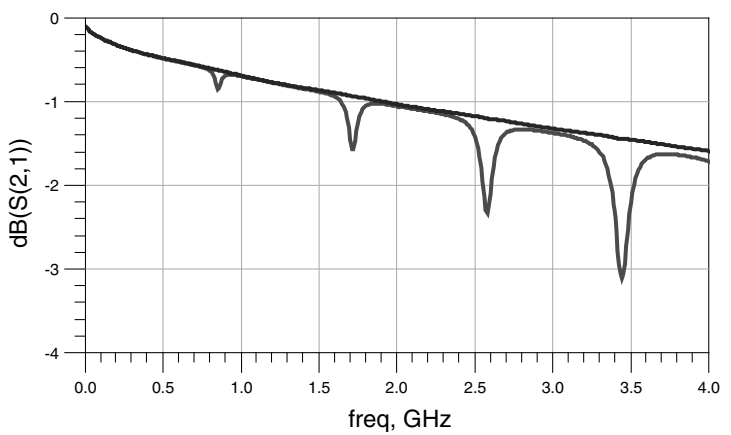


Figure 12-23 The insertion loss of a microstrip line when a floating adjacent microstrip line is very far away and closely spaced. When close, the high Q resonances of the floating line couples energy in narrow frequency bands. Simulated with Agilent ADS.

In the case of a single adjacent transmission line, the frequency of the dip is the resonant frequency of the quiet line. The width of the resonance is related to the Q of the resonator. By definition, the Q is given by:

$$Q = \frac{f_{res}}{FWHM} \tag{12-19}$$

where:

Q = the value of the quality of the resonance

f_{res} = the resonant frequency

FWHM = the full width, half minimum: the frequency width across the dip at half the minimum value

The higher the Q, the narrower the frequency dip. The depth of the dip is related to how tight the coupling is. A larger coupling and the dip will be deeper. Of course, as the dip gets deeper, the coupling to the floating line increases and its damping generally increases, causing the Q to decrease.

The resonator to which a signal line couples does not have to be another uniform transmission line, but can be the cavity made up of two or more adjacent planes. For example, when a signal transitions from one layer to another and its return plane also changes, the return current, transitioning between the return planes, can couple into the cavity, creating high Q resonant couplings. An example of this layer transition in a four-layer board is shown in Figure 12-24.

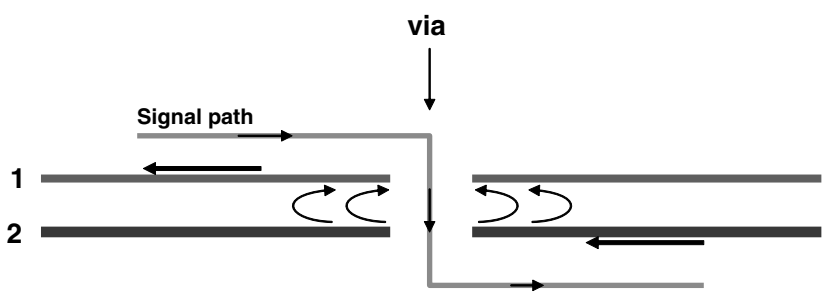


Figure 12-24 A signal transitions two reference planes, layer 1 and 2. The return current, flowing between the two planes couples into the plane cavity resonance.

The resonant frequencies of a cavity composed of two planes is related to the length of a side. It is given by:

$$f_{res} = n\frac{6}{\sqrt{Dk}}\frac{1}{Len} = n\frac{1.5\ GHz}{Len} \quad (12\text{-}20)$$

where:

f_{res} = the resonant frequency in GHz

Dk = the dielectric constant of the laminate inside the cavity

Len = the length of a side of the cavity in inches

n = the index number of the mode

For example, when the length of a side is about 1 inch, the resonant frequency with an FR4 laminate begins at about 1.5 GHz and increases from there for higher modes. In real cavities, the resonant frequency spectrum can be more complex due to cutouts in the planes.

In typical board level applications, the length of a side can easily be 10 inches, with a resonate frequency starting at about 150 MHz. In this range, the decoupling capacitors between power and ground planes can often help to suppress resonances, and they may not be clearly visible.

However, in multilayer packages, typically on the order of 1 inch on a side, the first resonances are in the GHz range, and the decoupling capacitors are not effective at suppressing these modes, since their impedance at 1 GHz is so high compared to the plane's impedance. When a signal line transitions from the top layer to a bottom layer, going through the plane cavity, the return current will excite resonances in the GHz range.

This can easily be measured using a one-port network analyzer. In Figure 12-25, the measured return loss of six different leads in a four-layer BGA is shown. In each case, the signal line is measured at the ball end, and the cavity end, where the die would be, is left open. Normally, the return loss should be 0 db, but at the resonant frequencies of the cavity, a considerable amount of energy is absorbed.

In this example, the return loss drops off very slowly from 0 dB, as frequency increases, due to the dielectric loss in the laminate. Above about 1 GHz, there appear very large, narrow bandwidth dips. These are the absorptions from coupling into the power and ground cavity of the package.

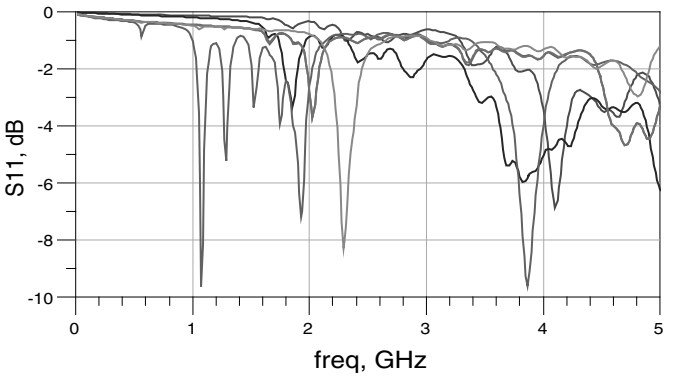


Figure 12-25 The measured return loss of six different leads in a BGA package, each open at the far end. The dips starting at 1 GHz are resonances in the cavity formed by the power and ground planes, as well as coupling into adjacent signal lines. Measured with an Agilent N5230 VNA and displayed with Agilent ADS.

If the return loss from a reflection at the open source is -10 dB, which is from a round-trip path, the drop in the one-way insertion loss might be on the order of -5 dB. When the magnitude is -5 dB, this means that only 50% of the signal amplitude would get through. This is a huge reduction in signal strength and would result in distortions of the signal as well as excessive cross talk between channels.

The return loss is really a measure of the absorption spectrum of the package, similar to an infrared absorption spectrum of an organic molecule. Infrared spectroscopy identifies the distinctive resonant modes of specific atomic bonds. In the same way, these high Q package resonances identify specific resonant modes of the package or component.

Not only can the plane cavities absorb energy but also other adjacent traces. The resonant absorptions of a package will often set the limit to its highest usable frequency. One of the goals in package design is to push these resonances to higher frequency or reduce the coupling of critical signal lines to the resonant modes. This can be accomplished in a number of ways:

- Don’t transition the signal between different reference planes.
- Use return vias adjacent to each signal via to suppress the resonance.
- Use low inductance decoupling capacitors to suppress the resonances.
- Keep the body size of the package very small.

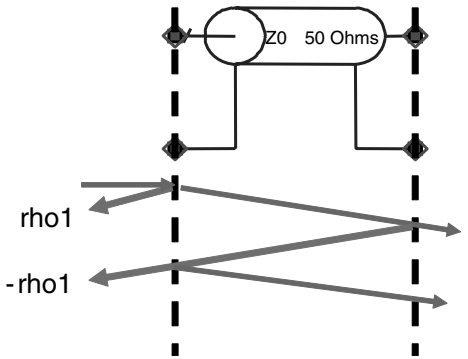


Figure 12-26 Multiple reflections from the interfaces through a non-50-Ohm, uniform transmission line.

12.11 Insertion Loss for Non-50-Ohm Transmission Lines

When the losses are small and coupling out of the line to adjacent lines is small, the dominant mechanism affecting the insertion loss is reflections from impedance discontinuities. The most common source of discontinuity is when the transmission line is different from the 50 Ohms of the ports. The impedance mismatch at the front of the line and the end of the line will give rise to a type of resonance, causing a distinctive pattern to the return and insertion loss.

Consider the case, as illustrated in Figure 12-26, of a short length of lossless transmission line. It has a characteristic impedance Z0, different from 50 Ohms, and a time delay, TD. The port impedances are 50 Ohms. As the sine wave hits the transmission line, there will be a reflection at the front interface, sending some of the incident sine wave back into port 1, contributing to the return loss. However, most of the incident sine wave will continue through the transmission line to port 2, where it will reflect from that interface.

The reflection coefficient for a signal traveling from port 1 into the transmission line, rho_1, is:

$$\text{rho}_1 = \frac{(Z_0 - 50\Omega)}{(Z_0 + 50\Omega)} \tag{12-21}$$

Once the signal is in the transmission line, when it hits port 2 or port 1, the reflection coefficient back into the line, rho, will be:

$$\text{rho} = \frac{(50\Omega - Z_0)}{(Z_0 + 50\Omega)} = -\text{rho}_1 \tag{12-22}$$

where:

rho_1 = the reflection coefficient from port 1 to the transmission line

rho = the reflection coefficient from the line into port 1 or port 2

Z_0 = the characteristic impedance of the transmission line

Whatever the phase shift of the reflection from the first interface, the phase shift from the reflection off the second interface will be the opposite. The phase shift propagating a round-trip distance down the interconnect and back is:

$$\text{Phase} = 2 \times \text{TD} \times \text{f} \times 360 \text{ degrees} \tag{12-23}$$

where:

Phase = the phase shift of the reflected wave, in degrees

TD = the time delay for one pass through the transmission line

f = the sine wave frequency

If the round-trip phase shift is very small, the reflected wave from the end of the line when it enters port 1 will have nearly equal magnitude but opposite phase as the reflected signal off port 2. Heading back into port 1, they will cancel and the net reflected signal into port 1 will be 0.

TIP At low frequency, the return loss of all lossless interconnects will always start at very little reflecting, or a very large, negative dB value.

If nothing reflects back into port 1, then everything must be transmitted into port 2 and the insertion loss of all lossless interconnects will start at 0 dB. This is due to the second reflected wave into port 2 having the same phase as the first wave into port 2, and they both add.

TIP At low frequency, the insertion loss of all lossless interconnects will always start at 0 dB.

As the frequency increases, the round-trip phase shift of the transmitted signal will increase until it is exactly half a cycle. At this point, the reflected signal from the front interface heading into port 1 will be exactly in phase with the reflected signal from the back interface, heading into port 1. They will add in phase, and the return loss will be a maximum. The return loss will be:

$$S11 \sim 2 \times rho_1 \tag{12-24}$$

If S11 is a maximum, the insertion loss will be a minimum. When the round-trip phase shift is 180 degrees, the second reflected wave heading back to port 2 will be 180 degrees out of phase with the first wave heading into port 2, and they will partially cancel out.

As frequency increases, the round-trip phase will cycle between 0 and 180, causing the return and insertion loss to cycle between minimum and maximum values.

Dips in return loss and peaks in insertion loss will occur when the round trip phase is a multiple of 360 degrees, or Phase = n × 360. This occurs when:

$$n \times 360 = 2 \times TD \times f \times 360 \text{ degrees} \tag{12-25}$$

or

$$f = \frac{n}{2} \times \frac{1}{TD} \tag{12-26}$$

The longer the time delay, TD, the shorter the spacing between frequency intervals for a 180 degree phase shift. An example of the return and insertion loss for two 30-Ohm transmission lines, with a TD of 0.5 and 0.1 nsec, is shown in Figure 12-27. The frequency interval between dips should be 1 GHz and 5 GHz, respectively.

A glance at the frequency interval between the ripples in the return or insertion loss can be a good indication of the physical length between the discontinuities of a transmission line interconnect. The TD of the interconnect is roughly:

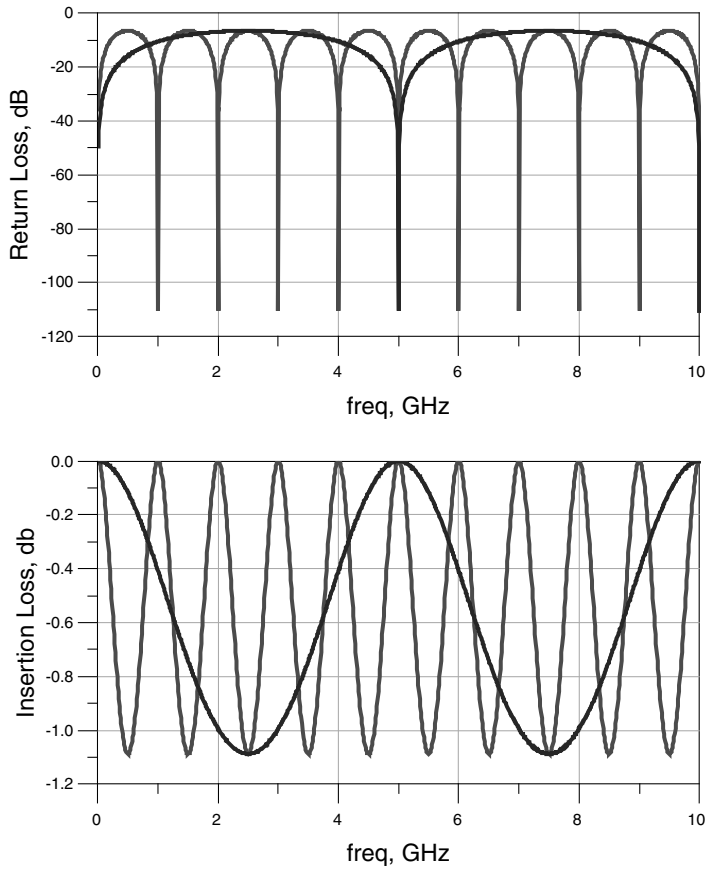


Figure 12-27 The return and insertion loss for two 30-Ohm, lossless, uniform transmission lines with a TD of 0.5 nsec and 0.1 nsec. The longer the time delay, the shorter the interval between dips. Simulated with Agilent's ADS.

$$TD = \frac{1}{2 \times \Delta f} \tag{12-27}$$

where:

TD = the interconnect time delay

Δf = the frequency interval between dips in the return loss or peaks in insertion loss

For example, in the above return loss plot, the frequency spacing between dips in one transmission line is 1 GHz. This corresponds to a time delay of the interconnect of about $1/(2 \times 1) = 0.5$ nsec.

TIP The best way to make a transparent interconnect is first to match the interconnect impedance to 50 Ohms. If you can't make the impedance 50 Ohms, then the next most important design guide is to keep it short.

Interposers between a semiconductor package and a circuit board are designed to be transparent. When their impedance is far off from 50 Ohms, the design guide to keep them transparent is to keep them short. This condition is that:

$$2 \times TD \times f \times 360 \text{ degrees} << 360 \text{ degrees} \tag{12-28}$$

If the wiring delay of an interconnect is roughly 170 psec/inch, then the maximum length of a transparent interposer is given by:

$$2 \times Len \times 170 \text{ psec/inch} \times f_{max} << 1 \tag{12-29}$$

or

$$Len << \frac{3}{f_{max}} \text{ and } f_{max} << \frac{3}{Len} \tag{12-30}$$

If we translate the << condition to be 10x, then this rough rule of thumb becomes:

$$Len < \frac{0.3}{f_{max}} \text{ and } f_{max} < \frac{0.3}{Len} \tag{12-31}$$

where:

Len = the interposer length in inches

f_{max} = the maximum usable frequency where the interconnect is still transparent, in GHz

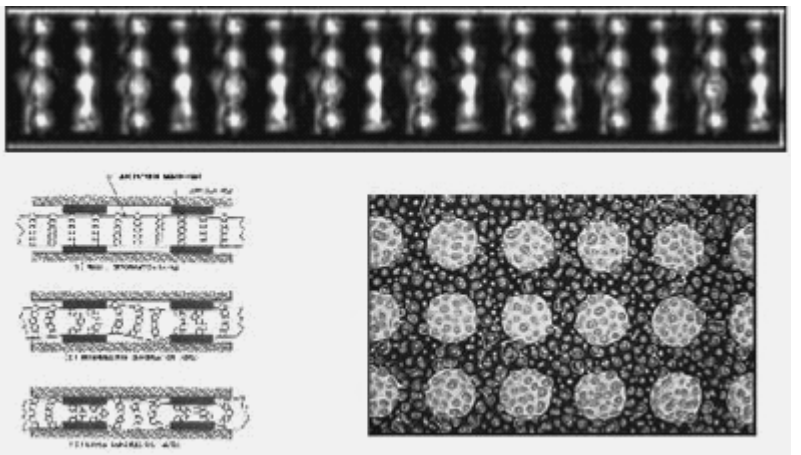


Figure 12-28 Cross section of the Pariposer interposer from Paricon, which is about 10 mil thick having a bandwidth in excess of 30 GHz.

For example, if the operating frequency is 1 GHz, an interposer or connector should be shorter than about 0.3 inches to still be transparent. Likewise, if an interposer is 10 mils long, it could have a usable bandwidth of 30 GHz, without any other special design conditions.

An example of a compliant interposer from Paricon, about 10 mils thick with a 30 GHz bandwidth is shown in Figure 12-28.

Of course, if the interconnect is designed as a controlled impedance path with a characteristic impedance near 50 Ohms, it can have a much higher bandwidth and be much longer.

12.12 Data-Mining S-Parameters

The interpretation of the S-parameter elements depends on the port assignments. For example, a semiconductor package could have 12 different traces, one end of each trace connected to a port, and every trace open at its far end. Or, there could be one net on a board with a fan out of 11 and a port at each end of the single net. Or, the port assignment could be as shown in Figure 12-29 with six different straight through interconnects in some proximity. These six different transmission lines could be grouped in pairs creating three different, differential channels.

The precise internal connections to the device will influence how to interpret each S-parameter. The most common case is with the six different through connections with port assignments as shown. Always remember that if the port assignments change, the interpretation of each specific S-parameter will change as well.

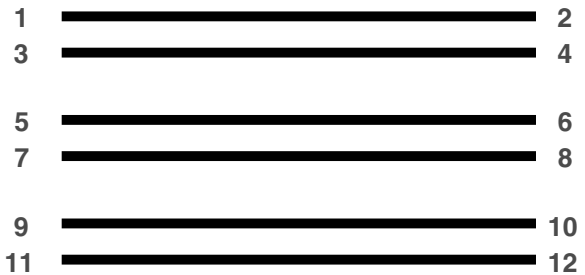


Figure 12-29 The index labeling for a 12-port device, which could be three differential channels. Not shown are the return paths, but they are assumed to be present and connected to the ports.

With 12 ports, there will be a total of 78 unique S-parameter elements, each having a magnitude and phase and each varying with frequency.

The diagonal elements are the return losses of each transmission line and have information about impedance changes in the interconnect. If the lines are all similar and symmetrical, all of the diagonal elements could be the same.

The six unique direct through signals, the insertion losses, S21, S43, S65, S87, S10,9, and S12,11, have information about the losses, the impedance discontinuities, and even resonances from stubs.

All the other S-parameter elements represent coupling terms, containing information about cross talk. For example, S51, the ratio of a sine wave coming out of port 5 to the sine wave going into port 1, is related to the near-end cross talk between lines two away. The S61 term has information about the far-end cross talk between lines far away.

Of course, the near-end noise between the first line and each of the adjacent signal lines will drop off with spacing. Figure 12-30 is an example of the simulated near-end noise from one line to five other adjacent lines, each 5 mils in width and spaced 7 mils to the adjacent trace, each being 10 inches long.

Though the near-end noise between adjacent lines, S31 may be –25 dB, the near-end noise to the line five lines away, S11,1, is less than –55 dB.

TIP Everything you ever wanted to know about the electrical behavior of a collection of interconnects is contained in their S-parameters.

So far, we have considered the signals into the ports as single-ended sine wave signals. Two other types of signals provide an alternative description of the

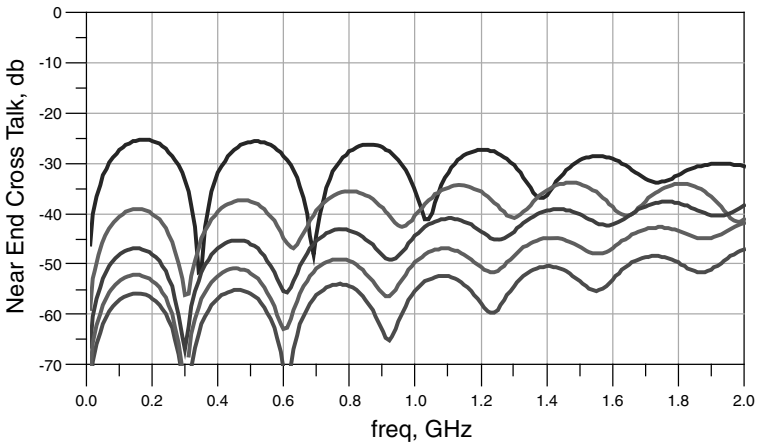


Figure 12-30 Near-end cross talk from one line to five adjacent microstrip lines, each 10 inches long, with a spacing between them of 7 mils. The farther away the line, the lower the near-end cross talk. Simulated with Agilent's ADS.

stimulus-response behavior, which can often provide valuable insight into the behavior of the interconnects. Depending on the specific question being asked, these other forms may offer a faster route to the correct answer.

The two other forms of the S-parameters are differential and time domain.

12.13 Single-Ended and Differential S-Parameters

Two independent transmission lines in proximity with coupling can be described in two equivalent ways. On the one hand, they are two independent transmission lines each with independent properties. For example, if we use the labeling scheme suggested above, where the connections are port 1 goes through to port 2 and port 3 goes through to port 4, each transmission line would have reflected elements, S11 and S33, and each would have a transmitted element, S21 and S43.

In addition, there would be cross talk between the two lines, the unique near end would be S31 and S42, and the unique far-end terms would be S41 and S32. The near- and far-end noise signatures would vary depending on the spacing, coupling lengths, and whether the topology were stripline or microstrip. All the electrical properties of these two transmission lines are completely described by these 10 unique S-parameter elements, across the frequency range.

These same two lines can also be described as one differential pair. No assumptions have to be made about the lines—this description as a differential

pair is a complete description. But, the words we use and the sorts of behavior we describe are very different for a single differential pair description compared to as two independent single-ended transmission lines with coupling.

TIP When the S-parameters are describing the differential properties of the interconnects, we refer to the S-parameters as the differential S-parameters, or the mixed-mode S-parameters or the balanced S-parameters. These terms are used interchangeably in the industry.

As differential S-parameters, we describe the four-port interconnect in terms of being a differential pair, in which case we refer to the ports as differential ports. With one differential pair and a differential port on either end, the only types of signals that exist are differential signals and common signals. Any arbitrary waveform going into either of the differential ports can be described by a combination of differential and common signals.

These signals are often called the differential mode and common mode signals. This is a bad habit to get into. There is no need to invoke the word *mode*. The signals entering the interconnect as stimuli or leaving the interconnect as responses are either differential or common signals. If you call them differential mode signals, it is too easy to confuse the signals with the even and odd propagation modes, which refer to the state of the interconnect. This is discussed in great detail in an earlier chapter.

TIP To avoid confusion, it is strongly recommended you never use the terms differential mode signals or common mode signals. There is no need. Always refer to signals as differential or common signals, and this will minimize confusion about mixed-mode S-parameters.

The differential S-parameters describe how these differential and common signals interact with the interconnect. With just two differential ports, one on either end of the differential pair, as shown in Figure 12-31, there are only three unique S-parameters with different index numbers, S11, S22, and S21, which describe how signals enter and come out of the differential pair.

These are defined in the conventional way. Each element describes the ratio of the sine wave coming out of one port to the sine wave going into another port. Each differential S-parameter has a magnitude and a phase.

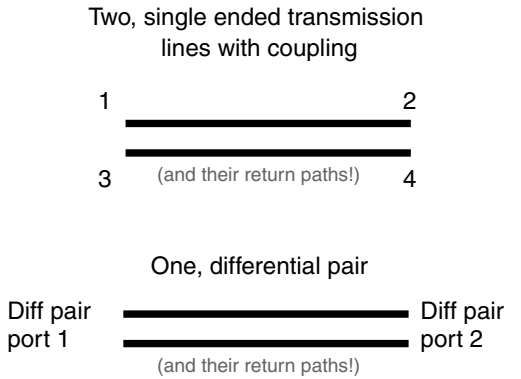


Figure 12-31 The same two transmission lines can be equivalently described as two single-ended lines with coupling, or one differential pair. As one differential pair, there is one differential port on each end.

However, we must keep track of what type of signals enter and come out of the differential pair. In a differential pair, there are only differential and common signals. These interact with the differential pair in four possible ways:

- A differential signal can enter a port and come out as a differential signal.
- A common signal can enter a port and come out as a common signal.
- A differential signal can enter a port and come out as a common signal.
- A common signal can enter a port and come out as a differential signal.

With single-ended S-parameters, each S-parameter describes the ratio of single-ended sine waves coming out of a port, compared with the single-ended amplitude and phase going into a port. With differential S-parameters, we have to include a labeling system to describe not only what port the sine waves enter and come out, but also what type of signal it is.

We use the letters D and C to refer to a differential or common signal. To designate a differential signal going in and a differential signal coming out, we use SDD. For a common signal in and common signal out, we use SCC. We also use the backward convention for the order of the type of signal, with the coming out port signal first, and the going in port signal second. For a differential signal in and a common signal out, we use SCD. And for a common signal in and a differential signal out, we use SDC.

Each differential S-parameter must contain information about the in and out ports and the in and out signals. By convention, we use the signal letters first, then

the port indices. For example, SCD21 would describe the ratio of the differential signal going in at port 1 to the common signal coming out at port 2.

The port impedances of the single-ended S-parameters is 50 Ohms. When two ports drive a differential signal, the outputs are in series and the differential port impedance for the differential signal is 100 Ohms.

When two ports drive a common signal, the single-ended ports are in parallel and the port impedance for the common signal is the parallel combination, or 25 Ohms. This means that a differential signal looking into one of the differential ports will always see a 100-Ohm termination impedance, while a common signal will always see the 25-Ohm common impedance port termination.

When displaying the mixed-mode S-parameters as a matrix, again by convention, we usually place the pure differential behavior in the upper-left quadrant and the pure common signal in the lower-right quadrant. The lower left is the conversion of differential into common signal, and the upper right is the conversion of common signal into differential signal. The mixed-mode matrix using this notation is shown in Figure 12-32.

This same matrix orientation can be used to display the 16 measured mixed-mode S-parameters. Using this formalism makes it much easier and less prone to

			Stimulus			
			Differential Signal		Common Signal	
			port 1	port 2	port 1	port 2
Response	Differential Signal	port 1	SDD11	SDD12	SDC11	SDC12
		port 2	SDD21	SDD22	SDC21	SDC22
	Common Signal	port 1	SCD11	SCD12	SCC11	SCC12
		port 2	SCD21	SCD22	SCC21	SCC22

Figure 12-32 Mixed-mode S-parameter labeling scheme for each matrix element. For mixed-mode S-parameters, the ports are always differential ports.

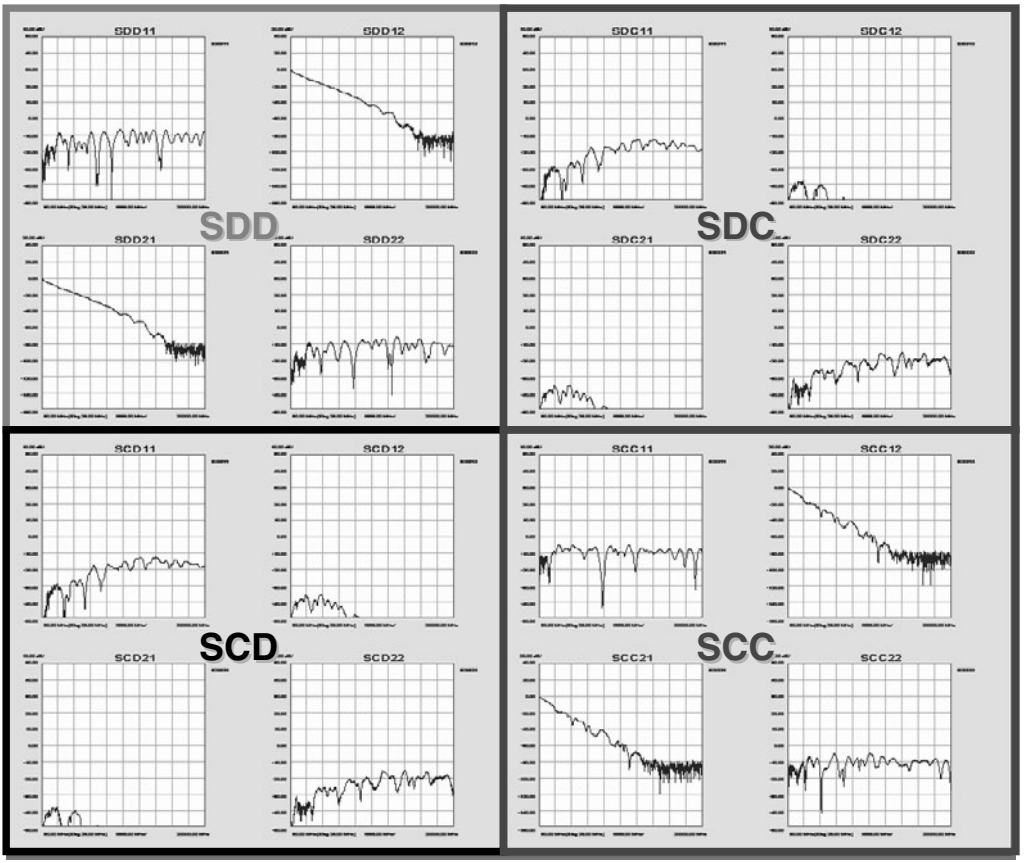


Figure 12-33 Measured mixed-mode S-parameters of a differential channel in a backplane, arranged in the same order as the matrix elements. Only the magnitude of each element is displayed, there is also a corresponding phase plot for each element, not shown. Measured with an Agilent N5230 VNA and displayed with Agilent's PLTS.

introducing errors. An example of the measured mixed-mode or differential S-parameters for a differential channel in a backplane is shown in Figure 12-33.

The CC terms contain information about how common signals are treated by the interconnect. The reflected common signal, SCC11 has information about the common impedance profile of the interconnect. The transmitted common signal, SCC21 describes how a common signal is transmitted through the interconnect. While the CC terms are important in the overall complete characterization of the differential pair, in most applications, the common signal properties of the interconnect are not very important.

12.14 Differential Insertion Loss

The DD terms contain information about how differential signals are treated by the interconnect. The reflected differential signal, SDD11, has information about the differential impedance profile of the interconnect. The transmitted differential signal, SDD21, describes how well differential signals are transmitted by the interconnect.

> **TIP** Since most applications of differential pairs are for high-speed serial links, the differential insertion loss is by far the most important differential S-parameter element. The phase has important information about the time delay and dispersion of the differential signal, and the magnitude has information about the attenuation from losses and other factors.

An example of the measured differential insertion loss, SDD21, of a typical backplane channel is shown in Figure 12-34.

It is dominated first by the conductor and dielectric losses. This gives rise to a general monotonic decrease in the differential insertion loss. In FR4, this is roughly –0.1 dB per inch per GHz for dielectric loss. In a 40-inch backplane channel, at 5 GHz, the differential insertion loss expected is about $-0.1 \times 40 \times 5 = -20$ dB, which is seen to be the case in the example above.

The second factor affecting SDD21 is impedance mismatches from connectors, layer transitions, and vias. These give rise to some ripples in the SDD21 plot.

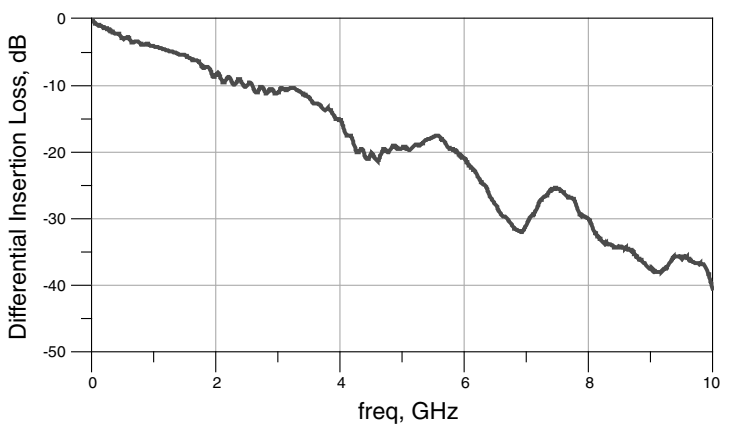


Figure 12-34 Measured SDD21 of a backplane trace. The general drop is due to losses. The ripples are due to impedance discontinuities from connectors and vias.

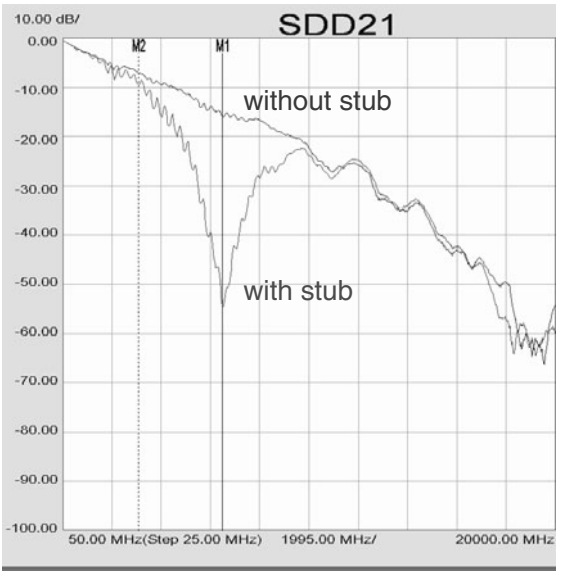


Figure 12-35 Measured differential return loss of two backplane channels, one with a 0.25-inch-long via stub and one with the via stub backdrilled. Measured with an Agilent N5230 VNA and displayed with Agilent's PLTS. Data courtesy of Molex Corporation.

Two other factors can sometimes dominate SDD21: resonances from via stubs and mode conversion.

Figure 12-35 is an example of the measured differential insertion loss, SDD21, for two different differential channels in a backplane. In one channel there is a long via stub, approximately 0.25 inches long. In the other channel this via stub has been removed by back drilling.

The sharp resonant dip at about 6 GHz is due to the *quarter wave stub resonance* and is a prime factor limiting the usable bandwidth of high-speed serial links.

The origin of the resonance is very easy to understand by referring to the illustration in Figure 12-36. In this illustration, the return path is not included but is present as planes on adjacent layers.

When the signal flows down the via to the connected signal layer, it splits. Part of the signal continues on the signal layer, and another part of the signal heads down to the end of the via stub, which is open.

When the signal hits the end of the via, it reflects and again splits at the layer transition. Some of the signal goes back to the source while the rest of it continues in the same direction as the initial signal but with a phase shift.

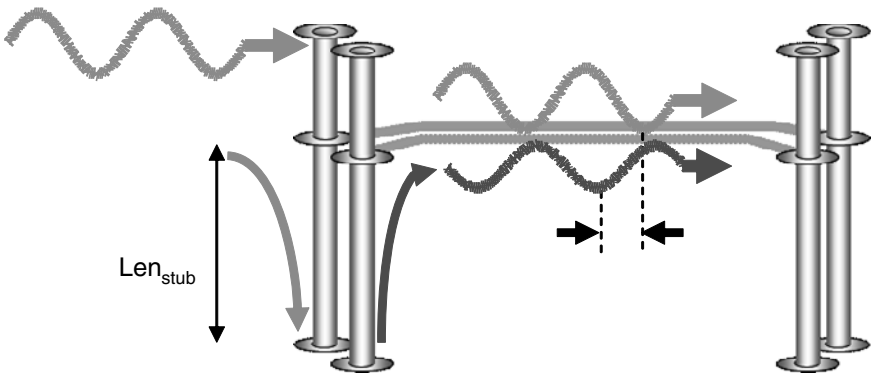


Figure 12-36 The insertion loss is due to the combination of the incident wave and reflected wave from the bottom of the via stub. When the stub length is 1/4 a wavelength, the two waves are 180 degrees out of phase and cancel out.

The phase shift of the reflected wave is the round-trip path length going down to the stub and back up to the signal layer. When this round-trip path length is half a wave, the two components—the initial signal and the reflected signal—heading to port 2 will be 180 degrees out of phase and offer the most cancellation. The condition for maximum cancellation, the dip in the differential insertion loss, is that the round-trip time delay for the stub is half a cycle or:

$$2\frac{\text{Len}_{\text{stub}}}{\text{v}} = \frac{1}{2}\frac{1}{\text{f}_{\text{res}}} \tag{12-32}$$

Or, for the case of Dk = 4:

$$\text{f}_{\text{res}} = \frac{1}{4}\frac{12 \text{ inches/nsec}}{\sqrt{\text{Dk}} \times \text{Len}_{\text{stub}}} = \frac{1.5}{\text{Len}_{\text{stub}}} \text{ GHz} \tag{12-33}$$

where:

Len_{stub} = the length of the stub in inches

v = the speed of the signal in inches/nsec

f_{res} = the resonant frequency for the stub in GHz

Dk = the dielectric constant of the laminate around the stub = 4

For example, if the stub is 0.25 inches long, the resonant frequency is 1.5 GHz/0.25 inches = 6 GHz, very close to what is observed.

The condition for maximum cancellation and the minimum differential insertion loss is that the round-trip length of a via stub be 1/2 of a wavelength, or that the one-way length be 1/4 of a wavelength. This is why this resonance is often called a *quarter wave stub resonance*.

As a rough guideline, for the large resonant absorption in the stub to not affect the high-speed signal the resonant frequency should be engineered to be a frequency that is at least 10x the bandwidth of the signal. Using the Nyquist frequency as the bandwidth of the signal, the bandwidth is:

$$BW = 0.5 \times BR \tag{12-34}$$

where:

BW = bandwidth of the signal in GHz

BR = the bit rate in Gbps

The condition for an acceptable via stub length is:

$$f_{res} > 10 \times BW \tag{12-35}$$

or

$$\frac{1.5}{Len_{stub}} > 5 \times BR \tag{12-36}$$

or

$$Len_{stub} > \frac{300}{BR} \tag{12-37}$$

where:

Len_{stub} = the maximum acceptable stub length in mils

BR is the bit rate in Gbps

TIP For example, to engineer the stub resonant frequency to be far above the Nyquist frequency of a 5 Gbps signal, so that no signal frequency components would see the resonant dip, the maximum length of any via stub in the channel should be shorter than 300/5 = 60 mils. This defines the design space for acceptable via stubs for 5 Gbps signals.

The via stub length can be engineered by restricting layer transitions between the top few layers and the bottom few layers, using thinner boards, backdrilling the long stubs, or using alternative via technologies such as blind and buried vias or microvias.

12.15 The Mode Conversion Terms

The off-diagonal quadrants of the differential S-parameter matrix are the most confusing, as they have information about how the differential pair interconnect converts differential signals into common signals and vice versa. They are also very important when it comes to hacking into the interconnect to determine possible root causes of behavior.

Even though it is incorrect to call the signal entering a port the differential mode signal, the industry has adopted a convention that is very confusing. These two off-diagonal quadrants are called the *mode conversion quadrants*, as they describe how one signal mode is converted into another signal mode.

The SCD quadrant describes how a differential signal enters the differential pair, but comes out as a common signal. It can enter at one port and come out that port or be transmitted to the other port.

TIP Only asymmetry between one line and the other line in a differential pair will convert some differential signal into common signal or vice versa. In a perfectly symmetrical differential pair, there is never any mode conversion. As long as what is done to one line is done to the other line, no matter how large the discontinuity, there will be no mode conversion and the SCD terms will be zero.

Any asymmetry between the two lines that make up the differential pair will cause mode conversion. This can be a length difference, a local variation in the dielectric constant between the two lines, a test pad on one line but not the other, a difference in the rise time of the drivers, a skew between the drivers of the two channels, a variation in the clearance holes in a via field, or a line width difference. How much signal appears in the SCD11 term compared to the SCD21 term depends on how much common impedance discontinuity the common signal encounters to reflect some of the common signal back to the originating port.

The same asymmetry that converts a differential signal into a common signal will convert a common signal into a differential signal. Mode conversion creates three possible problems for differential pairs used in high-speed serial link applications.

The first problem is that if there is significant mode conversion of the differential signal, the differential signal amplitude can drop. This drop might increase the bit error rate. An example of the impact on the differential signal from a small length difference between the two lines in a differential pair is shown in Figure 12-37.

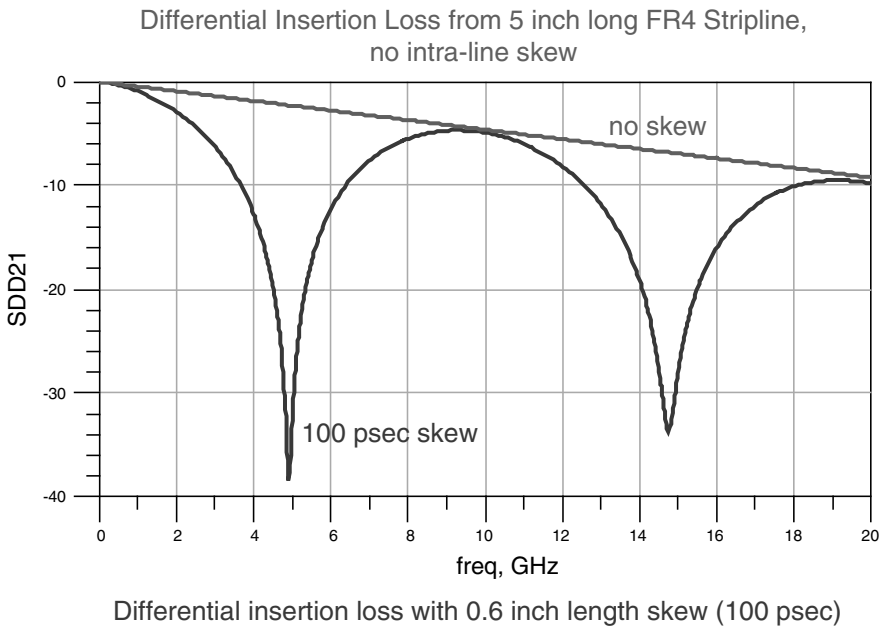


Figure 12-37 Differential insertion loss through a uniform differential pair, with no intraline skew and with 100 psec of skew between the two lines. The skew causes a dip in the differential signal when the skew is 1/2 of a cycle. Simulated with Agilent's ADS.

The second problem with mode conversion is that the common signal created might reflect from the unterminated ends of the differential pair, and each time it passes through the asymmetry, some of it might convert back into a differential signal but asynchronous with the data stream and distort the original differential signal. This would also increase the bit error rate.

The third problem with the common signal generated by mode conversion arises if the common signal gets out of the box and on an unshielded twisted pair cable. A pure differential signal on a twisted pair cable does not radiate EMI and has no problem passing an FCC or similar EMC certification test. However, if more than 3 microAmps of common signal at 100 MHz or higher frequency is created and gets out on one meter of twisted pair cable, it will radiate and cause a failure in an FCC test.

If the common impedance is 300 Ohms, it only takes about 1 mV of common signal generated from mode conversion, driving an unshielded, external twisted pair cable to cause an FCC failure. Depending on the differential signal swing, this might be on the order of 1% mode conversion.

Figure 12-38 shows an example of the measured mode conversion in three different, differential channels routed adjacent to each other in a backplane. In each case the converted common signal would exceed 1 mV. If any of this escaped onto external, unshielded twisted pair cable, it could cause an FCC failure. However, if this common signal stays inside the box, it will not contribute to EMI. It may still cause a problem in distorting the differential signal, though.

12.16 Converting to Mixed-Mode S-Parameters

Virtually all instruments only measure the single-ended S-parameters. To display the mixed-mode S-parameters, the single-ended S-parameters are mathematically transformed into the mixed-mode S-parameters. For all linear, passive interconnects, the mixed-mode S-parameters are just a linear combination of the single-ended S-parameters.

Exactly how the 10 unique single-ended S-parameters are transformed into the 10 unique differential S-parameters requires some complicated matrix math but is very straightforward to implement. The matrix math is summarized in a simple matrix equation:

$$S_d = M^{-1} S_s M \tag{12-38}$$

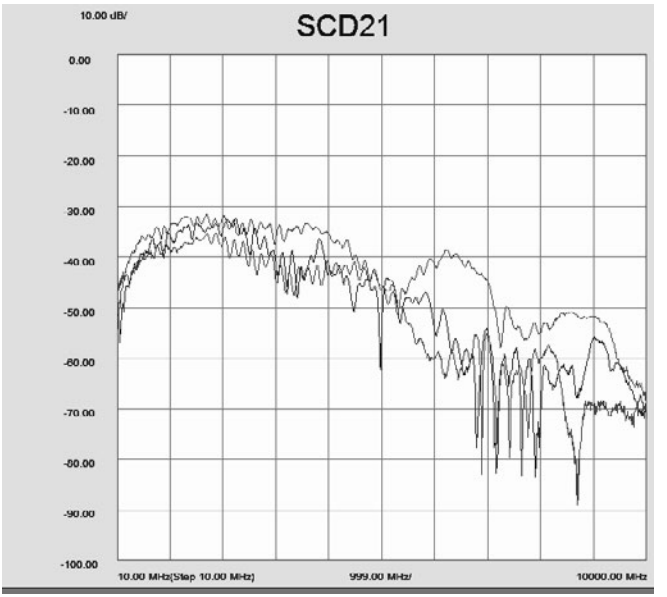


Figure 12-38 Measured mode conversion in the transmitted signal, SCD21, for three different channels in a backplane. The peak mode conversion is about -35 dB, which is slightly above 1%. If the differential signal were 100 mV, the common signal would above 1 mV, and if this appeared on unshielded twisted pair, would probably fail an FCC test. Measured with an Agilent N5230 VNA and displayed with Agilent's PLTS.

where:

S_d = the mixed-mode or differential S-parameter matrix

M = the transform matrix

S_s = the single-ended S-parameter matrix that is typically measured

The transform matrix relates how single-ended matrix elements combine to create the mixed-mode matrix. When the port assignments are as described in this chapter, the transform matrix is:

$$M = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \tag{12-39}$$

For example, the SDD11 element is given by:

$$SDD11 = 0.5 \times (S11 + S33 - 2 \times S31) \tag{12-40}$$

And the SDD21 element is given by:

$$SDD21 = 0.5 \times (S21 + S43 - S41 - S23) \tag{12-41}$$

This matrix manipulation allows us to take any single-ended S-parameter matrix of measured or simulated S-parameters and routinely transform it into the differential or mixed-mode elements. This is another reason why it is so important to use the port assignment listed above, as otherwise the transform matrix will change.

12.17 Time and Frequency Domains

For linear, passive interconnects, the same information contained in a individual S-parameter element can be displayed either in the frequency domain or in the time domain. After all, the frequency domain S-parameters describe how any sine wave voltage waveform is treated by the interconnect. We can synthesize any time domain waveform by combinations of different frequency components.

A linear system will have no interaction between the frequency components. If we know how each frequency component interacts with the interconnect, we could literally add up the right combinations of frequency components and evaluate any time domain waveform. This is essentially the inverse Fourier transform process.

There are two commonly used synthesized time domain waveforms that provide immediately useful information about an interconnect: a *step edge* and an *impulse response*. The reflected and transmitted behavior is interpreted slightly differently.

A step response is the same waveform used in a TDR and is often referred to as the TDR response. S11, when viewed in the time domain as a step response is identical to a TDR response and has information about the single-ended impedance profile of an interconnect. The S21 term, when viewed in the time domain as a step edge, has information about how the leading or falling edge of a signal is distorted by the interconnect. This is usually dominated by losses in the interconnect.

When there are multiple responses to evaluate in the time domain, we usually use the S-parameter notation for each matrix element to signify which ports

are involved, but instead of using the letter S, we use the letter T. T11 is the TDR response, and T21 is the TDT response.

> **TIP** Each domain has the same information, just displayed in a different format. They obviously look different visually, and each term has a different sensitivity to interconnect properties. Depending on the question being asked, one or the other format might be the right one to use to get to the answer faster.

For example, if the question is what is the characteristic impedance of the line, the T11 element would get us the answer the fastest. If the question is what are the losses in the line the S21 element would get us the answer the fastest.

This is illustrated in Figure 12-39, where the same measured insertion and return loss of a line in a microstrip pair is displayed in the frequency and time domains. Included are examples of two microstrip lines, one in a tightly coupled pair, and one in a weakly coupled pair.

> **TIP** Even though the information content in the time domain and the frequency domain is exactly the same, the information that is immediately available from the front screen is different. To get to the answer the fastest, it is important to have flexibility in converting the S-parameters between the time and frequency domains and between single-ended and differential domains.

It is much easier to interpret the impedance profile of an interconnect from the step edge time domain form of S11 than the frequency domain form. Likewise, the frequency domain form of S21 is a more sensitive measure of the interconnect losses than the step time domain response. Depending on the question being asked, one element in a particular domain might get you to the answer quicker than another. This is why having flexibility in data mining all the S-parameters is valuable.

The impulse time domain response is particularly useful when using the S-parameters as a behavioral model in circuit simulators, or to emulate the interconnect's response to any arbitrary waveform.

The impulse response is also sometimes called the Green's function of the interconnect. It describes how any, tiny, incident voltage is treated by the interconnect. If we know how the interconnect treats an impulse response, we can take any waveform in the time domain and describe it as a combination of successive impulse responses. Using a convolution integral, we can combine how each of the

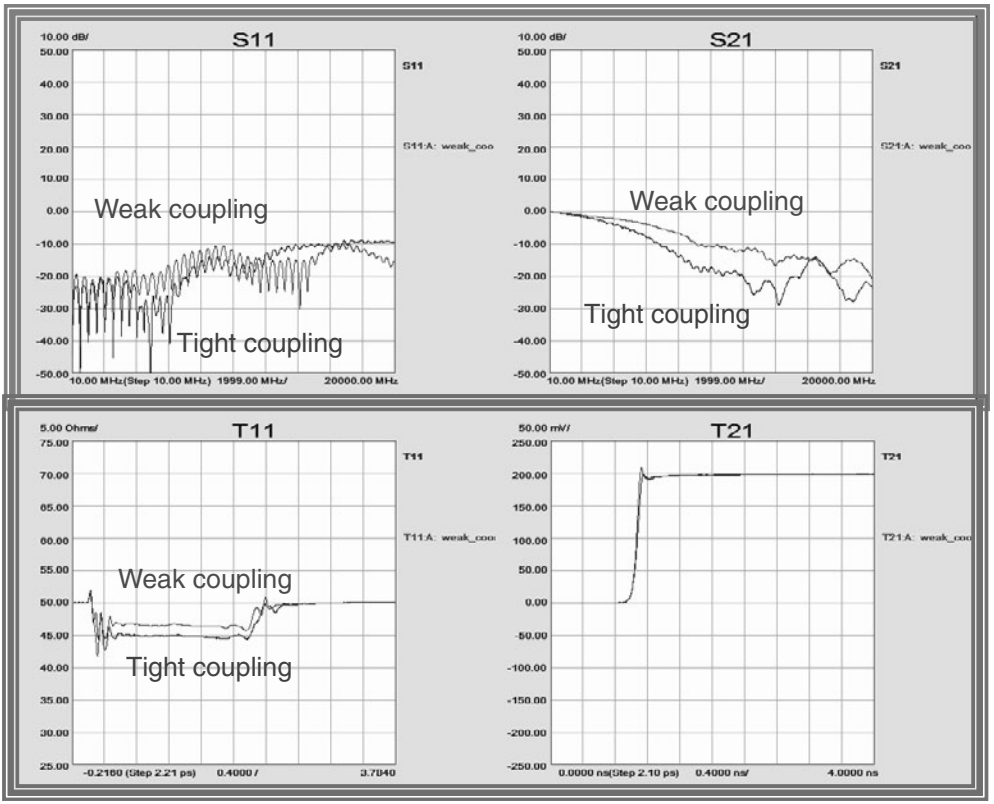


Figure 12-39 Measured single-ended return and insertion loss for a single line in a microstrip pair. One pair is tightly coupled, the other pair is loosely coupled. The exact same measured response is shown in the frequency domain and the time domain. Measured with an Agilent N5230 VNA and displayed with Agilent PLTS.

successive, scaled impulse responses is treated by the interconnect to get the output waveform.

This technique allows us to synthesize any arbitrary time domain waveform, convolve it with the impulse response, and simulate the output waveform. One application is to emulate eye diagrams of interconnects. This is diagrammed in Figure 12-40.

The SDD21 frequency domain waveform is converted into the time domain impulse response waveform. A pseudo random bit stream (PRBS) is synthesized using an ideal square wave, with some rise time. This time domain waveform is convolved with the impulse response of the interconnect. Essentially, the impulse

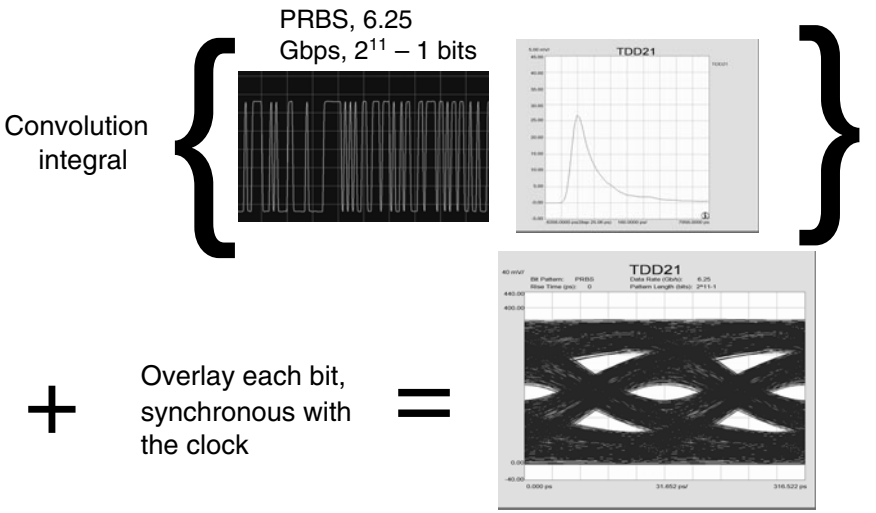


Figure 12-40 Process of simulating an eye diagram for an interconnect by convolving the impulse response with a synthesized pseudo random bit sequence and overlaying each bit synchronous with the clock. Measured with an Agilent N5230 VNA and simulated and displayed with Agilent's PLTS.

response says how each voltage point will be treated by the interconnect. The convolution integration takes each voltage point, multiplies it by the interconnect's response, and adds up all the time domain responses, walking along the time domain waveform.

The result is how the interconnect would treat the synthesized waveform. It is the simulation of the output real-time waveform. This is an example of using the S-parameters as a behavioral model of the interconnect. Its measured performance in the frequency domain is used to simulate the expected performance in the time domain without knowing anything about the internal workings of the interconnect. Many commercially available tools do this simulation automatically.

The simulated real-time, differential responses from the single-ended measurements can be used to generate eye diagrams. The clock waveform is used to slice out one, two, or three consecutive bits, and all the bits in the real-time response are superimposed with infinite persistence. The resulting superposition of all possible bit signals looks like an eye and is called an eye diagram. An example of two eye diagrams for a 3.125 Gbps and a 6.25 Gbps serial signal as they would appear through the same backplane channel based on S-parameter measurements of the channel is shown in Figure 12-41.

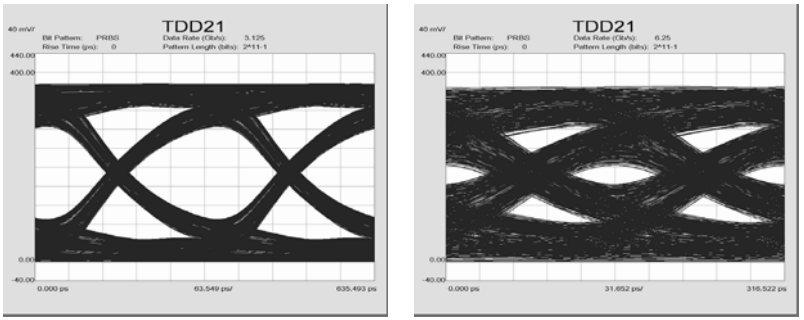


Figure 12-41 Emulated eye diagrams for a backplane channel at two different bit rates, based on the measured S-parameters. Measured with an Agilent N5230 VNA and simulated and displayed with Agilent's PLTS.

Two important features of an eye diagram can often predict bit error rates, the vertical opening, and the horizontal opening. The vertical opening is often called the *collapse* of the eye, and the horizontal opening is related to the bit period minus the deterministic jitter.

The center to center spacing of the crossovers is always the unit interval bit time. The width of the crossovers is a measure of the jitter. In the measured SDD21 response of an interconnect, the random jitter is related to the quality of the measurement instrument and is almost always negligible. All the jitter in a simulated eye diagram is deterministic in the sense that it is predictable from just the interconnect behavior. It arises from losses, impedance discontinuities, and other factors.

By synthesizing PRBS waveforms with different bit rates, a family of eye diagrams can be simulated to indicate the limits of performance of an interconnect.

12.18 The Bottom Line

1. As a universal description of any interconnect, even though they are new to the signal integrity field, S-parameters are quickly becoming an industry standard.
2. Each S-parameter is the ratio of an output sine wave to an input sine wave. It represents the behavior of the interconnect to sine waves across a defined frequency range.
3. The reflected S-parameters, S11, S22, have information about impedance discontinuities of the interconnect.

4. The transmitted S-parameters, S21, S43, have information about losses, discontinuities, and couplings to other lines.
5. Other terms may describe the cross talk between interconnects.
6. While the port impedance is generally 50 Ohms, the same S-parameter response can be transformed to any port impedance. If there is no compelling reason, 50 Ohms should always be used.
7. The single-ended S-parameters obtained in the frequency domain can be transformed into differential S-parameters or into the time domain.
8. S-parameters in the frequency domain are a measure of the overall, integrated, steady state response of the interconnect; while transformed in the time domain, S-parameters can provide spatial information about the properties of an interconnect.
9. Narrow dips in the return or insertion loss are indictors of coupling to high Q resonating structures.
10. The differential insertion loss describes the most important property of a differential channel.
11. Mode conversion is described by two of the differential S-parameter matrix elements, SCD11 and SCD21. They are useful in debugging the root cause of drops in the insertion loss.
12. Everything you ever wanted to know about the behavior of an interconnect is contained in the S-parameters, obtained either through measurement or by simulation.

CHAPTER 13

The Power Distribution Network (PDN)

The power distribution or delivery network (PDN) consists of all those interconnects from the voltage regulator module (VRM) to the pads on the chip and the metallization on the die that locally distribute power and return current. This includes the VRM itself, the bulk decoupling capacitors, the vias, the traces, the planes on the circuit board, the additional capacitors added to the board, the solder balls or leads of the packages, the interconnects in the packages mounted to the board, the wire bonds or C4 solder balls, and the interconnects on the chips themselves.

The primary difference between the PDN and signal paths is that there is just one net for each voltage rail in the PDN. It can be a very large net that can physically span the entire board and have many components attached.

TIP As we will see, the PDN is like an ecosystem. If one small part of the PDN were to change, the entire system performance can be affected. This makes generalizations very difficult.

13.1 The Problem

An example of a motherboard with all these interconnects is shown in Figure 13-1.

First and primary is to keep a constant supply voltage on the pads of the chips, and keep it within a narrow tolerance band, typically on the order of 5%.

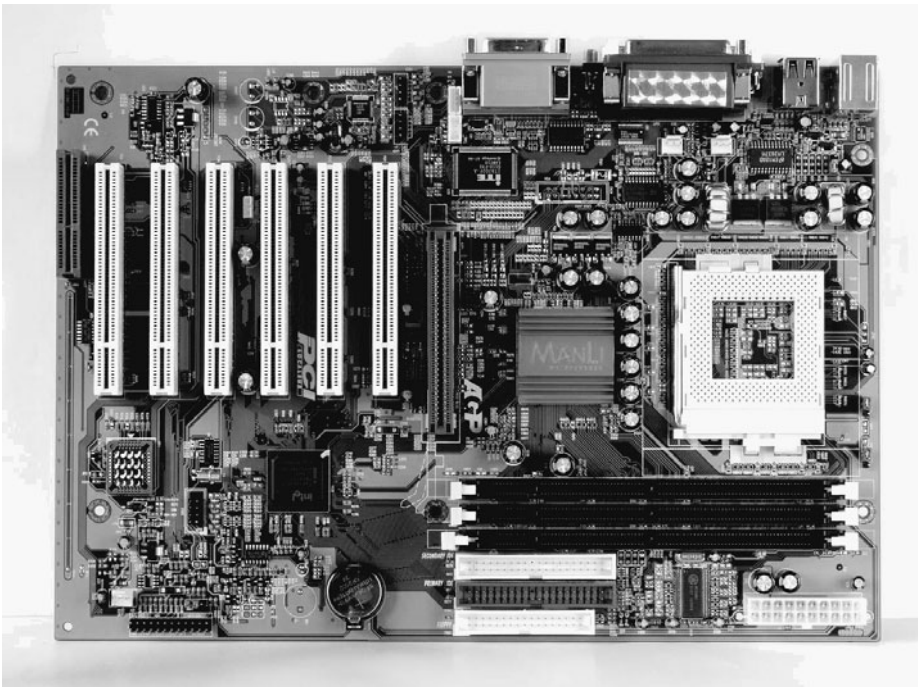


Figure 13-1 Example of typical motherboard showing all the interconnects of the PDN.

TIP The purpose of the PDN is threefold: Keep the voltage across the chip pads constant, minimize ground bounce, and minimize EMI problems.

This voltage has to be stable, within the voltage limits, from DC up to the bandwidth of the switching current, typically above 1 GHz.

Second, in most designs, the same PDN interconnects that are used to transport the power supply are also used to carry the return currents for signal lines. These interconnects must provide a low impedance return path for the signals.

The easiest way of doing this is by making the interconnects wide, so the return currents can spread out as much as they want, and by keeping the signal traces physically separated so that the return currents do not overlap. If these conditions are not met, the return current is constricted and the return currents from different signals overlap. The result is ground bounce, also called simultaneous switching noise (SSN) or just switching noise.

Finally, since the PDN interconnects are usually the largest conducting structures in a board, carry the highest currents, and sometimes carry high frequency noise, they have the potential of creating the most radiated emissions and causing failure of an EMC certification test. When done correctly, the PDN interconnects can mitigate many potential EMI problems and help prevent EMC certification test failures.

The consequence of not designing the PDN correctly is that there will be excessive noise on the voltage rails of the chips. This can cause a bit failure directly, or it can mean the clock frequency of the chip can't be met and timing errors result.

An example of the voltage noise on the pads of a processor chip is shown in Figure 13-2. In this example, the nominally constant 2.5 v rail to the core of the chip, referred to as Vdd, shows voltage noise of as much as 125 mV on some pads. As the Vdd supply drops, the propagation delay of the core gates will increase and timing problems can cause bit failures.

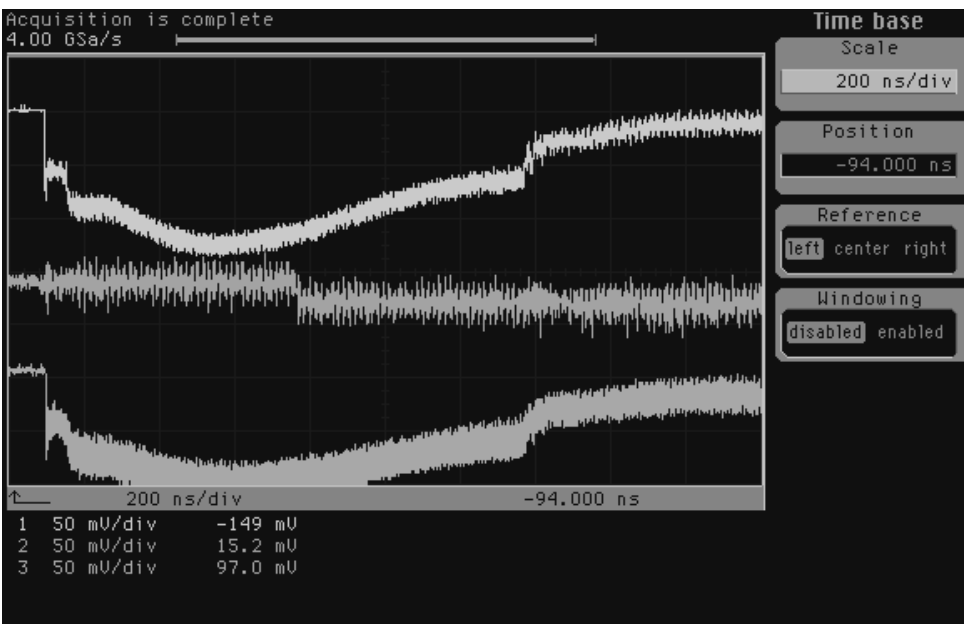


Figure 13-2 Example of the measured voltage between three different pairs of Vdd and Vss pads of a processor chip showing as much as a 125 mV drop. The initial step down is when the processor comes out of an idle state. The three traces are three different locations on the die. The precise shape is related to the microcode running on the processor.

13.2 The Root Cause

If the problem is a voltage drop or droop on the power supply rails on the pads of the chip, why not just use a "heftier" regulator, one that can supply a more rock stable voltage? Why not pay extra for a regulator with a 1% regulation or even 0.1% regulation? This way, the voltage from the regulator will be absolutely stable, no matter what, right?

What the chip cares about is the voltage on its pads. If there were no current flow in the PDN interconnects from the regulator pads to the chip pads, there would be no voltage drop in this path and the constant regulator voltage would appear as a constant rail voltage on the chip pads.

If there were a constant DC current draw by the chip, this DC current would cause a voltage drop in the PDN interconnects due to the series resistance of the interconnects. This is commonly referred to as the *IR drop*. As the current from the chip fluctuates, the voltage drop in the PDN would fluctuate and the voltage on the chip pads would fluctuate.

Now add not just the resistive impedance of the PDN, but also the complex impedance, including the inductive and capacitive qualities of the PDN interconnects. The impedance of the PDN, as seen by the pads on the chip, in general, is some impedance verses frequency, Z (f). This is diagrammed in Figure 13-3.

As fluctuating currents with some spectrum, I(f), pass through the complex impedance of the PDN, there will be a voltage drop in the PDN:

$$V(f) = I(f) \times Z(f) \tag{13-1}$$

where:

V(f) = the voltage amplitude as a function of frequency

I(f) = the current spectrum drawn by the chip

Z(f) = the impedance profile of the PDN as seen by the chip pads.

This voltage drop in the PDN means that the constant voltage of the regulator is not seen by the chip, but is changed. In order to keep the voltage drop on the chip pads less than the voltage noise tolerance, usually referred to as the ripple, given the chip current fluctuations, the impedance of the PDN needs to be below some maximum allowable value. This is referred to as the target impedance:

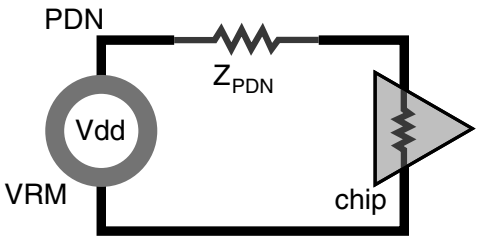


Figure 13-3 Diagram of the connections from the VRM, through the PDN to the chip pads and the voltage drop across the PDN interconnects due to the impedance of the PDN.

$$V_{ripple} > V_{PDN} = I(f) \times Z_{PDN}(f) \tag{13-2}$$

$$Z_{target}(f) = Z_{PDN}(f) < \frac{V_{ripple}}{I(f)} \tag{13-3}$$

where:

V_{ripple} = the voltage noise tolerance for the chip, in Volts

V_{PDN} = the voltage noise drop across the PDN interconnects, in Volts

I(f) = the current spectrum drawn by the chip, in Amps

Z_{PDN} (f) = the impedance profile of the PDN as seen by the chip pads, in Ohms

Z_{target} = the maximum allowable impedance of the PDN, in Ohms

As mentioned many times so far in this book, the most important step to solve a signal integrity problem is to identify the root cause of the problem. The root cause of rail collapse or voltage noise on the PDN conductors is that a voltage drop in the PDN interconnects is created by the chip's current flowing through the impedance of the PDN.

TIP If we want to keep the voltage stable across the pads of the chip, given the chip's current fluctuations, it means keeping the impedance of the PDN below a target value. This is the fundamental guiding principle in the design of the PDN.

13.3 The Most Important Design Guidelines for the PDN

The goal in designing the PDN interconnects is to bring the impedance below the target value from DC to high frequency. In general, this will be accomplished by following three important design principles. Though it may not always be possible to push these to the limit, it is always important to be aware of the directions to head, even if the real world constraints keep you from the ultimate path.

The three most important guidelines in designing the PDN are:

1. Use power and ground planes on adjacent layers, with as thin a dielectric as possible, and bring them as close to the surface of the board stack-up as possible.
2. Use as short and wide as possible surface trace between the decoupling capacitor pads and the vias to the buried power and ground plane cavity and place the capacitors where they will have the lowest loop inductance.
3. Use SPICE to help select the optimum number of capacitors and their values to bring the impedance profile below the target impedance.

Unfortunately, in the real world of practical product design, you may not always have the luxury of power and ground planes on adjacent layers or placed near the top of the board stack-up. There may be multiple voltage rails, and they may have odd and irregular shapes with many antipad clearance holes.

You may not be able to use as many capacitors as you think you need, nor place them in proximity to the devices they are decoupling. Even if you do the best you can, it will still be important to know if it is "good enough" before you build the product. The last place you want to find a problem is when you are making 100,000 units and finding that 1% of them are failing due to excessive ripple in the PDN.

The time to find this out is as close to the beginning of the design process as possible, and the only way to determine this is by using analysis tools that allow you to explore design space.

TIP It is essential to try to follow the three important design guidelines above, and at the same time, to use combinations of rules of thumb, approximations, and numerical simulation tools to predict the impedance profiles and voltage noise under typical and worst case conditions.

The most important principle to follow for cost-effective design is to add appropriate analysis as early in the design cycle as possible. This will reduce the surprises as the design progresses and result in a product with acceptable performance, at the lowest cost that works the first time.

13.4 Establishing the Target Impedance Is Hard

The first step in designing the PDN is to establish the target impedance. This must be done separately and independently for each voltage rail to all the chips on the board. Some designs may use as many as 10 different voltages. In each one, the target impedance may vary with frequency due to the specific current spectrum of the chip.

Suppose the current from the chip on one rail is a sine wave, with a peak to peak value of 1 A. The amplitude of the sine wave of current will be 0.5 A. This current from the chip is shown in Figure 13-4 in both the time domain and the frequency domain.

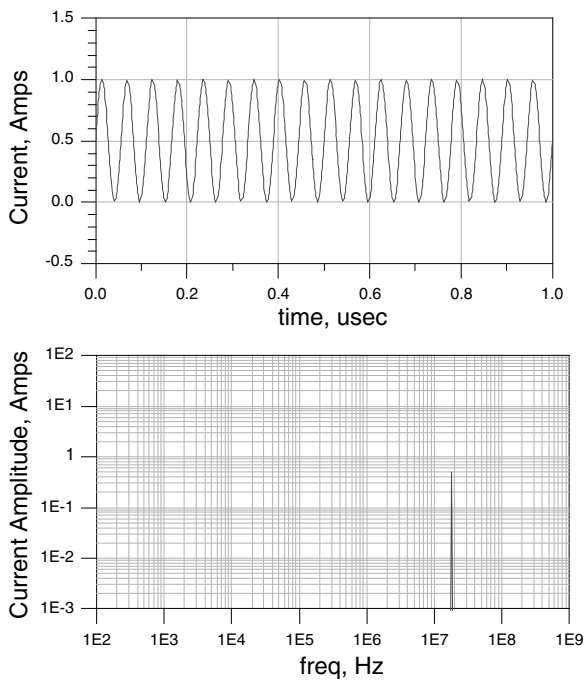


Figure 13-4 Example of the current waveform in the time domain (top) and the frequency domain (bottom) for a sine wave current draw.

When this frequency component of the current flows through the specific impedance profile of the PDN, a voltage noise will be generated in the PDN. An example of an impedance profile with the frequency component of the current, and the resulting time domain voltage noise across the chip pads, is shown in Figure 13-5.

When the sine wave current passes through an impedance that is too large, the voltage generated is above the ripple spec, which is typically +/– 5%, shown as the reference lines.

There is the potential for the current draw through a chip to be at almost any frequency from DC to above the clock frequency. This means that unless the precise current spectrum from the chip is well known, for all the possible microcode that could be running through it, we have to assume the peak current could be anywhere from DC to the bandwidth of the signals.

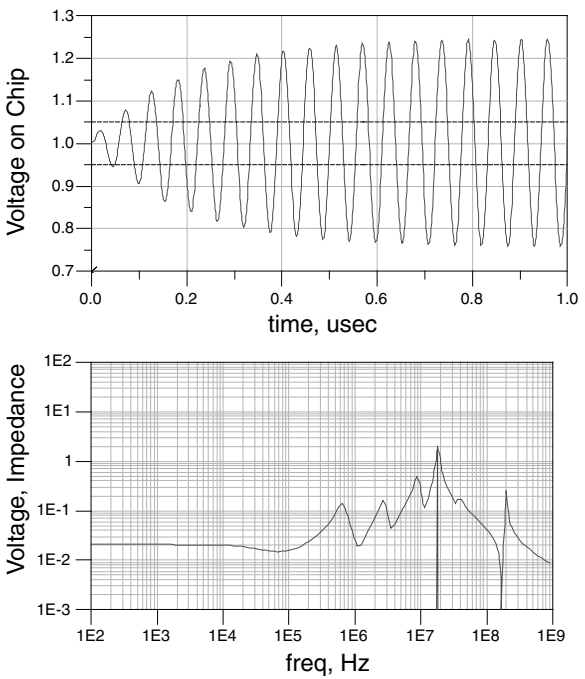


Figure 13-5 Voltage noise on the chip pads (top) as the sine wave current flows through the PDN impedance profile of simulated PDN (bottom). The spike in the PDN profile at about 20 MHz shows where the sine wave frequency component of the current is with respect to the PDN impedance peaks.

In a few rare cases, if it is known that the chip processing will have a high frequency fall off of the current draw above some frequency, it may be possible to put some constraints on the current spectrum. This should always be done, whenever possible.

While the current draw from a chip is rarely a pure sine wave, there are always sine wave frequency components in the current. The precise spectrum of the current amplitudes will interact with the impedance profile of the PDN completely independently of each other, but the resulting voltage waves will add together. Sometimes they can add and still meet the ripple spec, while at other times, they can add and exceed the ripple spec, depending on the precise overlap of current peaks and impedance peaks.

Figure 13-6 is an example of a 1 A peak to peak square wave current draw for two slightly different modulation frequencies. The square wave current will have sine wave frequency components at odd multiples of the square wave frequency. Above roughly the fifth harmonic, depending on the rise time, the amplitude of the sine wave harmonics will drop off much faster than 1/f. As the modulation frequency changes, the frequency distribution of the harmonics shifts and interacts differently with the PDN impedance profile.

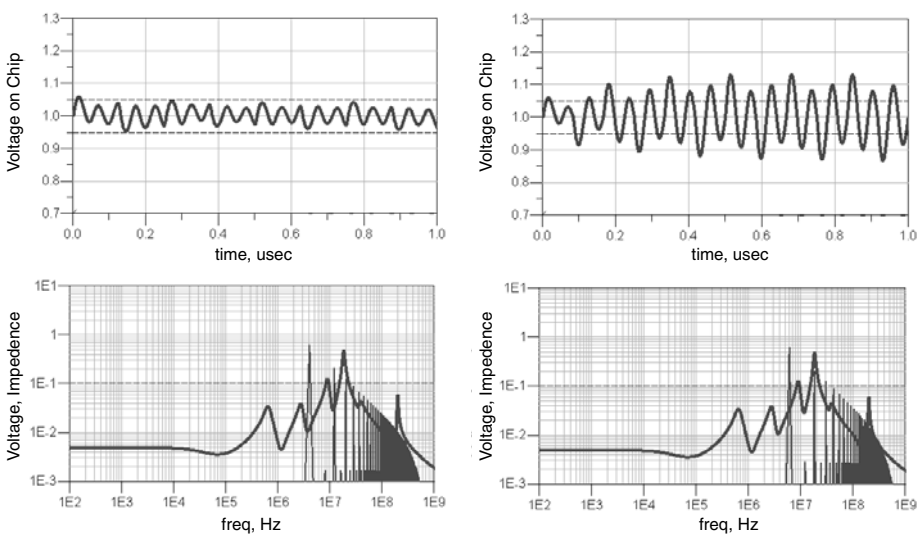


Figure 13-6 The resulting ripple noise when the same 1 A current draw has a slightly different modulation frequency, where one of the harmonic components overlaps an impedance peak of the PDN impedance profile.

A very slight frequency shift in the modulation of the current can mean the difference between acceptable performance and a failure. Unfortunately, the engineer designing the PDN has very little control of the current draw spectrum of the chip. It is whatever the chip is going to do, depending on its operation.

This means that unless there is precise information about the specific, worst case spectrum of the current draw of the chip, a conservative design must assume a worst case current that could occur at any frequency from DC to the bandwidth of the clock, which is a few times the clock frequency.

In practice, it is not the peak current but the maximum transient current that interacts with the higher frequencies of the PDN. If there is a steady state DC current draw from the chip, the sense lines of the VRM can usually compensate to keep the rail voltage close to the specified voltage value. It's when the current changes from the DC value, either increasing or decreasing, at frequencies above the response frequency of the VRM, that the current will interact with the PDN impedance.

The maximum impedance for the PDN, the target impedance, is established based on the highest impedance that will create a voltage drop still below the acceptable ripple spec. This is given by:

$$Z_{PDN} \times I_{transient} = V_{noise} < V_{dd} \times \text{ripple}\% \tag{13-4}$$

or

$$Z_{target} < \frac{V_{dd} \times \text{ripple}\%}{I_{transient}} \tag{13-5}$$

where:

V_{dd} = the supply voltage for a specific rail

$I_{transient}$ = is the worst case transient current

Z_{PDN} = the impedance of the PDN at some frequency

Z_{target} = the target impedance, the maximum allowable impedance of the PDN

V_{noise} = the worst case noise on the PDN

ripple% = the ripple allowed, assumed to be +/– 5% in this example

TIP The optimum PDN impedance should be below the target impedance, but not too far below the target impedance.

If the PDN impedance is kept below the target impedance at every frequency, the worst case voltage noise generated across it as the worst case, maximum transient current flows through it will be less than the ripple spec. If the PDN impedance is much below the target impedance, it means the PDN was overdesigned and costs more than it needs to.

TIP Whenever possible, the peak transient current should be used in estimating the target impedance. When the peak transient current is not available, it can be roughly estimated from the maximum current draw or from the power consumption of the chip.

While the worst case transient current is what is important, rarely is this provided in a spec sheet. Rather, it is the worst case peak current per rail that is on spec sheets. After all, this is important to estimate how large a voltage regulator module is needed. It must be capable of supplying the maximum current draw.

The peak current could be mostly DC with a 10% transient current, or it could be a very low quiescent current with most of the peak current being transient that would only last for a few microseconds. Without special knowledge of the behavior of the chips in each application, the conservative design has to plan for the worst case.

What fraction of the maximum current is transient? Obviously, it depends on the function of the chip. It could vary from 10% to 90% depending on the application. As a rough, general rule of thumb, the transient current can be estimated as being half of the maximum current:

$$I_{transient} \sim \tfrac{1}{2} \times I_{max} \tag{13-6}$$

where:

$I_{transient}$ = the worst case transient current from the chip

I_{max} = the maximum total current from the chip

Alternatively, the worst case power dissipation of the chip will always be provided in a chip's specs, since this is critical information when designing the thermal management approach for the package. It is not usually separated by voltage rail, so some assumptions would have to be made on the power consumption of each rail.

However, given the power consumption per voltage rail, the peak current draw of a chip can be estimated from:

$$I_{peak} = \frac{P_{max}}{V_{dd}} \tag{13-7}$$

And from this, the target impedance can be evaluated as:

$$Z_{target}(f) < \frac{V_{dd} \times ripple\%}{I_{transient}} = 2\frac{V_{dd} \times ripple\% \times V_{dd}}{P_{max}} \tag{13-8}$$

where:

I_{peak} = the worst case peak current in Amps

P_{max} = the worst case power dissipation in watts

V_{dd} = the voltage rail in volts

ripple% = the ripple spec in %

2 = comes from the transient current being ½ the peak current

For example, if the ripple spec is 5%, the target impedance is:

$$Z_{target}(f) = 0.1 \times \frac{V_{dd}^2}{P_{max}} \tag{13-9}$$

For example, in the case of a 1 volt rail and 1 watt power dissipation device, the target impedance would be about 0.1 Ohm:

$$Z_{target}(f) < 0.1 \times \frac{1^2}{1} = 0.1\Omega \tag{13-10}$$

Some chip vendors, especially FPGA vendors, also provide calculation tools that allow simple estimates of the current draw of specific voltage rails depending on the gate utilization. These can be used to estimate the target impedance specs of the rails. An example of the results of using one such analysis for the Altera Stratix II GX FPGA is shown in Figure 13-7.

Finally, the current requirements of the I/O voltage rails, typically referred to as either the Vcc or Vddq rails, can be estimated based on the number of gates that are switching.

If each output gate drives a transmission line with some characteristic impedance, then the load it sees, if only for a round-trip time, is the same as the characteristic impedance of the line it drives.

If n gates could switch simultaneously, the transient current draw could be:

$$I_{transient} = n\frac{V_{cc}}{Z_0} \tag{13-11}$$

And the target impedance of the VCC rail would be:

$$Z_{target}(f) < 0.05 \times \frac{1}{n}Z_0 \tag{13-12}$$

Power rail	Voltage (v)	ripple%	Max current (A)	Transient current amplitude (A)	Z_{target} (Ohms)
VCCT/R	1.2	2.5%	1.2	0.6	0.05
VCCH	1.5	2%	0.17	0.085	0.35
3.3 v Analog	3.3	3%	0.274	0.137	0.72
VCCP	1.2	2%	1.03	0.51	0.047

Figure 13-7 Example of a calculation of the target impedance of different voltage rails based on gate utilization of an Altera FPGA.

where:

$I_{transient}$ = the worst case transient current

n = the number of I/O that could switch simultaneously

V_{cc} = the voltage rail

Z_0 = the characteristic impedance of the transmission lines

For example, if the lines are all 50 Ohms and there are 32 bits switching simultaneously, the target impedance for the Vcc rail would be:

$$Z_{target}(f) < 0.05 \times \frac{1}{32} 50 = 0.08\Omega \quad (13\text{-}13)$$

Even with the peak current and the target impedance established, the current could be fluctuating at almost any frequency, due to the specific microcode or application running. This means that unless there is information to the contrary, it must be assumed this is the target impedance, flat from DC to very high frequency.

TIP The goal in designing the PDN is to keep the impedance of the PDN interconnects below this target value over a very wide bandwidth. A PDN above the target impedance may result in excessive ripple. A PDN impedance much below the target impedance may be overdesigned and more expensive than it needs to be.

When the impedance profile is kept below the target value, the worst case voltage rail noise will be less than the ripple spec. An example of a successful impedance profile is shown in Figure 13-8.

However, if there is a peak in the impedance profile that exceeds the target spec, and if the worst case current happens to fall on top of this peak impedance, there is the chance the ripple spec may be exceeded. This is shown in Figure 13-9.

Peak impedances in the PDN impedance profile are an important design feature to watch out for. Many aspects of PDN impedance design, especially the selection of capacitor values, are driven by the desire to reduce the peak impedances in the PDN.

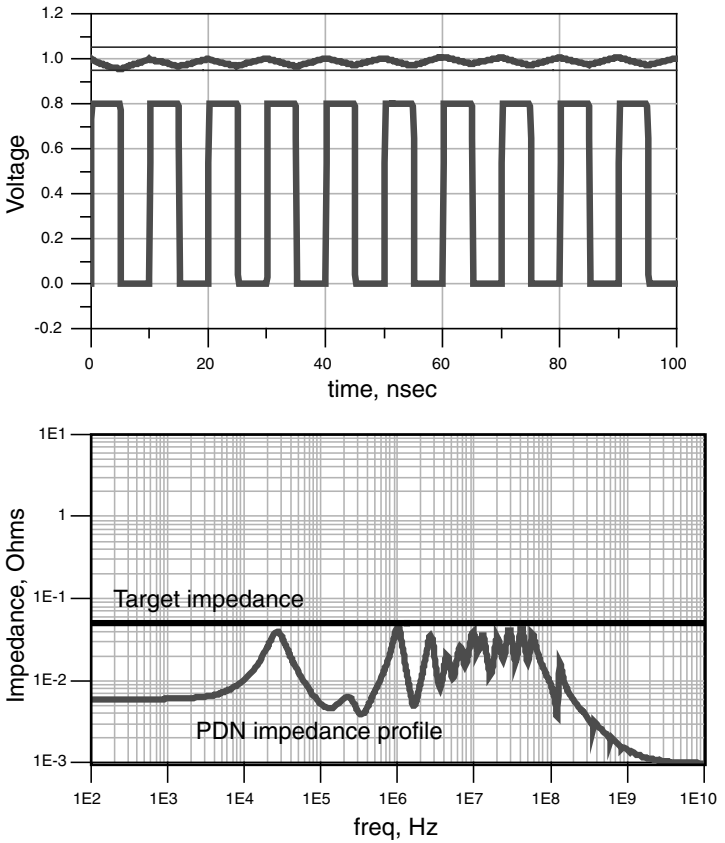


Figure 13-8 When the impedance profile (bottom) is below the target impedance, the worst case voltage noise (top) is below the ripple spec. The square wave is the current draw by the chip, while the flat curve is the voltage on the supply rail.

13.5 Every Product Has a Unique PDN Requirement

One of the greatest sources of confusion in PDN design is created by taking the PDN design features of one product and blindingly applying them to another product.

TIP Unlike the design of signal paths, where the design rules in one product can often be applied to other products of similar bandwidth, the behavior of the PDN depends on the interactions of all of its parts, and the goals and constraints vary widely from product to product.

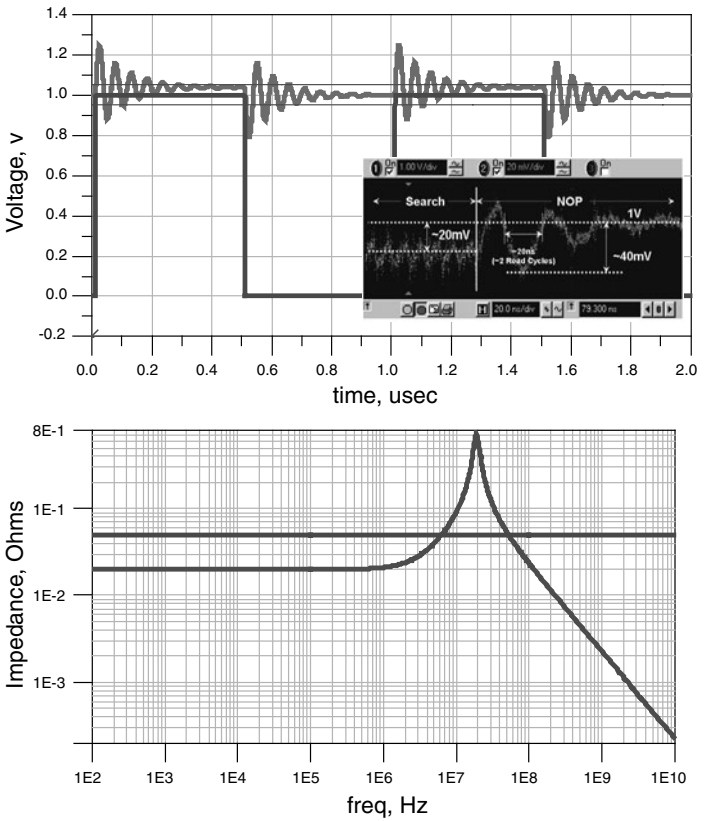


Figure 13-9 When the impedance exceeds the target spec (bottom) and the current's peak frequency hits this impedance peak, excess ripple can result (top). The square wave is the current draw through the chip. The ringing wave is the voltage on the supply rail. Inset is an example of the measured voltage noise on a PDN showing the typical ringing response of a peak in the impedance profile.

The PDN is one giant net, rather than a large number of individual nets, with only a small amount of local coupling between them. In this respect, the PDN net is like an ecosystem of interconnects. While it may be possible to suggest an optimized design from parts of it, the most cost-effective designs are based on optimizing the entire ecology of all the elements, across the entire frequency range.

The voltage level of the rails can vary from 5 volts to less than 1 v depending on the chip type and technology node. The ripple spec may be as large as 10% in some devices, or as low as 0.5% in others, such as phase locked loop (PLL) supplies or the analog to digital converter (ADC) reference voltage rails.

The current draw from chips can vary from more than 200 A in high-end graphics chips and processors, to as low as 1 mA for some low power micro controllers.

This means that the target impedance values can vary from below 1 mOhm for high-end chips to more than 100 Ohms. This is five orders of magnitude in impedance.

There could be as many as 10 different voltage rails in some designs, many sharing the same layer, while other designs may have just one voltage and ground. Some of the planes may be solid; others may be irregularly shaped and full of clearance holes.

TIP This wide variety of applications and board constraints means no one solution is going to fit all. Instead, each design must be treated as a custom design.

It is dangerous to blindly apply the specific features of one design to another design. However a general strategy can be followed to arrive at an acceptable impedance profile.

13.6 Engineering the PDN

It is remarkable that so complex a structure as the PDN interconnects can be partitioned in the frequency domain into just five simple regions. Figure 13-10 diagrams these five regions, based on what frequency range they can influence.

At the lowest frequency, the VRM dominates the impedance the chip sees looking into the PDN. Of course, the series resistance of the interconnects can also set a limit on the lowest impedance of the PDN, if it is larger than the VRM impedance. The VRM performance dominates from DC to about 10 kHz.

The next higher frequency regime, roughly in the 10 kHz to 100 kHz range, is dominated by the bulk decoupling capacitors. These are typically electrolytic and tantalum capacitors that provide a low impedance beyond where the VRM can go.

The highest frequency impedance is set by the on-die capacitance. This is the only feature in the PDN that the chip sees in the GHz regime. It generally has the lowest loop inductance associated with it and offers the lowest impedance at the highest frequency of any element in the PDN.

TIP Every chip interfaces to the board it is mounted to through some mounting inductance. Usually, this is dominated by the package, the board vias, and the spreading inductance of the via contacts into the power and ground planes of the board.

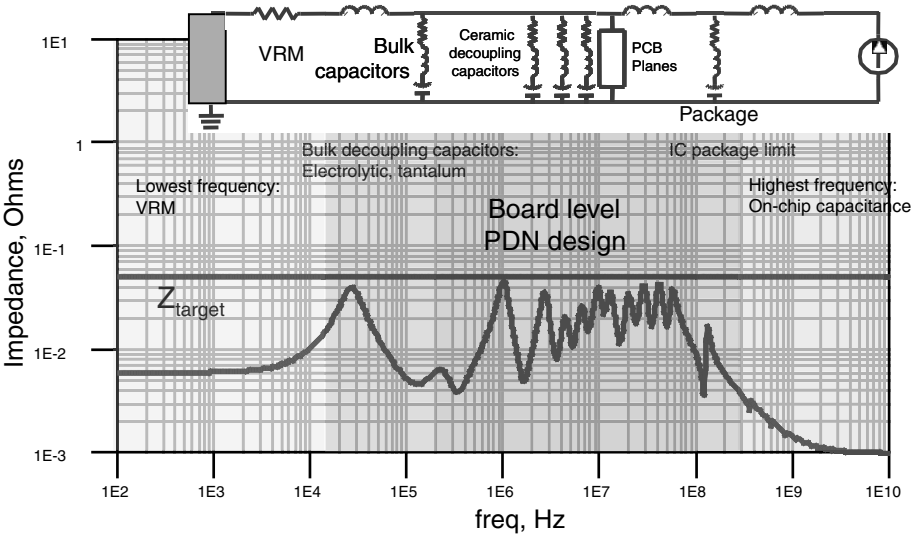


Figure 13-10 The five parts of the PDN separated by the frequency range they influence.

The PDN interconnects in the package are generally inductive. This means that at high frequency, they will act as a high impedance path. Even if the board was designed with an impedance of a dead short, the chip would be looking at this short through the chip attach and package attach inductance, and would see an impedance limited by these inductances.

The series inductance of the package's PDN will always limit the highest frequency at which the chip will see the board level PDN. This acts as a high-frequency limit to the board level PDN design. This means that above this package-limited frequency, the impedance the chip sees will be determined by the on-die capacitance and any capacitance in the package. This limit is generally in the 100 MHz range. Above this frequency, the impedance the chip sees is all about the package and the chip.

TIP The frequency region that board level PDN design can influence is roughly from the 100 kHz range up to about 100 MHz. This is where the planes of the board and the multilayer ceramic chip capacitors (MLCC) can play a role.

These capacitors, typically in sizes of 60 mils × 30 mils and referred to as 0603 or 40 mils × 20 mils and referred to as 0402, are called *chip capacitors* because they look like small "chips" of something on a board. Figure 13-11 is a

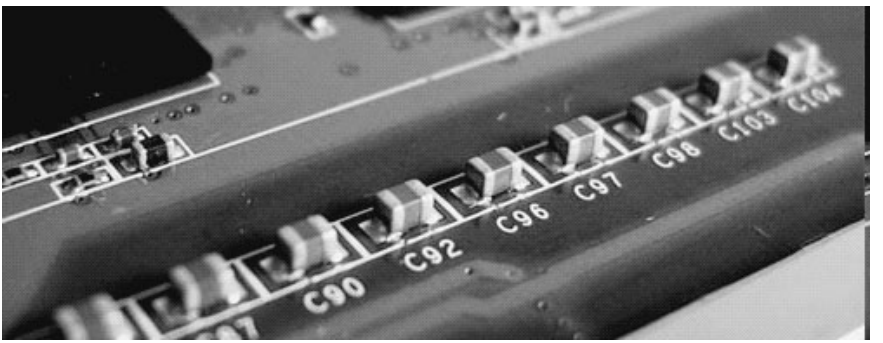


Figure 13-11 Close-up of typical 0402 MLCC capacitors mounted to a small memory board.

close-up of some typical multilayer ceramic chip (MLCC) capacitors on a small memory board.

13.7 The VRM

The low frequency impedance is set by the voltage regulator module. Regardless of the type of regulator, all VRMs have an output impedance profile. This can easily be measured using a two-port impedance analyzer.

An example of the measured impedance profile of a typical VRM is shown in Figure 13-12. In this example, the impedance looking into the output leads of the VRM was measured when the regulator was turned off and when it was turned on and providing regulation. In addition, the impedance of a simple two-capacitor model is superimposed.

This illustrates that when the regulator is off, the behavior seen at the output leads is almost exactly as predicted by a two-capacitor model, each capacitor being modeled as an RLC circuit.

This behavior corresponds to the two bulk decoupling capacitors associated with the VRM. The 910 uF capacitor is an electrolytic capacitor, while the 34 uF capacitor is a tantalum capacitor. This impedance profile is for the passive network of the leads and the two capacitors.

When the regulator is turned on, its output impedance drops by orders of magnitude at low frequency. This is exactly what is expected from a regulator. The output voltage is kept constant independent of the current load. A large change in current produces a small change in voltage, the behavior of a low impedance. However, we see that in the actual behavior of the VRM, this low impedance is maintained from DC only up to the kHz range.

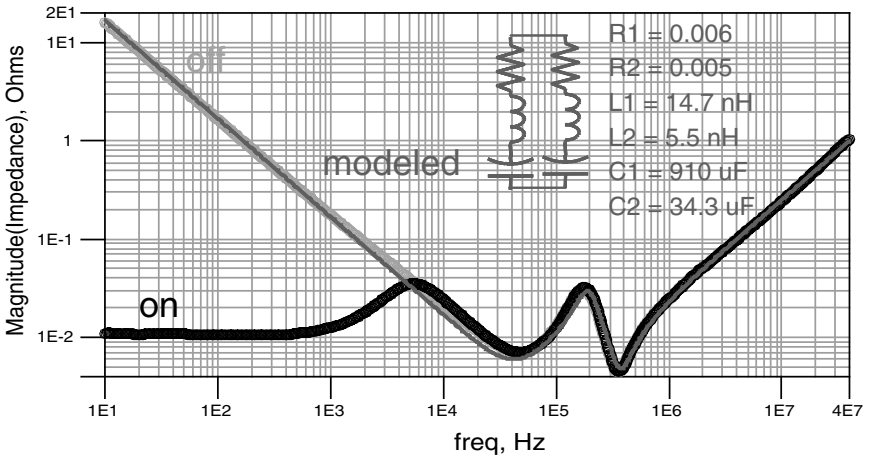


Figure 13-12 Measured impedance profile, from 10 Hz to 40 MHz, using an Ultimetrix Impedance Analyzer for a typical VRM, showing the impedance when on, when turned off, and the modeled impedance based on a two-capacitor model.

Above about 1 kHz, the impedance is seen to increase, until it matches the impedance of the bulk capacitor at about 4 kHz, at which frequency the impedance is brought down by the passive capacitor network on the regulator. Above about 1 kHz, the output impedance of the VRM is completely due to the passive capacitors, and the active regulation plays no role at all. Whether the regulator is on or off, the impedance is the same.

This is a slight exaggeration since the regulator actually fights with the capacitance of the passive network and when the regulator is turned on its impedance is actually higher than if it were literally turned off.

TIP The output impedance of most VRMs is low up to the kHz regime. Beyond this, what brings the impedance down is the bulk capacitors associated with the regulator.

The total amount of capacitance needed on a board in the form of electrolytic or tantalum capacitors can be estimated based on achieving the target impedance at the frequency where the VRM is no longer able to maintain the low impedance.

The capacitance is chosen so that its impedance at 1 kHz is less than the target impedance. The minimum capacitance needed is given by:

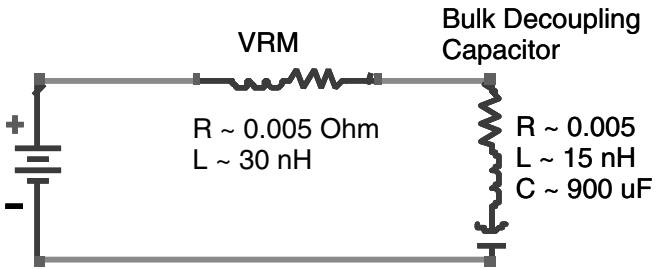


Figure 13-13 Typical equivalent circuit model of a VRM and bulk decoupling capacitor with typical values for each element.

$$C_{bulk} < \frac{1}{Z_{target} \times 2\pi \times 1\ kHz} = \frac{160\ uF}{Z_{target}} \quad (13\text{-}14)$$

where:

C_{bulk} = the minimum bulk capacitance needed, in uF

Z_{target} = the target PDN impedance in Ohms

1 kHz = the frequency at which the VRM is no longer able to provide low impedance

For example, if the target impedance is 0.1 Ohm, the minimum bulk capacitance needed is about 1600 uF. Of course, this is only a rough estimate but a good starting place. When it comes to establishing the actual target values of the capacitance, the interactions of the VRM effective inductance and capacitor's capacitance must be taken into account with a SPICE simulation.

The low-frequency model of a VRM can be easily approximated by a simple RL model with a voltage source. The equivalent circuit model of the VRM and bulk decoupling capacitor is shown in Figure 13-13.

This circuit can be used to optimize the capacitor value to keep the impedance below the target value at low frequency.

13.8 Simulating Impedance with SPICE

Simulating the impedance profile of different circuit models is essential in PDN design. Luckily most of the simple circuits that need to be analyzed can be simulated with free versions of SPICE that can be downloaded off the internet.

TIP The secret to using SPICE to perform impedance simulations is to build an impedance analyzer as a SPICE circuit. This is done with a single element in SPICE, a constant current AC current source.

This element is defined as a constant current sine wave source, outputting a sine wave of current with a constant amplitude. The output voltage of this element will be whatever it needs to be to always output a constant amplitude sine wave of current. The frequency of the current is set by the frequency of the frequency domain simulation. An example of a SPICE impedance analyzer is shown in Figure 13-14.

The amplitude of the constant current source is set to 1 A with a phase of 0. The voltage across the constant current source will depend on the impedance of whatever is connected across the leads. This voltage generated will be given by:

$$V = I(f) \times Z(f) = 1 \times Z(f) = Z(f) \tag{13-15}$$

where:

V = the voltage generated across the current source in volts

I(f) = the current from the source, a constant 1 A amplitude sine wave

Z(f) = the impedance of the device connected across the current source, in Ohms

We set the current amplitude to be exactly 1 A. This means the voltage generated across the current source is numerically the impedance in Ohms. The impedance of the circuit connected may vary with frequency. As the 1 A constant amplitude

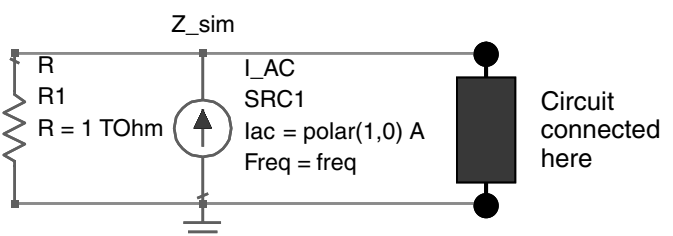


Figure 13-14 A SPICE impedance analyzer consisting of a constant current AC sine wave source.

sine wave flows through it, a voltage will be generated that is numerically equal to the impedance. The phase of the voltage will even track the phase of the impedance.

A large shunt resistor, in this case 1 TOhm, is connected across the current source. This is to keep SPICE from halting due to an error. SPICE wants to see a DC path to ground for all nodes. Without the resistor, an open across the constant current source could result in an infinite voltage, causing an error.

With this circuit the impedance of any circuit model can be simulated. It's actually the voltage across the current source that is simulated, but this is equal to the impedance of the circuit. The impedance of the two-capacitor model in the VRM was simulated using this SPICE impedance analyzer.

13.9 On-die Capacitance

The impedance at the highest frequency is established by the on-die decoupling capacitance. This arises from three general sources: the capacitance between the power and ground rail metallization, the gate capacitance from all the p and n junctions, and any added capacitance.

The largest component is from the gate capacitance distributed over the die. A typical CMOS circuit found by the millions on most chips, and for some chips by the billions, is shown in Figure 13-15. At any one time, one of the gates is on and the other is off.

This means that the gate capacitance of one of the gates, either the p channel or the n channel is connected between the power and ground rails on the die. The capacitance per area associated with the gate is simply approximated by:

$$\frac{C}{A} = \frac{8.85 \times 10^{-12} \frac{F}{m} \times Dk}{h} \tag{13-16}$$

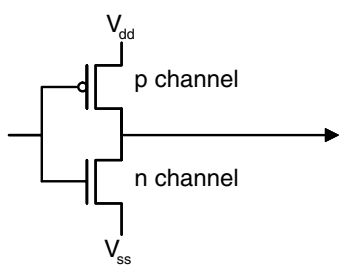


Figure 13-15 Typical CMOS circuit model for the transistors on a chip.

where:

C/A = the capacitance per area in F/m^2

Dk = dielectric constant of the oxide ~ 3.9 for SiO_2.

h = dielectric thickness in m

In general, the shorter the channel length, the thinner the gate oxide. As a rough rule of thumb, the gate oxide thickness is about 2 nm per 100 nm of channel length. However, below about 100 nm channel length, the scaling of h flattens out, due to higher leakage currents, but then the dielectric constant is increased with the use of "high Dk" gate insulator materials. This keeps the rule of thumb a good approximation even below 100 nm channel lengths.

For the 130 nm channel length node, the capacitance per area is about:

$$\frac{C}{A} = \frac{8.85 \times 10^{-12}\frac{F}{m} \times 3.9}{0.02 \times 130 \times 10^{-9}} = \frac{8.85 \times 10^{-12}\frac{F}{m} \times 3.9}{2.6 \times 10^{-9}} = 1.3\frac{uF}{cm^2} \tag{13-17}$$

Of course, not all the die is gate area. If we assume that 10% of the surface of the die is gate capacitance, then we see that as a rough rule of thumb the on-die decoupling capacitance on a 130 nm technology chip due to its p and n junctions is about:

$$\frac{C}{A} = 130\frac{nF}{cm^2} \tag{13-18}$$

As the technology node advances and the channel length decreases, the gate capacitance per area will increase, but the total gate area on a die will stay about the same. This means the capacitance per unit of die surface area will increase inversely with the technology node.

The capacitance of 65 nm chips is about 260 nF/cm^2. This estimate suggests that for a die that could be 2 cm × 2 cm, close to the largest mask size in volume production, at 65 nm channel length, the on-die decoupling capacitance could easily be in excess of 1,000 nF.

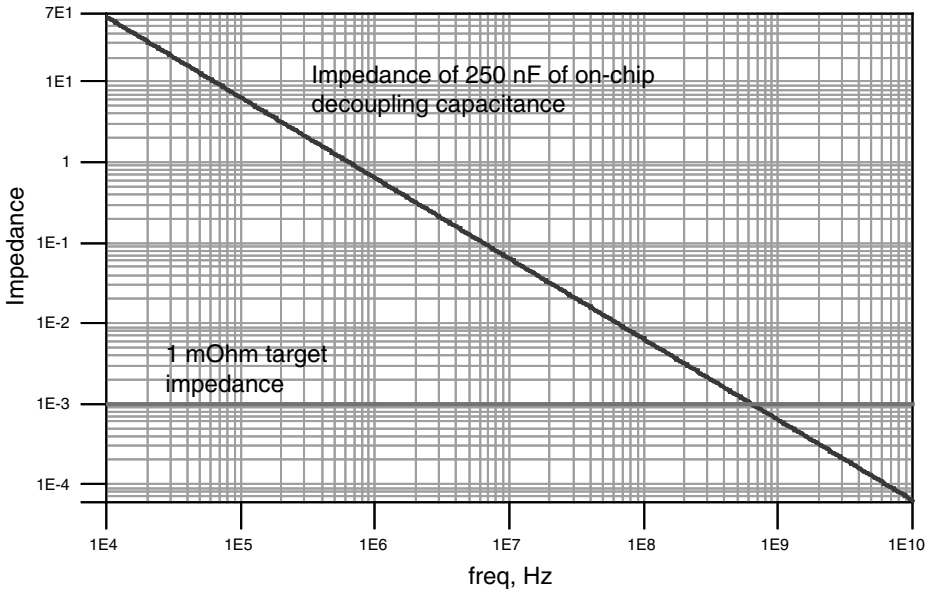


Figure 13-16 Impedance provided by 250 nF of on-chip decoupling capacitance, typical of a 65 nm, 1 cm on a side die.

A typical chip, only 1 cm × 1 cm, typical of many embedded processors, would have as much as 260 nF of capacitance. If the gate utilization on the die were larger, the on-die capacitance could be higher as well.

TIP At high frequency, it is the on-die capacitance that provides the low impedance.

The impedance profile of a capacitance that is 250 nF is shown in Figure 13-16. In this example, the on-die capacitance provides an impedance below 1 mOhm, at frequencies above 800 MHz. All high-frequency decoupling is provided by this mechanism.

If the target impedance were 10 mOhms, the on-die capacitance would provide significant decoupling for frequencies above about 100 MHz.

13.10 The Package Barrier

Between the pads on the chip and the pads on the circuit board is typically the IC package. Styles range from lead frame-based packages, to miniature circuit board-based packages to minimalist or chip-scale packages.

The loop inductance of the package leads in the power/ground distribution path is in series with the pads of the chip to the pads on the circuit board. This series inductance creates an impedance barrier. The impedance of an inductance is given by:

$$Z = 2\pi fL \tag{13-19}$$

where:

Z = the impedance, in Ohms

f = the frequency, in Hz

L = the inductance in H

For example, at 100 MHz, the impedance of a 0.1 nH inductor is about 0.06 Ohm. Even if the impedance of the PDN on the board was implemented as a dead short, the chip, looking through the package, would see a PDN impedance at 100 MHz of 0.06 Ohm. Of course, this is why on-die and on-package capacitance is so important.

Low-cost packages are often leaded, either as a stamped lead frame or as a two-layer printed circuit board. The loop inductance of adjacent leads is roughly about 20 nH/inch of length. For package leads 0.25 inches long, the loop inductance of a single power and ground lead pair can be as much as 5 nH. In the case of chip-scale packages, the loop inductance of a pair of leads may be on the order of 2 nH.

In multilayer BGA packages with at least four layers, a dedicated power and ground plane is often used. The loop inductance can be reduced to less than 1 nH per power and ground pair, limited by the roughly 50 mil total path length solder ball and its associated package via.

In small packages, there may be only a few power and ground pairs. In large BGA packages, there can be hundreds of pairs. This means the effective package lead inductance can vary from 1 nH to as low as 1 pH.

In addition to the package leads, there is also the loop inductance of the vias into the circuit board and the spreading inductance launching current into the power and ground planes of the board. When the package lead inductance is small, the board via and spreading inductance can limit the loop inductance as seen by the chip.

When the interactions of the on-die capacitance are added to the package inductance, the behavior is even more complicated. Figure 13-17 shows the

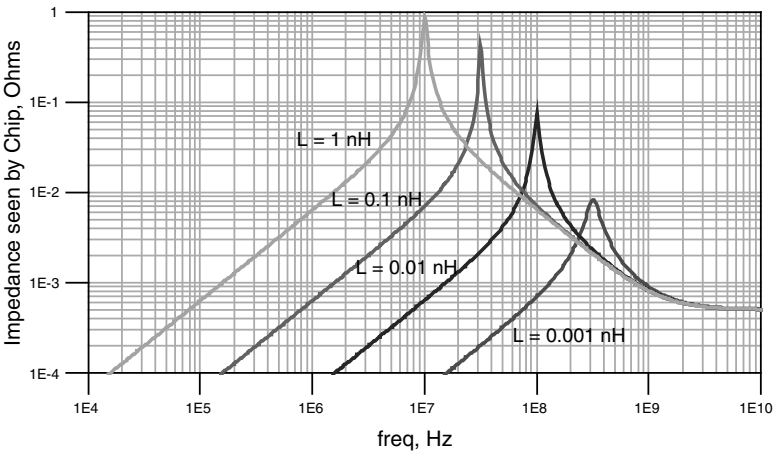


Figure 13-17 Impedance seen by the chip when the board is a dead short for different package lead inductances.

impedance profile the chip sees looking into a board that has a short for the PDN. The impedance profile is limited by the package inductance.

This suggests that no matter what the board level PDN does, it can never reduce the impedance the chip sees below the package lead impedance. When the package equivalent lead inductance is 0.1 nH, the board cannot influence the impedance the chip sees to below 10 mOhms at frequencies above 10 MHz.

Of course, in this example, there are large parallel resonance impedance peaks due to the interactions of the package inductance and on-die capacitance. Many times, these can be suppressed by using on-package decoupling capacitors.

For example, Figure 13-18 shows the reduction in peak impedance for the case of 0.1 nH of lead inductance with the addition of 10 different 700 nF capacitors, each with 50 pH of ESL mounted to the package.

TIP When establishing the design goals of the board level PDN, the high-frequency limit to where the board level impedance can be effective to the chip is set by the frequency at which the impedance from the combination of the package leads, board vias, and spreading inductance exceeds the target impedance.

This corresponds to:

$$Z_{target} < 2\pi\ L_{pkg}\ f_{max} \tag{13-20}$$

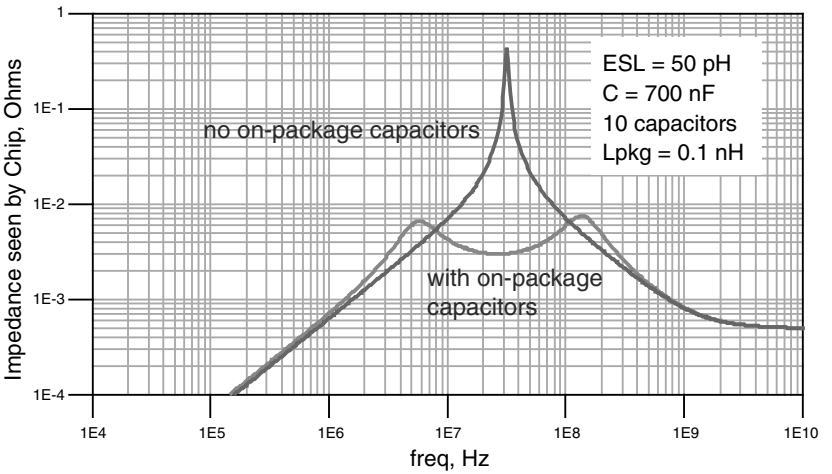


Figure 13-18 Suppression of package and on-die capacitance parallel resonances with on-package decoupling capacitors, as seen by the chip, if the board level impedance were a dead short.

where:

Z_{target} = the target impedance in Ohms

L_{pkg} = the equivalent lead inductance of all the PDN paths in the package

f_{max} = the highest useful frequency for the board level PDN

As a starting place, Figure 13-19 shows the map of the target impedance and package inductance for a specific maximum frequency of 100 MHz. If a product design falls below the line, for example, the target impedance is very low and the package lead inductance is very high, the maximum frequency for the board to be effective is below 100 MHz. In this case, the package severely limits the PDN performance.

If a design falls above the line, for example, the lead inductance is very low and the target impedance is high, the maximum frequency range for the board to still be effective is above 100 MHz.

As a rough rule of thumb, with about 20 nH/inch of loop inductance in a package lead and 0.05 inch of package lead length in a CSP package, the loop inductance per power and ground lead pair is about 1 nH. With 10 power and ground pin pairs in parallel, this is about 0.1 nH of equivalent lead inductance in a typical package that might have 10 pairs of PDN leads. If the target impedance were below about 0.6 Ohm, the board would not be effective much above 100 MHz.

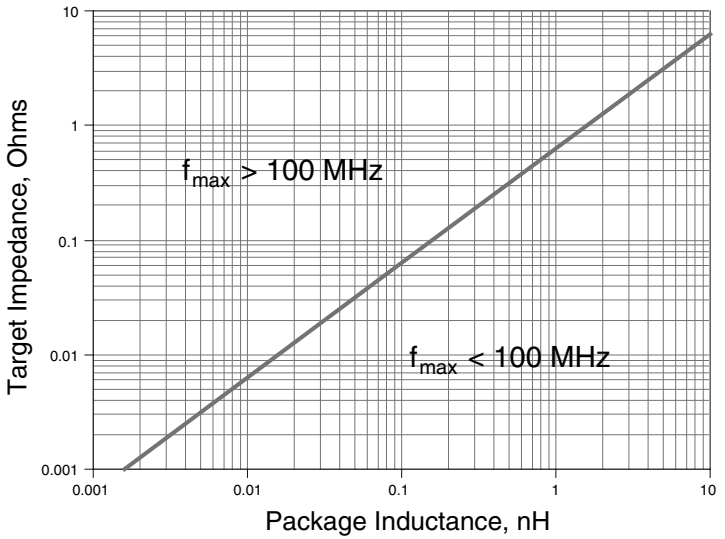


Figure 13-19 Map of the combination of target impedance and package lead inductance that has a maximum board level frequency limit of 100 MHz. If a design is above the line, the board level impedance will play a role at frequencies higher than 100 MHz, if the combination falls below the line, the board level impedance will play a role at less than 100 MHz.

Though it is difficult to generalize, sometimes an OK answer NOW is better than a good answer late. In general, the combination of packages and target impedance limits the board level impedance to be effective under about 100 MHz. This is why the board level PDN design goal is typically set to no higher than 100 MHz, unless there is other information to the contrary.

While it is possible to set the high-frequency limit higher, achieving higher frequency design limits is often much more expensive and should only be done when it is known to be important.

When on-package decoupling capacitors are provided, the maximum frequency the board level impedance can be effective is often lower than 100 MHz.

The lead inductance also acts as a filter to keep high-frequency noise from the chip's PDN off the board. When the core gates switch, the PDN rail voltage is kept low by the on-die capacitance. After all, if there is excessive noise on the chip's PDN pads, this will cause its own problem. Any voltage noise on the chip rails will be further filtered by the lead inductance before it gets to the board.

Figure 13-20 is an example of the simulated noise rejection from the chip pads to the board, for different package lead inductances and a board level impedance of 10 mOhms.

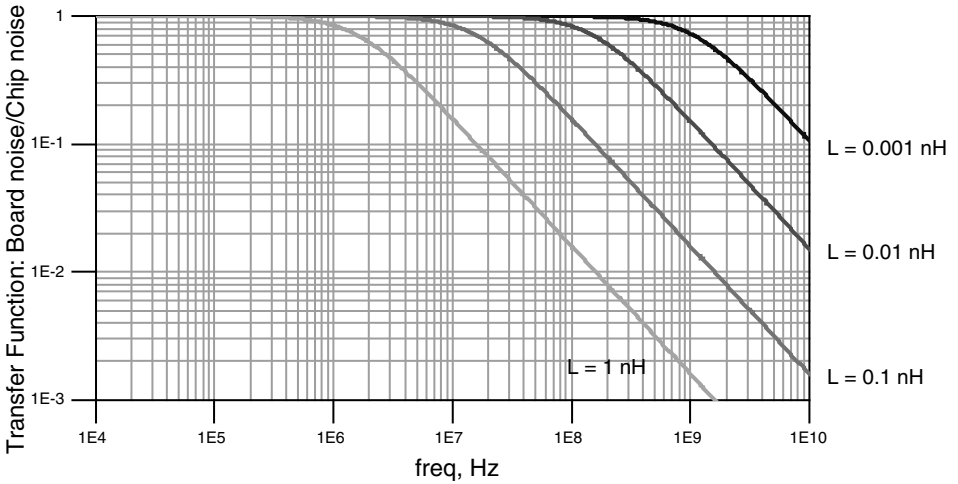


Figure 13-20 Relative noise injected onto the board from the chip pads for different package lead inductances. This is for the special case of the board level impedance at or below 10 mOhms.

When the target impedance is 10 mOhms, and the package lead inductance is 0.1 nH, the noise rejection is about 0.1 or –20 dB at 100 MHz. Less than 10% of the on-chip noise is coupled into the board. The higher the package lead inductance, the less on-chip noise gets on the board. This is why very little noise above about 100 MHz gets onto the board level PDN from the chip.

TIP In the absence of a complete package model including the PDN paths, it is difficult to do much more than roughly estimate the impact of the package on the PDN path.

13.11 The PDN with No Decoupling Capacitors

At low frequency, the VRM and the bulk decoupling capacitors provide the low impedance in the PDN. At high frequency, the on-die capacitance and on-package capacitance provide the low impedance to the PDN. We can see what the complete impedance profile might look like for this simple case using typical model parameter values.

Figure 13-21 is the simulated impedance profile for the case of power and ground planes in the board, with no added decoupling capacitors. It includes a simple VRM with bulk decoupling capacitor and 50 nF of on-die capacitance.

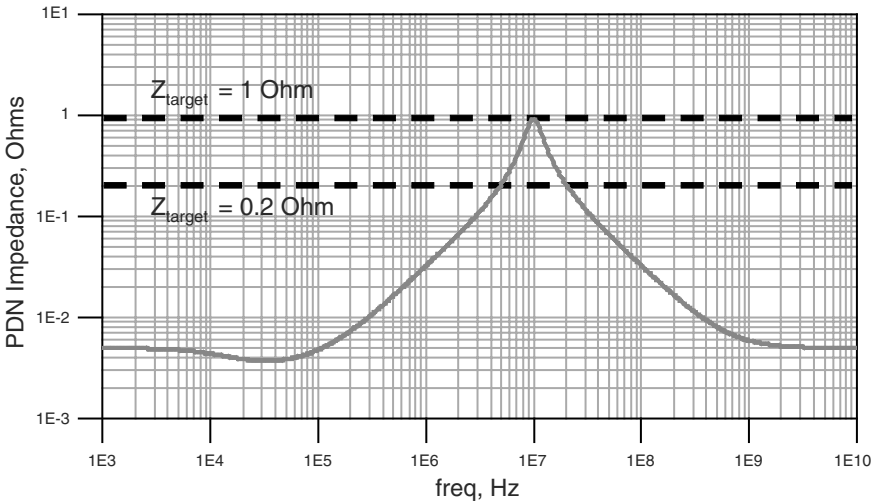


Figure 13-21 A typical impedance profile with just on-die capacitance and the VRM included.

If the target impedance were 1 Ohm, this board would work just fine, with no added decoupling capacitors. It would not matter how many or what value capacitors were added to the board, the PDN would still have acceptable noise. Even if the target value were as low as 0.2 Ohm, as long as the current spectrum did not have any worst case amplitude spikes in the 5 MHz to 20 MHz range, this board might work just fine.

This is why many boards work no matter what is done at the board level, because of the on-die capacitance and large bulk decoupling capacitors that are part of the VRM. This is also why it is sometimes reported that the decoupling capacitors were removed from the board and it worked just fine, thereby starting the myth that decoupling capacitors aren't so important.

It's just that there is no guarantee this condition will apply to your specific product application. Different chips with different current requirements and different on-die capacitance with different packages and the same board, can have very different performance.

TIP In order to have confidence in a PDN design, the board level designer must have information about the package model and the on-die capacitance, as well as the current spectrum of the chip.

While this information is important, it is also difficult to get from most semiconductor suppliers. We still have to design the board level decoupling in the absence of all the important information. In such cases, it's important to make some reasonable assumptions to base the board level design around.

TIP The two most common board level design assumptions are the package lead inductance will limit the frequency where the board level impedance is important to below 100 MHz, and the current draw and target impedance can be estimated based on the worst case power dissipation of the chips.

When the target impedance is 1 Ohm or above, the board design and decoupling capacitors may not play a very important role. However, to achieve target impedances below 1 Ohm requires careful selection of capacitors and their integration into boards to optimize their performance.

With the correct number, value, and implementation of decoupling capacitors and power and ground planes to connect them to the VRM and the package leads we can engineer the PDN impedance down below the mOhm range.

TIP Knowing the behavior of individual capacitors, combinations of capacitors, and how capacitors interact with planes will lay the foundation for the most cost-effective PDN designs.

13.12 The MLCC Capacitor

An ideal capacitor has an impedance that drops off inversely with increasing frequency, given by:

$$Z = \frac{1}{2\pi fC} \tag{13-21}$$

where:

Z = the impedance in Ohms

f = the frequency in Hz

C = the capacitance in F

For example, the impedance profile of four ideal capacitors is shown in Figure 13-22. It is easy to believe that if this is the behavior of a capacitor, then why

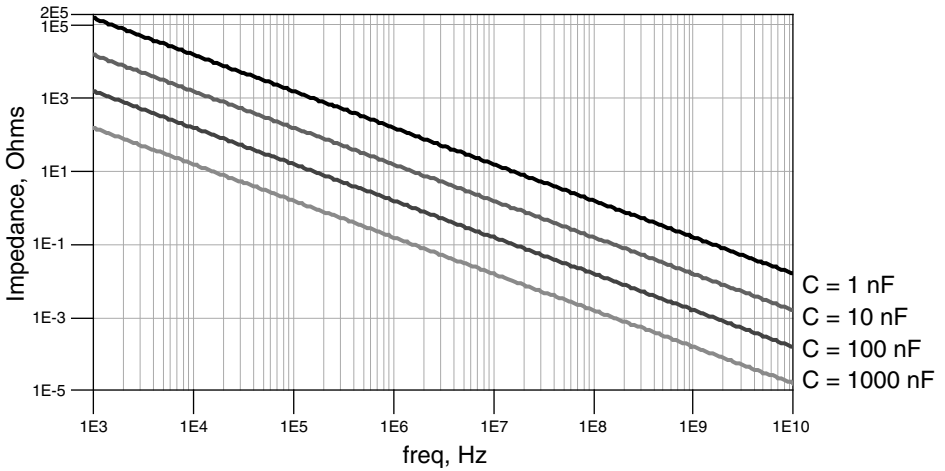


Figure 13-22 Impedance profile of ideal capacitors.

not just add a single, large capacitor to a board and use it to provide low impedance at ever higher frequencies?

The problem with this approach is that the behavior of a real capacitor is not quite the same as an ideal capacitor. An example of the measured impedance of a real 0603 capacitor is shown in Figure 13-23. While the impedance starts out like

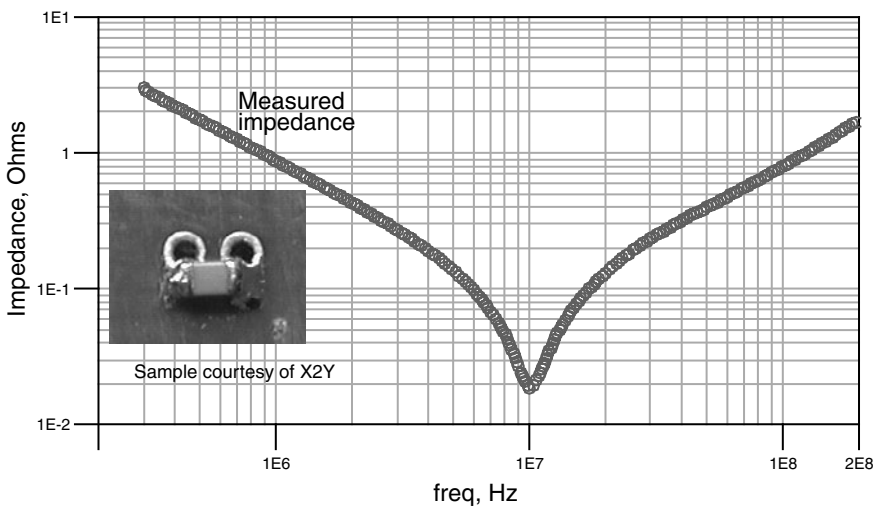


Figure 13-23 Measured impedance profile of an 0603 capacitor mounted to a test board.

an ideal capacitor, unlike an ideal capacitor, a real capacitor reaches a lowest impedance and then begins to increase in value.

A real capacitor can be approximated by a simple RLC circuit model to very high frequency. The simulated impedance of an ideal RLC circuit is an excellent match to this measured performance. Figure 13-24 shows the comparison of the measured and simulated impedance for the specific values of:

R = 0.017 Ohm

C = 180 nF

L = 1.3 nH

In this model, the R, L, and C parameter values are absolutely constant with frequency. They are each ideal elements. However, when connected together in series, the resulting impedance profile of the combination of ideal elements is remarkably close to the actual measured impedance of the capacitor.

TIP The fact that an ideal RLC circuit matches the behavior of a real capacitor makes this RLC circuit model incredibility useful for modeling real capacitors, even up to very high bandwidth, above 1 GHz.

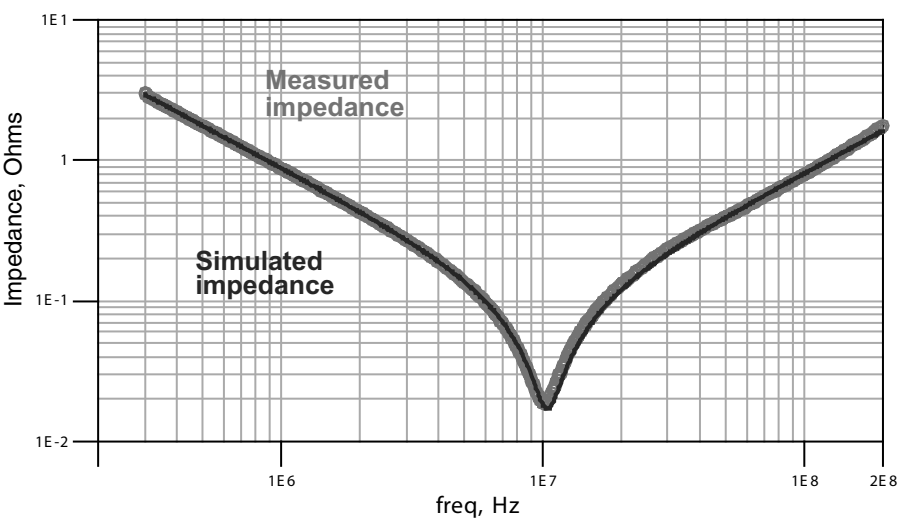


Figure 13-24 Comparing measured and simulated impedance of the 0603 MLCC capacitor.

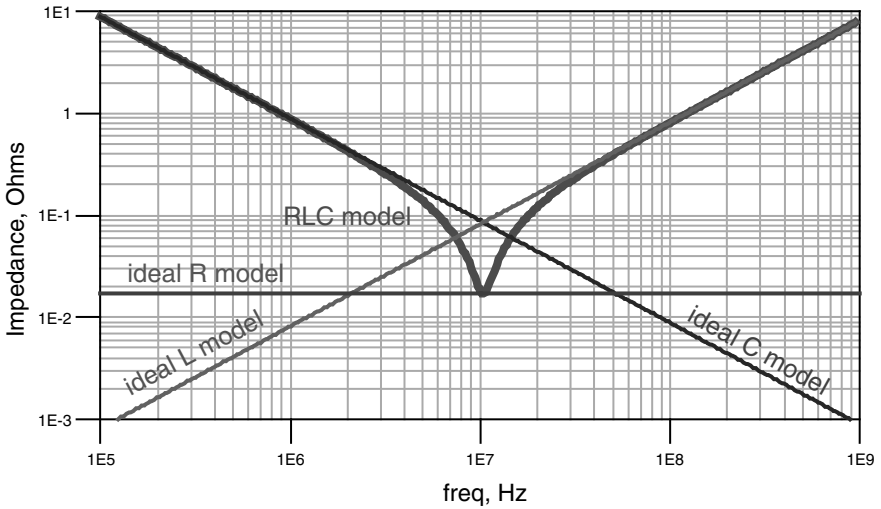


Figure 13-25 Impedance profile of the individual RLC elements that make up the RLC model.

The composite behavior of an RLC model is different than the behavior of any single element. These are compared in Figure 13-25.

At low frequency, the impedance of the RLC circuit is related to the ideal capacitance. At high frequency, the impedance of the RLC circuit is related to the ideal inductance. The lowest impedance of the RLC circuit is limited by the ideal resistance.

The frequency at which the impedance is the lowest is called the *self resonant frequency* (*SRF*) and is given by:

$$f_{SFR} = \frac{1}{2\pi}\frac{1}{\sqrt{L \times C}} = \frac{160\ \text{MHz}}{\sqrt{L \times C}} \qquad (13\text{-}22)$$

where:

f_{SRF} = the self resonant frequency in MHz

L = equivalent series inductance in nH

C = capacitance in nF

For example, for this real capacitor shown above, the self resonance frequency is estimated to be:

$$f_{SFR} = \frac{160\ \text{MHz}}{\sqrt{1.3\ \text{nH} \times 180\ \text{nF}}} = 10.4\ \text{MHz} \tag{13-23}$$

As can be seen in the example above, this is very close to the measured SRF of this capacitor.

Near the SRF, the impedance profile of the RLC circuit is not the same as the ideal L or C. It differs in a complicated way that also depends on the R value. This makes it difficult to perform simple analytical estimates, but can be easily simulated with any free version of SPICE, such as mentioned on www.beTheSignal.com.

TIP Above the SRF, the impedance is dominated by the inductance. Reducing the high-frequency impedance is about reducing the inductance. This is the most important engineering term to adjust in the selection of capacitors and their integration into the board.

The R is related to the series resistance of the metallization in the planes that make up the capacitors. The C is about the number of layers in the capacitor, the area of the internal planes, their separation, and dielectric constant.

13.13 The Equivalent Series Inductance

The L, often referred to as the *equivalent series inductance (ESL)*, is more about how the capacitor is mounted to the board or test fixture than the capacitor itself.

Even though many capacitor vendors offer an "intrinsic" inductance for their capacitor components, the inductance they provide is absolutely worthless and has no value in determining the performance of real capacitors. Instead, we will see how the ESL is affected by the mounting geometry of the capacitor.

Some capacitors are capable of achieving lower ESL for the same mounting features by nature of their design. It is not that they have lower intrinsic ESL but that they enable lower mounted inductance because of design features. For example, an X2Y capacitor, a type of interdigitated capacitor, will have a lower ESL under typical mounting conditions than an 0603. Figure 13-26 compares the measured impedance profile of the 0603 capacitor and an X2Y capacitor on the same board.

The impedance at low frequency for these two different capacitors is nearly the same, but their high-frequency impedance is very different. This is primarily due to the fact that the single X2Y capacitor with four terminals is really like four

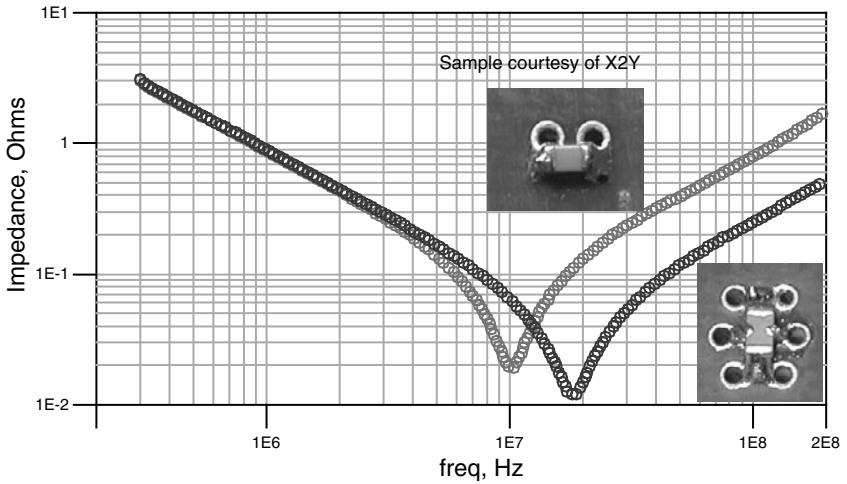


Figure 13-26 Measured impedance profiles of a conventional 0603 capacitor and an X2Y interdigitated capacitor on the same board. They have exactly the same value of capacitance, seen at low frequency, but very different ESL.

separate capacitors in parallel. The parallel combination of their loop inductances reduces the equivalent loop inductance of the whole capacitor. This can be a significant advantage in some designs.

The complete path of the power and return currents from the pads of the BGA package to the capacitor is shown in Figure 13-27. The ESL of the capacitor is, to first order, related to design features in this path.

The ESL associated with the capacitor and its path to the package can be divided into four regions:

1. The loop inductance of the surface traces and top of the plane's cavity
2. The loop inductance of the vias from the capacitor pads to the top of the plane cavity
3. The spreading inductance from the capacitor vias to the vias of the BGA
4. The loop inductance from the cavity under the package to the leads or solder balls of the package

TIP Different design techniques should be applied to each region in order to engineer the lowest ESL possible.

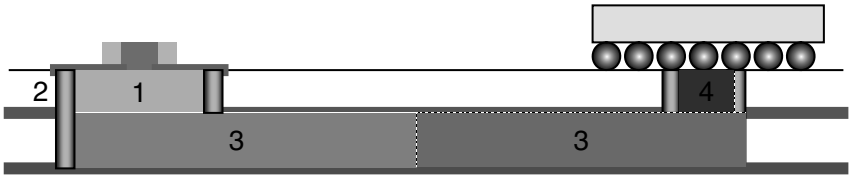


Figure 13-27 The ESL associated with a capacitor can be separated into four distinct regions.

When only a few capacitors are used on a board and the current distributions in the planes from the capacitors to the pins of the package do not substantially overlap, the ESL of each capacitor is the loop inductance of the entire path.

In this case, each capacitor behaves independently and it is possible to accurately simulate the impedance profile of combinations of the capacitors on the board using a simple SPICE model and simulation. The capacitors are independent.

However, when the current distributions overlap, such as when the capacitors are clustered in one region of the board, or many capacitors surround a package, the spreading inductance in the cavity of the power and ground cavity will be a complicated function of the location of the capacitors, their values, and the location of the package pins.

This is why it is useful to separate the ESL of a capacitor into the mounting inductance and the spreading inductance in the cavity. When the capacitors do not interact with each other, the cavity spreading inductance can be combined with the mounting inductance into the ESL. When the capacitors' spreading inductance interact, the only accurate way of estimating the impedance profile seen by the package is with a 3D simulator, which takes into account the current distribution of each capacitor. In this case, the location of the capacitors and the location of the power and ground pins in the package will be important.

TIP It's always a good practice to separate the mounting inductance and the cavity spreading inductance. They can be combined when needed into one number to estimate the ESL.

13.14 Approximating Loop Inductance

There are only a few geometries for which there are simple approximations for loop inductance. These are:

- Any uniform transmission line
- The special case of two round rods

- A pair of long, wide conductors with a thin dielectric between them
- The special case of edge to edge connections to planes
- Spreading inductance from a via to a distant ring
- Spreading inductance between two via contacts in a plane

The loop inductance of any uniform transmission line, assuming the signal and return paths are shorted at the far end, is given by:

$$L_{loop} = Z_0 \times TD = \frac{Z_0 \times Len}{v} \tag{13-24}$$

where:

L_{loop} = loop inductance nH

Z_0 = characteristic impedance in Ohms

TD = time delay of the transmission line in nsec

Len = length of the transmission line in inches

v = speed of light in the material, in inches/nsec

When the impedance of the line is 50 Ohms, such as a surface microstrip trace that is 10 mils wide and dielectric spacing in FR4 to the return path of 5 mils, the loop inductance is roughly:

$$L_{loop} = Z_0 \times TD = \frac{Z_0 \times Len}{v} = \frac{50 \times Len}{6} = 8.3\frac{nH}{in} \times Len \tag{13-25}$$

For a surface trace that is 0.2 inches long, the loop inductance of the surface trace can be as large as 1.7 nH.

This simple relationship suggests the two important design guidelines in order to engineer the lowest loop inductance possible for any structure that sort of looks like a uniform transmission line:

- Design as low a characteristic impedance as possible.
- Keep the lengths as short as possible.

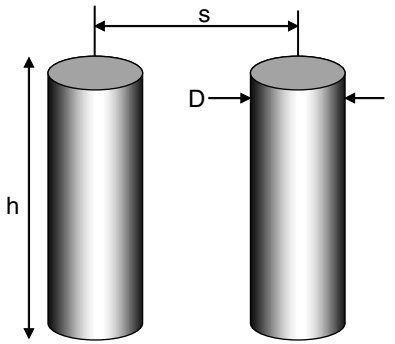


Figure 13-28 Geometry for two round rods, similar to two vias.

A special structure for which there is an analytical relationship between the geometry and loop inductance is two round rods, as illustrated in Figure 13-28.

The loop inductance from the end of one rod, down the rod, shorting across the end of the other rod and back again to the front is related to only the three geometry terms in Figure 13-28. If the length is increased, the loop inductance will increase. If the rods are brought closer together, their partial mutual inductance will help to cancel some of the total field lines and the loop inductance will be reduced. If the diameter of the rods is increased, the loop inductance will be decreased.

There are a number of analytical approximations for the loop inductance of these two rods. The simplest approximation is:

$$L_{loop} = 10 \times h \times \ln\left(\frac{2s}{D}\right) pH \tag{13-26}$$

where:

L_{loop} = the loop inductance in pH

h = the length of the rods in mils

s = center to center pitch of the rods in mils

D = the diameter of each rod in mils

For example, for 2 vias, 10 mils in diameter, on 50 mil centers, and 100 mils long, going through an entire board, the loop inductance is roughly:

$$L_{loop} = 10 \times 100 \times \ln\left(\frac{2 \times 50}{10D}\right) pH = 2300 \text{ pH} = 2.3 \text{ nH} \quad (13\text{-}27)$$

The uniform transmission line model gives the same constant loop inductance per length for the two rods, independent of the rod length. For the case of 10 mil via diameter and 50 mil centers, the loop inductance per length is roughly 23 nH/inch or 23 pH/mil. When the center to center pitch is 40 mils, typical of high density BGAs, the loop inductance per length is 21 nH/inch or 21 pH/mil.

TIP As a rough rule of thumb, if you want to carry around one value for the loop inductance of a pair of vias, a rough estimate is about 21 pH/mil. This is a reasonable estimate for the loop inductance contribution from vias.

When the two conductors that make up the loop are wide and closely spaced, such as with two plane segments shown in Figure 13-29, the loop inductance is approximated by:

$$L_{loop} = \left(32\frac{pH}{mil} \times h\right) \times \frac{Len}{w} pH \quad (13\text{-}28)$$

where:

L_{loop} = the loop inductance between the planes in pH

Len = the length of the planes, in inches

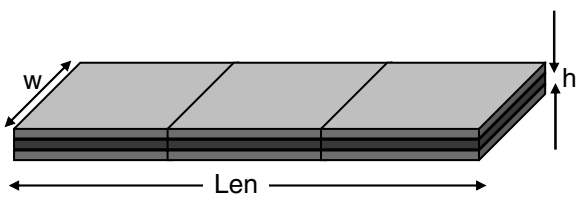


Figure 13-29 Geometry for the loop inductance of two plane segments.

w = the width of the planes, in inches

h = the thickness between the two planes in mils

For example, if the planes are 2 inches long and 0.5 inches wide, with 4 mils between them, the loop inductance would be:

$$L_{loop} = \left(32\frac{pH}{mil} \times 4\right) \times \frac{2}{0.5} = 512\ pH = 0.5\ nH \tag{13-29}$$

When the length of the trace is equal to the width, the structure looks like a square and the ratio of Len/w is always 1. The loop inductance of this square section of plane is the first part of the equation and is called the *loop inductance* per square, or the *sheet inductance*:

$$L_{square} = \left(32\frac{pH}{mil} \times h\right) \tag{13-30}$$

Any square piece of a pair of planes has the same loop inductance. The thinner the dielectric between them, the lower the sheet loop inductance.

This approximation assumes the currents flow in a uniform sheet down the top trace and back to the bottom, uniformly distributed along the both sheets. When the contacts are spread along the edge of the strip, this is a good approximation. However, in via contacts to planes the current does not flow uniformly. Instead, it spreads out from sources and constricts into sinks. An example of the current flow map in a plane between two via contacts is shown Figure 13-30.

The spreading inductance in planes is the most important property of planes and is discussed in detail in earlier chapters. It contributes to the additional loop inductance between point contacts in planes over their sheet inductance, when contacts are at vias, rather than at an edge of the plane.

The narrow contact regions of vias increase the current density and increase the local loop inductance. In general spreading inductance is complicated to calculate and usually requires a 3D field solver, as the current flow is difficult to calculate by approximation.

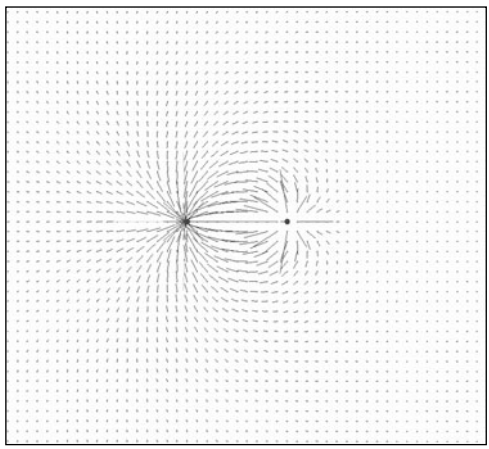


Figure 13-30 Current flow pattern in the top plane from a via source point to a via sink point into the bottom plane. Simulated with HyperLynx.

There is one special case for which there is an accurate approximation for spreading inductance. This is the case of current flowing from a central ring contact to an outer, symmetrical ring contact, where it flows into the bottom plane and then reverses back, constricting to an inner ring contact on the bottom plane. This is diagrammed in Figure 13-31.

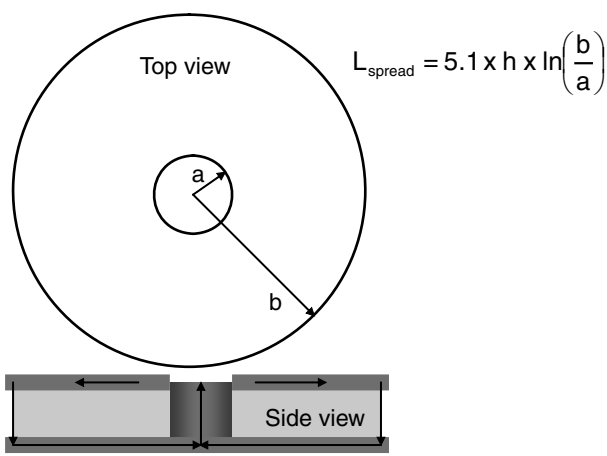


Figure 13-31 Inner and outer contact regions on the top plane, with similar regions on the bottom plane. Spreading inductance calculation is the loop inductance from the top contact point, radially outward to the edge, down the edge, and back in to the center contact.

In this geometry, the loop spreading inductance is:

$$L_{spread} = 5.1 \times h \times \ln\left(\frac{b}{a}\right) pH \tag{13-31}$$

where:

L_{spread} = the loop spreading inductance between the planes in pH

a = the radius of the inner contact region in inches

b = the radius of the outer contact region in inches

h = the thickness between the two planes in mils

This assumes the current is flowing from the center via contact to the bottom plane and returns to the inside edge of the clearance hole, and the clearance hole is just slightly larger than the via contact diameter in the bottom plane. For example, if the inner radius is 5 mils, corresponding to a 10 mil diameter via, and the outer radius is 1 inch, corresponding to the perimeter of a package, and the dielectric thickness between the planes is 10 mils, the loop spreading inductance is:

$$L_{spread} = 5.1 \times 10 \times \ln\left(\frac{1}{0.005}\right) = 270 \text{ pH} \tag{13-32}$$

This relationship of spreading inductance has the same form as the sheet loop inductance of a path, if we use as the number of squares:

$$n = \frac{1}{2\pi} \ln\left(\frac{b}{a}\right) \tag{13-33}$$

then,

$$L_{spread} = \left(32 \frac{pH}{mil} \times h\right) \times n \text{ pH} \tag{13-34}$$

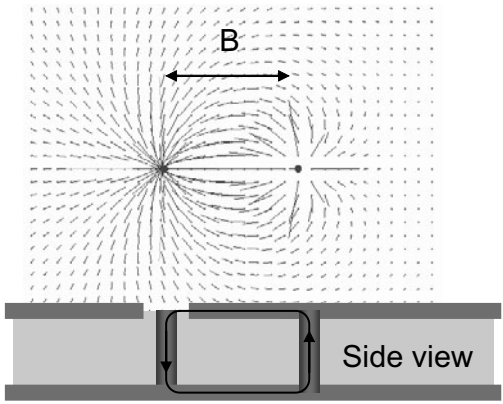


Figure 13-32 Spreading current from one via contact to another via contact in a pair of planes. There is spreading loop inductance between the two locations.

For typical cases, b/a could be on the order of 100, and the number of squares is of order 1.

In the special case of the current flow between the via contacts to a buried plane pair from a capacitor and BGA pins on the surface of a board, the loop inductance is more complicated to calculate. There are no exact analytical equations that describe this loop spreading inductance. However, by making a few assumptions, a simple approximation can be developed for the loop inductance in a pair of planes with round contact points.

Figure 13-32 illustrates the example of two via contacts positioned a distance B apart in a pair of planes and the current flow between them, spreading out and constricting in the planes.

The spreading loop inductance between these two via contacts is given by:

$$L_{via-via} = 21 \times h \times \ln\left(\frac{B}{D}\right) pH \qquad (13\text{-}35)$$

where:

$L_{via-via}$ = the loop spreading inductance in the planes between the two via contacts in pH

h = the dielectric thickness between the vias in mils

B = the distance between the via centers in mils

D = the diameter of the vias in mils

For example, if the via diameters are 10 mils and they are spaced 1 inch apart, in a pair of planes with h = 10 mils, the spreading inductance in the planes between the contacts is about:

$$L_{via-via} = 21 \times 10 \times \ln\left(\frac{1000}{10}\right) = 967 \text{ pH} \sim 1 \text{ nH} \quad (13\text{-}36)$$

The contribution of the spreading inductance in the planes between the vias can be as much as 1 nH. The thinner the dielectric, the lower the spreading inductance.

TIP The lower spreading inductance in ultra thin laminates is the real reason they offer a performance advantage over conventional FR4 for the dielectric between power and ground planes. The higher capacitance plays little role as there is far more capacitance in the on-die capacitance than in the power and ground planes.

If the connections between the capacitors and the pads of the package can be routed in planes with a cavity thickness that is not 4 mils but 1 mil or 0.5 mils, the spreading inductance in this path can be reduced from 0.4 nH with 4 mils down to 0.05 nH with half a mil thick dielectric. An example of the cross section of a board with half a mil thick dielectric in the power and ground planes is shown in Figure 13-33.

The predicted values of this approximation can be compared to the results predicted by a 3D field solver. Figure 13-34 shows the estimates of these approximations to the simulated via to via spreading inductance using the HyperLynx PI tool for two planes separated by 3 mils.

These various approximations can be used to roughly estimate the impact of physical design features and the resulting ESL of a capacitor mounted to a board. Using these approximations, we can explore design space to determine general design guidelines to follow.

TIP Since each design is custom, care must be taken when applying an observation for one case and blindly applying it to another without putting in the numbers.

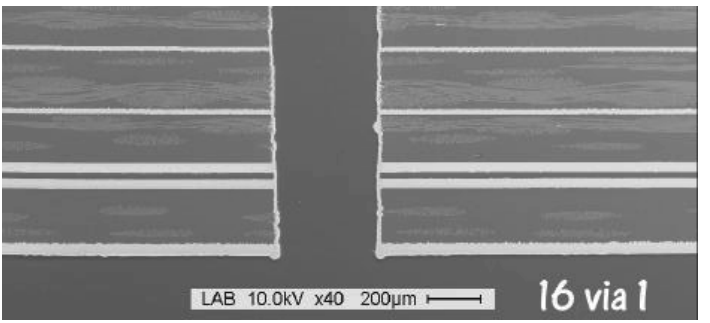


Figure 13-33 Cross section of a board with a 0.5-mil-thick layer of DuPont Interra HK04 laminate between the power and ground planes, close to the bottom surface of the board.

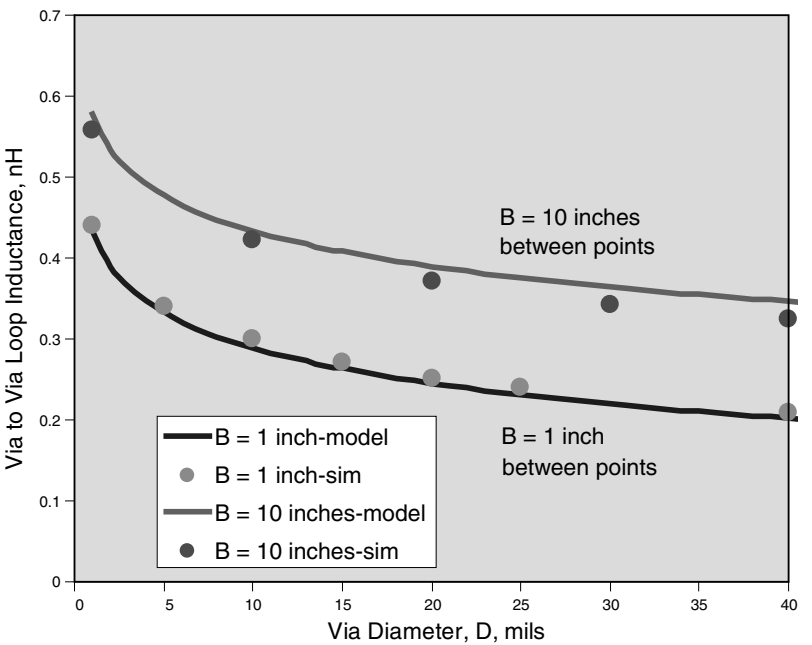
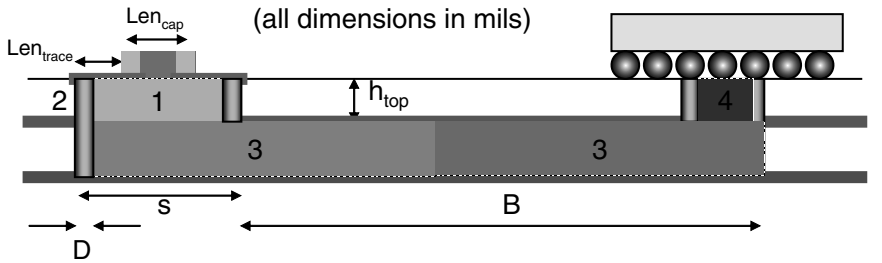


Figure 13-34 The comparison of the approximation (solid lines) and the simulated loop inductance (single dots) using HyperLynx for the case of a pair of planes separated by 3 mils.

13.15 Optimizing the Mounting of Capacitors

The three most useful approximations for loop inductance are summarized in one place in Figure 13-35.



- Capacitor trace inductance $$L_{trace} = 32 \times h_{top} \left(\frac{2 \times Len_{trace}}{w_{trace}} + \frac{Len_{cap}}{w_{cap}} \right) pH$$
- Via pair loop inductance $$L_{vias} = 10 \times h_{top} \times \ln\left(\frac{2s}{D}\right) pH$$
- Spreading inductance $$L_{spread} = 21 \times h_{planes} \times \ln\left(\frac{B}{D}\right) pH$$

Figure 13-35 Summary of the three approximations to estimate the ESL of a capacitor.

These approximations describe the important design trade-offs. If you want to reduce the loop inductance associated with the traces from the pads of the capacitor to the vias, there are three important design knobs to adjust:

- Keep the depth to the top of the power/ground cavity thin.
- Use wide surface traces.
- Keep the length of the surface traces short.

To reduce the inductance of the vias, there are three design knobs to adjust:

- Keep the depth to the top of the power/ground cavity thin.
- Use large diameter vias.
- Keep the via pitch as close as possible.

To reduce the spreading loop inductance in the planes there are three knobs to adjust:

- Keep the dielectric thickness of the power/ground cavity thin.

- Use large diameter vias.
- Place the capacitor close to the package it is decoupling.

While these are important design guidelines to be aware of, some are more important than others.

TIP The terms that affect the total loop inductance the most should always be optimized first.

They are:

- Keep the depth to the top of the power/ground cavity thin.
- Keep the dielectric thickness of the power/ground cavity thin.
- Use wide surface traces.
- Keep the length of the surface traces short.

The other design features are of second and third order importance and can sometimes be a distraction from the first order concerns. In general, the only way to know what is important is to put in the numbers for specific cases. Integrating these approximations in a spreadsheet allows us to easily explore design space and identify what is really important and what is not.

In the first example, shown in Figure 13-36, three cases are explored. In each case an 0603 capacitor is supplying current to one power and ground pin pair in a BGA package, located some distance away. The vias are 13 mil in diameter. This estimate is for the ESL of the capacitor as though it were not interacting with other capacitors. Case 1 is the starting place, with long and narrow surface traces. The total ESL is found to be about 6.1 nH.

In case 2, the surface traces are shortened and widened. The resulting ESL is 3.7 nH. Finally, in case 3, the capacitor is moved closer to the package and the cavity thickness is decreased. The resulting loop inductance is reduced to 1.8 nH.

TIP This example clearly shows that in typical cases, the loop inductance of the vias is negligible. In most typical cases, especially with thick spacing between the planes, the spreading inductance can be as significant as the surface trace inductance. By careful design of the stack-up, it is possible to routinely achieve less than 2 nH loop inductance.

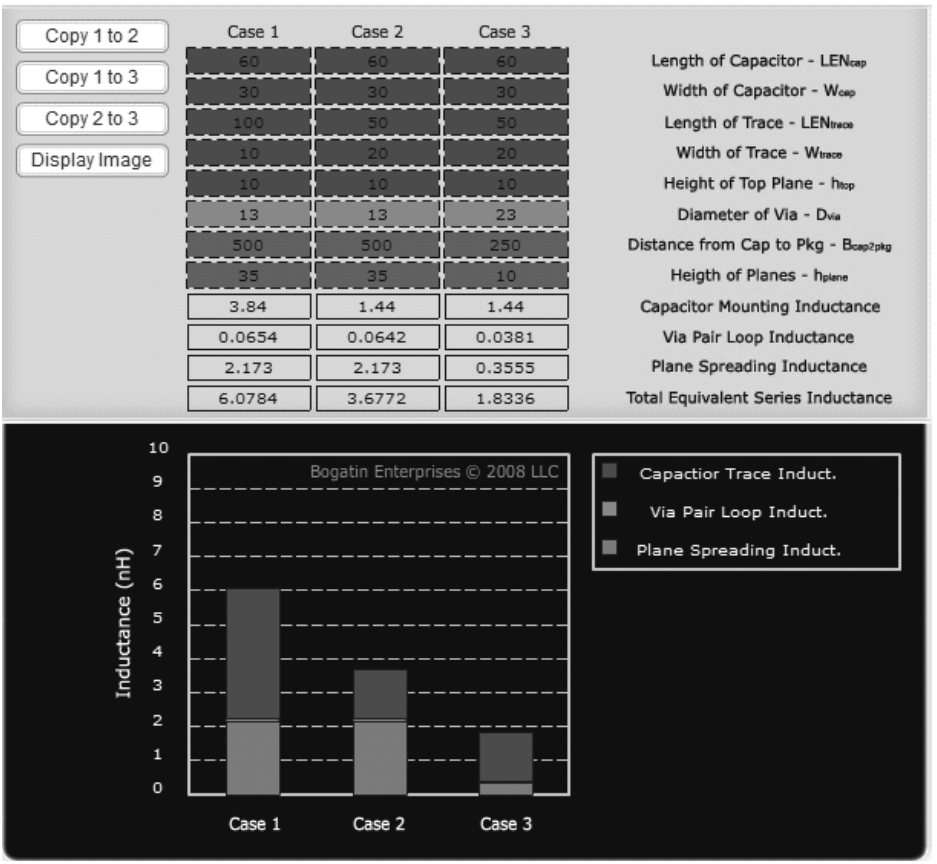


Figure 13-36 Analysis of three typical mounting geometries for an 0603 capacitor, analyzed with an online tool at www.beTheSignal.com.

Surprisingly, the board stack-up plays a significant role in the ESL of the capacitor, in two respects. By moving the top of the cavity closer to the capacitor, the loop inductance of the capacitor and the surface traces is reduced. By making the dielectric thickness of the cavity between the power and ground planes thinner, the spreading inductance is reduced. Adjusting these two design features can bring the ESL from 6 nH to 1 nH in some cases.

Both of these design features are first order and linear in the thickness. Changing the via diameter and moving the capacitor closer to the BGA are log-dependent factors and of only slight, second or third order importance.

If the surface trace length is also reduced and widened, ESL values as low as 0.5 nH can be achieved. An example of three similar cases is shown in Figure 13-37.

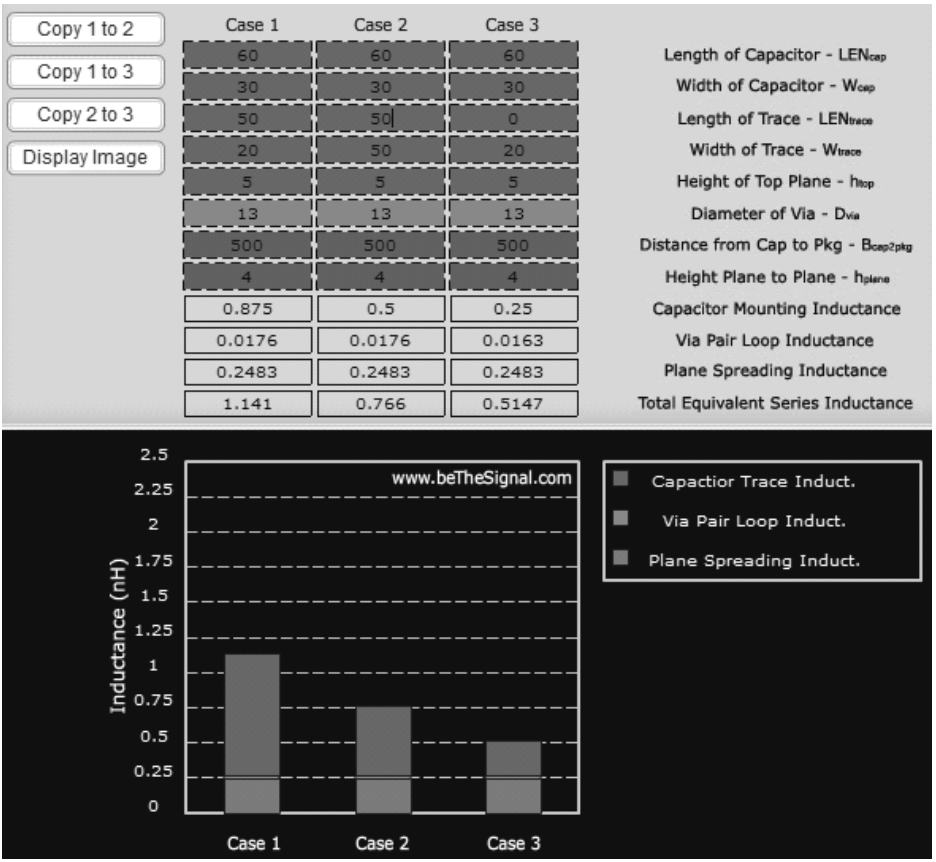


Figure 13-37 Three examples of thin cavity, close to the surface with three different surface traces resulting in an ESL as low as 0.5 nH, analyzed with an online tool at www.beTheSignal.com.

This model can also be used to assess important design questions such as is it better to add capacitors under the BGA or on the same surface as the BGA? Figure 13-38 illustrates the two options.

Of course, the most common answer to all signal integrity questions is "it depends," and the only way to answer an "it depends" question is by putting in the numbers.

The right place to put the capacitor is where it will have the lowest loop inductance. Clearly, if the total board thickness is thin, the via loop inductance will be low. If the cavity is far from the surface, and thick, the loop inductance of the capacitor on the top will be high. It is possible to find a combination where the

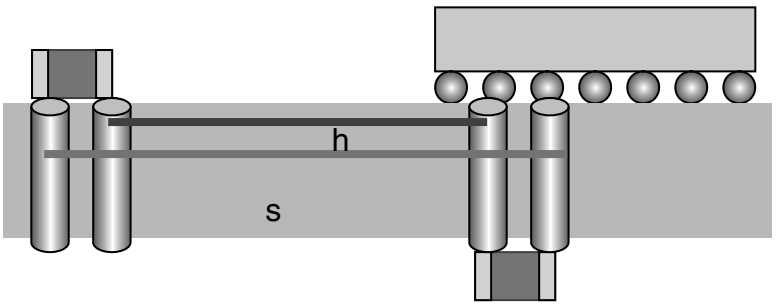


Figure 13-38 Where should the capacitor go: on the same surface as the BGA or directly under the BGA?

top surface capacitor has a much higher loop inductance than the bottom surface capacitor.

However, if the board is thick and the cavity is near the top surface and the cavity is thin, the capacitor on the bottom will have the higher loop inductance. Figure 13-39 summarizes three cases. It shows that placing the capacitor on the bottom can have a loop inductance on the order of 2 nH.

If it is possible to achieve lower loop inductance by placing capacitors on the top surface; this is preferred, but as a general rule, if there is the option of doing both, both locations should be used, especially when many capacitors are used in low impedance applications. When many capacitors are placed around the periphery of the package, their currents can overlap and the cavity spreading inductance can increase. Placing some of the capacitors under the BGA minimizes the increase in cavity spreading inductance.

TIP The combination of short, wide surface traces, or via in pad technologies, and thin dielectric between the power and ground planes close to the surface, can result in typical ESL values from 0.5 to 2 nH. By going to extreme efforts and utilizing interdigitated capacitors, it is possible to achieve loop inductances below 0.5 nH.

If the capacitor mounting inductance is known, based on the design constraints, it will be possible to predict the impedance profile of a collection of capacitors using a 3D field solver. If the mounting inductance changes, as from a stack-up change or a surface mounting design change, the loop inductance will change and the impedance profile of the collection of capacitors will change. This is why every PDN design is custom.

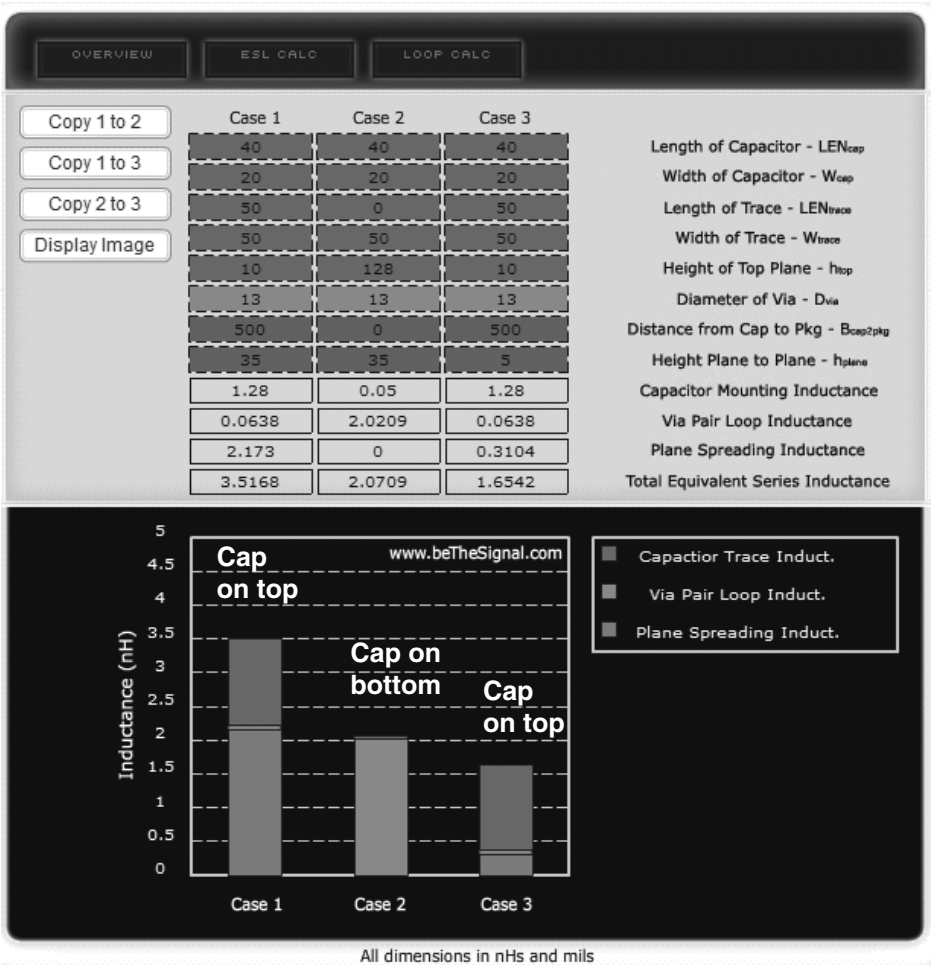


Figure 13-39 Analysis of capacitors on the top and on the bottom of the board. Analyzed with an online tool at www.beTheSignal.com.

TIP The PDN impedance profile of the combination of capacitors depends very much on the details of the board stack-up, capacitor mounting geometry, and location on the board.

13.16 Combining Capacitors in Parallel

The strategy to engineer the PDN impedance profile is to select the right number and value of capacitors to keep the peak impedance below the target value from where the VRM and bulk capacitors no longer provide low impedance, up to about 100 MHz.

When multiple, identical capacitors are connected in parallel, the resulting impedance matches the behavior of an RLC circuit, but the circuit elements values are different.

The equivalent R, L, and C of n capacitors in parallel are:

$$C_n = nC \tag{13-37}$$

$$ESR_n = \frac{1}{n}ESR \tag{13-38}$$

$$ESL_n = \frac{1}{n}ESL \tag{13-39}$$

where:

C_n = the equivalent capacitance of n identical real capacitors in parallel

C = the capacitance of each individual capacitor

n = the number of identical capacitors in parallel

ESR_n = the equivalent series resistance of n identical real capacitors in parallel

ESR = the equivalent series resistance of each individual capacitor

ESL_n = the equivalent series inductance of n identical real capacitors in parallel

ESL = the equivalent series inductance of each individual capacitor

Figure 13-40 is an example of the impedance profile of multiple identical capacitors in parallel, showing the same general RLC profile, but with lower impedance at all frequencies. We are approximating the problem by assuming the capacitors are independent and their currents do not overlap. The SRF stays the same; it's the entire impedance profile that scales lower. This is one way of decreasing the impedance profile of a capacitor: Add more of them in parallel.

However, if the two capacitors have a different value of capacitance or ESL, when they are added in parallel, the behavior is not so simple. Figure 13-41 shows the impedance profile of two different capacitors, with the same ESL and the same ESR.

The behavior of the two capacitors in parallel has the same low impedance dips at the self resonant frequencies of the individual RLC models. The larger

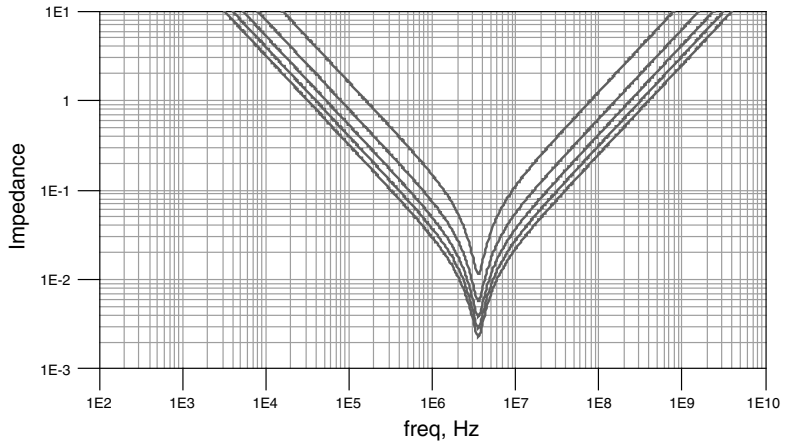


Figure 13-40 Impedance profile of 1, 2, 3, 4, and 5 identical capacitors added in parallel. With each additional capacitor, the impedance decreases at all frequencies.

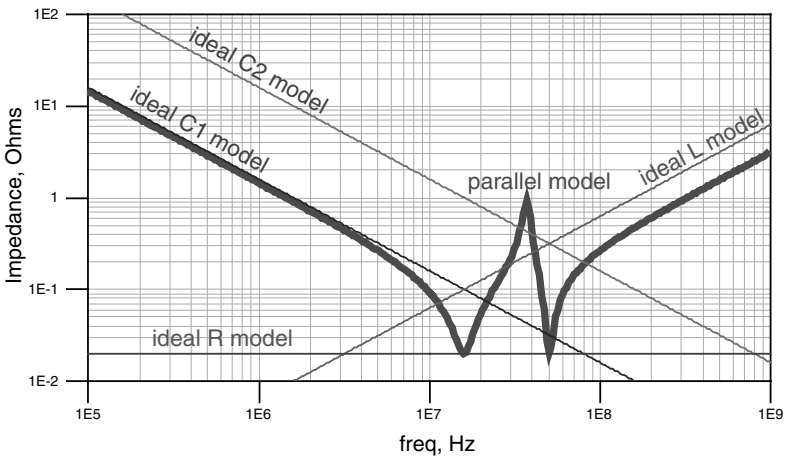


Figure 13-41 The impedance profile of two RLC circuits in parallel, with the same R and L values but different C values. Superimposed is the impedance of the two ideal capacitors and the ideal L and ideal R of both capacitors.

capacitor has the lower SRF. The smaller capacitor has the higher SRF. They each occur when the impedance of the ideal capacitor matches the impedance of the ideal inductance associated with each capacitor. The SRF seen in the parallel combination of capacitors is the same as each individual capacitor's.

In addition, there is a new feature between the self resonant frequencies, a peak in the impedance, called the *parallel resonant peak*, which occurs at the *parallel resonant frequency (PRF)*.

The value of the PRF is difficult to calculate accurately, as it depends on the ESL of the larger capacitor, the C of the smaller capacitor, and the ESR of both of them. If the SRF values are far apart, the PRF is roughly related to:

$$\mathrm{PRF} \approx \frac{1}{2\pi} \frac{1}{\sqrt{C_2 \times \mathrm{ESL}_1}} = \frac{160\ \mathrm{MHz}}{\sqrt{C_2 \times \mathrm{ESL}_1}} \tag{13-40}$$

where:

PRF = the parallel resonant frequency, in MHz

C_2 = the capacitance of the smaller capacitor in nF

ESL_1 = the equivalent series inductance of the first capacitor, in nH

For example, if ESL_1 = 2 nH and C_2 = 10 nF, then the PRF is:

$$\mathrm{PRF} \approx \frac{160\ \mathrm{MHz}}{\sqrt{10 \times 2}} = 36\ \mathrm{MHz} \tag{13-41}$$

However, when the SRF values are within a factor of 10 of each other, the impedance profile of the parallel combination is distorted from the impedance of the ideal L, and the PRF is a more complicated function of the elements, it can more easily be found from a SPICE simulation.

TIP The PRF is one of the most important features of parallel combinations of capacitors because it denotes where there are peaks in the impedance. When few numbers of capacitors are used, it's the parallel resonant impedance that always sets the limit to the PDN performance and must be engineered to lower values.

The peak impedance at the PRF is roughly related to:

$$Z_{\mathrm{peak}} \sim \frac{L_1}{C_2}\left(\frac{1}{R_1 + R_2}\right) \tag{13-42}$$

where:

Z_{peak} = the peak impedance at the PRF in Ohms

L_1 = the equivalent series inductance of the larger capacitor

C_2 = the capacitance of the smaller capacitor

R_1 = the equivalent series resistance of the larger capacitor

R_2 = the equivalent series resistance of the smaller capacitor

This is only approximate and is less accurate as the SRF of each capacitor is brought closer together. However, it points out the important ways of engineering a reduction in the peak impedance:

- Reduce the ESL of the larger capacitor.
- Increase the capacitance of the smaller capacitor.
- Increase the ESR of both capacitors.

TIP Unfortunately, rarely is there a low-cost option to purchase capacitors with larger ESR. This is not a cost-effective option. However, it is possible to select capacitors not only for their capacitance value but also for their ESR value.

The ESR of a capacitor is related to the structure of the parallel plates that make it up and the metallization between the layers. In general, the higher the capacitance, the more plates in parallel and the lower the ESR. Looking at the specifications of a variety of 0402 capacitors can provide a simple generalization for the series resistance of capacitors by capacitor value. Figure 13-42 shows the plotted ESR for various capacitor values, taken off the AVX data sheets.

From the specified ESR it is possible to derive a simple empirical relationship between the ESR and the capacitance of a capacitor. It is given by:

$$ESR \approx \frac{180\ m\Omega}{2.5^{\log(C)}} \tag{13-43}$$

where:

ESR = the equivalent series resistance of the capacitor in mOhms

C = the capacitance of the capacitor in nF

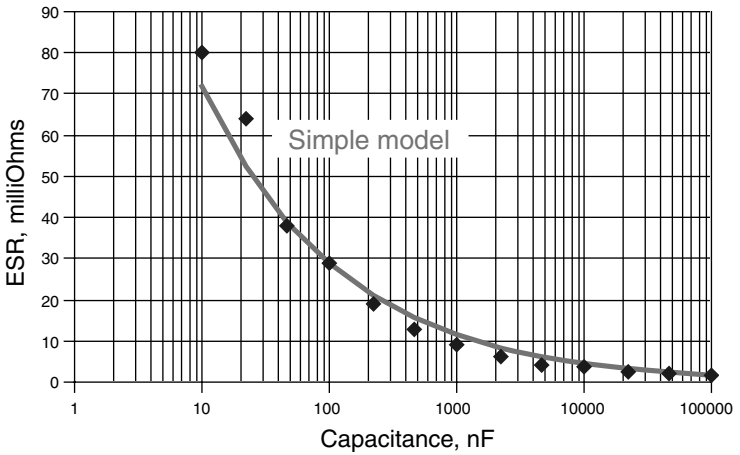


Figure 13-42 ESR and capacitance for 0402 capacitors, taken from AVX data sheets of capacitors.

This simple model is compared with the specified values of ESR in the figure above, and the agreement is seen to be very good.

This suggests that it may be possible to select for higher ESR and lower parallel resonant peak heights if smaller value capacitors are used. This is especially true when one of the capacitors is the power and ground cavity's capacitance.

Another important design feature to engineer to decrease the peak impedance value is decreasing the ESL of the larger capacitor or increasing the capacitance of the smaller capacitor. Figure 13-43 shows the impact on the peak impedance as the ESL of the larger capacitor is changed, from 10 nH down to 0.1 nH.

In this example, the larger capacitor is 100 nF and the smaller one is 10 nF with an ESL of 3 nH. As the ESL of the larger capacitor is reduced from 10 nH, the peak impedance at the PRF decreases until the SRF of the larger capacitor matches the SRF of the smaller capacitor, in which case there is no peak.

TIP Reducing the ESL is a significant method of reducing peak impedances.

Unfortunately, due to the complex interactions of the circuit elements, it is not possible to do a simple and accurate analytical analysis of the features of the impedance profile of multiple capacitors. This is especially true as more capacitors are added. Instead, it is critical to use SPICE for this analysis. Luckily there are many free versions of SPICE readily available from the internet that can routinely perform this sort of analysis. For examples of these tools, visit beTheSignal.com.

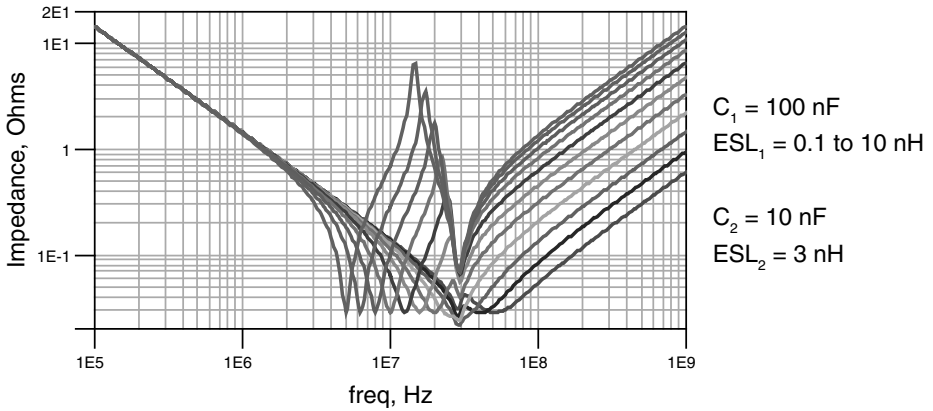


Figure 13-43 Impedance profile of a 100 nF and a 10 nF capacitance with ESL of 3 nH in parallel while changing the ESL of the larger capacitor from 10 nH down to 0.1 nH. As the ESL is reduced, the peak impedance drops.

In addition to reducing the ESL of the capacitors to bring the peak impedance values down, there is another way of reducing the peak impedances: Add more capacitors. These can be identical capacitors, or they can be different value capacitors. Both approaches can work.

13.17 Engineering a Reduced Parallel Resonant Peak by Adding More Capacitors

When two capacitors with different SRFs are added in parallel, they create a parallel resonant peak impedance between their self resonant dips. The peak impedance can be reduced by adding a third capacitor with an SRF between them. What is the optimum value of the SRF of the third capacitor?

The optimum SRF of the third capacitor to give the lowest peak impedance depends on the capacitance values, the ESL values, and the ESR values of all three capacitors. It is difficult to evaluate without a SPICE simulation. There are two obvious algorithms to choose from: Select the third capacitor so that its SRF matches the PRF, or select its SRF so that it is midway between the two other capacitors' SRFs.

The choice depends on the ESL values of the capacitors, their ESR values, and how far apart are the capacitances. Consider the simple case of two capacitors, a 10 nF and a 100 nF, each with the same ESL of 3 nH. When combined in parallel, they have a PRF at 21 MHz.

Option 1 is to add a capacitor with its SRF at 21 MHz. The capacitance value is given by:

$$C_3 = \left(\frac{160}{f_{peak}}\right)^2 \frac{1}{ESL} = \left(\frac{160}{21}\right)^2 \frac{1}{3} = 19.3 \text{ nF} \quad (13\text{-}44)$$

where:

C_3 = the capacitance of the third capacitor to be added, in nF

ESL = 3 nH, assumed to be the same for each of the three capacitors

21 = the SRF required, to match the PRF, in MHz

Option 2 is to select the value of the third capacitor so its SRF is midway (on a log scale) between the SRFs of the other two capacitors. Since their ESL values are the same, this translates to a capacitance of the third capacitor that is the geometric mean of the other two:

$$C_3 = \sqrt{C_1 C_2} = \sqrt{100 \times 10} = 33 \text{ nF} \quad (13\text{-}45)$$

Figure 13-44 shows the resulting simulation of the original combination of two capacitors and the impedance profile of the three capacitors, with the value of the third capacitor chosen based on the two options.

This example illustrates that when the ESL of each capacitor is the same, the lowest peak impedance can be engineered by choosing a third capacitor that is the geometric mean of the other two capacitors. This is why it is often recommended to select capacitor values spread over a decade scale. If the values are distributed uniformly in a log scale, they will provide the lowest peak impedance.

The only way to know the optimum capacitor values that result in the lowest peak impedances, given the ESL and ESR values, is with a SPICE simulation. When the SRF values for the two capacitors are far apart, using a third capacitor with an SRF near the PRF may be a better choice.

TIP Whenever different value capacitors are brought together in parallel, there will always be parallel resonant peaks that must be managed. This will occur at low frequency with the bulk capacitors, and at high frequency with the capacitance of the planes and the on-die capacitance.

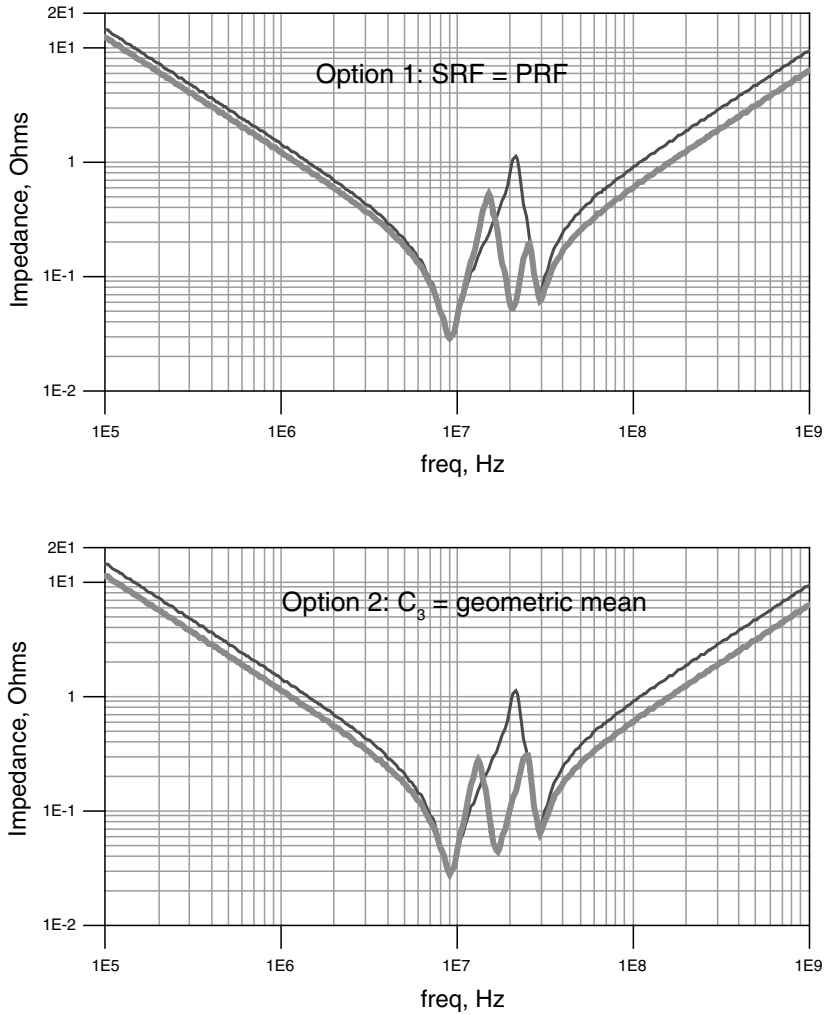


Figure 13-44 Comparing the two options for selecting the third capacitor value. The approach that gives the lowest peak impedance is option 2: Choose the geometric mean.

13.18 Selecting Capacitor Values

Many application notes recommend that all you need to do is add three capacitors per power and ground pin in the package. Half of them recommend using all three capacitors the same value, and the other half of them recommend using different capacitor values. Which is right? The only way to tell is to put in the numbers.

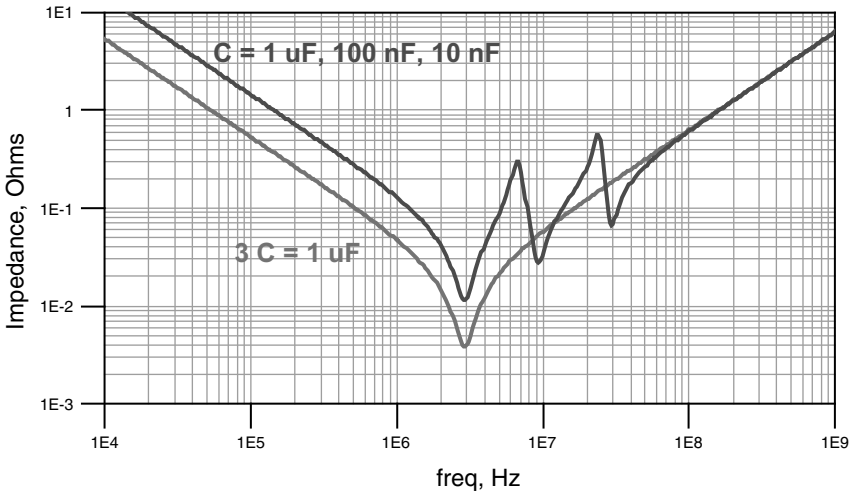


Figure 13-45 Impedance profile for the case of three capacitors, each with a different value, and each with the same value.

Figure 13-45 compares the impedance profile for three capacitors, each with the same value of 1 uF and three capacitors with values of 1 uF, 0.1 uF, and 0.01 uF. In each case, the ESL is the same 3 nH, and the ESR is chosen based on the specified values from the manufacturer.

At first glance, the conclusion would be that the three capacitors all with the same value give the lowest impedance, but, at 100 MHz, the impedance limit they both set is the same at about 0.6 Ohm. However, this analysis neglects the two important effects: at the low frequency end, the interaction with the VRM and bulk capacitor, and at the high frequency end, the interactions with the planes of the board.

The resulting impedance profile for these two sets of capacitors, a VRM with a bulk decoupling capacitor and the planes for a board 5 inches on a side and 4 mils thick is shown in Figure 13-46.

TIP At the low frequency, the interactions of the bulk capacitor and the small ceramic capacitors cause the impedance peak at about 1 MHz. The peak impedance is primarily due to the inductance of the bulk capacitor and the capacitance of the MLCC capacitors.

The way to drop this down is to reduce the inductance of the bulk capacitors. In this example, it was assumed to be 15 nH, typical of an electrolytic capacitor. If this can't be reduced by design, one way it can be dropped is by adding more

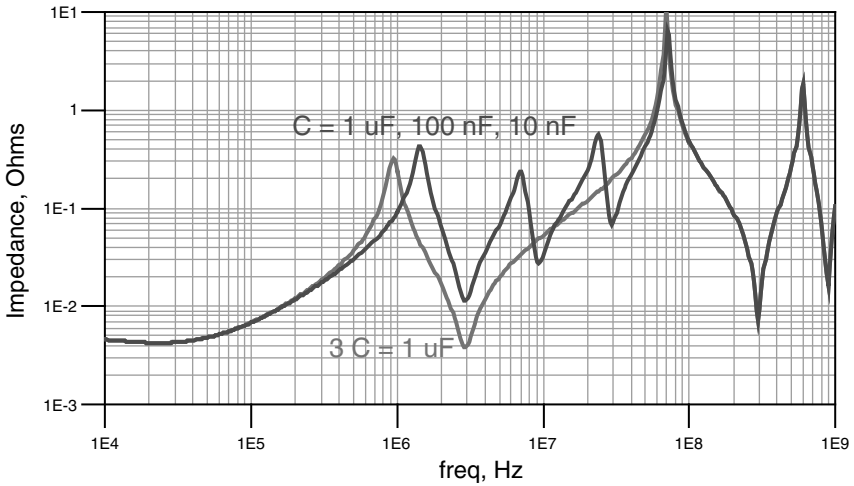


Figure 13-46 The parallel resonances at the boundaries cause peak impedances in both combinations of capacitors.

capacitors in parallel. As long as their SRF is lower than the peak impedance at roughly 1 MHz, their ESL in parallel with the electrolytic capacitor will reduce the peak impedance.

The minimum capacitance needed can be found from a simple estimate. If we assume we use a tantalum capacitor, with an ESL on the order of 5 nH, in order to have a SRF < 1 MHz, the condition is:

$$C_3 = \left(\frac{160}{f_{PRF}}\right)^2 \frac{1}{ESL} = \left(\frac{160}{1}\right)^2 \frac{1}{5} \sim 5 \text{ uF} \tag{13-46}$$

By adding an additional capacitor of more than 5 uF and less than 5 nH ESL, the peak impedance at low frequency can be reduced. The precise value of the capacitance is not important; the ESL value is.

The new impedance profile with a 10 uF tantalum capacitor added is shown in Figure 13-47.

The high-frequency impedance peak is created by the interactions of the capacitance of the planes and the ceramic capacitors. The capacitance in a pair of planes is:

$$C_{planes} = 0.225 \times Dk \frac{A}{h} \tag{13-47}$$

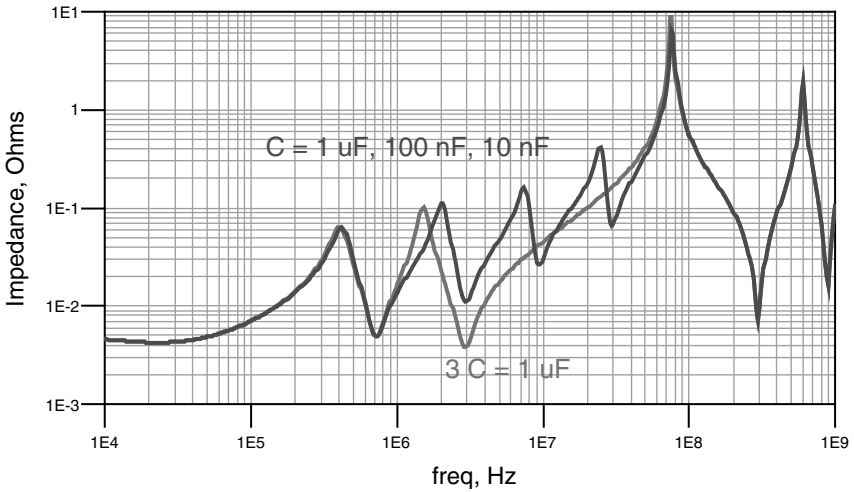


Figure 13-47 Impedance profile with an additional 10 uF bulk decoupling capacitor added with an ESL of 5 nH, reducing the peak impedance at low frequency to under 0.1 Ohm.

where:

C_{planes} = the capacitance in the planes, in nF

Dk = the dielectric constant of the laminate materials, typically 4 for FR4

A = the area of the planes in square inches

h = the dielectric thickness in mils

For example, in this case with A = 5 inches × 5 inches = 25 square inches and h = 4 mil and Dk = 4, the capacitance of the planes is:

$$C_{planes} = 0.225 \times 4\frac{25}{4} = 5.6 \text{ nF} \tag{13-48}$$

The parallel resonant frequency is expected at roughly:

$$f_{PRF} = \frac{160 \text{ MHz}}{\sqrt{\frac{1}{n}ESL \times C_{planes}}} = \frac{160 \text{ MHz}}{\sqrt{\frac{1}{3}3 \times 5.6}} = 67 \text{ MHz} \tag{13-49}$$

The simulated PRF is 70 MHz.

The peak impedance at the PRF is related to the inductance of the capacitors and the capacitance of the planes. In this case, the inductance of the three capacitors is identical, independent of the value of their capacitance. This is why the peak impedance is exactly the same whether we use three capacitors of the same value or three different values.

This peak impedance limits the PDN impedance at the board level to about 10 Ohms. If no worst-case current amplitudes are near 70 MHz, this impedance peak may not be a problem. But if you are designing the PDN assuming you need a target impedance below 10 Ohms, this peak impedance needs to be brought down.

There are six ways of reducing the peak impedance for frequencies below 100 MHz:

- Increase the capacitance in the planes a lot to push its SRF to very low frequency.
- Decrease the capacitance of the planes so the PRF is well above 100 MHz.
- Reduce the inductance of the decoupling capacitors.
- Increase the ESR of the capacitors.
- Adjust a capacitor value so its SRF is closer to the PRF.
- Add an additional capacitor with an SRF near the PRF.

We rarely can adjust the capacitance of the planes. It is what it happens to be. One of the reasons all PDN designs are custom is that the plane capacitance will vary depending on the board size and stack-up. This will shift the PRF over a wide range of frequencies.

TIP Reducing the ESL of the bulk capacitors should always be at the top of the list of actions. Everything should always be done to reduce the ESL. Selecting lower value capacitance capacitors may provide higher ESR and more damping.

If we are limited to using just three capacitors, it may be possible to find a value of one of the capacitors so that its SRF is closer to the PRF of the planes. This would reduce the peak impedance by a closer SRF and by a higher ESR, increasing damping.

We would want to adjust the third capacitor, C3, so that its SRF is close to the PRF. The condition is:

$$SRF = PRF = \frac{160\ MHz}{\sqrt{ESL \times C_3}} = \frac{160\ MHz}{\sqrt{\frac{1}{3}ESL \times C_{planes}}} \tag{13-50}$$

This reduces to:

$$C_3 = \frac{1}{3}C_{planes} = \frac{1}{3}5.6\ nf = 1.9\ nf \tag{13-51}$$

Figure 13-48 shows the impedance profile of the capacitors, VRM, and planes with the third capacitor changed from 10 nF to 2 nF.

TIP It is counterintuitive that by decreasing the capacitance of one of the capacitors the impedance profile actually improves. By optimizing the capacitor value, we reduce the peak impedance from 10 Ohms down to 2.5 Ohms. This would reduce the PDN noise by 4x in the 50 MHz to 100 MHz range.

Which is better, three capacitors all with the same value or three capacitors with different values? If you randomly select the three capacitor values, or blindly

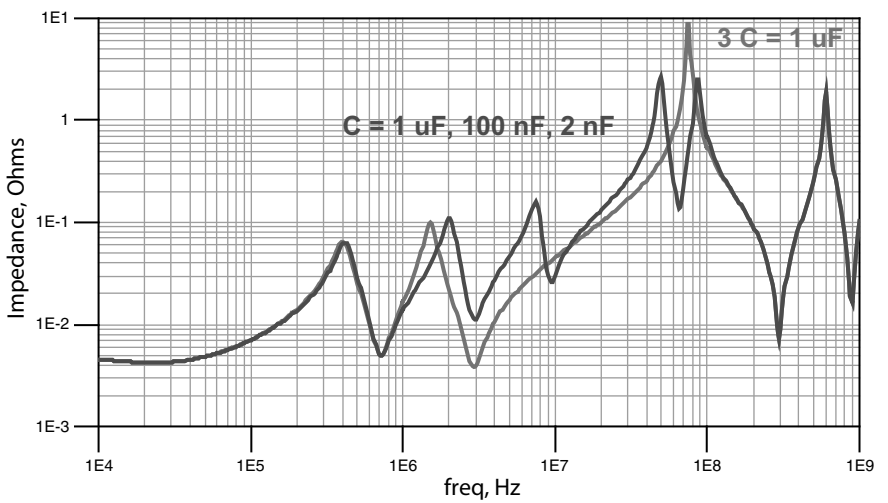


Figure 13-48 Impedance profile with the 10 nF capacitor changed to 2 nF with a lower peak impedance.

use values of 1 uF, 0.1 uF, and 0.01 uF, then it probably doesn't matter which approach you choose. They each have the same chance of success or failure. However, if you can optimize the capacitor value to minimize the peak impedance at the PRF with the plane's capacitance, using different value capacitors results in a lower impedance profile.

In this example, the lowest impedance that could be obtained below 100 MHz using just three capacitors, even if their values are optimized, is still limited to about 2 Ohms. This can be dramatically improved with more capacitors.

13.19 Estimating the Number of Capacitors Needed

In the absence of more detailed information, the goal of the board level PDN design is to engineer the impedance below the target value up to about 100 MHz, or roughly where the package limits the impedance the chip will see, which could be at lower frequency.

At the low frequency end, the ESL and number of bulk capacitors can be adjusted to keep the peak impedance below the target value as it interacts with the ceramic decoupling capacitors.

At the high frequency end, the absolute lowest impedance a collection of capacitors can theoretically have is set by their parallel combination of equivalent series inductance. The best case is if that there is no parallel resonance with the plane's capacitance and all the inductances are in parallel. The design condition is that:

$$Z_{capacitors} < Z_{target} \text{ at } F_{max} \tag{13-52}$$

where:

$Z_{capacitors}$ = the impedance of the capacitors in parallel in Ohms

Z_{target} = the target in impedance in Ohms

F_{max} = the highest frequency where the board level impedance can play a role

If the impedance of the capacitors at the high frequency end is all due to the parallel combination of inductances and they all have the same value of ESL, this condition translates to:

$$2\pi F_{max}\left(\frac{ESL}{n}\right) < Z_{target} \tag{13-53}$$

where:

Z_{target} = the target in impedance in Ohms

F_{max} = the highest frequency where the board level impedance can play a role in GHz

ESL = the equivalent series inductance of each capacitor, in nH

n = the number of capacitors needed in parallel to meet the target impedance

This establishes the theoretical minimum number of capacitors needed in parallel to meet this impedance target as:

$$n > 2\pi F_{max}\left(\frac{ESL}{Z_{target}}\right) \tag{13-54}$$

For example, if the target impedance is 0.1 Ohm and F_{max} is 100 MHz and each capacitor has 2 nH of ESL, then the theoretical minimum number of capacitors needed is:

$$n > 2\pi \times 0.1\left(\frac{2}{0.1}\right) = 13 \tag{13-55}$$

TIP In order to reduce the number of capacitors needed, regardless of their value, the ESL must be reduced. This is why ESL is such an important number.

Figure 13-49 shows how the theoretical minimum number of capacitors varies based on the ESL and the target impedance.

The minimum number of capacitors possible to achieve a target impedance is a good figure of merit to evaluate how well optimized a design might be. In the

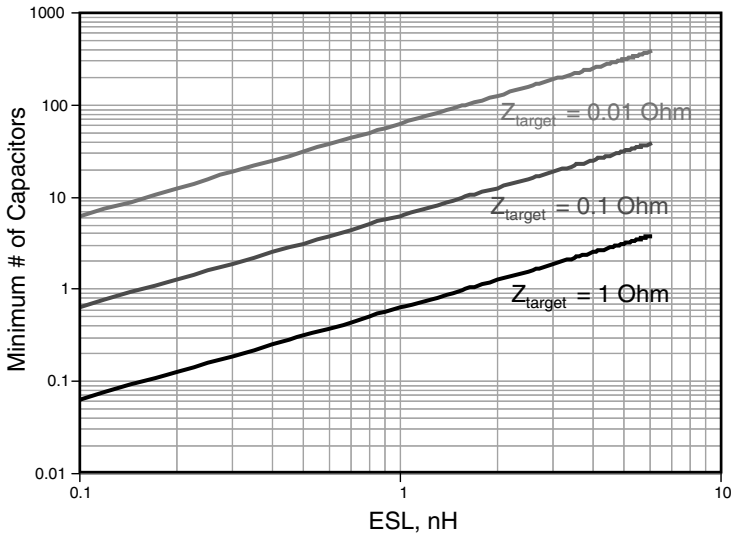


Figure 13-49 Minimum number of capacitors required to achieve a target impedance at 100 MHz based on an ESL value.

example in the last section, we achieved a target impedance of 2 Ohms with three capacitors, each with an ESL of 3 nH. The theoretical minimum number that could be used, as seen from the chart above is one. Using three to get there is not very efficient, due to the complication of the PRF of the planes.

13.20 How Much Does a nH Cost?

The cost of the small ceramic decoupling capacitor is almost negligible. Its largest direct cost is in the assembly operation and in the indirect costs of more vias to drill, surface real estate taken up, potential of blocked routing channels, and impact on the board layer count.

Using a rough estimate of $0.01 for the total direct material cost and assembly cost per capacitor, we can estimate how much a nH is worth. For every fraction of a nH reduction in the ESL, fewer capacitors need to be used and there is a direct savings.

The cost per nH is derived from the expression above for the theoretical minimum number of capacitors needed:

$$\text{TotalCost} = \$0.01 \times n = \$0.01 \times 2\pi F_{max}\left(\frac{ESL}{Z_{target}}\right) \tag{13-56}$$

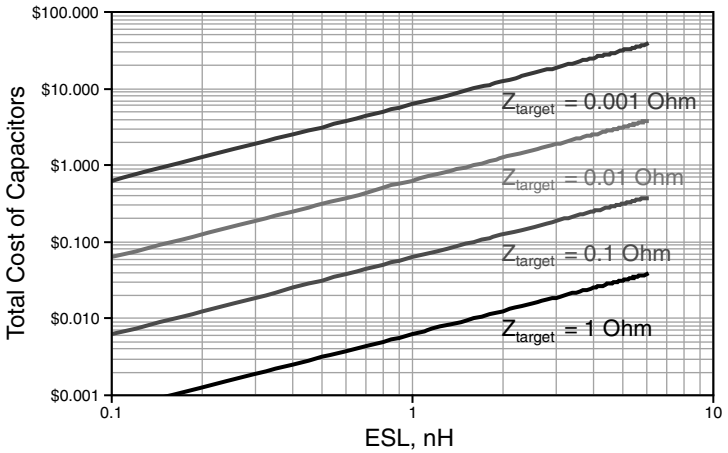


Figure 13-50 Total cost of all the capacitors assuming $0.01 per, as the ESL is reduced.

This is displayed in Figure 13-50 for different ESL and target impedance values for a maximum frequency where the board plays a role of 100 MHz.

The cost per nH can be estimated based on:

$$\frac{TotalCost}{ESL} = \frac{\$0.01 \times 2\pi F_{max}}{Z_{target}} = \frac{0.006}{Z_{target}} \frac{\$}{nH} \tag{13-57}$$

where:

TotalCost/ESL = the cost per nH in $/nH

F_{max} = the highest frequency the board level impedance can play a role in GHz, assuming ~ 0.1 GHz

Z_{target} = the target impedance in Ohms

TIP This is a simple result. It suggests that the lower the target impedance, the more expensive every nH becomes, and the more valuable a reduction in the ESL becomes.

For example, with a target impedance of 0.01 Ohm, the cost per nH is $0.6/nH. If the mounted inductance is 2 nH, the total cost of the capacitors is about $1.20

for the board. If the ESL can be reduced from 2 nH to 1 nH just by changing the surface traces or bringing the power and ground plane cavity closer to the board surface, the cost savings from the reduction in the number of capacitors used would be $0.60, with no sacrifice in performance. If this were a high volume board, with 1 million units per month, the cost savings would be $600k per month, or $7.2 million per year.

The contribution of spreading inductance for the planes is roughly:

$$L_{via-via} = 21 \times h \times \ln\left(\frac{B}{D}\right) pH \tag{13-58}$$

where:

$L_{via–via}$ = the loop spreading inductance in the planes between the two via contacts in pH

h = the dielectric thickness between the vias in mils

B = the distance between the via centers in mils

D = the diameter of the vias in mils

For a typical case of B = 1 inch and D = 10 mils, the spreading inductance in the planes is roughly about:

$$L_{via-via} = 21 \times h \times \ln\left(\frac{1}{0.01}\right) pH \sim 0.1 \times h \text{ nH} \tag{13-59}$$

When a conventional thickness of 4 mils is used, the spreading inductance contribution is on the order of 0.4 nH per capacitor. If an ultra thin laminate, such as the DuPont Interra HK04 material, with a thickness of 0.5 mils, were used, the spreading inductance would be on the order of 0.05 nH. This is a reduction of about 0.35 nH. This translates into a potential cost reduction of:

$$\text{CostReduction} = \frac{0.006}{Z_{target}} \frac{\$}{\text{nH}} \times 0.35 \text{ nH} = \frac{\$0.002}{Z_{target}} \tag{13-60}$$

When the cost reduction is greater than the cost premium for the thinner dielectric, using the more expensive, ultra thin laminate becomes a reduction in the total cost of ownership. The premium is rated as an extra price per square foot of board area, or:

$$\text{CostReduction} > \text{Premium} \times \text{area} \tag{13-61}$$

$$\frac{\$0.002}{Z_{target}} > \text{Premium} \times \text{area} \tag{13-62}$$

$$\frac{\$0.002}{\text{Premium}} > Z_{target} \times \text{area} \tag{13-63}$$

If the premium were about \$3/sq. foot extra cost, then the condition for total cost reduction with a thin laminate would be:

$$\frac{\$0.002}{\$3} = \$0.0007 > Z_{target} \times \text{area} \tag{13-64}$$

where:

CostReduction = the cost reduction in the number of capacitors not needed, in \$

Z_{target} = the target impedance in Ohms

Premium = the added cost per square foot of thin laminate over conventional laminate, in \$/sq. ft.

area = the area of the board surface in the specific application in sq. ft.

If the area is described in square inches, this relationship becomes:

$$\$0.1 > Z_{target} \times \text{area} \tag{13-65}$$

This suggests that if the board area is 10 square inches, a thinner laminate is a cost reduction if the target impedance is lower than 0.01 Ohm.

13.21 Quantity or Specific Values?

To first order, the impedance at high frequency of a collection of capacitors is related to the parallel combination of their inductance. However, if a parallel resonance with the capacitance of the board planes exists near the maximum frequency where the board impedance plays a role, it will artificially increase the impedance profile of the capacitors. In this regime, the impedance of the collection of capacitors can be brought down by carefully selecting the values of capacitor to "sculpt" the impedance profile and compensate for the parallel resonance.

The impact from the parallel resonance on the impedance profile for a collection of capacitors is illustrated in Figure 13-51 for the specific case of:

Z_{target} = 0.1 Ohm

F_{max} = 0.1 GHz

ESL = 2 nH

n = 13 capacitors

A = 65 square inches and 6.5 square inches

In this example, the theoretical minimum number of capacitors needed to achieve the 0.1 Ohm at 0.1 GHz is:

$$n > 2\pi F_{max}\left(\frac{ESL}{Z_{target}}\right) = 2\pi 0.1\left(\frac{2}{0.1}\right) = 13 \qquad (13\text{-}66)$$

When the PRF is close to the F_{max}, it artificially increases the impedance profile of the capacitors and the planes' capacitance. This increase can be more than a factor of 2 or 3.

However, if the PRF can be engineered to be a higher frequency, as by decreasing the area of the planes, for example, the parallel resonance does not interact with the impedance of the capacitors near the F_{max} and the impedance can be close to the theoretical impedance of the n inductors in parallel. In this case, to achieve the target impedance at the F_{max} with the minimum number of capacitors does not depend on the specific value of capacitors, just their number and their ESL.

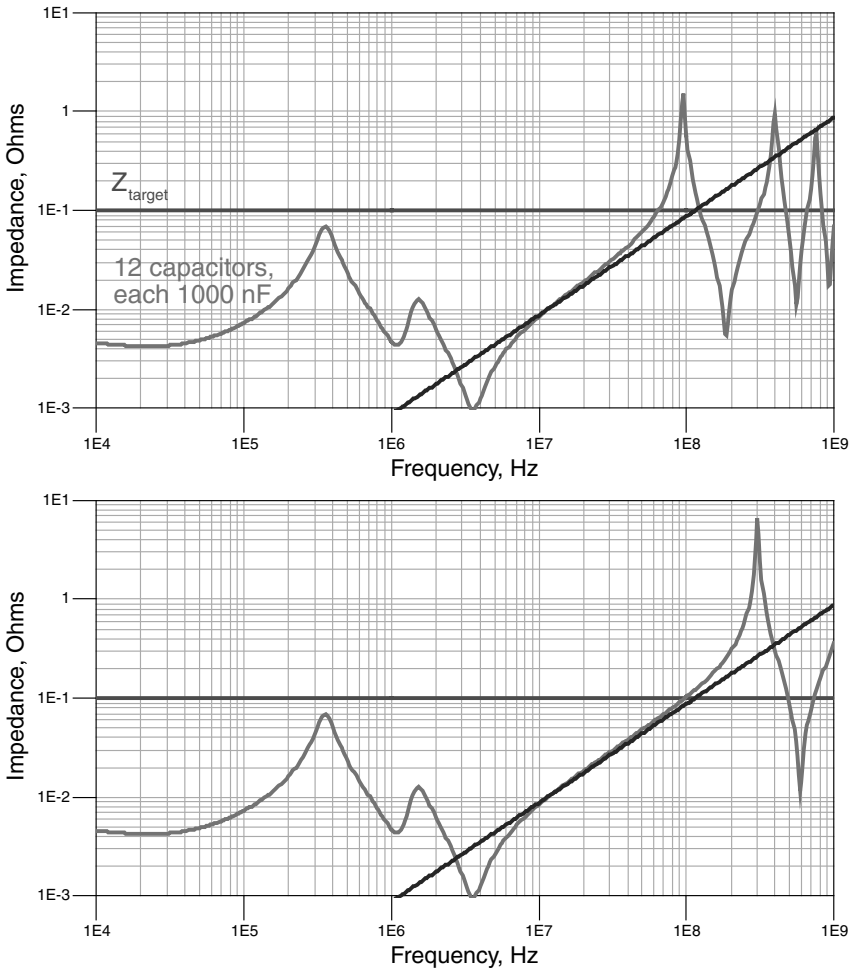


Figure 13-51 Impedance profile of 13 capacitors interacting at high frequency with the 65 and 6.5 square inches of board capacitance, compared to the impedance of the ideal inductance of 13 capacitors. Top: when PRF = 100 MHz; bottom: when PRF = 3 × 100 MHz.

The PRF of the capacitors' inductance interacting with the planes' capacitance is given by:

$$\text{PRF} = \frac{160\ \text{MHz}}{\sqrt{\frac{1}{n}(\text{ESL}\ C_{\text{planes}})}} = \sqrt{\frac{h}{\frac{1}{n}\text{ESL} \times A}}\,160\ \text{MHz} \tag{13-67}$$

where:

PRF = the parallel resonant frequency, in MHz

n = the number of capacitors in parallel

ESL = the equivalent series inductance of each capacitor, in nH

C_{planes} = the capacitance of the planes in nF

h = the dielectric thickness between the planes, in mils assuming Dk = 4

A = the area of the planes in square inches

The goal is to engineer conditions so that the PRF is pushed to frequencies above the maximum frequency. To first order, this would suggest:

- Large n
- Thicker h
- Small ESL
- Small A

However, the dielectric thickness also affects the ESL. Increasing h will increase ESL. Given the importance of lower ESL, thinner h is usually better. To push the PRF to a high enough frequency where it is not interacting with the impedance of the capacitors, it needs to be at least 3x higher than the max frequency.

The condition for the specific values of the capacitors to not be significant is roughly approximated, as illustrated above to be:

$$PRF > 3 \times F_{max}$$

$$\sqrt{\frac{h}{\frac{1}{n}ESL \times A}}\,160\ MHz > 3 \times F_{max} \tag{13-68}$$

In addition, if the number of capacitors is adjusted to meet the target impedance at the maximum frequency, a further condition is:

$$Z_{target} = \frac{1}{n}ESL \times 2\pi F_{max} \tag{13-69}$$

These two relationships can be combined to result in the condition where the impedance at the F_{max} is independent of the specific values of the capacitors selected as:

$$\frac{h}{Z_{target}A} > 56 \times F_{max} \tag{13-70}$$

where:

n = the number of capacitors in parallel

Z_{target} = the target impedance in Ohms

A = the area of the planes in square inches

F_{max} = the highest frequency where the board level impedance is important, in GHz

ESL = the equivalent series inductance of each capacitor, in nH

h = the dielectric thickness between the planes, in mils assuming Dk = 4

For the best conventional case, of h = 4 mils, and the typical F_{max} of 0.1 GHz, this condition reduces to:

$$Z_{target}A < 0.7 \tag{13-71}$$

where:

Z_{target} = the target impedance in Ohms

A = the area of the planes in square inches

TIP This suggests that to engineer a condition where the values of the capacitors are not important and where it is still possible to use the theoretical minimum number of capacitors, the area of the power planes should be kept to a minimum and the target impedance is low.

In general, the area of the planes is always less than the area of the board. When split planes are used, the actual board area can be more than a factor of

three larger than the power planes. If well engineered, the area of the power plane can be kept to a minimum to support all the capacitors and connections to the VRM. When power planes are mixed on signal layers as copper fill areas, the small power planes are sometimes referred to as a *copper puddle*.

TIP Whenever split planes are used, it is always important to keep signal layers from crossing underneath them, as this has the potential of generating noise in the power plane and excessive coupling between adjacent signal lines.

In many applications with dedicated power planes, the plane area and board area may be close.

This design space is mapped in Figure 13-52.

For this specific case of a dielectric thickness of 4 mils, if the target impedance is 0.1 Ohm, then as long as the area of the planes is less than roughly 7 square inches, the selection of capacitor values is not important. They can all be the same value.

If, however, the plane area is larger than 7 square inches, the capacitance of the board will push the PRF close to 0.1 GHz and increase the impedance of the combination of capacitors. In this case, to use the minimum number of capacitors, their value is important and should be selected to "sculpt" the impedance profile.

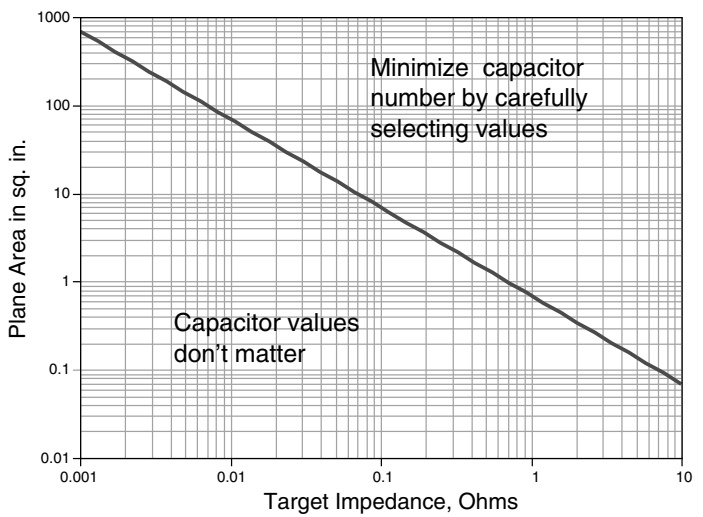


Figure 13-52 Design space of when the capacitor values selected are important and when they are not important, for the special case of dielectric thickness of 4 mils.

This design space of target impedance higher than 0.1 Ohm and plane area larger than 7 square inches encompasses many boards. In this regime, to use the fewest number of capacitors and achieve the lowest cost, the precise values of capacitors should be carefully selected. Their values and number are chosen to bring the peak impedances below the target value up to the maximum frequency.

This is why for many common board applications, using a distribution of capacitor values will enable the lowest impedance with the fewest capacitors, rather than all the same value of capacitance. Of course, the minimum number of capacitors is based on using the right distribution of values.

If the plane area can be kept to less than 2 square inches, then all designs with a target impedance less than 0.3 Ohm can use the same value capacitors and the minimum number of capacitors. This is a rather small plane area and not very common. However, the package is about this size.

TIP The package can act as a small board. If enough capacitors are added to the package, the impedance may be reduced to the level where the impedance of the board is not important.

While there is rarely enough on-package decoupling capacitance to supplement all the board level needs, it is possible to add all the necessary capacitors using a small size interposer, rather than in the package. Such an alternative approach is available from Teraspeed Consulting Group.

An example of a small island of low impedance power and ground provided in the PowerPoser is shown in Figure 13-53. This small board fits underneath the package and has multiple layers of thin laminate with all the very low inductance decoupling capacitors to bring the impedance below the target value from low frequency to very high frequency.

By using an interposer, very thin dielectric layers near the surface can be used without paying a large-area price penalty, and low inductance capacitors can be placed in proximity to the package. The area of the planes can be kept small so the parallel resonant frequency is well above the package limit and all capacitor values can be the same.

13.22 Sculpting the Impedance Profiles: The Frequency Domain Target Impedance (FDTI) Method

When the capacitance in the power and ground planes causes a parallel resonance at a low enough frequency to affect the impedance, the fewest number of capacitors can be used to build the impedance profile by optimizing their values. In this

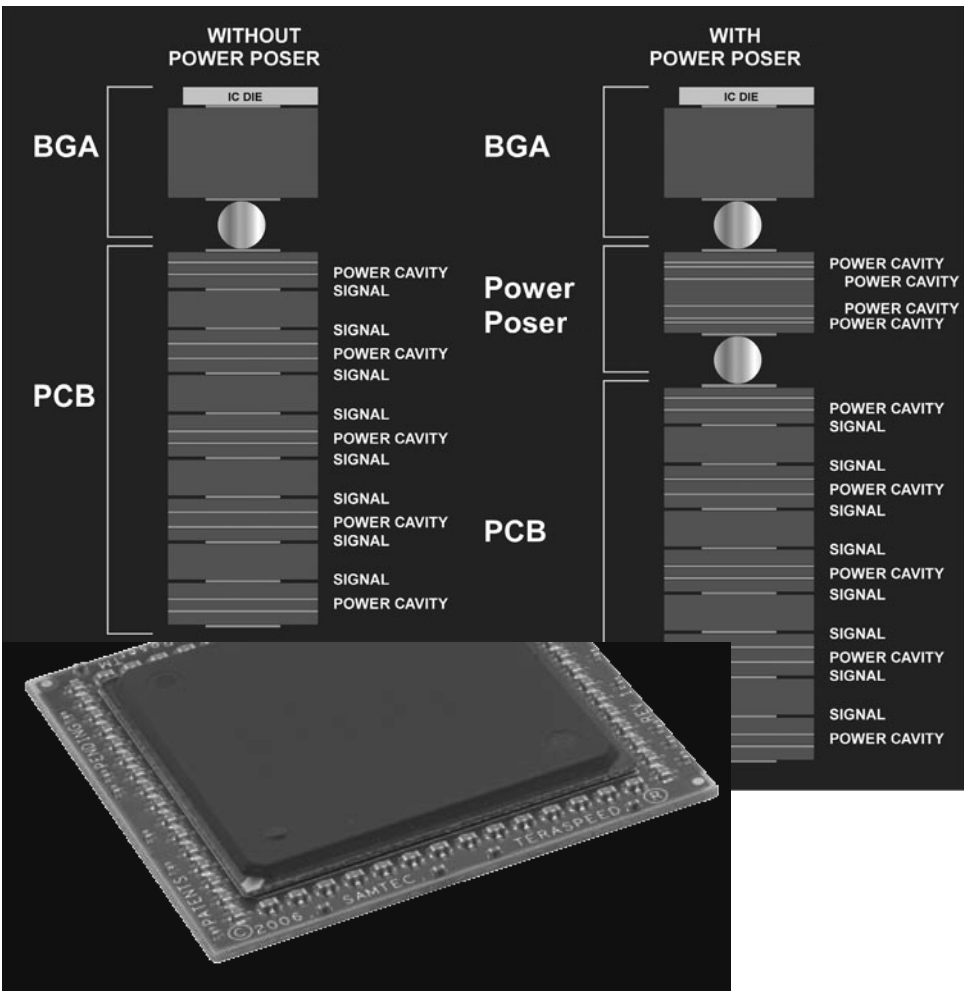


Figure 13-53 The role of the PowerPoser in a board stack-up and a close-up view of an FPGA chip mounted on a PowerPoser.

design regime, it is possible to minimize the number of capacitors used to achieve the target impedance by careful selection of their values. This process is called *sculpting* the impedance profile.

TIP No matter what, it is always important to use the lowest possible ESL for all decoupling capacitors. This will always result in the fewest number of capacitors and the lowest-cost system.

The precise number and optimized values of capacitors needed, will depend on:

- The bulk decoupling capacitors associated with the VRM
- The capacitance in the board
- The target impedance
- The maximum frequency
- The ESL of each capacitor

The combination of these terms varies dramatically from product to product so it is not possible to give one capacitor distribution that will always work. However, the methodology can be applied to many designs.

This methodology was pioneered by Larry Smith while he was at Sun Microsystems and has been termed the *Frequency Domain Target Impedance (FDTI)* method.

The process leverages the simulated impedance profile of a collection of capacitors including their ESL and ESR, including the capacitance in the planes at high frequency and the bulk capacitors associated with the VRM at the low frequency end.

Capacitor values are selected from the values available from vendors. Not all values are available, rather the common values are in decade steps of multiples of 1.0, 1.5, 2.2, 3.3, 4.7, and 6.8. When the ESL of each capacitor in a collection is the same, the minimum parallel resonant impedance peak is obtained when each capacitor value is the geometric mean of the capacitor on either side of it.

The optimum distribution would be using values of decade multiples of 1, 2.2, and 4.7, for example. The largest capacitance value of an 0402 capacitor easily available is about 1 uF. When higher values are needed, a 1206 capacitor can provide as much as 100 uF.

The lowest value capacitor needed is about 1/3 × the capacitance in the planes. The board capacitance is on the order of 10 nF for a board with 40 square inches, for example. The smallest value capacitor needed would be about 2.2 nF.

The selection of possible values might range from 1 uF to 2.2 nF and include nine different values from which to select:

1000 nF

470 nF

220 nF

100 nF

47 nF

22 nF

10 nF

4.7 nF

2.2 nF

A typical set of parameters might be:

- 50 square inches of board area with ~ 20 nF of capacitance.
- Target impedance is 0.1 Ohm.
- ESL of each capacitor is 2 nH.
- Maximum frequency is 0.1 GHz.

In this example, the board area of 50 square inches and target impedance of 0.1 Ohm puts the design in the upper part of the design space where the value of the capacitors matters. Of course, if the target impedance was low enough or the board capacitance small enough, it would not matter what values capacitor were used, and they could all be 1 uF. In this case, parallel resonances would not play a role.

The theoretical minimum number of capacitors required to meet the target value for the above condition is:

$$n > 2\pi F_{max}\left(\frac{ESL}{Z_{target}}\right) = 2\pi 0.1\left(\frac{2}{0.1}\right) = 13 \tag{13-72}$$

where:

n = the minimum number of capacitors needed

F_{max} = the highest frequency for board level impedance in GHz

ESL = the equivalent series inductance of the capacitor, including the mounting inductance and some of the cavity spreading inductance in nH

Z_{target} = the target impedance in Ohms

Starting at the low frequency end, the largest capacitor value is selected and simulated. Enough capacitors of each value are added to bring the peak impedance

below the target value. The next capacitor is added and enough quantity of this value until the target impedance is reached. Specific capacitor values are skipped, especially at the low frequency end, where the low target impedance can be reached without them. The ESR for each capacitor value is used in the simulation.

Figure 13-54 shows an impedance profile based on the following capacitor selection with a total of 14 capacitors:

C	n
470	1
100	1
47	1
22	1
10	3
4.7	3
2.2	4
Total	14

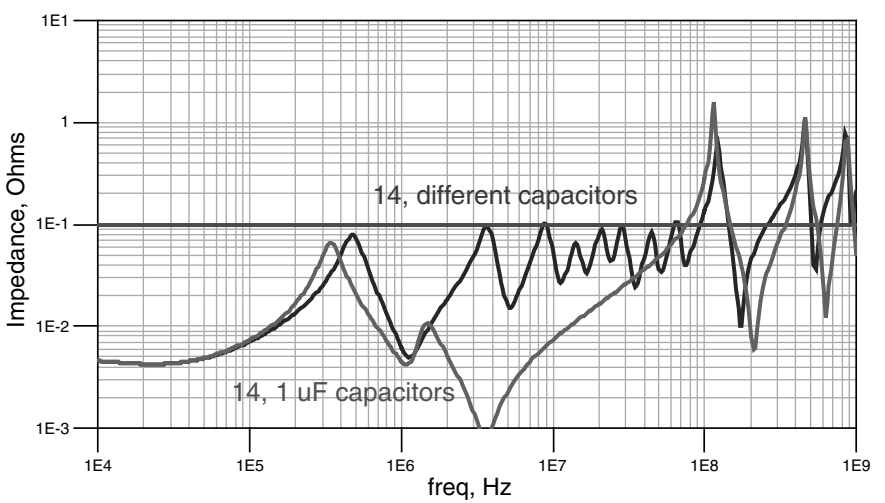


Figure 13-54 Impedance profiles for two distributions of 14 capacitors. One distribution uses all the same value of 1 uF capacitor. The other distribution was chosen to sculpt the profile. At 100 MHz, the sculpted profile meets the target impedance, but the other distribution does not.

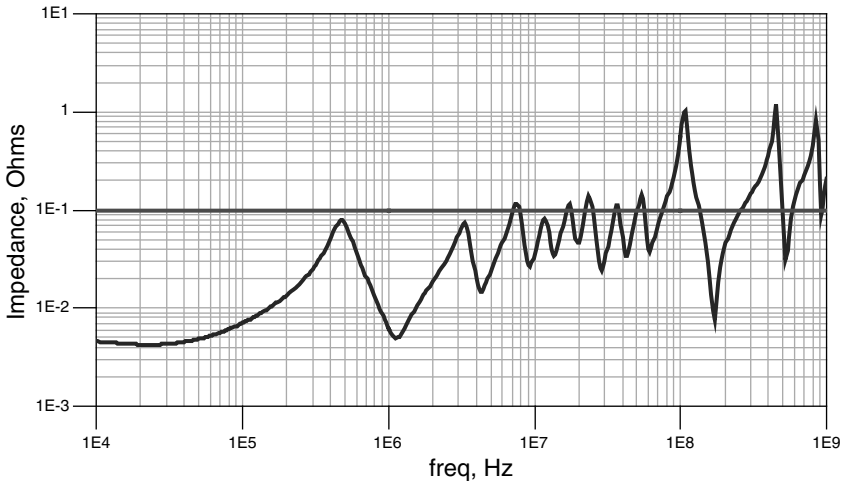


Figure 13-55 Impedance profile for the same distribution of capacitor values as above, but each with an ESL of 3 nH rather than 2 nH.

In this example, the impedance profile meets the target impedance up to 100 MHz, using 14 capacitors. This is close to the theoretical minimum of 13. However, using 14 capacitors all with the same value of capacitance is not able to achieve the same low impedance. If all the same value capacitors were used, more than 14 would be required. It would not be as low a cost as using the FDTI method.

TIP Of course, if any of the initial conditions of this specific problem were changed, for example, if the ESL was not 2 nH but really 3 nH, this combination would no longer work.

Figure 13-55 shows the impedance profile for these capacitors with an ESL of 3 nH, exceeding the target impedance at a number of frequencies.

This again is another example of how important reducing the ESL is for capacitors, and how custom the selection of capacitors becomes when so many system parameters affect the impedance profile.

TIP Of course, there are many right distributions. The most cost-effective solutions use a total number of capacitors that is close to the theoretical minimum.

Another example of a sculpted impedance profile for a target impedance of 0.05 Ohm is shown in Figure 13-56. In this case, the theoretical minimum with an ESL of 2 nH is 26. This distribution used 33, slightly above the theoretical minimum. The capacitor values used were:

C	n
1000 nF	1
470 nF	1
220 nF	1
100 nF	1
47 nF	2
22 nF	3
10 nF	5
4.7 nF	6
2.2 nF	13
Total	33

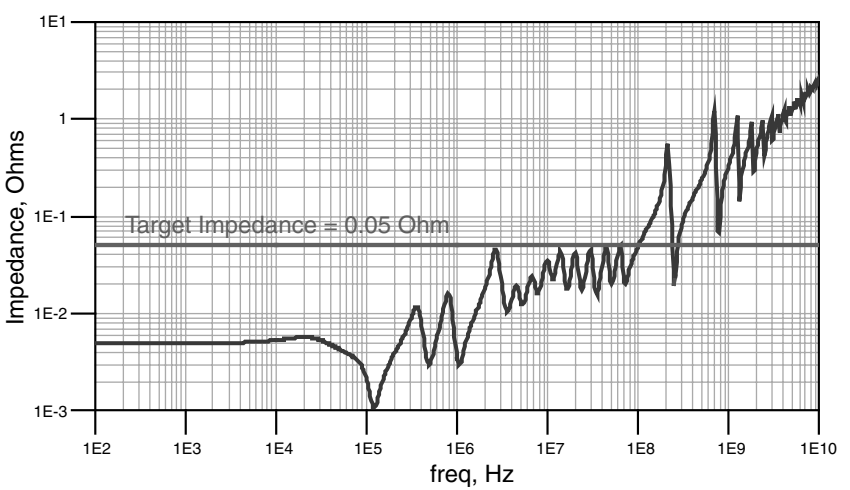


Figure 13-56 Impedance profile for a target impedance of 0.05 Ohm using 33 capacitors.

13.23 When Every pH Counts

The four most important design "habits" to follow are:

1. Use as short and wide a surface trace that is consistent with the assembly design rules. In other words, use as few squares of surface interconnect between the capacitors and the vias as possible.
2. Place the capacitors in proximity to the package, some below it on the bottom side of the board and some on the same layer, to avoid saturating the spreading inductance with all peripheral capacitors.
3. When power and ground planes are on adjacent layers, use the thinnest dielectric that does not cost extra. This is usually 2.7 to 4 mils depending on the vendor.
4. When possible, place the power and ground cavity as close to the surface of the board as possible.

There is rarely a reason not to do these, and they will always result in lower ESL. The cost impact for an ESL reduction is:

$$\frac{\text{TotalCost}}{\text{ESL}} = \frac{0.006}{Z_{\text{target}}} \frac{\$}{\text{nH}} \tag{13-73}$$

When the target impedance is 0.1 Ohm or higher, this is less than $0.06 per nH per voltage rail on a board. There usually isn't enough cost savings potential to justify paying extra for low inductance features.

However, when the target impedance is 0.001 Ohm, the cost savings is $6/nH for each voltage rail on the board. Every pH reduction is a ½¢ cost reduction.

TIP The lower the mounting loop inductance of each capacitor, the fewer capacitors are required to achieve the low target impedance at the high frequency. Everything should be done that is free to reduce the ESL of all decoupling capacitors.

As pointed out above, sometimes, paying extra for thinner dielectric between the power and ground planes is worth the extra cost of the lower spreading inductance.

In addition, alternative capacitor technologies can offer a lower mounted ESL than conventional capacitors. Most capacitors are designed with their terminals

along their long axis. The capacitor body is a minimum of two squares. Even with via in pad, there are still two squares of surface trace.

An alternative design uses *reverse aspect ratio* capacitors, with terminal pads along the long side of the capacitor. Mounted onto a board, these capacitors could be implemented with as low as 0.5 square. An example of these capacitors is shown in Figure 13-57.

If the depth in the board stack-up to the power and ground cavity is 5 mils below the top surface, the sheet inductance of surface traces would be about 32 pH/mil × 5 mil = 160 pH/sq. The trace loop inductance of a via-in-pad, best case standard capacitor is about 320 pH, while for a reverse aspect ratio capacitor, the best case could be as low as 160 × 0.5 = 80 pH. This is a reduction of 240 pH.

Interdigitated capacitor technologies can offer even lower capacitance. These are constructed as multilayer ceramic capacitors with multiple interleaved terminals on each end. Examples are illustrated in Figure 13-58.

An IDC effectively is multiple capacitors in parallel, with the ESL of each current path in parallel. The equivalent ESL of the four capacitors in one IDC would be ¼ the ESL of any one of them. In addition, since the current flow is in opposite directions in adjacent capacitor segments and they are in close proximity, the effective ESL of each one is further reduced. An IDC can have less than 20% the ESL of a conventional capacitor.

Another type of IDC is provided by X2Y Attenuators. These are multilayer ceramic capacitors with alternating plates coming out to each of four different electrode terminals. An example of these four-terminal capacitors and their internal structure is shown in Figure 13-59.

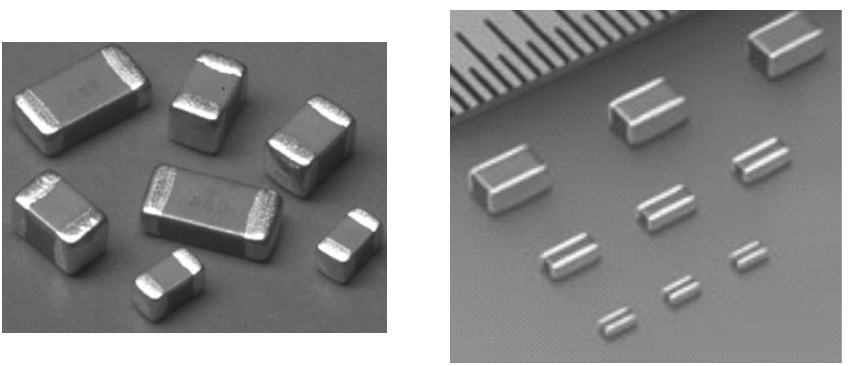


Figure 13-57 Capacitor technologies. Left: conventional aspect ratio capacitors with minimum number of surface trace squares of n = 2. Right: reverse aspect ratio capacitors available from AVX, with a minimum n = 0.5.

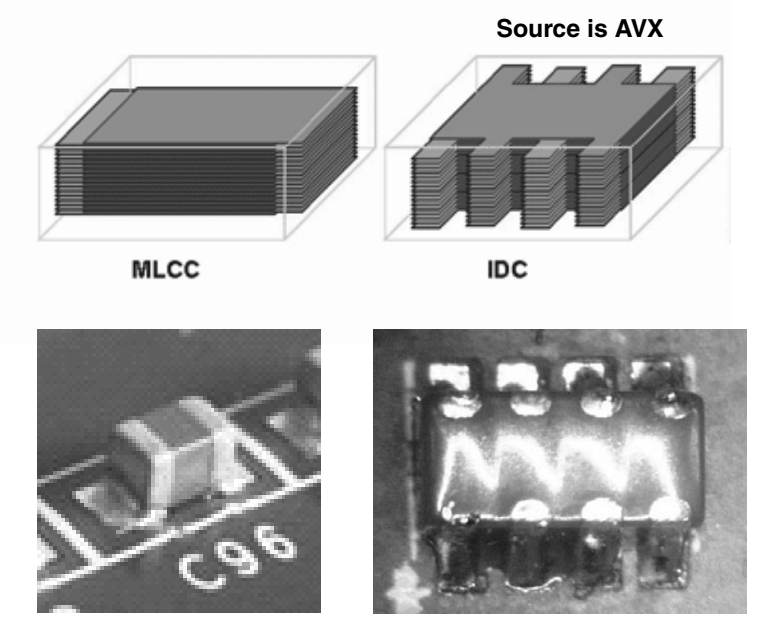


Figure 13-58 Comparison of conventional MLCC capacitors and interdigitated capacitors, IDC.

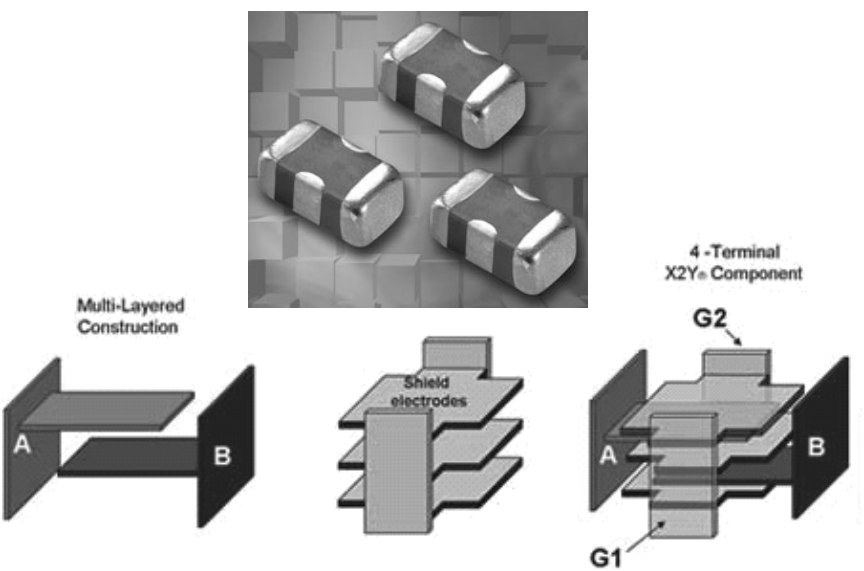


Figure 13-59 Examples of X2Y capacitors and their internal structure, as described by X2Y Attenuators.

The A and B plates are both connected in parallel to the power plane while the central two G1 and G2 plates are connected in parallel to ground. In this configuration, the capacitor behaves like four capacitors in parallel, with current flows illustrated in Figure 13-60.

While there are similar performance advantages using either IDC capacitor technology, an advantage of the X2Y capacitors is the ease of integration with conventional through-hole circuit board technology.

An 0805 IDC with four terminals on a side will have a pad footprint with 20 mil centers between pads. This is difficult to connect to circuit boards using conventional through-hole technology and requires via in pad. An 0805 X2Y capacitor can use via holes on 40 mil centers, as shown above, and use conventional through-hole technology, leaving a routing channel through the holes available for multiple 5-mil-wide tracks.

TIP The combination of IDC capacitors, minimum surface trace size, and cavity close to the surface can enable a total ESL from the capacitor to the package pin of less than 250 pH.

However, if the surface traces are long and the cavity is not near the surface, the same X2Y capacitor can show an ESL that is more than 1 nH, 3× larger than it needs to be, just from very slight design variations. Figure 13-61 is an example of

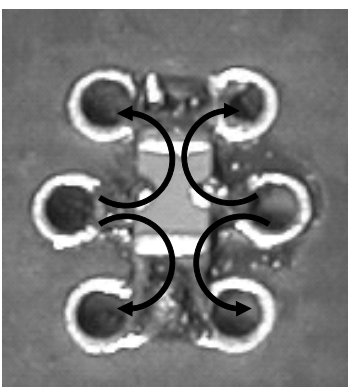


Figure 13-60 X2Y capacitor mounted to a circuit board with the top and bottom terminals tied to power and the two central terminals tied to ground. It behaves like four capacitors in parallel.

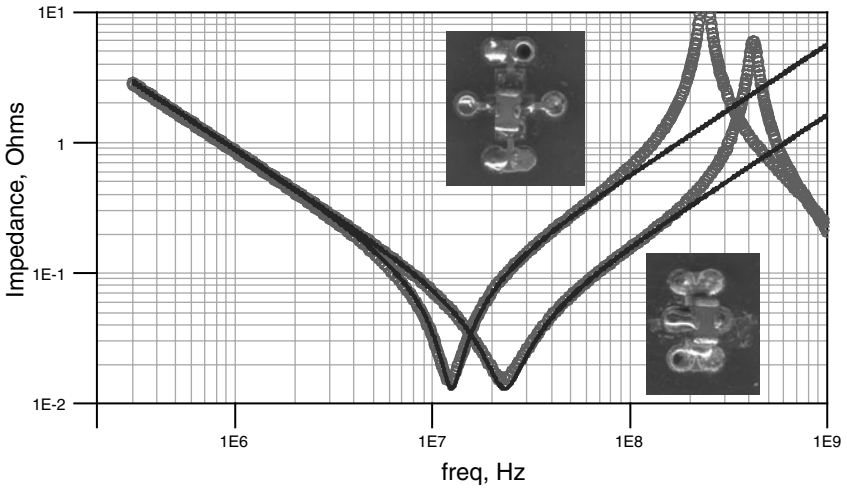


Figure 13-61 Measured impedance profile for two X2Y capacitors mounted to a test board with slightly different conditions. The measured data is compared to the simulated impedance of an ideal RLC circuit. The insets show the capacitor for each measurement.

the measured impedance profile of two X2Y capacitors with the comparison of an RLC model.

For each capacitor RLC model, the same value of C = 180 nF was used in the model, the same value of R = 0.013 Ohm was used, but two different inductance values were used. For the best case, L = 260 pH gave an excellent fit, while for the other case, L = 900 pH was the best fit. The peaks near 200 and 300 MHz are the parallel resonances with the circuit board to which the capacitors were mounted.

TIP When every pH counts, the right capacitors and the optimum mounting inductance can make all the difference.

13.24 Location, Location, Location

At low frequency, below the parallel resonant frequency of the capacitors' ESL and the planes' capacitance, the planes will interact with the capacitors as a lumped element. However, when the length of an edge of the planes is comparable to a fraction of the wavelength, the resonant behavior of the board will show up in the impedance profile.

When the probe point is located at the edge of the board, the first resonant frequency will be when:

$$Len = \frac{1}{2}\lambda = \frac{1}{2\sqrt{Dk}}\frac{c}{f_{res}} = \frac{1}{2}\frac{v}{f_{res}} = \frac{3}{f_{res}} \quad (13\text{-}74)$$

$$f_{res} = \frac{3}{Len} \quad (13\text{-}75)$$

where:

Len = the length of an edge of the board, in inches

λ = the wavelength of light where the first resonance shows up, in inches

Dk = dielectric constant of the laminate between the planes

c = speed of light in air, 12 inches/nsec

f_{res} = the resonant frequency in GHz

v = the speed of light in the material, assumed v = 6 inches/nsec, for FR4

For example, if the board is 10 inches on a side, the first resonance will be about 300 MHz. If the probe point is in the middle of the board, the first resonance will be at twice this frequency, or 600 MHz.

Figure 13-62 is an example of the simulated impedance profile of a 10 inch × 10 inch bare board, probed in the center, showing the capacitive behavior at low frequency, its self resonant frequency, and the onset of board resonances at 600 MHz.

While the board will appear as a lumped capacitor below the resonant frequency, and a simple SPICE simulation will accurately reflect the impedance profile of the capacitor on the board, the spreading inductance the capacitor sees between its location on the board and the device it is decoupling, will depend on location.

The farther away the capacitor is from the package it is decoupling, the higher its total ESL due to the spreading inductance. When the spreading inductance is small compared to the mounted inductance of the capacitor, the location is not important. Changing the capacitor position will change the spreading inductance, but this will have minimal impact on the total ESL of the capacitor.

TIP However, when the spreading inductance is a significant fraction of the total capacitor's ESL, the location will have a significant impact on the ESL of the capacitor, and moving the capacitors closer to the device is important.

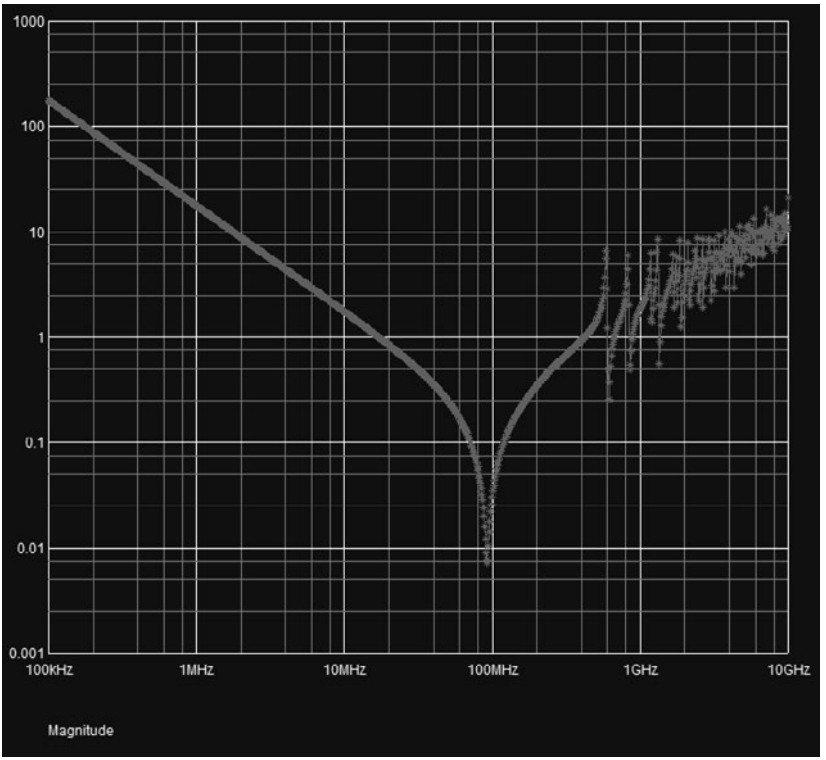


Figure 13-62 Impedance profile of a bare board showing the onset of board resonances at 600 MHz. Simulated with HyperLynx 8.0.

The condition for when location is important is based on the amount of spreading inductance compared with the mounted inductance of the capacitor.

The condition is:

When spreading inductance ~ mounting inductance, location matters.

When spreading inductance << mounting inductance, location does not matter.

For a via about 10 mil in diameter and a capacitor 1 inch away, the spreading inductance is roughly:

$$L_{via-via} = 21 \times h \times \ln\left(\frac{B}{D}\right) pH = 21 \times h \times \ln\left(\frac{1}{0.01}\right) pH = 100 \times h \text{ pH} \quad (13\text{-}76)$$

When h is thin, spreading inductance is small and location is important only for very low values of mounting inductance capacitors.

When h is thick, spreading inductance is larger. When the mounting inductance of capacitors is low, location can be important. Due to the very rough approximate nature of the spreading inductance estimate, when the spreading inductance is on the same order as the mounting inductance, it is appropriate to use a 3D field solver tool to estimate the impact on the impedance profile of the mounted capacitors from location.

Figure 13-63 is an example of the impedance profile of two board configurations. In each case, the same four capacitors are mounted to the board, close together. The mounting inductance of each is 5 nH, so the four of them have an equivalent inductance of 1.25 nH.

In the first example, the cavity thickness is 30 mils. The spreading inductance is on the order of 3 nH, large compared to the mounting inductance of the four capacitors. The impedance profile of these four capacitors is simulated when

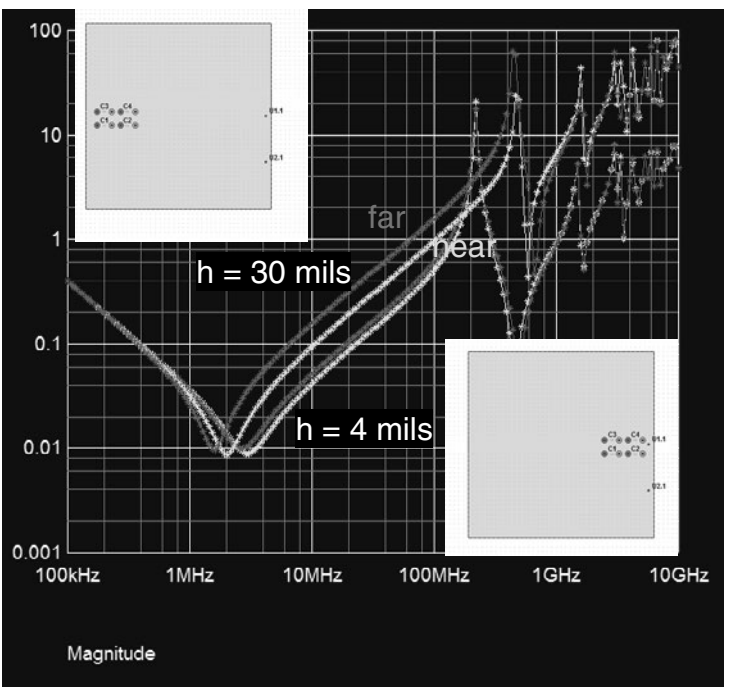


Figure 13-63 Simulated impedance of four capacitors mounted to a 30 mil thick cavity and a 4-mil-thick cavity located close to the package pin and far from the package pin. Simulated with HyperLynx 8.0.

they are located far from the package pin, and when they are close. The large difference in impedance in these two positions shows the impact from location on the total ESL.

In the second example, the cavity thickness is 4 mils and the spreading inductance is on the order of 0.4 nH. This is small compared to the 1.25 nH inductance of the capacitors. When the position of the capacitors is changed from near to far, there is little change in the simulated impedance. The spreading inductance is not an important contributor to the capacitor's inductance and location is not important.

In low impedance designs, where the ESL has been optimized to less than 0.25 nH for each capacitor, the spreading inductance can be a significant contributor and should be taken into account with a 3D field solver. The spreading inductance will increase the contribution to the inductances of the individual capacitors in a complicated way that depends on the position of the capacitors and the location of the package pins. It can only be analyzed with a 3D field solver.

13.25 When Spreading Inductance Is the Limitation

For a given target impedance and maximum frequency, the total, maximum allowable series inductance that could be in the path, including the capacitors and plane spreading inductance must be:

$$L_{max} < \frac{Z_{target}}{2\pi F_{max}} \tag{13-77}$$

where:

L_{max} = the maximum allowable series inductance, in nH

F_{max} = the maximum frequency where board level impedance is important, in GHz

For example, if the target impedance is 0.01 Ohm and the maximum frequency is 100 MHz, the maximum allowable series inductance before it dominates the impedance of all the capacitors is:

$$L_{max} < \frac{Z_{target}}{2\pi F_{max}} = \frac{0.01}{2\pi \times 0.1} = 0.016 \text{ nH} = 16 \text{ pH} \tag{13-78}$$

If the total series inductance to the capacitors exceeds 16 pH, the PDN impedance will be higher than the target impedance at the highest frequency the board is effective. If the vias in the board from the package to the planes and the spreading inductance in the planes from the package pins to the capacitors is a large fraction of this inductance, the spreading inductance will limit the impedance of the board.

If the capacitors are uniformly distributed around the package with the power and ground pins also distributed around the perimeter of the package, as shown in Figure 13-64, the spreading inductance in the planes can be estimated.

The spreading inductance in the planes is approximately:

$$L_{spread} = 5.1 \times h \times \ln\left(\frac{b}{a}\right) pH \tag{13-79}$$

where:

L_{spread} = the spreading inductance in the planes in pH

h = the dielectric thickness between the planes in mils

b = the distance to the capacitors in inches

a = the radius of the power/ground pins in the package in inches

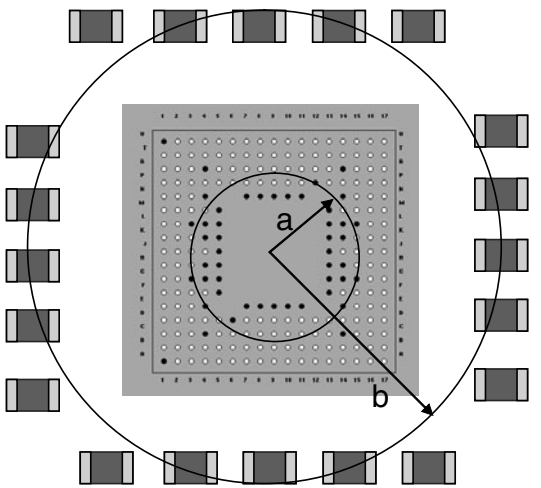


Figure 13-64 Estimating the best case spreading inductance in the planes as current flows from the distribution of capacitors to the pins in the package.

For example, if h = 4 mils, b = 1 inch, a = 0.25 inches, then the spreading inductance is:

$$L_{spread} = 5.1 \times h \times \ln\left(\frac{b}{a}\right) = 5.1 \times h \times \ln\left(\frac{1}{0.25}\right) = 28 \text{ pH} \quad (13\text{-}80)$$

It is possible that the equivalent inductance of the vias in the board from the package pins may also add to this limiting inductance. If the power and ground cavity is 10 mils below the top surface, there will be about 210 pH per power and ground via. For 10 power and ground pin pairs, this is an equivalent via inductance of 21 pH, comparable to the spreading inductance. The combination may almost double the limiting inductance when the package pins look into the board to the capacitors.

TIP The vias in the board to the cavity and the spreading inductance in the cavity to the capacitors can easily dominate the impedance of the collection of capacitors when the target impedance is below 0.05 Ohm.

Based on the root cause of the spreading inductance there are only a few board level design features that can be engineered to reduce it.

1. Use thinner dielectric layers between the planes.
2. Use multiple planes in parallel.
3. Move the power and ground cavity closer to the top of the board.
4. Spread the capacitors farther out around the perimeter of the package. This has a limit once the capacitors are uniformly distributed around the perimeter.
5. Mount some of the capacitors under the package on the bottom of the board.

There are some design features that can be implemented in the package design if planned ahead of time. These would include:

1. Mount some of the capacitors within the ring of the power and ground pins of the package.
2. Spread the power and ground pins of the package to the outer perimeter of the package.

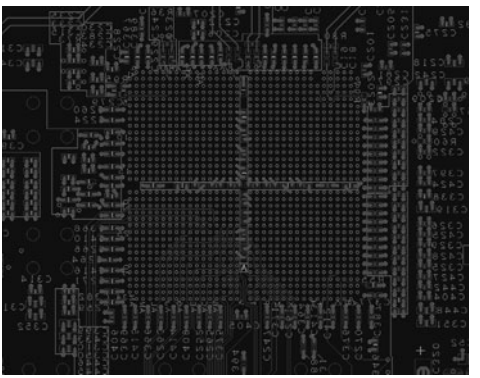


Figure 13-65 Package attach footprint of an Altera Stratix II GX FPGA showing the capacitor attach pads around the periphery and inside the BGA footprint.

3. Add decoupling capacitors on the package.

4. Use more power and ground pin pairs in parallel.

Figure 13-65 shows an example of adjusting the chip attach footprint of the package to allow decoupling capacitors inside the power/ground ring of pads.

When spreading inductance limits the series inductance, the worst thing to do would be clustering the capacitors, or clustering the power and ground pins into the core region of the package.

Because of the 3D nature of the current flows in the planes, the only way to accurately estimate the series inductance contribution from the planes given the capacitor locations and power pin locations is with a 3D field solver.

13.26 The Chip View

Through most of this chapter, we have considered the impedance profile seen by the board. Once the board level impedance is designed, what impact does this have on the chip-package combination?

In the following example, the conditions are:

- Target impedance is 0.01 Ohm.
- F_{max} is 100 MHz.
- Board area is 25 square inches.
- h is 4 mils.

- ESL per capacitor = 1 nH.
- On-die capacitance of 250 nF.

This combination puts us in the regime where the precise capacitor values don't matter and they can all be the same value. However, where possible, the lowest value of capacitors should be used so that the capacitors have the highest possible ESR and contribute to damping of parallel resonances. The theoretical minimum number of capacitors needed is:

$$n > 2\pi F_{max}\left(\frac{ESL}{Z_{target}}\right) = 2\pi 0.1\left(\frac{1}{0.01}\right) = 63 \quad (13\text{-}81)$$

The maximum allowable spreading inductance to the package would be:

$$L_{max} < \frac{Z_{target}}{2\pi F_{max}} = \frac{0.01}{2\pi \times 0.1} = 0.016 \text{ nH} = 16 \text{ pH} \quad (13\text{-}82)$$

As we saw above, it would be impossible to achieve this with a conventional BGA package and 4 mil thick layers with the capacitors on the top surface. The capacitors would have to be attached under the BGA, and the laminate thickness would have to be reduced or multiple planes added to the board. Only a 3D field solver would be able to confirm that these modifications provided low enough spreading inductance.

Given these modifications, the board level decoupling can be achieved with 63 0402 1 uF capacitors each with an ESL below 1 nH.

What will the chip see? Figure 13-66 shows the impedance profile from the board level and the chip level for this condition.

The low frequency impedance is optimized by designing the VRM and bulk decoupling capacitors. The board impedance is set by adding 63 MLCC capacitors to achieve the 0.01 Ohm impedance up to 100 MHz. This meets the condition for an optimized board level PDN design.

However, as viewed by the chip pads, looking through the package into the board, the on-die capacitance and the series package pin inductance and capacitor

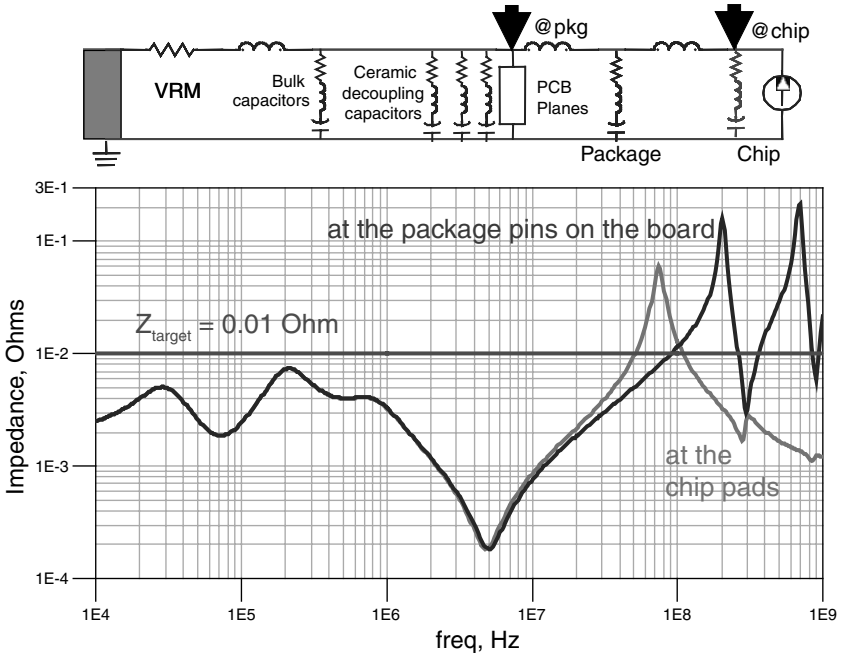


Figure 13-66 Impedance profile as seen at the package pins on the board and the chip pads.

equivalent inductance create a parallel resonance that results in a large impedance peak near 80 MHz in this case.

As with any parallel peak impedance, it is composed from a capacitance sloping down and an inductance sloping up. An important term that affects the peak impedance is the equivalent series resistance of the parallel resonant circuit. This is the combination of the ESR of the parallel combination of the capacitors and the series resistance in the package leads and on the die itself.

When the equivalent series resistance is low, the peak impedance can be high. For capacitors near 1 uF, their ESR is on the order of 10 mOhms. With 63 in parallel, the equivalent resistance is reduced to about 0.0002 Ohm, well below the target impedance level. If 10 nF capacitors were used, their ESR would be about 70 mOhms each with a total series resistance about 0.001 Ohm. This is better than using all higher values of capacitors, but still not enough to damp out the parallel resonances. The series resistance of the package traces will be below 0.001 Ohm. This results in a series resistance that might be dominated by the on-die metallization.

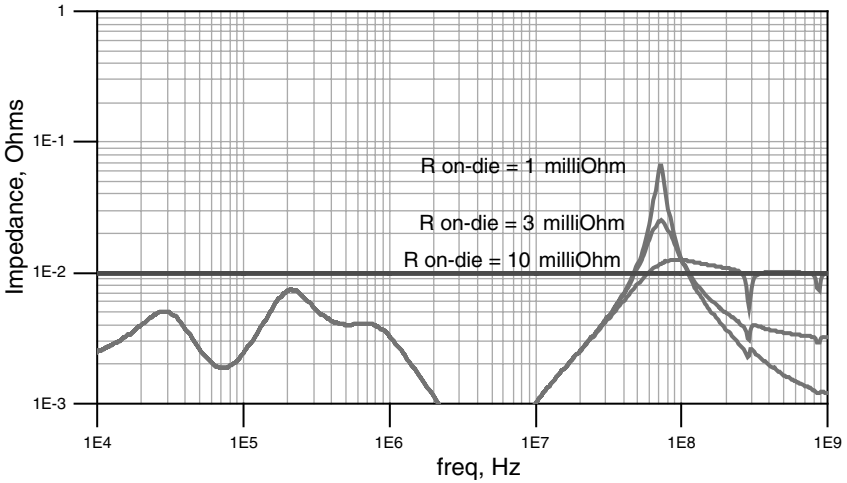


Figure 13-67 The impact on the impedance profile as seen by the chip, from different on-die series resistance in the PDN.

T I P The resistance from on-die interconnects is an important term that can be used to engineer a lower peak impedance.

Figure 13-67 shows the impact different on-die resistance can have. The resistance is changed from 1 milliOhm, to 3 milliOhms, and 10 milliOhms. It has to be lower than 10 mOhms to meet the target impedance spec, but some resistance on-die is a good thing.

Even if the on-die resistance is optimized, the peak impedance at the parallel resonance will still be above the target. This is where on-package decoupling capacitors can play a role in suppressing the peak impedance.

Figure 13-68 shows the impact on the impedance profile the chip sees looking into the PDN as the number of on-package decoupling capacitors is increased from none to 10. Each has an ESL of 0.1 nH, which is typically implemented using IDC capacitors.

In this example, when 10 IDC capacitors, each with an ESL of 0.1 nH, is added to the package, the impedance profile seen by the chip is as shown in Figure 13-69, meeting the 0.01 Ohm target impedance from DC to very high bandwidth.

13.27 Bringing It All Together

The most important feature of the PDN is how low it keeps the voltage noise on the pads of the chip. This is fundamentally related to the current draw of the chip

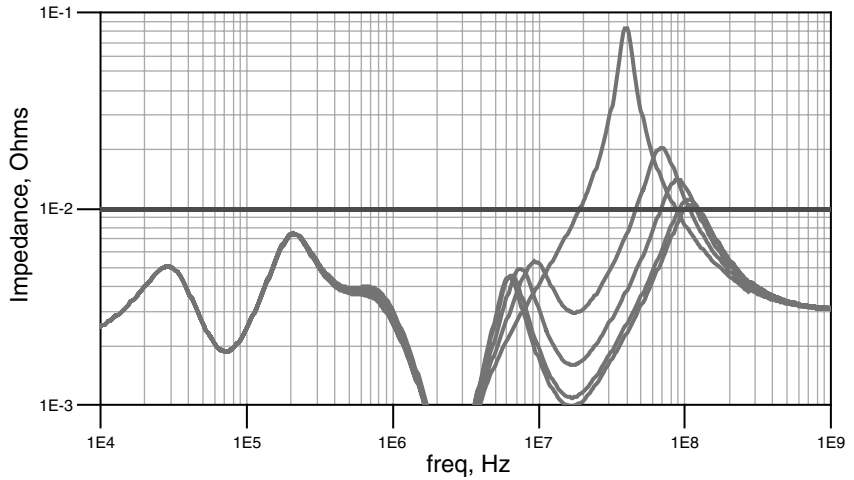


Figure 13-68 The impedance profile seen by the chip as more on-package decoupling capacitors are added, using none, 3, 6, 8, and 10.

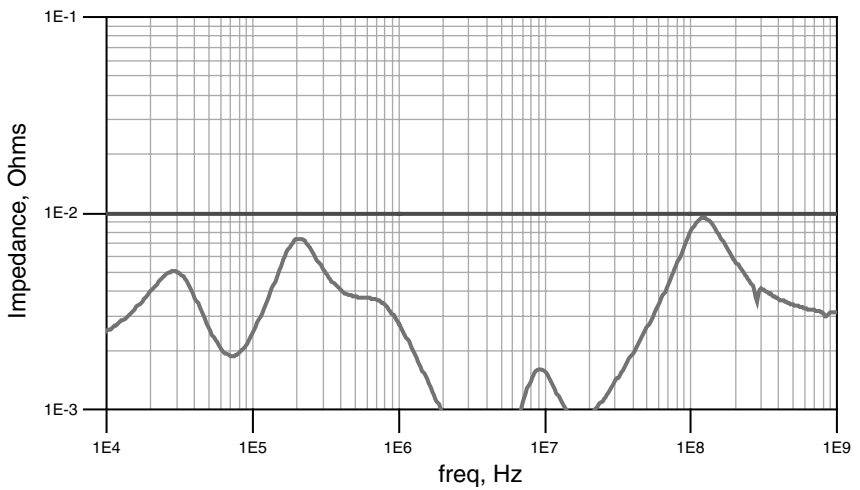


Figure 13-69 Impedance profile as seen by the chip-pads, meeting the 0.01 Ohm target impedance.

and the impedance profile from the chips' pads to the VRM. Unfortunately, this depends very strongly on factors that are often outside the control of the board level designer, such as:

- Current spectrum drawn by the chip operations
- On-die capacitance

- On-die resistance
- Chip attach inductance
- Package attach inductance
- On-package decoupling capacitors

In the highest performance systems, where the target impedance is in the milliOhm range, such as for high-end processors, servers, and graphics chips, system co-design, simultaneously optimizing the chip features, package features, and board features, is critically important for a robust, cost-effective design.

TIP Those companies that implement system level PDN design will end up being the most successful.

But for most designs, there is nothing the board level designer can do about the chip or the package. It is limited by the semiconductor provider. While it is always important to continue to push the envelope and ask for the chip and package models to incorporate in the board level PDN design, it is rarely possible to get all the important information required.

The board level designer is faced with having to follow the guideline that "sometimes an OK answer NOW is better than a good answer late," and base the board level PDN design on reasonable assumptions.

If enough on-package decoupling capacitance is provided, the requirements for on-board decoupling can be dramatically reduced. For example, in the example above, the 10 on-package decoupling capacitors provide an equivalent inductance of $1/10 \times 100$ pH = 10 pH. This interacts with the on-die capacitance to keep the peak impedance below the target impedance in the 100 MHz region and above.

This inductance is on the chip side of the interconnects between the package and the board. The equivalent package-attach inductance to the board could be large and not affect the high frequency impedance the chip sees. In fact, it could be as large as 0.1 nH. This is illustrated in Figure 13-70.

If the package inductance is 0.1 nH, the highest frequency the board can affect the impedance the chip sees is:

$$F_{max} = \frac{Z_{target}}{2\pi L_{pkg}} = \frac{0.01}{2\pi \times 0.1} = 0.016 \text{ GHz} = 16 \text{ MHz} \quad (13\text{-}83)$$

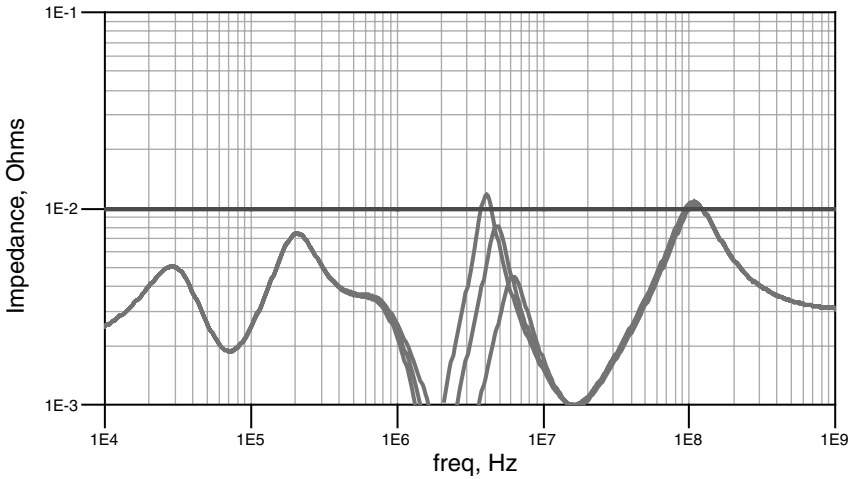


Figure 13-70 Impedance profile the chip sees with on-package capacitors providing the low impedance at high frequency and changing the package-attached inductance from 0.05 nH, to 0.1 nH, and 0.15 nH.

where:

Z_{target} = the target impedance in Ohms

L_{pkg} = the inductance of the package attach to the circuit board in nH

F_{max} = the highest frequency the board level impedance can affect the chip-pad impedance, in GHz

The use of package level decoupling capacitors and higher package attach inductance can mean that the requirements for the board level decoupling are reduced. If the maximum target impedance is no longer 100 MHz but 16 MHz, then fewer capacitors are required to meet the same impedance level. Of course, the same considerations about parallel resonance peak impedances at the interfaces must be taken into account. Figure 13-71 shows that using the lower ESL capacitors on-package can reduce the number of board level capacitors required from 63 to 15 and still meet the target impedance.

TIP When there are on-package capacitors, the board level requirements can be reduced. Even then, selecting the capacitors is not about the capacitor value, but about the ESR and ESL they provide.

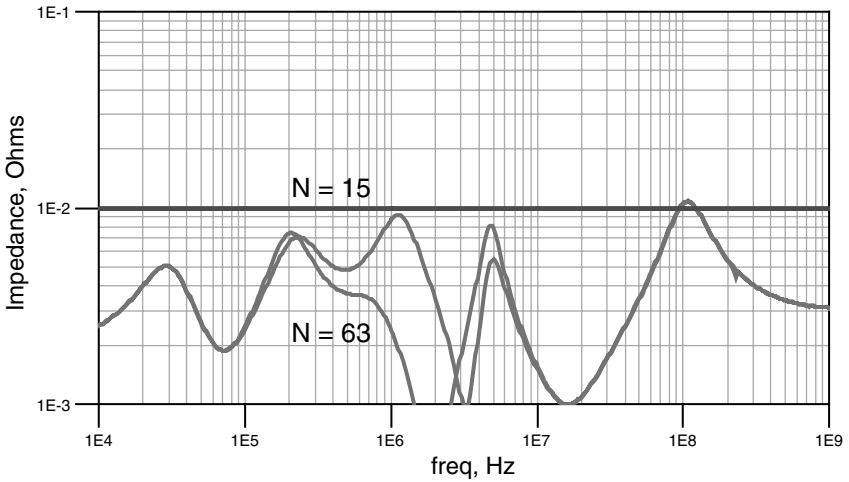


Figure 13-71 Impedance profile chip sees with on-package capacitors and reducing the board level capacitors from 63 to 15 while still maintaining the impedance target.

When semiconductor vendors incorporate on-package decoupling capacitors, they often specify the recommended board level decoupling requirements in terms of how many capacitors and what value should be added to the board. This is rarely useful information, as the number of capacitors to add to the board is not about their capacitance, but about their ESL.

13.28 The Bottom Line

1. PDN design is confusing, contradictory, complicated, and complex. This is fundamentally due to the complex interactions of many features that are outside the control of the board designer, and often poorly documented.
2. The goal of the PDN is to provide a low impedance from the chip pads to the VRM.
3. The VRM and bulk decoupling capacitors provide low impedance at low frequency.
4. Do everything possible to integrate the bulk capacitors with low loop inductance.
5. The chip and package design have the most influence on the impedance in the 100 MHz and above region. The most important design guidelines are

add more on-die decoupling capacitance, keep the chip attach inductance low, and add low inductance decoupling capacitors to the package.

6. The most important starting place in PDN design is establishing a target impedance. This can be approximated based on the worst case power consumption of the device.

7. The package lead inductance and circuit board via to the power/ground cavity will fundamentally limit the high-frequency design limit for the board level impedance.

8. At the board level, everything possible should be done to reduce the loop inductance from the capacitors to the packages.

9. The most important design guidelines are to use power and ground planes on adjacent layers, with thin dielectric, placed close to the surface of the board, use short and wide surface traces between the capacitors and their vias to the cavity, and place the capacitors in proximity to the package on the top surface, and when spreading inductance is saturated, place some directly beneath the packages on the bottom of the board.

10. The most important terms that influence the number of capacitors to use is their ESL and the maximum frequency for the board level decoupling.

11. When the board planes' capacitance creates a parallel resonance with the decoupling capacitors, at a low frequency, you can use the fewest number of capacitors by carefully selecting their values using SPICE to simulate the impedance profile.

12. For very low impedance designs, the values of the capacitors are not as important as their ESL and the number used. All the same value capacitors can work well.

13. When parallel resonant peaks are involved the lowest capacitance capacitors will have the highest ESR and provide some damping to reduce the peak impedance.

14. For very low impedance, on-package capacitors are essential. Their use may decrease the on-board capacitor requirements.

15. For the very lowest impedance PDN designs, co-design of the chip, package, and board level PDN will provide the best cost-performance solutions.

APPENDIX A

100 General Design Guidelines to Minimize Signal-Integrity Problems

Never follow a rule blindly. Always understand its purpose and put in the numbers to evaluate its cost/benefit for your specific design.

0. Always use the longest rise time you can.

A.1 Minimize Signal-Quality Problems on One Net

Strategy: Keep the instantaneous impedance the signal sees constant throughout its entire path.

Tactics:

1. Use controlled-impedance traces.

2. Ideally, all signals should use the low-voltage planes as their reference planes.

3. If different voltage planes are used as signal references, there should be tight coupling between them. Do so by using the thinnest dielectric you can afford and multiple, low-inductance decoupling capacitors between the different voltage planes.

4. Use a 2D field solver to calculate the stack-up design rules for the target characteristic impedance. Include the effects of solder-mask and trace thickness.
5. Use series termination for point-to-point topologies, whether single or bidirectional.
6. Terminate both ends of the bus in a multidrop bus.
7. Keep the time delay of stubs less than 20% of the rise time of the fastest signals.
8. Place the terminating resistors as close to the package pads as possible.
9. Don't worry about corners unless 10 fF of capacitance is important.
10. Follow the return path of each signal and keep the width of the return path under each signal path at least as wide, and preferably at least 3 times as wide, as the signal trace.
11. Route signal traces around rather than across return-path discontinuities.
12. Avoid using engineering change wires in any signal path.
13. Keep all nonuniform regions as short as possible.
14. Do not use axial-lead terminating resistors for system rise times less than 1 nsec. Use SMT resistors and mount them for minimum loop inductance.
15. When rise times are less than 150 psec, do everything possible to minimize the loop inductance of the terminating SMT resistors or consider using integrated or embedded resistors.
16. Vias generally look capacitive. Minimizing the capture pads and increasing the antipad clearance diameter will help make the via look transparent.
17. Consider adding a little capacitance to the pads of a low-cost connector to compensate for its typically higher inductance.
18. Route all differential pairs with a constant differential impedance.
19. Avoid all asymmetries in a differential pair. Whatever you do to one trace, do the same to the other.
20. If the spacing between the traces in a differential pair has to change, adjust the line width to keep a constant differential impedance.
21. If a delay line is to be added to one leg of a differential pair, add it near the beginning of the trace and keep the traces uncoupled in this region.

22. It is okay to change the coupling in a differential pair as long as the differential impedance is maintained.

23. In general, route differential-pair traces with as tight a coupling as practical.

24. Decide on edge- versus broadside-coupled differential pairs, based on routing density, total board-thickness constraints, and ability of the fab vendor to maintain tight laminate thickness control. Performance-wise, they can be equivalent.

25. For any board-level differential pairs, there will be significant return current in the planes, so avoid all discontinuities in the return path. If there is a discontinuity, do exactly the same thing to each line in the pair.

26. Worry about terminating the common signals only if the common-mode rejection ratio of the receiver is poor. Terminating the common signals will not eliminate the common signal, just minimize its ringing.

27. If losses are important, use as wide a signal trace as possible, and never use a trace of less than 5 mils.

28. If losses are important, keep traces as short as possible.

29. If losses are important, do everything possible to minimize all capacitive discontinuities.

30. If losses are important, engineer the signal-vias to have a 50-Ohm impedance, which usually means do everything possible to decrease the barrel size, decrease the capture-pad size, and increase the antipad-clearance holes.

31. If losses are important, use as low a dissipation-factor laminate as you can afford.

32. Consider using pre-emphasis and equalization if losses are important.

A.2 Minimize Cross Talk

Strategy: Minimize mutual capacitance and mutual inductance between signal and return paths.

Tactics:

33. For microstrip or stripline transmission lines, keep the spacing between adjacent signal paths at least twice the line width.

34. Minimize any discontinuities in the return path the signals might cross over.

35. If you have to cross a gap in the return path, only use differential pairs. Never cross a gap with single-ended signals routed close together.

36. For surface traces, keep the coupled lengths as short as possible and use as much solder mask as practical to minimize far-end cross talk.

37. If far-end cross talk is a problem, add a laminate layer to the top of the surface traces to make them embedded microstrip.

38. For long, coupled lengths where far-end cross talk may be a problem, route the traces in stripline.

39. If you can't keep the coupling length less than the saturation length, changing the coupling length will have no impact on the near-end cross talk, so don't worry about decreasing coupling length.

40. Use the lowest dielectric constant laminate you can afford so the dielectric spacing to the return planes can be kept to a minimum for the same target characteristic impedance.

41. In a tightly coupled microstrip bus, the deterministic jitter can be reduced by keeping the spacing at least as wide as twice the line width or by routing timing-sensitive lines in stripline.

42. For isolations in excess of –60 dB, use stripline with guard traces.

43. Always use a 2D field solver to evaluate whether you need to use a guard trace.

44. If you do use a guard trace, make it as wide as will fit and use vias to short the ends to the return path. Add additional shorting vias along the length if it is free and easy to do so. They are not as critical as the two on the ends.

45. Minimize ground bounce by making the return paths in any packages or connectors as short and as wide as possible.

46. Use chip-scale packages rather than larger packages.

47. Minimize ground bounce in the power return path by bringing the power plane as close to the return plane as possible.

48. Minimize ground bounce in the signal return paths by bringing the signal path as close to the return path as acceptable, consistent with matching the impedance of the system.

49. Avoid using shared return paths in connectors and packages.

50. When assigning leads in a package or connector, reserve the shortest leads for the ground paths and space the power and ground leads uniformly among the signal paths, or closest to those signal paths that will carry a lot of switching current.

51. All no-connect leads or pins should be assigned as ground-return connections.

52. Avoid using resistor single inline packages (SIPs) unless there are separate return paths for each resistor.

53. Check the film to verify that antipads in via fields do not overlap and there is a well-defined web between clearance holes in the power and ground planes.

54. If a signal changes reference planes, the reference planes should be as closely spaced as you can afford. If you use a decoupling capacitor to minimize the impedance of the return path, its capacitance value is immaterial. Select it and design it in for lowest loop inductance.

55. If many signal lines switch reference planes, space the signal path vias as far apart as possible, rather than clustering them all in the same location.

56. If a signal switches reference layers, and the planes are the same voltage level, place a via between the return planes as close to the signal via as possible.

A.3 Minimize Rail Collapse

Strategy: Minimize the impedance of the power-distribution network.

Tactics:

57. Minimize the loop inductance between the power and ground paths.

58. Allocate power and ground planes on adjacent layers with as thin a dielectric as you can afford.

59. Get the lowest impedance between the planes by using as high a dielectric constant between the planes as you can afford.

60. Use as many power- and ground-plane pairs in parallel as you can afford.

61. Route the same currents far apart and opposite currents close together.

62. Place each power via as close as practical to a ground via. If you can't get them at least within a pitch equal to their length, there will be no coupling and no value in proximity.

63. Route the power and ground planes as close as possible to the surface where the decoupling capacitors are mounted.

64. Use multiple vias to the same power or ground pad, but keep the vias as far apart as possible.

65. Use vias as large in diameter as practical when routing to power or ground planes.

66. Use double-bonding on power and ground pads to minimize the loop inductance of the wire bonds.

67. Use as many power and ground connections from the chip as you can afford.

68. Use as many power and ground connections from the package as you can afford.

69. Use chip-interconnect methods that are as short as possible, such as flip-chip rather than wire-bond.

70. Use package leads as short as possible, such as chip-scale packages rather than QFP packages.

71. Keep all surface traces that run between the pads of the decoupling capacitors and their vias as short and wide as possible.

72. Use a total amount of bulk-decoupling capacitance to take over from the regulator at low frequency.

73. Use a total number of decoupling capacitors to reduce the equivalent inductance at high frequency.

74. Use as small a body size for a decoupling capacitor as you can afford and minimize the length of all connections from the capacitor pads to the power and ground planes.

75. Place as much decoupling capacitance as you can afford on the chip itself.

76. Place as many low-inductance decoupling capacitors as you can afford on the package.

77. Use differential pairs for I/Os.

A.4 Minimize EMI

Strategy: Reduce the voltage that drives common currents, increase the impedance of the common current paths, and shield and filter as a quick fix.

Tactics:

78. Reduce ground bounce.

79. Keep all traces at least five line widths from the edge of the board.

80. Route traces in stripline when possible.

81. Place the highest-speed/highest-current components as far from the I/O connections as possible.

82. Place the decoupling capacitors in proximity to the chips to minimize the spread of high-frequency-current components in the planes.

83. Keep power and ground planes on adjacent layers and as close together as possible.

84. Use as many power- and ground-plane pairs as you can afford.

85. When using multiple power- and ground-plane pairs, recess the power planes and then stitch shorting vias between the ground planes along the edges.

86. Use ground planes as surface layers, where possible.

87. Know the resonant frequency of all packages and change the package geometry if there is an overlap with a clock harmonic.

88. Avoid signals switching different voltage reference planes in a package. This will drive package resonances.

89. Add ferrite filter sheets to the top of packages if they might have a resonance.

90. Minimize any asymmetries between the lines in each differential pair

91. Use a common-signal-choke filter on all differential pair connections

92. Use a common-signal-choke filter around the outside of all peripheral cables.

93. Filter all external I/O lines to use the longest signal rise time that is tolerable for the timing budget.

94. Use spread-spectrum clock generator to spread the first harmonic over a wider frequency range and decrease the radiated energy within the bandwidth of the FCC test.

95. When connecting shielded cables, try to keep the shield as an extension of the enclosure.

96. Minimize the inductance of the shielded cable connections to the enclosure. Use a coaxial connection right from the end of the cable and to the enclosure.

97. Equipment bays should not penetrate the integrity of the enclosure.

98. Only interconnects need to break the enclosure integrity.

99. Keep aperture diameters small, significantly smaller than a wavelength of the lowest frequency radiation that might leak. More and smaller holes are better than fewer and larger holes.

100. The most expensive rule is the one that delays the product ship date.

APPENDIX B

100 Collected Rules of Thumb to Help Estimate Signal-Integrity Effects

A rule of thumb should be used only when an OK answer *right now* is more important than a good answer later.

A rule of thumb is only meant to be a very rough approximation. It is designed to help feed our intuition and to help find a quick answer, with very little effort. It should always be the starting place for any estimate. It can help us distinguish between a 5 or a 50. It can help us see the big picture, early in the design phase. In the balance between quick and accurate, a rule of thumb is quick. It is not meant to be accurate.

Of course, you can't use a rule of thumb blindly. It must be coupled with an understanding of the principles and good engineering judgement.

When accuracy is important, such as in a design sign-off where being off by a few percent may have a $1M impact, why use anything other than a verified numerical simulation tool?

The following is a collection of the most useful rules of thumb, separated by the chapter in which they are introduced:

B.1 Chapter 2

1. The rise time of a signal is ~ 10% the clock period and ~ $1/10 \times 1/F_{clock}$. For example, a 100-MHz clock has a rise time of about 1 nsec.

2. The amplitude of the n^{th} harmonic of an ideal square wave is ~ $2/(\pi n)$ times the magnitude of the clock voltage. For example, the first harmonic amplitude of a 1-v clock signal is about 0.6. The third harmonic is about 0.2.

3. The bandwidth, BW, and rise time, RT, of a signal are related by BW = 0.35/RT. For example, if the rise time is 1 nsec, the bandwidth is 350 MHz. If the bandwidth of an interconnect is 3 GHz, the shortest rise time it can transmit is about 0.1 nsec.

4. If you don't know the rise time, the bandwidth of a signal is roughly five times the clock frequency. For example, if the clock frequency is 1 GHz, the bandwidth of the signal is about 5 GHz.

B.2 Chapter 3

5. The resonant frequency of an LC circuit is 5 GHz/sqrt(LC) with L in nH and C in pF. For example, a package lead and its return path might have a loop self-inductance of 7 nH. Its capacitance might be about 1 pF. The frequency at which it would ring would be about 2 GHz.

B.3 Chapter 4

6. An axial lead resistor looks like an ideal resistor up to about 400 MHz. An SMT 0603 resistor looks like an ideal resistor up to about 2 GHz.

7. The ESL of an axial lead resistor is about 8 nH. The ESL of an SMT resistor is about 1.5 nH.

8. The resistance per length of a 1-mil-diameter gold wire bond is about 1 Ohm/inch. For example, a 50-mil-long wire bond has a resistance of about 50 milliOhms.

9. A length of 24 AWG wire has a diameter of about 20 mils, and a resistance per length of about 25 milliOhms/foot.

10. The sheet resistance of 1-ounce copper is about 0.5 milliOhm/square. For example, a trace 5 mils wide and 1 inch long has 200 squares and would have a series resistance of 200 × 0.5 = 100 milliOhms = 0.1 Ohm.

11. The skin-depth effects for 1-ounce copper begin at about 10 MHz.

B.4 Chapter 5

12. The capacitance of a sphere 1 inch in diameter is about 2 pF. For example, a pigtail cable hanging off a board a few inches long will have a capacitance to the floor of about 2 pF.

13. The capacitance of a pair of plates the size of a penny, with air between the faces, is about 1 pF.

14. When the plates of a capacitor are as far apart as the plates are wide, the fringe fields contribute just as much capacitance as the parallel-plate fields. For example, if we estimate the parallel-plate capacitance for a microstrip with line width of 10 mils and dielectric thickness of 10 mils, to be 1 pF/inch, the actual capacitance will be about twice this value, or 2 pF/inch.

15. If we don't know anything else about the material except that it is an organic laminate, a good estimate for its dielectric constant is 4.

16. For a 1-watt chip, the amount of time, in sec, a decoupling capacitance, in F, will provide the charge with less than 5% voltage droop is C/2. For example, if there is a decoupling capacitance of 10 nF, it will provide decoupling for only 5 nsec. If we require 10 microseconds of decoupling, we need 20 uF of capacitance.

17. The capacitance available in the power and ground planes of a typical circuit board when the separation is 10 mils is 100 pF/square inch and it scales inversely with the dielectric thickness. For example, in a 10-mil separation board, the total board area available for decoupling an ASIC may be only 4 square inches. The capacitance would be 0.5 nF. It would provide decoupling for about 0.2 nsec.

18. The effective dielectric constant in a 50-Ohm microstrip is 3, when the bulk dielectric constant is 4.

B.5 Chapter 6

19. The partial self-inductance of a round wire, 1 mil in diameter, is about 25 nH/inch or 1 nH/mm. For example, the partial self-inductance of a via, 1.5 mm long, is about 1.5 nH.

20. The loop self-inductance of a 1-inch-diameter round loop, made from 10-mil-thick wire about the size of your finger and thumb held together in a circle, is 85 nH.

21. The total inductance per length for a section taken out of a 1-inch-diameter loop is about 25 nH/inch or 1 nH/mm. For example, if a package lead is part of a loop and is 0.5 inch long, it has a total inductance of about 12 nH.

22. When the center-to-center separation of a pair of round rods is 10% of their length, the partial mutual inductance will be about 50% of the partial self-inductance of either one. For example, if we have two wire bonds, 1 mm long on 0.1-mm centers, the partial self-inductance of either one is about 1 nH, while their partial mutual inductance will be about 0.5 nH.

23. When the center-to-center separation between two round rods is about equal to their length, the partial mutual inductance between them is less than 10% of the partial self-inductance of either one. For example, if we space 25-mil-long vias more than 25 mils center-to-center, there is virtually no inductive coupling between them.

24. The loop inductance of an SMT capacitor is roughly 2 nH including the surface traces, vias, and capacitor body. Heroic efforts are required to reduce this below 1 nH.

25. The loop inductance per square of a pair of planes is 33 pH per square per mil of spacing. For example, if the dielectric spacing is 2 mils, there is 66 pH/square of loop inductance between the planes.

26. The larger the via diameter, the lower its spreading inductance. A 25-mil-diameter via will have a spreading inductance of about 50 pH.

27. The loop inductance of a pair of planes having a field of clearance holes with 50% open area will increase by 25%.

28. The skin depth in copper is 2 microns at 1 GHz and increases with the square root of frequency. For example, at 10 MHz, the skin depth is 20 microns.

29. In a 50-Ohm transmission line of 1-ounce copper, the loop inductance per length is constant at frequencies above about 50 MHz. This means characteristic impedance is constant above 50 MHz.

B.6 Chapter 7

30. The speed of electrons in copper is about the speed of an ant, or 1 cm/sec.

31. The speed of a signal in air is about 12 inches/nsec. The speed of a signal in most polymer materials is about 6 inches/nsec.

32. The wiring delay, 1/v, in most laminates is about 170 psec/inch.

33. The spatial extent of the signal is the rise time times the speed, or RT × 6 inches/nsec. For example, if the rise time is 0.5 nsec, the spatial extent of the edge as the signal propagates down a board is 3 inches.

34. The characteristic impedance of a transmission line varies inversely with the capacitance per length.

35. All 50-Ohm lines in FR4 have a capacitance per length of about 3.3 pF/inch. For example, if a BGA lead is designed as 50 Ohms and is 0.5 inch long, it has a capacitance of about 1.7 pF.

36. All 50-Ohm lines in FR4 have an inductance per length of about 8.3 nH/inch. For example, if a connector is designed for 50 Ohms and is 0.5 inch long, the loop inductance of the signal and return path is about 4 nH.

37. A 50-Ohm microstrip in FR4 has a dielectric thickness about half the line width of the trace. For example, if the line width is 10 mils, the dielectric thickness will be 5 mils.

38. A 50-Ohm stripline in FR4 has a plane-to-plane spacing about twice the line width of the signal trace. For example, if the line width is 10 mils, the spacing between the two planes will be 20 mils.

39. The impedance looking into a transmission line will be the characteristic impedance for a time shorter than the round-trip time of flight. For example, if a 50-Ohm line is 3 inches long, all drivers with a rise time shorter than 1 nsec will see a constant 50-Ohm load during the transition time when driving the line.

40. The total capacitance in a section of transmission line with a time delay of TD is $C = TD/Z_0$. For example, if the TD of a line is 1 nsec and the characteristic impedance is 50 Ohms, there is 20 pF of capacitance between the signal and return paths.

41. The total loop inductance in a section of a transmission line with a time delay of TD is $L = TD \times Z_0$. For example, if the TD of a line is 1 nsec and the characteristic impedance is 50 Ohms, there is 50 nH of loop inductance between the signal and return paths.

42. If the width of the return path in a 50-Ohm microstrip is equal to the signal line width, the characteristic impedance is 20% higher than the characteristic impedance when the return path is infinitely wide.

43. If the width of the return path in a 50-Ohm microstrip is at least three times the signal line width, the characteristic impedance is within 1% of the characteristic impedance when the return path is infinitely wide.

44. Trace thickness will decrease the characteristic impedance of a line by about 2 Ohms per mil of thickness. For example, from 1/2-ounce copper to 1-ounce copper, the thickness increases by 0.7 mil. The impedance of the line would decrease by about 1 Ohm.

45. Solder mask on top of a microstrip will decrease its characteristic impedance by about 2 Ohms per mil of thickness. For example, a 0.5-mil-thick solder mask will decrease the characteristic impedance by about 1 Ohm.

46. For an accurate lumped-circuit approximation, you need at least 3.5 LC sections per spatial extent of the rise time. For example, if the rise time is 1 nsec, in FR4, the spatial extent is 6 inches. You would need at least 3.5 LC sections per 6 inches, or about one section every two inches of trace for an accurate approximation.

47. The bandwidth of a single-section LC model is 0.1/TD. For example, if the time delay of a transmission line is 1 nsec, using a single LC section to model it would be accurate up to a bandwidth of about 100 MHz.

B.7 Chapter 8

48. If the time delay of a transmission line is less than 20% of the rise time of the signal, you may not have to terminate the line.

49. In a 50-Ohm system, an impedance change of 5 Ohms will give a reflection coefficient of 5%.

50. Keep all discontinuities shorter, in inches, than the rise time, in nsec. For example, if the rise time is 0.5 nsec, keep all impedance discontinuities, such

as a neck down to pass through a via field, less than 0.5 inch long and it may be acceptable.

51. A capacitive load at the far end will increase the rise time of the signal. The 10–90 rise time is about 100 × C psec, for C in pF. For example, if the capacitance is 2 pF, typical of the input-gate capacitance of a receiver, the RC limited rise time would be about 200 psec.

52. If the capacitance of a discontinuity is less than 0.004 × RT, it may not cause a problem. For example, if the rise time is 1 nsec, capacitive discontinuities should be less than 0.004 nF, or less than 4 pF.

53. The capacitance, in fF, of a corner in a 50-Ohm line is twice the line width, in mils. For example, if the line width of a 50-Ohm trace is 10 mils, a 90-degree bend would have a capacitance of about 20 fF. It might cause reflection problems for rise times of 0.02 pF/4 = 5 psec.

54. A capacitive discontinuity will add a delay time to the 50% threshold point of about $0.5 \times Z_0 \times C$. For example, if the capacitance is 1 pF in a 50-Ohm line, the delay adder will be about 25 psec.

55. If the inductance, in nH, of a discontinuity is less than 10 times the rise time, in nsec, it may not cause a problem. For example, if the rise time is 1 nsec, the maximum inductive discontinuity that may be acceptable is about 10 nH.

56. An axial lead resistor with a loop inductance of about 10 nH may contribute too much reflection noise for rise times less than 1 nsec. Switch to surface-mount resistors.

57. To compensate for a 10-nH inductance, a 4-pF capacitance in a 50-Ohm system is required.

B.8 Chapter 9

58. At 1 GHz, the resistance of a 1-ounce copper trace is about 15 times the resistance at DC.

59. At 1 GHz, the attenuation from the resistance of an 8-mil-wide trace is comparable to the attenuation from the dielectric, and the attenuation from the dielectric will get larger faster with frequency.

60. The low-loss regime, for lines 3 mils or wider, is all frequencies above about 10 MHz. In the low-loss regime, the characteristic impedance and signal

speed are independent of the loss and of frequency. There is no dispersion due to loss in typical board-level interconnects.

61. A –3-dB attenuation is a drop to 50% of the initial power level and a drop to 70% of the initial amplitude.

62. A –20-dB attenuation is a drop to 1% of the initial power level and a drop to 10% of the initial amplitude.

63. When in the skin-depth regime, the series resistance per inch of a signal- and return-path line, of width w in mils, is about 8/w × sqrt(f), with f in GHz. For example, a line 10 mils wide, has a series resistance of about 0.8 Ohm/inch and it increases with the square root of frequency.

64. In a 50-Ohm line, the attenuation from the conductor is about $36/wZ_0$ in dB per inch. For example, if the line width is 10 mils in a 50-Ohm line, the attenuation is 36/(10 × 50) = 0.07 dB/inch.

65. The dissipation factor of FR4 is about 0.02.

66. In FR4, the attenuation from the dielectric is about 0.1 dB/inch at 1 GHz and increases linearly with frequency.

67. At 1 GHz, a 50-Ohm line in FR4 with a line width of 8 mils has the same conductor loss as the dielectric loss at 1 GHz.

68. The bandwidth of an FR4 interconnect, Len inches long, when limited by dissipation factor is 30 GHz/Len. For example, a 50-Ohm line, 10 inches long, has a bandwidth of 3 GHz.

69. The shortest rise time that can be propagated by an FR4 interconnect is 10 psec/inch × Len. For example, a signal propagating down a 10-inch length of 50-Ohm line in FR4 will have a rise time of at least 100 psec.

70. The rise-time degradation from losses will be important in FR4 laminates if the interconnect length, in inches, is greater than 50 times the rise time, in nsec. For example, if the rise time is 200 psec, worry about losses when line lengths are greater than 10 inches.

B.9 Chapter 10

71. In a pair of 50-Ohm microstrip transmission lines, with spacing equal to the line width, the coupling capacitance between the signal lines is about 5%.

72. In a pair of 50-Ohm microstrip transmission lines, with spacing equal to the line width, the coupling inductance between the signal lines is about 15%.

73. The saturation length for near-end noise in FR4 is 6 inches for a 1-nsec rise time and it scales with the rise time. For example, if the rise time is 0.5 nsec, the saturation length is 3 inches.

74. The loaded capacitance of a trace is constant and independent of the proximity of other traces nearby.

75. The near-end cross talk, for 50-Ohm microstrip traces with spacing equal to the line width, is 5%.

76. The near-end cross talk, for 50-Ohm microstrip traces with spacing equal to the line width, is 2%.

77. The near-end cross talk, for 50-Ohm microstrip traces with spacing equal to the line width, is 1%.

78. The near-end cross talk, for 50-Ohm stripline traces with spacing equal to the line width, is 6%.

79. The near-end cross talk, for 50-Ohm stripline traces with spacing equal to the line width, is 2%.

80. The near-end cross talk, for 50-Ohm stripline traces with spacing equal to the line width, is 0.5%.

81. In a pair of 50-Ohm microstrip traces,with spacing equal to the line width, the far-end noise is 4% × TD/RT. If the time delay of the line is 1 nsec and the rise time is 0.5 nsec, the far-end noise is 8%.

82. In a pair of 50-Ohm microstrip traces, with spacing equal to twice the line width, the far-end noise is 2% × TD/RT. If the time delay of the line is 1 nsec and the rise time is 0.5 nsec, the far-end noise is 4%.

83. In a pair of 50-Ohm microstrip traces, with spacing equal to three times the line width, the far-end noise is 1.5% × TD/RT. If the time delay of the line is 1 nsec and the rise time is 0.5 nsec, the far-end noise is 3%.

84. There is no far-end noise in stripline or fully embedded microstrip.

85. In a 50-Ohm bus, stripline or microstrip, to keep the worst case near-end noise below 5%, keep the spacing between the lines more than twice the line width.

86. In a 50-Ohm bus, with a spacing equal to the line width, 75% of all the cross talk on any victim line is from the trace on either side of the victim line.

87. In a 50-Ohm bus, with a spacing equal to the line width, 95% of all the cross talk on any victim line is from the nearest two traces on either side of the victim line.

88. In a 50-Ohm bus, with a spacing equal to twice the line width, 100% of all the cross talk on any victim line is from the trace on either side of the victim line. You can ignore the coupling from all other traces in the bus.

89. For surface traces, separating the adjacent signal traces sufficiently to add a guard trace will often reduce the cross talk to an acceptable level, eliminating the need for a guard trace. Adding a guard trace with the ends shorted can reduce the cross talk by about 50%.

90. For stripline traces, using a guard trace can reduce the cross talk to less than 10% of that without the guard trace.

91. To keep the switching noise below an acceptable level, keep the mutual inductance less than 2.5 nH times the rise time, in nsec. For example, if the rise time is 0.5 nsec, the mutual inductance should be less than 1.3 nH for acceptable switching-noise cross talk, due to coupling between only two signal/return-path pairs.

92. For a connector or package that is limited by switching noise, the maximum usable clock frequency is 250 MHz/($n \times L_m$), for a mutual inductance between pairs of signal and return paths, in nH, and a number of simultaneous switching lines, n. For example, if there are four pins that share the same return path and their mutual inductance is about 1 nH between pairs, the maximum usable clock frequency for the connector will be 250 MHz/4 ~ 60 MHz.

B.10 Chapter 11

93. In LVDS signals, the common signal component is more than two times the differential signal component.

94. With no coupling, the differential impedance in a differential pair is twice the single-ended impedance of either line.

95. In a pair of 50-Ohm microstrip lines, the single-ended characteristic impedance of one line is completely independent of the proximity of the adjacent line, as long as the other line is tied low or tied high.

96. In the tightest coupled differential microstrip, a line width equal to the spacing, the differential impedance drops only about 10% from the differential impedance when the traces are far apart with no coupling.

97. In a broad-side coupled differential pair, the trace-to-trace separation must be at least larger than the line width in order to have the possibility of getting as high as a 100-Ohm differential impedance.

98. The FCC class B requirement is a far-field strength less than about 150 microV/m at 100 MHz and 3-meters distance.

99. A highly coupled differential pair will have 30% less differential-signal cross talk than a weakly coupled pair, from a closely spaced single-ended aggressor.

100. A highly coupled differential pair will have about 30% more common-signal cross talk than a weakly coupled pair, from a closely spaced single-ended aggressor.

This page intentionally left blank

APPENDIX C

Selected References

Anderson, E.M. *Electric Transmission Line Fundamentals*. Reston, VA: Reston Publishing Company, Inc., 1985.

Archambeault, B. *PCB Design for Real World EMI Control*. The Netherlands: Kluwer Academic Publishers, 2002.

Bakoglu, H.B. *Circuits, Interconnects, and Packaging for VLSI*. Reading, MA: Addison-Wesley, 1990.

Bennett, W.S. *Control and Measurement of Unintentional Electromagnetic Radiation*. Hoboken, NJ: John Wiley and Sons, 1997.

Buchanan, J.E. *Signal and Power Integrity in Digital Systems*. Columbus, OH: McGraw-Hill Book Company, 1995.

Chipman, R.A. *Transmission Lines*. Schaum's Outline Series. Columbus, OH: McGraw-Hill Book Company, 1968.

Dally, W.J., and Poulton, J.W. *Digital Systems Engineering*. Cambridge, England: Cambridge University Press, 1998.

Gardial, F. *Lossy Transmission Lines*. Norwood, MA: Artech House, 1987.

Grover, F.W. *Inductance Calculations*. Mineola, NY: Dover Publications, 1973.

Hall, S.H., Hall, G.W., and McCall, J.A. *High Speed Digital System Design*. Hoboken, NJ: John Wiley and Sons, 2000.

Itoh, T. *Planar Transmission Line Structures*. Piscataway, NJ: IEEE Press, 1987.

Johnson, Howard, and Graham, Martin. *High Speed Digital Design*. Upper Saddle River, NJ: Prentice Hall, 1993.

Konsowski and Helland et al. *Electronic Packaging of High Speed Circuitry.* Columbus, OH: McGraw-Hill, 1997.

Mardiguian, Michel. *Controlling Radiated Emissions by Design*. The Netherlands: Chapman and Hall, 1992.

Martens, L. *High Frequency Characterization of Electronic Packaging*. The Netherlands: Kluwer Academic Publishers, 1998.

Ott, Henry. *Noise Reduction Techniques in Electronic Systems*. Hoboken, NJ: Wiley-Interscience, 1988.

Paul, Clayton. *Introduction to Electromagnetic Compatibility.* Hoboken, NJ: Wiley-Interscience, 1992.

Poon, Ron. *Computer Circuits Electrical Design*. Upper Saddle River, NJ: Prentice Hall, 1995.

Rosenstark, Sol. *Transmission Lines in Computer Engineering*. Columbus, OH: McGraw-Hill, 1994.

Skilling, H.H. *Electric Transmission Lines*. Melbourne, FL: Krieger Publishing Company, 1979.

Smith, D. *High Frequency Measurements and Noise in Electronic Circuits*. New York: Van Nostrand Reinhold, 1993.

Tsaliovich, A. *Cable Shielding for Electromagnetic Compatibility.* The Netherlands: Chapman and Hall, 1995.

Wadell, Brian. *Transmission Line Design Handbook*. Norwood, MA: Artech House, 1991.

Walker, C. *Capacitance, Inductance and Crosstalk Analysis*. Norwood, MA: Artech House, 1990.

Walsh, J.B. *Electromagnetic Theory and Engineering Applications*. New York: The Ronald Press Company, 1960.

Williams, Tim. *EMC for Product Designers*. Burlington, MA: Newnes Press, 1992.

Young, B. *Digital Signal Integrity.* Upper Saddle River, NJ: Prentice Hall, 2000.

INDEX

2D field solvers
  calculating characteristic impedance, 265–270
  calculating differential impedance, 494–495
  calibrating, 264
  capacitance, 143–146
  Maxwell capacitance matrix, 417–423
  parasitic extraction tool, 34
2D modeling tool, 39
3D field solver, 34, 164, 166, 185
3M C-Ply, 137–138

A

A_{cond}, 380, 579–580
A_{diel}, 579–580
AC current, 193–194, 366
AC simulation, 27, 102–105
active elements, 102
active equalization, 403
active net. *See* aggressor net
aggressor net, 405–410, 427–429, 447, 453–464, 466–469, 471, 517, 539–542
Agilent
  86100 DCA TDR, 10
  54120 TDR, 395
  Advanced Design System (ADS), 103
  N5230 VNA, 563
  PLTS, 565
Alloy 42, 119, 154, 202–204
alloys, 154, 202–204
Altera Stratix II GX FPGA, 627
alumina, 134
aluminum, 119
AN-905 National Semiconductor Application Note, 493–494
analytical approximations, 30–34
angular frequency, 49–50, 94, 97, 346, 357–359, 364, 368, 382
Ansoft
  High Frequency Structure Simulator (HFSS), 25
  Q3D field solver, 164
  SI2D field solver, 145
  2D Extractor, 144
antipads, 401, 620
approximations
  analytical, 30–34
  capacitance per length, 138–143
  characteristic impedance, 262–265
  loop inductance, 175–176, 652–661
  microstrip, 141, 264–265
  parallel plate, 347–348
  resistance of interconnects, 115–117
  stripline, 142, 264–265
area
  cross-section, 116, 120
  of the loop, 174–176
  planes, 136, 359, 690–692

asymmetric stripline, 213
asymmetries, 481, 524, 529–531, 537, 605–607
attenuation, 3, 60–61, 231–232, 342
  decibels (dB), 371–376
  in lossy lines, 376–390, 400–402
  napiers, 372
  S-parameters (scattering parameters), 579–580
AVX, 671–672, 700–701
AWG (American Wire Gauge) reference, 120
axial-lead resistors, 108–109, 330–331

B

backward cross-talk coefficient, 440, 519
backward-moving current, 410, 432, 436–439, 453
balanced S-parameters. *See* differential S-parameters
balanced transmission lines, 213–214
Ball grid arrays (BGAs), 8, 219, 267, 309–310, 587–588, 664–666, 710–711
bandwidth
  clock frequency, 68–70
  estimating, 69–70
  interconnect, 74–78, 390–396
  lumped-circuit model, 274–275
  of a measurement, 70–72
  model, 72–74
  rise time and, 43, 58–63
  signals, 67
barium titanate, 134, 137, 154
BCB, 134
Bell, Alexander Graham, 373–374
Bell (Bel) scale, 374
Berkeley Nuclear Corp (BNC), 139, 230
Bismaleimide triazine (BT), 134, 143, 353, 393
bit error rate, 339–340, 399, 606–607, 613
blind vias, 400, 605
bounce diagrams, 292–294
BR, 604–605
branch, 6–7, 282, 309
broadside-coupled differential stripline, 480, 498, 500–501
bulk capacitance, 632–635, 676, 679, 681
bulk conductivity, 202, 349–351, 359
bulk resistivity, 118–119, 342–343, 344
buried vias, 400, 605
bus noise, 454–457

C

C_{even}, 548–549
C_{load}, 547–549
C_{loaded}, 419
C_{odd}, 547–548
C_{planes}, 677–680, 688–689
C11, 414, 422, 442, 464–465, 487–491, 547
C12, 414, 427–428, 433–434, 464, 486–491, 547–550
C-Ply laminates, 13–14, 137–138
calibrating 2D field solvers, 264
capacitance
  approximations, 130–131
  coax cable, 139–140
  corners, 320
  decoupling, 135–138, 179, 181, 187, 631
  dielectric constant, 133–134
  displacement current, 129–130
  effective dielectric constant, 146–149
  of an ideal capacitor, 89–91, 99
  input-gate, 311–313
  insulating dielectric, 129
  measurement of, 127–129
  negative values of, 420–423
  parallel plate approximation, 131–132
  per length, 138–143, 320–321, 327
  planes, 677–681
  of a sphere, 130–131
  2D field solvers, 143–146
  voltage and, 89–91, 127–128
capacitance matrix, 413–423, 486, 546–550
  See also Maxwell capacitance matrix; SPICE (Simulation Program with Integrated Circuit Emphasis)
capacitive coupling, 9
capacitive delay adders, 317–319
capacitive discontinuities, 314–317, 324
capacitively coupled currents, 431

capacitor(s)
chip, 632
current flow through, 128–130
DC blocking, 528–529
decoupling, 644–646
electrolytic, 631–634, 676–677
ESL, 649–652
estimating needs, 681–683
interdigitated capacitors (IDCs), 254, 650–651, 700–702, 713
location effects, 703–707
loop inductance in, 180
MLCC, 632, 646–650, 701, 711
mountings, 661–667
mutual capacitor, 411–412
in parallel, 667–673
reverse aspect ratio, 700
and SRF, 673–675
tantalum capacitors, 631–634, 677
values, 675–681, 687–692
cavity spreading inductance, 652, 666, 695
center-to-center separation, 140, 165–166, 170–173, 176–177, 654
characteristic impedance
approximations, 262–265
calculating, 265–270
coax cable, 230
even-mode, 520, 546–550
of free space, 231
frequency variation, 278–279
imaginary part, 367–368
lossy transmission lines, 367–369
magnitude, 368–369
matrix, 547–553
odd-mode, 520
proximity effect, 488
real part, 367–368
single-ended, 486, 489, 490, 520
transmission lines, 227–233, 385–390
2D field solvers and, 265–270
chip capacitors, 632
chip-scale packages (CSPs), 8, 11, 642
chokes, 15, 82, 538–541, 545
circuit model. *See* models and modeling
circuit theory, 26, 101–105, 362
circuit topology, 99, 114–115, 190, 528
circular loop inductance, 175
clearance holes, 187–188
clock cycle, 44
clock frequency, 16–20
bandwidth and, 68–70
increasing, 1–3
and signal integrity problems, 1–2
and signal rise times, 3
clock period, 68–70
coax cable, 139–140, 230, 262–263
cobalt, 154, 202
coils, 539
collapse, 340, 399, 613
common impedance, 524–528, 535–536
common impedance, 508–511
common mode signals, 597
common signals, 478–482
chokes, 538–541
common signal component, 481–482
component in LVDS, 479
conversion to, 529–534
EMI and, 534–539
modes and, 508–513, 605–608
pi and tee topologies, 525–528
S-parameters (scattering parameters), 596–600
terminating, 522–529, 522–533
compensation, 331–334
complex dielectric constant, 356–358, 385, 391–393
conductivity, 118, 196, 202, 348
controlled-impedance board, 228, 282
controlled-impedance interconnects, 230
controlled-impedance lines, 227–230
convolution integral, 610–612
coplanar, 146, 212–213, 480
copper, 119, 122, 124, 137, 200, 343
electron velocity in, 214–215
skin depth, 195–196, 198–200
surface smoothness, 381
copper puddle, 691
corners, 319–323, 335

cosine waves, 96–97
coupled transmission lines, 413–417, 451–452
coupling, 9, 10
  cross talk, 407–409, 411
  differential impedance and, 482–493
  differential pairs, 541, 544–546, 551
  high Q resonators and, 585–588
  inductive coupling, 9, 29
  models, 411
cross talk, 10–11
  active net/quiet net, 405–409, 431
  capacitive coupling, 9
  capacitively coupled currents, 431–435
  coupling, 407–409, 411
  design principle, 41
  dielectric constant, 464–465
  differential pairs, 539–542
  equivalent circuit model, 411–413
  far-end, 409–411, 426, 441–451
  fringe fields, 407–408
  guard traces, 457–464
  impedance and, 82–84
  induced voltage, 160
  inductance matrix, 424–425
  inductively coupled currents, 9, 435–437
  Maxwell capacitance matrix, 417–423
  near-end, 409–411, 426, 429, 438–451, 463, 520
  NEXT and FEXT, 409–411, 438
  noise budget, 440
  noise sources, 9–11
  quiet net, 405
  reduction strategies, 473
  saturation length, 429–432, 438–439, 454
  signal swing, 405, 440
  simulating, 451–457
  SPICE capacitance matrix, 413–417
  superposition, 406–407
  switching noise, 409, 469–473
  timing, 466–469
  2D field solvers, 417–423
  in uniform transmission lines, 425–431, 473
current distributions
  in a differential pair, 496–502
  skin depth and, 193–201
current flow (direction), 168–170, 436–437
CyanateEster, 353, 393

D

damping, 586, 679, 711–712
data mining, 594–596
DC-blocking capacitor, 528–529
DC (direct current), 48, 56, 92–93, 193, 200, 349, 365
decibels (dB), 49, 75–76, 371–376
decoupling capacitance, 135–138, 179, 181, 187, 631
decoupling capacitors, 644–646, 699–703
delay adders, 233, 312–313, 317–320, 324, 331–332
Delphi Electronics, 38–39
delta I noise, 9, 16, 93, 161, 189, 196, 435–437, 469–471, 475
design methodology, 22–24
diagonal elements
  converting, 421–422
  impedance matrix, 551
  inductance matrix, 424, 548
  Maxwell capacitance matrix, 420
  SPICE capacitance matrix, 416
diamond, 134
dielectric constant, 217–218, 385
  capacitance, 133–134, 146
  complex, 356–358, 385, 391–393
  cross talk, 464–465
  dissipation factor and, 352–353, 355–359
  Dk, 133, 580, 637–638, 677–678, 704
  effective dielectric constant, 146–149
  frequency and, 134
  imaginary part, 356–358, 385
  real part, 356–359, 382–383, 385
dielectric-loss shunt resistance, 361–362
dielectric material, 134, 143, 184, 444–446, 448–450, 464
dielectric spectroscopy, 352
difference voltage, 476, 482, 511

differential impedance, 482–486, 493–496, 506–508, 526–527
differential insertion loss, 600–606
differential mode, 506, 553, 597, 605
differential mode signals, 597
differential noise, 540–542
differential pairs
  characteristic impedance matrix, 550–553
  common impedance, 508–511
  common signal component, 481–482
  coupling, 482–493, 544–546
  cross-sections, 480–481
  cross talk, 539–542
  current distribution, 496–502
  difference voltage, 476, 482, 511
  differential impedance, 482–486, 493–496, 506–508, 526–527
  differential signaling, 475–480
  EMI, 476, 534–539
  far-end cross talk, 516–518
  modes, 504
  odd and even modes, 502–511, 513–518, 546–550
  pair makeup, 480
  return path gaps, 542–544
  signal velocity, 513–518
differential S-parameters, 596–600, 607–609
differential signals, 475–480, 505
  conversion to common signals, 529–534
  pi and tee topologies, 525–528
  terminating, 522–529
dipoles
  dielectric materials and, 134
  dipole moment, 352
  dissipation factor and, 351–354, 385
  reorientation, 349
Discrete Fourier Transform (DFT), 51–55
dispersion, 370–371, 601
displacement current, 129–130
dissipation factor, 342, 351–359, 367, 382–383, 385, 402, 580–581
distortions, 7, 83, 300–301, 320, 529
domains. *See* frequency domain; time domain
double bonding, 172
DriClad (IBM), 353, 393
DTDR (differential TDR), 521–523
DUT (device under test), 35
dV/dt, 16, 129, 311–312, 315, 346, 355, 426–428

E

eddy currents, 204–206, 245–247, 435–437
edge-coupled microstrip, 480, 496–499, 502–505, 509–511, 543–544
edge-coupled stripline, 480, 493–494
effective dielectric constant, 146–149
effective inductance, 156, 167–174, 192
electrolytic capacitors, 631–634, 676–677
electromagnetic compatibility (EMC) test, 607, 617
electromagnetic fields, 24–25
electromagnetic interference (EMI), 3, 14–15, 41, 68, 82–84, 534–539
electromagnetic interference (EMI) certification tests, 535–537
electron velocity, 214–215
embedded microstrip, 148–149, 213, 444–445, 450
emissions, 15
equalization, 402–403
equivalent electrical circuit model. *See* models and modeling
equivalent inductance, 172, 189–192, 709, 715
equivalent series inductance (ESL), 181, 189–190, 254, 649–652
  See also capacitor(s); PDN (power distribution network)
equivalent series resistance (ESR), 254, 668–674, 679, 711–712
Ethernet, 476
euctectic Pb/Sn solder, 119
even mode, 411, 502–506, 511–518
even-mode characteristic impedance, 520, 546–550
even-mode impedance, 411, 508–511, 520–522, 552
even-mode voltage components, 511–513

external self-inductance, 193, 200–202
eye diagrams, 337–340, 342, 399–402, 612–613

F

f (frequency) designation, 49–50
F (repeat frequency), 44, 52–55, 64–66, 70
f_{res}, 272, 586–587, 603–604, 704
"fab" (chip-fabrication facility), 19–20
fall time, 18, 45, 300
false triggering, 5, 8
far-end cross talk, 10, 426, 513–518
  coefficient, 442–444, 451
  decreasing, 448–451
  differential pairs, 516–519
  FEXT, 409–411, 438, 442–444
  TD and, 514
far-end termination, 305
far-end Thevenin termination, 305
far-field electric-field strength, 14, 536
FCC (Federal Communications Commission) EMI tests, 20, 535–536, 607
FDTI (frequency domain target impedance) method, 692–698
field solvers. *See* 2D field solvers
ferrite core, 539
ferrites, 15, 82, 202
ferromagnetic metals, 154, 202
50-Ohm line, 143, 261
50 Ohms, 231–235
first-order model, 101
forward-moving current, 434–437, 460
.FOUR command, 53
four-point probe, 122–123
Fourier Transforms
  Discrete Fourier Transform (DFT), 51–55, 58, 66–67
  Fast Fourier Transform (FFT), 51–53
  Fourier Integral (FI), 51–53
  inverse, 57, 609
FPGAs, 11, 69, 627, 693, 710
FR4, 14, 60–61, 134, 137, 219, 353, 393
  bulk conductivity, 350
  dielectric constant, 134
  dissipation factor, 383
FR5, 581–583
free space, 130–132, 231
frequency-dependent losses, 337–340, 386
frequency domain, 25–27, 43
  bandwidth and, 48
  converting the spectrum into the waveform, 57
  EMI and, 49
  impedance and, 48
  lossy transmission lines and, 49
  S-parameters (scattering parameters), 609–613
  simulation, 27, 102–105
  sine waves, 46–51
  VNA and, 49
frequency ranges, 632
fringe fields, 132, 145, 268–269, 407–408, 422, 466, 469
fully embedded microstrip, 149, 445, 450
FWHM, 586

G

G_L (shunt conductance), 362–365, 367–368, 377–379
gaps in return paths, 542–544
gate capacitance, 637–638
gauge chart, wire, 120
GETek, 134, 353, 393
GHz, 19, 63
Gigabit Ethernet, 22
GigaTest Labs Probe Station, 37
glass, 134, 218, 353, 393
gold, 119–120
Grasso, Charles, 186
Green's function, 610
ground bounce, 3, 9–10, 168–173, 184
  PDN, 616–617
  reducing, 469–470
  return path, 249, 253
  switching noise and, 169, 170
ground line. *See* return path
ground planes, 82, 135–138, 184–188
guard traces, 457–464

H

hacking interconnects, 559, 605
harmonics, 55–60, 66, 623
Henry (Weber/Amp), 155
high-density interconnects (HDI), 11
high-permeability materials, 202–204
high-speed regime, 3
high-speed serial links, 21, 339, 342, 606
humidity, 353

I

I_{comm}, 535–536
I_{peak}, 626
I_{max}, 625
I_{trans}, 286
$I_{transient}$, 624–628
IBIS (input/output buffer interface spec), 27–29
ideal circuit elements, 29, 48, 85, 86–94, 99, 347
ideal differential pair, 519–520
ideal distributed transmission line, 258–259, 270–271
ideal lossy transmission lines, 360–367, 397–398
ideal resistor, 85, 88–89, 99, 108, 347, 398
ideal RLC circuit, 48
ideal square wave, 54–60, 65–66, 391
ideal transmission line, 88, 106, 209–210, 257, 270–274, 519–520
IM Intermark, 15
image currents, 204–205
impedance profile, sculpting, 687, 691–698
impedance, 5–9, 32, 81–86
  analyzer, 35, 71
  common, 508–511, 524–528, 535–536
  controlled, 227–230
  differential impedance, 482–486, 493–496, 506–508, 526–527
  discontinuities, 6–7, 281–284, 307, 589
  even-mode, 411, 508–511, 520–522, 552
  in the frequency domain, 94–99
  i/o connectors, 15
  ideal capacitor, 89–92, 647
  ideal inductor, 92–94, 99
  ideal resistor, 88–89, 108
  inductance, 640–642
  input, 234–238
  instantaneous, 5, 36, 223–228, 235–236, 281
  interconnect, 16, 81–82
  matrix, 550–553
  measuring, 234–235
  models, 99–101
  odd-mode, 411, 506–508, 520–522, 528–529, 551
  odd-mode characteristic impedance, 520, 546–550
  PDN, 621–629, 645
  peak, 628, 641, 670–681, 712–716
  port, 559, 573–574, 599
  real versus ideal circuit elements, 86–88
  shunt impedance, 315–316
  single-ended, 40, 484–486, 497–501
  single-ended characteristic impedance, 489–491, 520
  source, 290–292
  target, 618, 621–629
  traces, 7
  transmission line, 233–238, 282–284, 385–390
  2D field solver, 265–270
  See also characteristic impedance; SPICE (Simulation Program with Integrated Circuit Emphasis)
impulse response waveforms, 609–612
in-phase currents, 355–359
induced current, 204–206, 245–247, 435–437
induced voltage, 159–160
inductance, 92, 99, 152–161
  circular loop, 175
  common definition, 152
  decreasing, 175
  effective, 167–174, 192
  equivalent, 172, 189–192, 709, 715
  ferromagnetic metals, 154
  ground bounce, 168–173
  loop mutual, 188–189, 192, 413, 471

inductance (*continued*)
loop self-inductance, 167, 173–178, 192, 197, 424–425
magnetic-field lines, 152–161
mounting, 631, 650–652, 684, 695, 703–706
partial, 156, 161–166, 192
partial mutual, 166, 171, 192
partial self-inductance, 161–165, 169, 174, 176, 190, 192, 197
per length, 207, 227, 260, 366, 440
per square of planes, 183–184, 656–661
power integrity and, 178–183
principles of, 152–161
self-inductance, 156–158, 192–193, 201
spacing rules and, 170–173
spreading, 186, 653, 657, 662, 685, 704–710
total, 156
types of, 192
voltage and, 158–161
Webers, 153–154
See also loop inductance; mutual inductance
inductance matrix, 424–425, 547–550
inductive coupling, 9, 29
inductively coupled currents, 9, 435–437
Infiniband, 21
input-gate capacitance, 311–313
input impedance, 234–238
insertion losses, 75, 386–390
differential insertion loss, 600–606
magnitude of, 577–582
in non-50-Ohm lines, 589–594
simulating, 567–570
instantaneous impedance, 5, 36, 223–228, 235–236
See also transmission lines
instantaneous voltage, 311, 428–429
insulators, 133–134
integrated passive devices (IPDs), 107–109
Intel processor chips, 16–18
inter-symbol interference (ISI), 8, 337–340, 396
interconnect(s), 334
alloys, 154
approximation of resistance, 115–117
bandwidth, 74–78, 390–396
capacitance and, 128
capacitance per length approximations, 138–140
dielectric constants of insulators, 133–134
eye diagram, 612
impedance, 16, 81–82
Maxwell's Equations and, 113–114
models and, 28
PDN, 632
permeability, 202
propagation delay, 219
speed of light in, 218
substrates, 121
transparent, 569–573
interdigitated capacitors (IDCs), 254, 650–651, 700–702, 713
internal self-inductance, 193–194, 200–201
interposer, 593–594, 692
intrinsic rise time, 76–78
inverse Fourier Transforms, 57, 609
ionic motion, 349
IPC approximations, 141–143, 145–146, 264–267
IR drop, 618
iron, 154, 202
ITRS (International Technology Roadmap for Semiconductors), 17

J

jitter, 338–340, 613

K

k_b, 438–440, 451
k_f, 443–451, 519
Kapton, 134, 353
kink of field, 216–217
knee frequency, 64
Kovar, 119, 154, 202–203

L

L_{odd}, 548
L_{pkg}, 641–642, 715–716

L_{self}, 32, 424
L_{spread}. *See* loop spreading inductance
L_{via}, 387–388, 659–660, 685, 705
laminate materials, 393
LaPlace's Equation, 143, 417–418
  See also Maxwell's Equations
lattice diagrams. *See* bounce diagrams
lead, 119
leading edge of signal, 219
leakage current, 347, 350
Lentz's law, 435–436
line drivers, 240
loaded line, 325–327
loop inductance, 13–14, 156, 167–173, 192
  approximations, 175–176, 652–661
  capacitors, 180
  matrix, 547–550
  PDN, 178–183, 653
  per length, 176–178, 258, 548, 655
  per square of planes, 183–184, 656–661
  of planes and vias, 184–188
  power-distribution network and, 178–183
  radius and, 176–177
  transmission lines and reflections, 327–331
  of two rods, 177
loop mutual inductance, 188–189, 192, 413, 471
loop self-inductance, 167, 173–178, 192, 197, 424–425
loop spreading inductance, 658–659, 662–663, 685, 708–709
loss angle, 352, 357–359, 385
lossless transmission lines, 278–279, 340–342
lossy transmission lines, 360–367
  analyzing, 337–338
  attenuation, 338–339, 342, 376–390, 400–402
  characteristic impedance, 367–369
  conductor loss, 342–346
  dielectric loss, 346–351
  eye diagrams, 399–402
  frequency dependence, 338–339
  intersymbol interference (ISI), 337, 339–340, 396, 399
  measured properties of, 385–390
  modeling, 360–367
  signal velocity, 369–371
  time domain behavior, 397–400
low-loop inductance decoupling capacitors, 178
low-loss regime, 366–371
lumped circuit elements, 88, 106
lumped circuit model, 270–278, 411–412, 546–547
LVDS (low-voltage differential signals), 476–479, 523, 541

M

magnetic-field lines, 157–158, 160, 167–170, 193, 202–205, 619
magnitude, 95–96, 368–369, 577–582
matrix
  algebra, 52
  impedance, 550–553
  loop inductance, 547–550
  matrix elements, 562–567
  Maxwell capacitance matrix, 417–423
  notation, 414
  transform matrix, 607–609
  See also SPICE (Simulation Program with Integrated Circuit Emphasis)
Maxwell, James Clerk, 129
Maxwell capacitance matrix, 417–423
Maxwell's Equations, 24–26, 33, 113–114, 143, 216–217
Mears, James, 493
Mentor Graphics Hyperlynx, 2
MHz, 3–8, 17–20, 63, 66, 679–681
microseconds, 63, 180
microstrip
  approximations, 141, 266
  edge-coupled, 480, 496–499, 502–505, 509–511, 543–544
  embedded, 148–149, 213, 444–445, 450
  interconnects, 141–146, 149
  lines, 458–459
  traces, 450–451
microvias, 605

mixed-mode S-parameters. *See* differential S-parameters
MLCC capacitor, 632, 646–650, 701, 711
mode conversion, 605–607
mode conversion quadrants, 605
models and modeling
  bandwidth and, 72–74, 100–101, 106–110
  capacitively coupled currents, 431
  circuit topology, 99, 114–115, 190, 528
  cost and accuracy, 106
  creating models from calculation, 30–35
  first-order model, 101
  ideal transmission lines in, 106, 109
  introduction and basics, 105–110
  lossy transmission lines, 360–367
  lumped circuit model, 270–278, 411–412, 546–547
  real versus ideal circuit elements, 86–88
  simplicity in, 105–106
modes of the pair, 504
molybdenum, 119
monotonically, 328
Moore's Law, 20
mounting inductance, 631, 650–652, 684, 695, 703–706
mutual capacitance, 489
mutual capacitor, 411–412
mutual-field line rings, 157–158, 160, 167, 170, 205
mutual inductance, 10–11, 41, 156–158, 171, 192, 411–412, 489

N

n-section lumped circuit model, 270–278, 411–412, 546–547
nanoseconds, 44, 63
Napier, John, 372
Napiers (nepers), 364, 372–373, 377–378
near-end cross talk, 10, 426, 463
  coefficients, 438–441, 520
  NEXT, 409–411
  transmission lines, 429–433, 438–441
neck downs, 300, 306–308, 481
negative capacitance, 420–423
NelcoN6000SI, 353, 393
net inductance. *See* effective inductance
nets. *See* aggressor net
network theory, 259–260
NEXT, 40, 409–411
nH (nanoHenry), 155, 158, 176, 683–687
nickel, 119, 154, 202–203
noise, 3–4
  budget, 178, 301, 440
  cross talk, 9–11, 41
  design principles, 41
  differential, 540–542
  EMI, 13–15, 41
  rail-collapse, 11–13, 41, 184
  signal quality, 4–9, 41
  simultaneous switching output (SSO), 9–10
  switching noise, 9–10
  undershoot, 281
  voltage, 617–618
  See also attenuation; ground bounce; near-end cross talk; reflections; switching noise
nonmonotonicity, 3, 328
nonuniform transmission lines, 212–213, 228, 473
numerical simulations, 30–31, 34–35
Nyquist frequency, 604–605

O

OC (optical-carrier), 21, 476
odd mode, 502–506, 511–513
odd-mode characteristic impedance, 520, 546–550
odd-mode impedance, 411, 506–508, 520–522, 528–529, 551
odd-mode time delay, 411, 520
off-diagonal elements, 416, 420–424, 548, 551
ohmmeter, 234–235
Ohms, 117–118, 122
on-die capacitance, 637–639
orthonormal functions, 47
out-of-phase currents, 355–359
overshoot, 328

P

P_{max}, 626
package barrier, 639–644
package inductance, 631–632, 640–643, 715–716
parallel plate approximation, 131–132, 135, 347–348
parallel resonant frequency (PRF), 670–675
parallel resonant peak, 670–674
parameter values. *See* parasitic values
parasitic extraction tool, 34, 114
parasitic values, 87, 99
Paricon interposer, 594
partial inductance, 156, 161–166, 192
partial mutual inductance, 161–175, 191–192
partial self-inductance, 161–165, 169, 174, 176, 190, 192, 197
passive devices. *See* passive elements
passive elements, 24, 27–29, 102, 107–109
pattern-dependent noise. *See* inter-symbol interference (ISI)
PCBs, 121
PCI-express, 21
PDN (power distribution network), 12–13, 615–617
  capacitor location effects, 703–707
  capacitor mountings, 661–667
  capacitor values, 675–681, 687–692
  capacitors and SRF, 673–675
  capacitors in parallel, 667–673
  chip capacitors, 632
  decoupling capacitors, 644–646, 699–703
  design guidelines, 620
  equivalent series inductance (ESL), 650–652, 681–687
  equivalent series resistance (ESR), 668–673, 694, 711, 716
  FDTI (frequency domain target impedance) method, 692–698
  five parts of, 631–633
  frequency ranges, 632
  interconnects, 632
  IR drop, 618
  loop inductance and, 178–183
  loop inductance approximations, 652–661
  MLCC capacitor, 632, 646–650, 701, 711
  on-die capacitance, 637–639
  package barrier, 639–644
  parallel resonant frequency (PRF), 670–675
  primary functions, 616–617
  ripple, 618–619, 623–630
  spectrum, 618–619, 621–624
  SPICE simulations, 635–637
  spreading inductance, 704–710
  target impedance, 619, 621–629
  voltage, 178, 617–619
  VRM (voltage regulator module), 633–635
peak impedance, 628, 641, 670–681, 712–716
peak resonant frequencies, 272
permeability of free space, 183, 217, 231
permeability of the conductor, 202–204
permittivity, 130–132, 136, 139–141, 217, 231, 250–251, 351, 359
pF, 130–132, 261, 311–317, 534
pH, 699–703
phase, 51, 96
phase shift, 94–98
Pi-termination, 525–528
pitch, 165–166, 232, 448, 451, 465, 499, 654
planes
  capacitance, 677–681
  with clearance holes, 187–188
  ground, 82, 135–138, 184–188
  loop inductance, 183–184
  power, 135–137, 690–691
polyethylene, 134, 139, 353
Polyimide/glass, 134, 353, 393, 465
polymers, 133, 217–218, 352–353
port assignment labeling, 561–562
port impedance, 559, 573–574, 599
port indices, 598–599
power bounce, 3
power planes, 135–137, 690–691
power supply voltage, 11
PowerPoser, 693–698
PQFP (plastic quad flat pack), 25, 117
PRBS pseudo random bit stream, 611–613
pre-emphasis, 402–403

prepreg material, 450
printed circuit board (PCB), 2
propagation delay, 40, 82, 219, 227
proximity effect, 488

Q

Q (resonance), 585–588
Q3D field solver, 184
quad flat pack (QFP), 212
quarter wave stub resonance, 602–605
quartz, 134
quiet loop, 189, 470–471
quiet net. *See* victim net

R

R_{equiv}, 525–528
R_{shunt}, 361–362
radians per second measurement, 49, 94
radiated emissions, 15, 68, 535, 537–539, 578
radiated field levels, 535–539
radio frequency (rf), 46, 59
rail collapse
  and C-Ply laminates, 13–14
  design principle, 41
  impedance and, 82–84
  noise, 11–13, 41, 184
  PDN, 178–183
rail droop. *See* rail collapse
reactance, 367, 370
real versus ideal circuit elements, 86–88
reflection coefficient, 283, 460–461, 521
  transmission lines and reflections, 284–290, 293, 295, 300, 304, 307, 311–312, 589–593
reflections, 3, 224
  capacitive end terminations, 311–313
  short-series transmission lines, 306–309
  See also transmission lines and reflections
relative dielectric constant. *See* dielectric constant
reorientation of electric dipoles, 349–351
repeat frequency (F), 44, 52–55, 64–66, 70
resistance
  bulk resistivity, 117–119
  conductivity, 118
  of interconnects, 115–117
  measurement, 117
  Ohms-length measurement, 118
  of 1-mil-diameter wire bond, 116
  physical basis, 113–115
  resistance per length, 119–121
  sheet resistance, 121–124
resistor component technologies, 108–109
resistor impedance, 88–89, 108
resonances in the planes, 256
return current distribution, 496–502
return losses, 36, 386, 558, 562–565, 567–569, 571–573, 596
return path, 139, 174, 177, 197, 210–214, 475
  gaps, 328, 542–544
  ground bounce, 249, 253
  transmission lines, 241–256
reverse aspect ratio capacitors, 700
RG58 coax cable, 139, 230
RG174 coax cable, 230
ribbon cable, 177–178, 198
ringing, 3, 7, 67, 282, 293
ripple, 178, 180, 574, 581, 601, 618–619, 622–630
rise time (RT), 219, 338–339, 518, 528–530
  bandwidth and, 43, 58–62
  and clock period estimates, 18, 68–70
  degradation, 395–398, 400–402
  distortion magnitude, 7
  intrinsic, 76–78
  laminate materials, 393
  10–90, 18, 45
  20–80, 45
  units of measurement, 63
  unterminated lines and, 304
RLC circuit, 48, 100–102, 105, 648–649, 668–669, 703
rod over plane, 139–140, 263–264
RogersRF35, 353, 393

S

S_d, 607–608
S_{dB}, 560
S_s, 607–608

S11, 36, 386, 558, 562–565, 567, 596
S21, 75, 386–390
  differential insertion loss, 600–606
  magnitude of, 577–582
  phase of, 574–577
  polar plot, 582–583
  simulating, 567–570
S22, 564–567, 597, 613
S31, 563, 584, 595–596
S32, 596
S41, 563, 584, 596
S42, 596
S43, 584, 595–596
S-parameters, 36
  attenuation, 579–580
  behavioral model, 555–556
  common signals, 596–600
  cross talk, 582–589
  data mining, 594–596
  dB and magnitude, 560, 571, 575, 588
  differential, 596–609
  dissipation factor, 580–582
  insertion loss, 36, 75, 386–390, 567–570, 574–583
  insertion loss, non-50-Ohm lines, 589–594
  magnitude, 560, 562–565, 577–582
  matrix elements, 562–567
  mode conversion, 605–607
  phase, 560–561, 563, 574–577
  port, 559, 562
  port assignment labeling, 561
  port impedance, 559, 573–574, 599
  port indices, 598–599
  return loss, 36, 386, 558, 562–565, 567–569, 571–573, 596
  reverse notation, 562
  simulators, 569–570
  sine waves, 559–560, 568–569, 576
  single-ended, 596–600
  time and frequency domains, 609–613
  transform matrix, 607–609
  transparent interconnect, 569–573
  two-port device, 564–566
  unique elements, 566–567
saturation length, 429–432, 439, 518
scalar network analyzer, 558
SCD11, 599, 606
SCD21, 599, 606–608
schematic, 26, 29, 99
SCSI bus, 476
sculpting the impedance profile, 687, 691–698
second-order model, 73–74, 100–101, 108, 165
self-field line rings, 157–159, 167–170, 194
self-inductance, 156–158, 192
self-resonant frequency (SRF), 181, 649–650, 668–675, 677–680
serial ATA, 21
series loop inductance, 327–331
series resistance, 49, 200, 342, 360–362, 365–368, 371, 378
sheet inductance, 183–184, 186, 656–661, 700
sheet resistance, 122–123
shield radius, 499–500
shielded twisted pair, 499–500, 537–538
shielding, 15
short-series transmission lines, 306–309
short-stub transmission lines, 309–311
shorting vias, 460–463
shunt conductance (G_L), 362–365, 367–368, 377–379
shunt impedance, 315–316
shunt resistance, 360–362
SIA (Semiconductor Industry Association), 17–18
Siemens, 118, 196
signal path, 138–139, 174, 177, 197, 210–214
signal-to-noise ratio (SNR), 49, 72
signal velocity, 369–371, 513–518
signal-voltage swing, 306, 329, 405, 440, 479, 483
significant waveforms, 63–66
silver, 119
silver-filled epoxy, 119
silver-filled glass, 119
silver plating, 204
simulations
  AC simulations, 27
  behavioral simulators, 24, 27

simulations (*continued*)
circuit simulators, 24, 26–27
electromagnetic simulators, 24–25
Maxwell's Equations, 24–25
SPICE, 26–27
transient, 26–27
simultaneous switching noise (SSN), 9, 189, 616–617
simultaneous switching output (SSO), 9–10
sine waves
features, 49–55, 94
in frequency domain, 46–50
phase, 49, 94
ratio, 95
S-parameters, 559–560, 568–569, 576
in time domain, 50
single-ended characteristic impedance, 483–484, 489–491, 494, 500–501, 520, 546
single-ended impedance, 40, 484–486, 497–501
single-ended NEXT, 40
single-ended S-parameters, 596–600
single-ended signaling, 476, 491
single-ended transmission lines, 413, 480, 497, 520, 539–542, 596–598
skew, 8–9, 481, 529–533, 606
See also time delay (TD)
skin depth, 193–201, 342–346, 352, 365–366
Smith, Larry, 694
SMT resistors, 107–109, 330–331
solder mask, 265–270, 509–510
source impedance, 290–292
source-series termination, 304–305, 334, 452–453, 460
spatial extent of leading edge, 219–220, 277
spectroscopy, 352, 588
spectrum, 66, 68
PDN (power distribution network), 618–619, 621–624
of a repetitive signal, 53–55
of a waveform, 51–53
to waveform conversion, 57
speed of light, 216–218, 234
sphere capacitance, 130–131
SPICE (Simulation Program with Integrated Circuit Emphasis)
capacitance matrix, 413–418, 421–423, 547–549
circuit theory and, 101–105
.FOUR command, 53
history, 26–27
impedance simulations, 635–637
transmission lines and reflections, 293–295
spreading inductance, 186, 653, 657, 662, 685, 704–710
square wave, ideal, 55–66
step edge, 72, 295, 557–558, 609–610
stripline
approximations, 266–268
broadside-coupled differential, 480, 498, 500
edge-coupled differential, 480, 493–494
interconnects, 141–143
stub resonance, 602–605
stubs, 309–311, 401, 602–605
Sun Microsystems, 12–13
superposition, 406–407
switching noise, 3, 9–10, 161, 469–473, 616–617
loop mutual inductance and, 471
reducing, 189, 469
See also ground bounce

T

tangent of the loss angle, 352, 354, 357–359, 385
tantalum capacitors, 631–634, 677
target impedance, 618, 621–629
TDA Systems IConnect, 37, 71
TD_{even}, 514, 549
TD_{odd}, 514, 548
TDR (time domain reflectometer), 10, 35–37, 71–72
DTDR (differential TDR), 521–523
measuring reflections, 295–299
response, 609
TDT (time domain transmitted) wave, 557–558, 610

tee-termination, 525–529
Teflon, 134, 353, 394, 581–582
terminating resistors, 29, 107–108, 304–306
terminating the line, 304
termination topologies, 525–529
terminations, 301–306, 311–313, 522–529
time delay (TD), 529–530
  cross talk and, 466–469, 514
  delay adders, 233, 312–313, 317–320, 331–332
  interconnect length and, 218–219
  return paths, 241
  transmission lines, 259–261, 271–272, 275–276, 301–304, 307–309
  unloaded lines, 325–327
  See also skew
time domain, 25–26
  clock frequency and, 44
  clock period and, 44–45
  frequency domain and, 43
  rise time and, 44–45
  S-parameters (scattering parameters), 609–613
  simulation, 102
  sine waves, 51
  waveform, 45, 53, 59, 66, 609–612
timing, 3, 8–9, 466–469
tin, 119
total inductance. *See* effective inductance
trace resistance, 121
trace spacing, 457–464
trade-off analysis, 30
transfer function, 644
transform matrix, 607–609
transient current, 624–628
transient simulation, 26, 102
transistors, 45
transmission coefficient, 284, 287
transmission lines
  balanced line, 213–214
  characteristic impedance, 227–233
  coupled, 413–417, 451–452
  driving, 238–241
  electron velocity, 214–215
  first-order model, 257–262
  ground line, 210
  ideal transmission line, 106, 209–210, 270–274, 519–520
  impedance, 233–238
  input impedance, 234–238
  kink of field, 216–217
  lossless, 278–279
  lumped circuit model, 270–278
  nonuniform, 212–213, 228, 473
  return paths, 210, 241–256
  speed, 214–219
  time delay, 218–219, 259–261, 271–272, 275–276, 301–304, 307–309
  unbalanced line, 213–214
  uniform, 212–214, 425–431, 473, 574–582
  zeroth-order model, 224–227
  See also cross talk; differential pairs; lossless transmission lines; lossy transmission lines; transmission lines and reflections
transmission lines and reflections
  bounce diagrams, 292–294
  capacitive delay adders, 317–319
  capacitive discontinuities, 314–317, 324
  capacitive end terminations, 311–313
  compensation, 331–334
  corners, 319–323
  delay adders, 317–319, 324, 331–332
  impedance discontinuities, 281–284
  incident voltage, 285–288
  loaded lines, 325–327
  loop inductance, 327–331
  negative voltage, 289
  noise budget, 301
  reflected voltage, 283, 288–289
  reflection coefficient, 284–289, 295, 312
  short-series, 306–309
  short-stub lines, 309–311
  source impedance, 290–292
  source resistance, calculating, 292
  SPICE, 293–296
  TDR, 295–299
  terminations, 301–306

transmission lines and reflections (*continued*)
transmission coefficient, 287
transmitted voltage, 285–287
unintentional discontinuities, 299–301
unterminated lines, 301–304
vias, 323–324, 401
transparent interconnect, 569–573
tungsten, 119
twin rods, 139–140, 263
twinax, 480
twisted pairs, 15, 230, 232–233, 480, 498, 539

U

unbalanced line, 213–214
undershoot noise, 281, 314–315, 317
uniform transmission lines, 212–214, 425–431, 473, 574–582
unshielded twisted pair, 500, 537, 607–608
unterminated lines, 301–304

V

V_a, 438–439, 442–443, 446, 470–471, 519
V_b, 438–439, 519
V_{cc}, 12, 45, 173, 627–628
V_{comm}, 478–479, 511, 535–536
V_{dd}, 211, 617, 619, 624, 626
V_{ddq}, 627
V_{diff}, 477–479, 482, 484, 511
V_{even}, 511–512, 517–519
V_f, 442–446, 517–519
$V_{incident}$, 36, 283, 521, 576
V_{noise}, 161, 189, 624
V_{odd}, 511–512, 515–519
V_{PDN}, 619
$V_{reflected}$, 36, 283, 521
V_{ripple}, 619
vector, 558
vector-network analyzer (VNA), 35–36, 39, 386, 558–559
velocity, 369–371, 385–390, 429, 513–518
vias, 172–173
blind, 400, 605
buried, 400, 605
corners and, 319–324
loop inductance, 184–187, 661
L_{via}, 387–388, 659–660, 685, 705
microvias, 605
shorting, 460–463
transmission lines and reflections, 323–324, 401
victim net, 9, 405, 427–429, 453–463, 466–468, 472, 541–542
voltage
capacitance and, 89–91, 127–128
difference, 476, 482, 511
drop, 11–12, 92–93, 168–169, 178–179
ground bounce, 168–169
incident, 285–288
inductance and, 158–161
instantaneous, 311, 428–429
negative, 289
noise, 617–618
PDN and, 178
polarity, 92, 160
rails, 135, 615, 620, 626–631
ratios, 377
reflected, 283, 288–289
transmitted, 285–287
voltage regulator module (VRM), 615, 619, 624, 631–635, 644–646, 711
volume resistivity, 119
Vss, 45, 173, 617

W

waveforms
conversion, 57
frequency domain, 94
impulse, 557
impulse response, 609–612
significant, 63–66
step edge, 72, 295, 557–558, 609–610
time domain, 45, 53, 59, 66, 609–612
wavelets, 46
Webers, 153–155
wire bond, 73, 115–116, 120, 165–172
wire material composition, 154
www.beTheSignal.com, 672

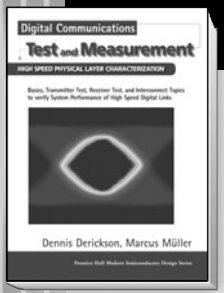
X

XAUI, 22
X2Y Attenuators, 647, 650–651, 700–703

Z

Z_{diff}, 484–486, 494, 507
Z_{even}, 508–509, 519, 524–527, 549, 552
Z_{odd}, 507, 519, 524–527, 548, 552
Z_{PDN}, 619, 624
Z_{peak}, 670–671
Z_{target}, 619, 621–629
Zcap, 315–316
Zcapacitors, 681
zeroth harmonic, 56–60
zeroth-order model, 224–227, 242, 260

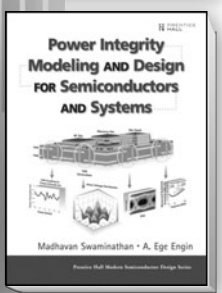
SIGNAL INTEGRITY LIBRARY FROM PRENTICE HALL PROFESSIONAL



◆ **DIGITAL COMMUNICATIONS TEST AND MEASUREMENT**
High-Speed Physical Layer Characterization

Dennis Derickson and Marcus Müller, editors • ©2008, 0-13-220910-1

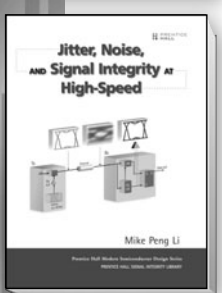
This book focuses on the important test and measurement techniques used for improving the performance of high-speed communications hardware. The main areas of coverage are waveform analysis, bit error ratio measurements, and jitter analysis of high-speed signals.



◆ **POWER INTEGRITY MODELING AND DESIGN FOR SEMICONDUCTORS AND SYSTEMS**

Madhavan Swaminathan and A. Ege Engin • ©2008, 0-13-615206-6

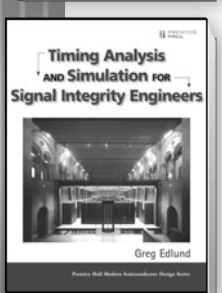
This is the first book written on Power Integrity Modeling. It covers everything from the basics to advanced concepts. The focus of the book is on being able to model the voltage fluctuations on advanced semiconductor packages and boards, especially on system applications.



◆ **JITTER, NOISE, AND SIGNAL INTEGRITY AT HIGH-SPEED**

Mike Peng Li • ©2008, 0-13-242961-6

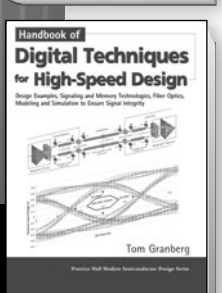
This book presents the latest theory and analysis methods for jitter, noise, and bit error (JNB) and signal integrity (SI) for high-speed devices and systems. It starts from basic statistical signaling and linear system theory and then applies them in solving the JNB and SI problems for both link components and overall systems.



◆ **TIMING ANALYSIS AND SIMULATION FOR SIGNAL INTEGRITY ENGINEERS**

Greg Edlund • ©2008, 0-13-236504-9

Chip designers live and die by timing. This is the first book devoted to giving signal integrity engineers a practical understanding of how to do timing analysis for signal integrity.



◆ **HANDBOOK OF DIGITAL TECHNIQUES FOR HIGH-SPEED DESIGN**
Design Examples, Signaling and Memory Technologies, Fiber Optics Modeling, and Simulation to Ensure Signal Integrity

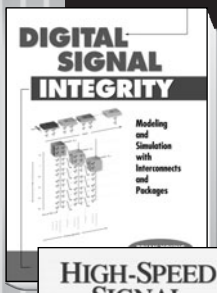
Tom Granberg • ©2004, 0-13-142291-X

Encompasses more than just signal integrity. Senior design engineers and staff engineers will benefit by developing a robust skillset and having a desk-side design companion 24/7. Provides detailed technical information on high-speed device families, applications, and more than 30 EDA tools that other books omit.

PRENTICE HALL

For more information, please visit InformIT.com/ph

SIGNAL INTEGRITY LIBRARY FROM PRENTICE HALL PROFESSIONAL

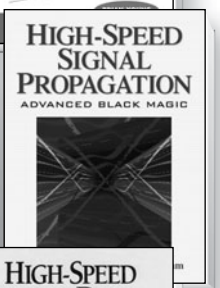


◆ DIGITAL SIGNAL INTEGRITY

Modeling and Simulation with Interconnects and Packages

Brian Young • ©2001, 0-13-028904-3

Offers a strong mathematical foundation for every technique. Contains all the theory and practical methods needed to accurately model and simulate high-speed interconnections and predict real-world performance. In-depth reviews of inductance, capacitance, resistance, single and multiconductor transmission lines, advanced modeling with modal decompositions, generalized termination schemes, crosstalk, and differential signaling issues.

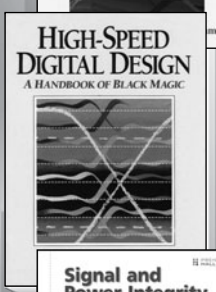


◆ HIGH-SPEED SIGNAL PROPAGATION

Advanced Black Magic

Howard Johnson and Martin Graham • ©2003, 0-13-084408-X

Advanced-level reference for digital designers interested in the upper limits of speed and distance on bandwidth-limited communication channels. Extensive treatment of PCB traces, copper cables, fiber, differential signaling, and jitter. Explains the principles of skin effect, surface roughness, dielectric loss, and non-TEM modes of propagation using both frequency-domain and time-domain modeling techniques. A companion to the classic *High-Speed Digital Design*.

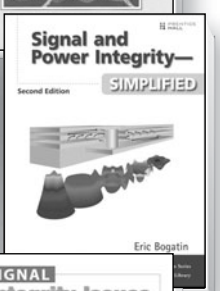


◆ HIGH-SPEED DIGITAL DESIGN

A Handbook of Black Magic

Howard W. Johnson • ©1993, 0-13-395724-1

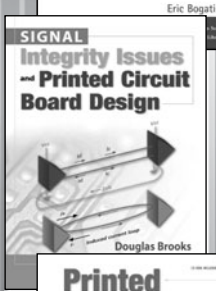
Broadest coverage of topics for digital board-level designers including transmission lines, terminations, crosstalk, power distribution, chip packaging, heat dissipation, probes, connectors, cables, and clock oscillators. Appropriate for readers with any level of technical background. With over 50,000 copies sold this is the gold standard against which other books in the field are measured.



◆ SIGNAL AND POWER INTEGRITY — SIMPLIFIED, SECOND EDITION

Eric Bogatin • ©2010, 0-13-234979-5

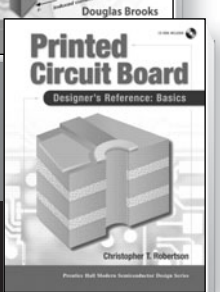
An introduction to the world of signal integrity. Provides a solid foundation on the principles used in signal integrity. Illustrates the use of modern simulation and measurement tools. The #1 guide to signal integrity, updated with all-new coverage of power integrity, high-speed serial links, and more.



◆ SIGNAL INTEGRITY ISSUES AND PRINTED CIRCUIT BOARD DESIGN

Douglas Brooks • ©2003, 0-13-141884-X

Our technician-level signal integrity book. Written in a plain spoken, non-rigorous manner for the non-degreed designer, community college or technical institute student, or recent engineering graduate, especially those with little EMC or high-speed background. Covers printed circuit board level interconnects.



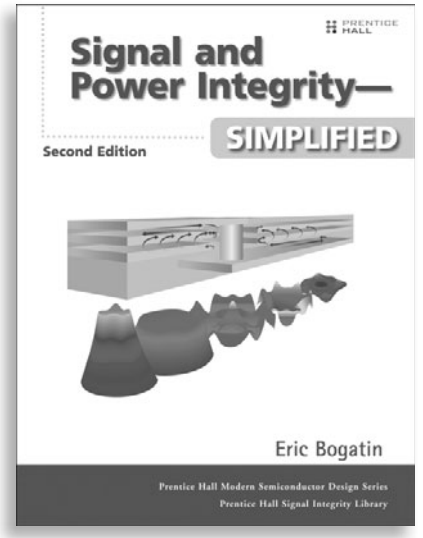
◆ PRINTED CIRCUIT BOARD DESIGNER'S REFERENCE

Chris T. Robertson • ©2004, 0-13-067481-8

For beginners and seasoned PCB designers, this guide contains both basic information as well as advanced tools. Nearly every aspect of basic board design is covered, including DFM. The accompanying software provides the everyday tools that are left out of most layout programs. PCB Designer's Reference provides the fundamentals of "why" and "how" a board is designed, laying the groundwork for signal integrity issues. Robertson and Brooks make an excellent two-book package.

PRENTICE HALL

For more information, please visit InformIT.com/ph



FREE Online Edition

Your purchase of ***Signal and Power Integrity—Simplified, Second Edition*** includes access to a free online edition for 45 days through the Safari Books Online subscription service. Nearly every Prentice Hall book is available online through Safari Books Online, along with more than 5,000 other technical books and videos from publishers such as Addison-Wesley Professional, Cisco Press, Exam Cram, IBM Press, O'Reilly, Que, and Sams.

SAFARI BOOKS ONLINE allows you to search for a specific answer, cut and paste code, download chapters, and stay current with emerging technologies.

Activate your FREE Online Edition at www.informit.com/safarifree

- **STEP 1:** Enter the coupon code: UBVUNCB.
- **STEP 2:** New Safari users, complete the brief registration form. Safari subscribers, just log in.

If you have difficulty registering on Safari or accessing the online edition, please e-mail customer-service@safaribooksonline.com

Safari Books Online

Addison Wesley · Adobe Press · Alpha · Cisco Press · FT Press · IBM Press · lynda.com · Microsoft Press · New Riders · O'Reilly · Peachpit Press · Que · Redbooks · Sams · SAS Publishing · Sun microsystems · Wharton School Publishing · Wiley