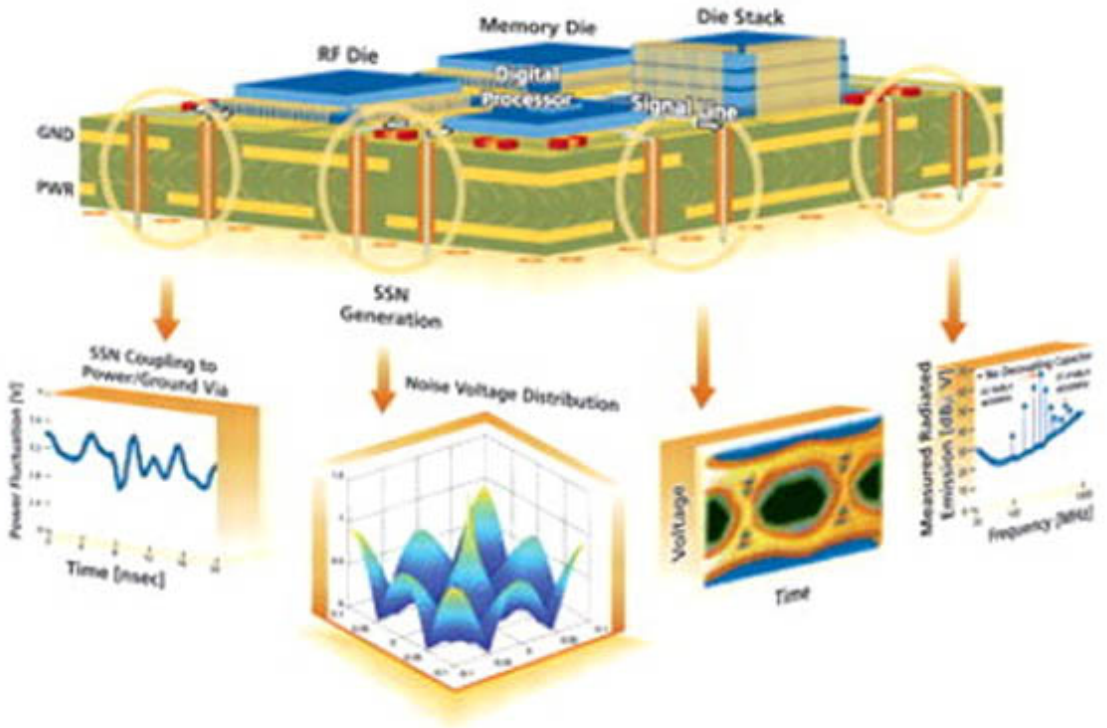
PRENTICE HALL

Power Integrity Modeling and Design for Semiconductors and Systems



Madhavan Swaminathan • A. Ege Engin

Prentice Hall Modern Semiconductor Design Series
Prentice Hall Signal Integrity Library

POWER INTEGRITY MODELING AND DESIGN FOR SEMICONDUCTORS AND SYSTEMS

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 Integrity: Simplified

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 P. 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: Simulation and Formal Method-Based Approaches

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

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: System-on-Chip Design, Third Edition

Bob Zeidman
Verilog Designer's Library

Power Integrity Modeling and Design for Semiconductors and Systems

Madhavan Swaminathan
A. Ege Engin

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

Visit us on the Web: www.prenhallprofessional.com

Library of Congress Cataloging-in-Publication Data

Swaminathan, Madhavan.
Power integrity modeling and design for semiconductors and systems / Madhavan Swaminathan, A. Ege Engin.
p. cm.
Includes index.
ISBN 0-13-615206-6 (hardback : alk. paper) 1. Semiconductors—Protection. 2. Integrated circuits—Protection. 3. Semiconductors—Design and construction. 4. Integrated circuits—Design and construction. I. Engin, A. Ege. II. Title.
TK7871.85.S935 2007
621.3815'2—dc22 2007034155

ISBN-13: 978-0-13-615206-4
ISBN-10: 0-13-615206-6

Text printed in the United States on recycled paper at Courier in Westford, Massachusetts.
First printing, November 2007

Editor-in-Chief: Mark Taub
Acquisitions Editor: Bernard Goodwin
Managing Editor: John Fuller
Project Editor: Elizabeth Ryan
Copy Editor: Carol Lallier
Indexer: Richard Evans
Proofreader: Amy Finch
Cover Designer: Alan Clements
Composition: codeMantra

This book is dedicated to my father, K. Swaminathan, who taught me perseverance, and to my mother, Rajam Swaminathan, who taught me patience, two qualities I have learned to cherish.

—Madhavan Swaminathan

To my mother and father, Tanju and Burhan Engin.

—A. Ege Engin

This page intentionally left blank

CONTENTS

Preface xiii

Acknowledgments xvii

About the Authors xxi

Chapter 1

Basic Concepts 1

1.1 Introduction 1

1.1.1 Functioning of Transistors 1

1.1.2 What Are the Problems with Power Delivery? 4

1.1.3 Importance of Power Delivery in Microprocessors and ICs 5

1.1.4 Power Delivery Network 6

1.1.5 Transients on the Power Supply 8

1.2 Simple Relationships for Power Delivery 10

1.2.1 Core Circuits 10

1.2.2 I/O Circuits 14

1.2.3 Delay Due to SSN 15

1.2.4 Timing and Voltage Margin Due to SSN 16

1.2.5 Relationship between Capacitor and Current 17

1.3 Design of PDNs 17

1.3.1 Target Impedance 20

1.3.2 Impedance and Noise Voltage 22

1.4 Components of a PDN 24

1.4.1 Voltage Regulator 24

1.4.2 Bypass or Decoupling Capacitors 28

1.4.3 Package and Board Planes 37

1.4.4 On-Chip Power Distribution 42

1.4.5 PDN with Components 45

1.5 Analysis of PDNs 45

1.5.1 Single-Node Analysis 48

1.5.2 Distributed Analysis 55

1.6 Chip-Package Antiresonance: An Example 61

1.7 High-Frequency Measurements 65

1.7.1 Measurement of Impedance 66

1.7.2 Measurement of Self-Impedance 68

1.7.3 Measurement of Transfer Impedance 70

1.7.4 Measurement of Impedance by Completely Eliminating Probe Inductance 70

1.8 Signal Lines Referenced to Planes 71
1.8.1 Signal Lines as Transmission Lines 72
1.8.2 Relationship between Transmission-Line Parameters and SSN 74
1.8.3 Relationship between SSN and Return Path Discontinuities 75
1.9 PDN Modeling Methodology 77
1.10 Summary 79

Chapter 2

Modeling of Planes 83
2.1 Introduction 83
2.2 Behavior of Planes 84
2.2.1 Frequency Domain 84
2.2.2 Time Domain 86
2.2.3 Two-Dimensional Planes 88
2.3 Lumped Modeling Using Partial Inductances 89
2.3.1 Extracting the Inductance and Resistance Matrices 90
2.4 Distributed Circuit-Based Approaches 94
2.4.1 Modeling Using Transmission Lines 94
2.4.2 Transmission Matrix Method (TMM) 97
2.4.3 Frequency-Dependent Behavior of Unit-Cell Elements 104
2.4.4 Modeling of Gaps in Planes 113
2.5 Discretization-Based Plane Models 117
2.5.1 Finite-Difference Method 117
2.5.2 Finite-Difference Time-Domain Method 128
2.5.3 Finite-Element Method 132
2.6 Analytical Methods 133
2.6.1 Cavity Resonator Method 133
2.6.2 Network Representation of the Cavity Resonator Model 135
2.7 Multiple Plane Pairs 138
2.7.1 Coupling through the Vias 141
2.7.2 Coupling through the Conductors 154
2.7.3 Coupling through the Apertures 158
2.8 Summary 169

Chapter 3

Simultaneous Switching Noise 175
3.1 Introduction 175
3.1.1 Methods for Modeling SSN 175
3.2 Simple Models 177
3.2.1 Modeling of Output Buffers 180

3.3 Modeling of Transmission Lines and Planes 185
3.3.1 Microstrip Configuration 186
3.3.2 Stripline Configuration 189
3.3.3 Conductor-Backed Coplanar Waveguide Configuration 205
3.3.4 Summary of Modal Decomposition Methods 207
3.4 Application of Models in Time-Domain Analysis 209
3.4.1 Plane Bounce from Return Currents 209
3.4.2 Microstrip-to-Microstrip Via Transition 217
3.4.3 Split Planes 222
3.5 Application of Models in Frequency-Domain Analysis 226
3.5.1 Stripline between a Power and a Ground Plane 226
3.5.2 Microstrip-to-Stripline Via Transition 228
3.5.3 Reduction of Noise Coupling Using Thin Dielectrics 231
3.6 Extension of M-FDM to Incorporate Transmission Lines 233
3.6.1 Analysis of a Complex Board Design 236
3.7 Summary 239

Chapter 4

Time-Domain Simulation Methods 243
4.1 Introduction 243
4.2 Rational Function Method 244
4.2.1 Basic Theory 244
4.2.2 Interpolation Schemes 246
4.2.3 Properties of Rational Functions 252
4.2.4 Passivity Enforcement 257
4.2.5 Integration in a Circuit Solver 283
4.2.6 Disadvantages 291
4.3 Signal Flow Graphs 295
4.3.1 Causality 296
4.3.2 Transfer-Function Causality 296
4.3.3 Minimum Phase 296
4.3.4 Delay Extraction from Frequency Response 300
4.3.5 Causal Signal Flow Graphs 302
4.3.6 Computational Aspects in SFG 303
4.3.7 Fast Convolution Methods 307
4.3.8 Cosimulation of Signal and Power Using SFGs 312

4.4 Modified Nodal Analysis (MNA) 317
4.4.1 What Is MNA? 317
4.4.2 Frequency Domain 318
4.4.3 Time Domain 320
4.4.4 MNA Formulation with *S*-Parameters 322
4.5 Summary 327

Chapter 5

Applications 333
5.1 Introduction 333
5.2 High-Speed Servers 334
5.2.1 Core PDN Noise 336
5.2.2 I/O PDN Noise 345
5.2.3 Summary 349
5.3 High-Speed Differential Signaling 349
5.3.1 Test Vehicle Description 350
5.3.2 Plane Modeling 352
5.3.3 Modeling of Master and Slave Islands 358
5.3.4 Rational Function Modeling 361
5.3.5 Modal Decomposition and Noise Simulation 361
5.3.6 Summary 364
5.4 Analysis of IC Packages 365
5.4.1 Simulation of a Multilayered Package Using M-FDM 366
5.4.2 Causal Simulation of HyperBGA Package 368
5.4.3 Summary 372
5.5 Extraction of Dielectric Constant and Loss Tangent 372
5.5.1 Problem Definition 373
5.5.2 Corner-to-Corner Plane-Probing Method 378
5.5.3 Causal Model Development 386
5.5.4 Summary 391
5.6 Embedded Decoupling Capacitors 392
5.6.1 Embedded Individual Thin- or Thick-Film Capacitors 394
5.6.2 Why Embed Individual Capacitors 395
5.6.3 Design of an Embedded Thick-Film Capacitor Array 395
5.6.4 Integration of Embedded Capacitors into IBM Package 400
5.6.5 Embedded Planar Capacitors 404
5.6.6 Summary 415

5.7 Electromagnetic Bandgap (EBG) Structures 415
5.7.1 Basic Theory 416
5.7.2 Response of EBG Structures 417
5.7.3 Dispersion-Diagram Analysis 420
5.7.4 Modification of M-FDM Using Fringe and Gap Fields 424
5.7.5 Scalable Design of EBG Structures for Power Plane Isolation 430
5.7.6 Digital–RF Integration 434
5.7.7 ADC Load-Board Design 436
5.7.8 Issues with EBG Structures for Digital Systems 439
5.7.9 Summary 442
5.8 Future Challenges 443

Appendix A **451**
A.1 Multiport Networks 451
A.2 Matrix Representation of Transmission Lines 453
A.3 Spectrum of Digital Signals 454

Appendix B Software list **459**

Index **461**

This page intentionally left blank

PREFACE

During my (M.S.) undergraduate days in a little town called Tiruchirapalli in Southern India, we used to have frequent voltage and current surges that knocked out all the electrical equipment such as fans and lights in our rooms. Frustrated, my friend once remarked, "We are *powerless* to solve the *current* problem." Of course, he meant this in jest, but little did I realize that his statement would become the theme of my research for many years. Although my area of specialty is semiconductors and computer systems, the issues related to power haven't changed.

Power represents the major bottleneck in modern semiconductors and systems. With transistor scaling over the last two decades, Moore's law has enabled the integration of millions of transistors within an integrated circuit. With lower gate capacitance and lower voltage, faster transistors have become available with each new generation of computers. However, increased transistor integration has resulted in an increase in the current supplied to the integrated circuit, thereby increasing power. Managing the *transient current* supplied to the integrated circuit at gigahertz frequencies is one of the biggest challenges faced by the semiconductor industry. With lowering of the supply voltage to the transistors, dynamic variation in the power supply due to current transients is becoming a major bottleneck. The dynamic variation of the supply voltage, also called power supply noise, delta I noise, or simultaneous switching noise, is the subject of this book.

Managing power integrity is the process by which the variations on the power supply of the transistors can be maintained within a specified tolerance

value. Noise on the power supply can have a direct influence on the speed of an integrated circuit, and hence supplying clean power is a very important element in the design of a computer system. A power distribution network consists of interconnections in the chip, package, and board that include decoupling capacitors, ferrite beads, DC–DC converters, and other components. Both the package and board form a very critical part of the power distribution network, which is the focus of this book.

The book covers two aspects of power distribution: design and modeling, with an emphasis on modeling. The book is organized into five chapters, which cover basic and advanced concepts. All chapters contain several examples to illustrate the concepts, some of which can be reproduced using the software provided. These examples can also be used to evaluate the accuracy and speed of several commercial tools that are available today.

Chapter 1, "Basic Concepts," is for engineers and students who are entering the field of power integrity. The basic concepts are covered in this chapter, which includes a discussion on the fundamentals of power supply noise, its role in the speed of a computer system, the parasitics that produce it, and its effect on jitter and voltage margin for high-speed signal propagation. A power distribution network is best designed in the frequency domain, and the reasons for this are discussed in this chapter. The entire book is based on the parameter called *target impedance*, which can be used to evaluate the properties of a power distribution network. This parameter, developed in the mid-1990s, provides an elegant method of analysis, which can be used to understand the role of various components in the response of a power distribution network. The target impedance is therefore explained in detail in Chapter 1, with examples that can be reproduced using a circuit simulator such as Simulated Program with Integrated Circuit Emphasis (Spice). The concept of target impedance is used to promote better understanding of the placement of decoupling capacitors. The components of a power distribution network consist of several voltage regulator modules, decoupling capacitors, package and board interconnections, planes, and on-chip interconnections, each of which are explained in this chapter. Planes represent a very critical part of modern power distribution networks. Their frequency behavior can either reduce power supply noise or increase it by a large amount. Hence, a fundamental understanding of plane behavior and its effect on advanced power distribution networks is necessary. The entire book is centered around planes from both a modeling and design standpoint. The fundamental behavior of planes is covered in Chapter 1, with a focus on standing waves, their frequency of occurrence, capacitive and inductive

behavior, and use of decoupling capacitors to minimize their effect. The interaction between components of a power distribution is as important as the components themselves. For example, a surface-mount device (SMD) capacitor can interact with the via inductance, causing the self-resonance frequency to shift to a lower frequency; the chip can interact with the package, causing an antiresonance; or the power supply noise can couple into a signal line, causing excessive jitter. The basics associated with such phenomena are covered in Chapter 1. Finally, a methodology is presented that centers on frequency domain analysis initially followed by time domain analysis. The authors believe that this is the optimum way for analyzing and designing advanced power distribution networks.

A power distribution network containing suitably designed planes, signals well referenced to planes, and decoupling capacitors appropriately placed on planes will always result in minimum power supply noise. Planes are therefore the focus of Chapter 2, "Modeling of Planes," which covers the various methods available for plane modeling. Some of these methods are used by commercial tools today. This chapter, which requires some background in numerical modeling, provides a survey of modeling methods along with examples that are useful to a designer and can be used to evaluate commercial tools for accuracy and speed. The in-depth numerical formulations can be reproduced in MATLAB and hence are useful to both students and application engineers who are interested in power integrity modeling. Since Maxwell's equations have been converted into circuit representations, we believe that the numerical formulations in this chapter are easier to understand. The modeling methods are separated into lumped element modeling and distributed modeling methods, each covered in detail. The chapter starts with modeling a plane pair and then explains modeling of multilayered planes. The coupling effects in multilayered planes, which include field penetration concepts, aperture coupling, and wraparound currents, are discussed, and the plane modeling methods are compared from a qualitative standpoint. This comparison, along with the rest of the chapter, allows an engineer to benchmark commercial tools.

Signals from the output of a driver are propagated on signal line interconnections. However, the driver requires voltage and current to function, and these are supplied by the power distribution network. The signal and power interconnections therefore have to be coupled, with noise on one producing noise on the other. Hence, managing both signal and power integrity requires an understanding of the coupling mechanism between the signal lines and planes. Chapter 3, "Simultaneous Switching Noise," requires little understanding of numerical methods. The

entire chapter is based on circuit-level implementations using a concept called *modal decomposition*, which allows the separation of signal lines from the power distribution network so that each can be analyzed separately and later combined for analysis. Simple Spice models can be used to capture modal decomposition using coupling coefficients and controlled current or voltage sources. The important concept to understand in this chapter is the role of return currents—a concept that every power integrity engineer must understand for minimizing noise.

Chapter 4, "Time-Domain Simulation Methods," describes methods for converting a frequency response into a Spice subcircuit. Also called *macromodeling,* this is a new area of time-domain simulation that is ripe for research. We include this chapter in the book because a few commercial electronic design automation (EDA) vendors have started developing tools in this area. The purpose of Chapter 4 is to enable an engineer or student to better understand the issues involved. The early part of the chapter is easy to follow; it requires some mathematics background and is therefore targeted at designers who use commercial tools. Several examples illustrate simple concepts that can be reproduced using MATLAB. The latter part of the chapter is intense and is mainly intended for people working in the numerical modeling area. The purpose of this chapter is to provide an introduction to the issues involved and possible solutions.

In Chapter 5, "Applications," all of the issues discussed in Chapters 1 to 4 are linked to real-world examples. Several examples from companies such as Sun Microsystems, IBM, Oak Mitsui, National Semiconductor, Cisco, DuPont, Panasonic, and Rambus are provided. These applications cover both design and modeling aspects of power integrity. Each example was chosen carefully to ensure that a specific aspect of power integrity is addressed.

The best part of the book is that it reproduces some of the examples using the software provided. We hope that through this software, some of the subtle effects related to power integrity, which are only discussed in research papers, can be reproduced and appreciated by a larger community.

Madhavan Swaminathan (M. S.)
A. Ege Engin (A. E. E.)

ACKNOWLEDGMENTS

The power integrity work discussed in this book started in the mid-1990s when I had the opportunity to collaborate with Larry Smith and Istvan Novak of Sun Microsystems. This collaboration over a 5-year period laid the foundation for the work described in this book. I am forever grateful to both Larry and Istvan for providing me insights into the power distribution issues, making available several test vehicles, and providing access to system-level measurements. Both Larry and Istvan are my good friends, and I continue to consult with them.

Around 2000, I started working with James Libous of IBM through an SRC project. Work over the next several years resulted in some very interesting ideas in power integrity modeling. In research, not every idea results in a useful solution. However, without trying out new ideas, advancements are not possible. I am forever indebted to Jim for providing me unconditional support for my research without raising any doubts whatsoever on the validity of the solutions. Through his mentorship and leadership of the SRC project, I had a number of students work on very interesting problems related to power integrity, most of which have been captured in this book.

My move to Georgia Tech in 1994 is what started all of my research on power integrity. I was recruited by Professor Rao Tummala, a former IBM Fellow who is currently a chaired professor at Georgia Tech, and Professor Roger Webb, the former Chair of Electrical and Computer Engineering at Georgia Tech. Through their mentorship and guidance, I was able to start a research program and I am forever thankful to them.

My strength has always been my students. I have challenged them in research and have often made unreasonable demands, but my students have always delivered. This book is a result of their hard work over a 10-year period. I would like to acknowledge the support of all my students and in particular the following graduate students who made a direct contribution to this book: Nanju Na, Jinseong Choi, Sungjun Chun, Joongho Kim, Sunghwan Min, Rohan Mandrekar, Jifeng Mao, Jinwoo Choi, Vinu Govind, Krishna Bharath, Abdemanaf Tamabawala, Krishna Srinivasan, Lalgudi Subramaniam, Prathap Muthana, Tae Hong Kim, Ki Jin Han, Janani Chandrasekhar, and Bernard Yang. Some of them have graduated, but I continue to interact with them.

I have been fortunate to have had several visiting scholars who worked with me over several years in the area of Power Integrity. I am grateful to the work done by Hideki Sasaki (NEC), Takayuki Watanabe (Shizuoka University), Yoshitaka Toyota (Okayama University), and Kazuhide Uriu (Panasonic) in this area.

Every chapter in this book has been carefully reviewed by experts in the field. I would like to acknowledge the following people who, through their careful review and feedback, have ensured the quality of this book: Chapter 1, Tawfik Arabi (Intel) and Jose Scutt-Aine (University of Illinois); Chapter 2, Eric Bogatin (Bogatin Enterprises) and Barry Rubin (IBM); Chapter 3, Flavio Canavero (University of Torino) and Bob Ross (Teraspeed); Chapter 4, Ram Achar (Carleton University) and Rohan Mandrekar (IBM); Chapter 5, Mahadevan Iyer (Georgia Tech) and Dan Amey (DuPont).

I have wanted to write a book on power integrity for several years. It all became possible after my coauthor, Ege Engin, joined my research program in 2005. Through his hard work and meticulous planning, we were able to complete this book. My thanks go to him.

A very important element of this book is the software. User-friendly software is always very difficult to develop. I would like to thank Andy Seo, who spent several nights developing the graphical user interface for the Sphinx software program, described in Appendix B. I am also very thankful to Sunghwan Min, who developed BEMP, described in Chapter 4 and Appendix B.

Finally, I would like to acknowledge the support of my wife Shailaja and daughter Sharanya. This book was written as a result of research over a 10-year period. During this time, I visited and worked with several companies, academic institutions, and research organizations and traveled more than a million miles. The most affected by these trips have been my wife and daughter. Without their loving support, this book would not have been possible.

Madhavan Swaminathan

Let me begin by expressing my thanks to the many reviewers of this book, whose names are given above. I started working on power integrity during my Ph.D., while I was working as a research engineer at the Fraunhofer Institute for Reliability and Microintegration in Berlin, Germany. Among my many former colleagues, I would especially like to thank Uwe Keller and Umberto Paoletti for the inspiring discussions on computational electromagnetics, and Ivan Ndip for the discussions on high-speed design. My Ph.D. advisor, Professor Wolfgang Mathis of the University of Hannover, has shown me new ways of looking at power-integrity modeling from a circuit point of view. I would like to thank him and my coadvisor, Professor Herbert Reichl of the Fraunhofer Institute, both of whom have always supported me.

After I completed my Ph.D., I moved to Georgia Tech to work with Professor Madhavan Swaminathan, who encouraged me to coauthor this book. It is a pleasure for me to acknowledge his mentorship and to thank him for it. I would also like to thank Professor Rao Tummala from the Packaging Research Center at Georgia Tech for his continuous support.

Finally, I would like to thank my wife, Asuman, and my son, Anka, for their love and patient support.

A. Ege Engin

This page intentionally left blank

About the Authors

Madhavan Swaminathan received his B.E. in electronics and communication from Regional Engineering College, Tiruchirapalli, in 1985, and his M.S. and Ph.D. in electrical engineering from Syracuse University in 1989 and 1991. He is currently the Joseph M. Pettit Professor in Electronics in the School of Electrical and Computer Engineering and deputy director of the Packaging Research Center, Georgia Tech. He is also the cofounder of Jacket Micro Devices, a company specializing in RF modules for wireless applications. Before joining Georgia Tech, he worked on packaging for supercomputers for IBM. Swaminathan has written more than 300 publications, holds 15 patents, and has been honored as an IEEE Fellow for his work on power delivery.

A. Ege Engin received his B.S. and M.S. in electrical engineering from Middle East Technical University, Ankara, Turkey, and from the University of Paderborn, Germany. From 2001 to 2004, he was with the Fraunhofer Institute for Reliability and Microintegration in Berlin. During this time, he also received his Ph.D. from the University of Hannover, Germany. He is currently a research engineer in the School of Electrical and Computer Engineering and an assistant research director of the Packaging Research Center at Georgia Tech. He has more than 50 publications in refereed journals and conferences in the areas of signal and power integrity modeling and simulation.

This page intentionally left blank

CHAPTER 1

Basic Concepts

Power delivery is a major challenge in present-day systems. This challenge is expected to increase in the next decade as systems become smaller and new materials are introduced into packages and boards. As devices scale and more transistors are integrated into a single integrated circuit, the power and current levels are expected to increase with a corresponding decrease in the voltage. With gigabit signals being propagated through the package and board, the ability to supply clean power to the transistor circuits becomes very critical. In addition, electromagnetic interference levels have to be kept low in the system to manage coupling and crosstalk.

In this chapter, the basics of power delivery are described. Along with a description of the components of a power delivery network (PDN), the analysis methodology of such networks is described with examples.

1.1 Introduction

1.1.1 Functioning of Transistors

Integrated circuits (ICs) such as microprocessors, field programmable gate arrays, memory devices, and other application-specific ICs contain transistors. Transistors are multiterminal switches that can be turned on or off on the basis of a control signal. The on or off position of the switch determines the current flowing through the device. In complementary metal oxide semiconductor (CMOS) field effect transistor (MOSFET) technology (which is the most popular technology used to

design microprocessors), two types of transistors are used, namely the NMOS (n-channel) transistor and the PMOS (p-channel) transistor. The detailed operation of these devices can be obtained from [1]. In this book, for simplicity, we will assume that both transistors are three-terminal devices that can be represented using switches, as shown in Figure 1-1. The three terminals are called the gate, source, and drain. By applying a voltage between the gate and source, the current through the transistor (from drain to source for NMOS and reverse for PMOS) can be turned on or off. The NMOS transistor is called as a normally open switch, since a gate voltage has to be applied to pass current through the transistor. Hence, if a binary 0 (logic level low) signal exists at the gate, the switch is OFF, and when a binary 1 (logic level high) signal is available at the gate, the switch turns ON. The reverse is true for the PMOS transistor, since a binary 1 level at the gate turns OFF the current, while a binary 0 level at the gate allows current to pass, that is, the switch is ON. Hence, the PMOS transistor is called as the normally closed switch. The drain and gate terminals of the NMOS and PMOS transistors can be connected together to form an inverter, which is one of the basic building blocks in any IC. We limit ourselves to the discussion of such an inverter in this section.

Figure 1-2 shows the inverter circuit. The gate connection is called as the input node, and the drain connection is called as the output node. The output node is connected to the input node of the following transistor circuit. Since the gate of

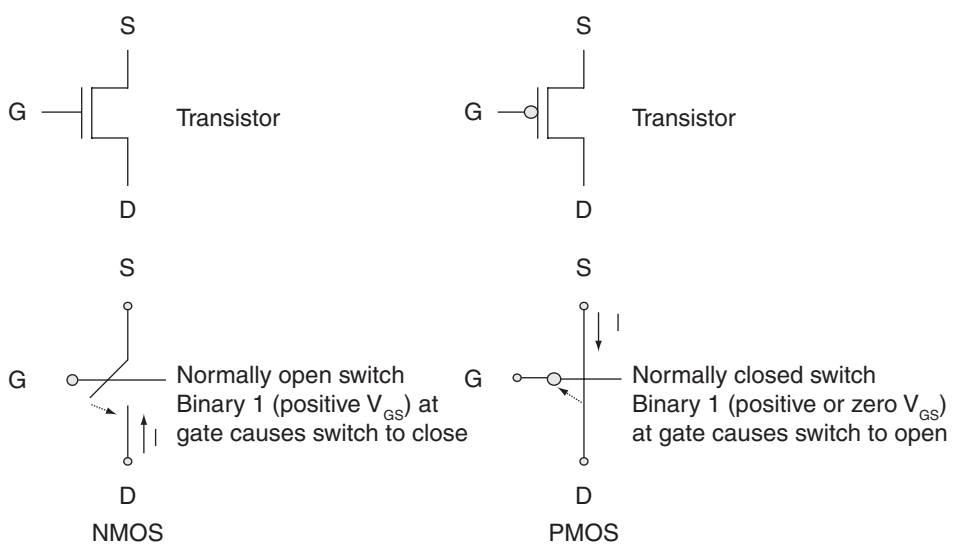


Figure 1-1 NMOS and PMOS transistors represented as switches.

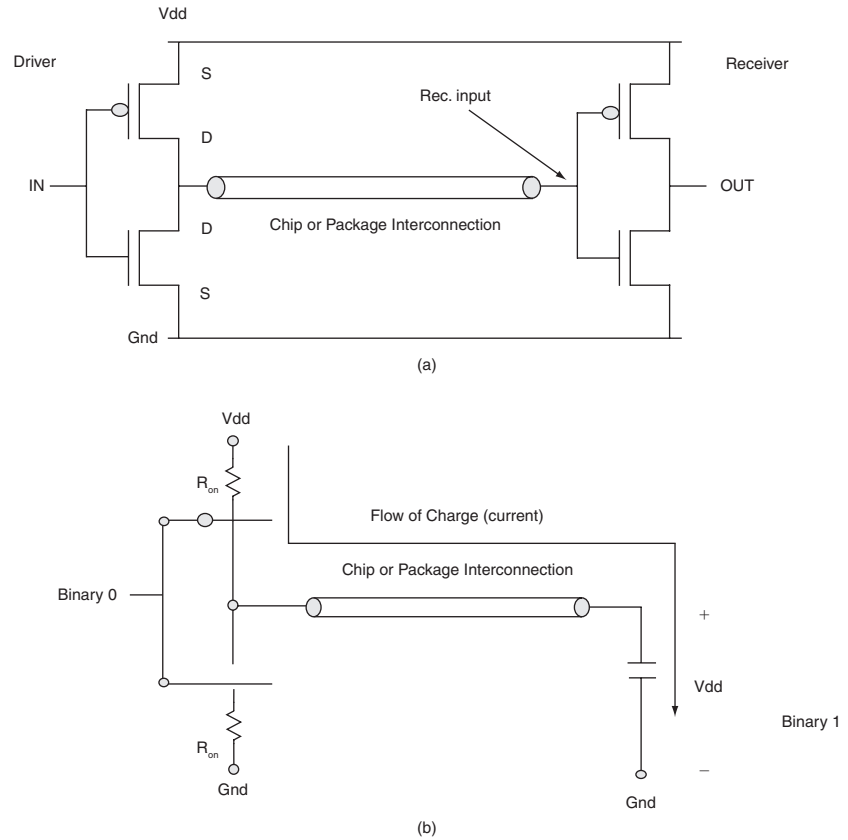


Figure 1-2 (a) Driver (inverter) connected to a receiver (inverter). (b) Input capacitance of receiver being charged to Vdd.

the transistors acts as a capacitor (formed between the metal-oxide-semiconducting substrate), the inverter (also called the driver) is used to charge and discharge the input capacitance of the succeeding stage. The capacitor must be charged to reach the binary 1 voltage level. Similarly, discharging a capacitor to 0 voltage requires the removal of charge. The inverter circuit must be connected to a power supply (shown as Vdd and Gnd terminals), which provides the ability to charge and discharge a capacitor node within the IC. In Figure 1-2, a wire (interconnection) is used between the two inverters to act as a conduit for the charge, and R_{on} is the on resistance of the transistor. The speed at which the circuit operates determines how quickly charge can be either supplied or removed from the capacitor through the switches. A PDN in a system provides the interconnection framework to make this happen, supplying the transistors with sufficient voltage and current for them to switch states.

1.1.2 What Are the Problems with Power Delivery?

The power supply (which is the source of the voltage and current) is typically bulky and cannot be connected directly to the Vdd and Gnd terminals of the IC. Therefore, wires (interconnections), which have resistance and inductance in them, are used to establish this connection. The current flowing through these wires creates both a DC drop (not shown) and time-varying fluctuation of the voltage across the Vdd and Gnd terminals of the IC (shown in Figure 1-3), which is detrimental to the transistors in the IC. Hence, a suitable PDN must be created between the power supply and the IC, such that the voltage is well regulated for the required current to be supplied to the transistors over a required time period. The voltage fluctuation across the Vdd and Gnd terminals of the transistors can cause the following problems with the transistors:

- Reduction in voltage across the power supply terminals of the IC that slows down the transistor or prevents the transistor from switching states.
- Increase in voltage across the power supply terminals of the IC, which creates reliability problems.
- Leakage of the voltage fluctuation into a quiet transistor, as shown in Figure 1-3, causing incorrect switching of quiet transistor circuits at the far end of a communication path along with crosstalk from neighboring signal lines.
- Timing margin errors caused by degraded waveforms at the output of the drivers.

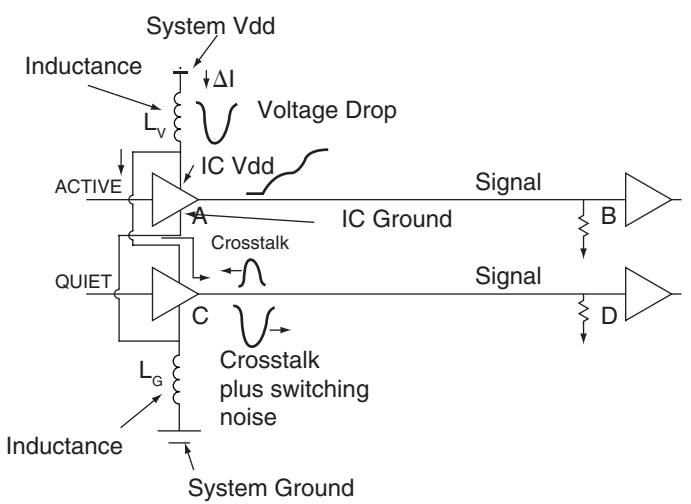


Figure 1-3 Voltage fluctuation [2].

The voltage fluctuation across the power supply of the IC is called power supply noise, delta I noise, or simultaneous switching noise (SSN), since it occurs only during the switching of the transistors.

1.1.3 Importance of Power Delivery in Microprocessors and ICs

Given the voltage fluctuations across the power supply of a transistor, it is helpful to understand how a microprocessor operates (as an example at a holistic level) and the impact of voltage fluctuations on microprocessor performance.

A microprocessor consists of millions of CMOS transistors interconnected through wires in a very complex fashion. The microprocessor speed can be limited by the gate (or transistor) delays, interconnect (or wire) delays, or both. The inverse of the gate delay (frequency) is proportional to the gate voltage. For a gate-dominated circuit, a 1% drop in the gate voltage results in nearly a 1% drop in frequency. The interconnect delay, however, is a very weak function of voltage. An important relationship exists between the operating voltage of the microprocessor and its speed (measured as frequency) around the nominal voltage of the microprocessor. This relationship is shown in Figure 1-4 for a 64-bit Scalable Processor Architecture (SPARC) microprocessor [3]. Around the nominal voltage of 1.6 V, the relationship between frequency and voltage is almost linear. As the graph shows, a reduction in voltage reduces the operating frequency of the microprocessor, while an increase in voltage increases its frequency. This important relationship is true in most microprocessors, and we use this example to explain the impact of power supply fluctuations on the operating frequency of the microprocessor. In reality, the relationship between processor performance and voltage is more complex and depends not only on the magnitude of power supply noise but also on the frequency of the noise.

Consider Figure 1-5, which assumes a linear relationship between the frequency of a microprocessor (along the y-axis) and voltage (along the x-axis), as in Figure 1-4. In Figure 1-5, F_{MAX} is the maximum operating frequency of the microprocessor. Any voltage above 1.65 V causes reliability problems and is shown as the reliability wall. Any voltage that falls within the reliability wall causes the dielectric breakdown of the gate oxide in the MOSFET due to excessive electric field. Hence, the power supply voltage cannot exceed 1.65 V for this example. Let's assume initially that the operating voltage of the microprocessor is 1.55 V. According to the graph, the nominal F_{MAX} for the microprocessor is 720 MHz. However, voltage variations on the power supply cause the voltage to vary plus or minus 100 mV around the nominal voltage. On the high side, a voltage of

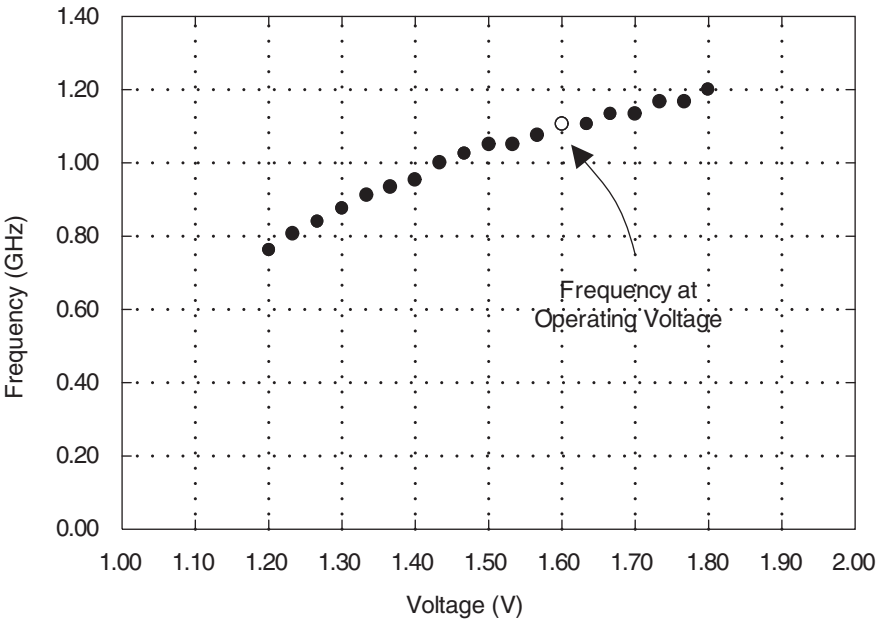


Figure 1-4 Frequency–voltage relationship for the 64-bit SPARC V9 microprocessor. The chip is built using a 150-nm technology that has seven layers of aluminum interconnects. It is designed to operate at 1.0 GHz with a supply voltage of 1.6 V. The temperature is maintained at 60°C [3].

1.65 V (1.55 V + 100 mV) is below the maximum allowed voltage of 1.65 V, which ensures no reliability problems. On the low side, the power supply voltage reduces to 1.45 V (1.55 V – 100 mV). At 1.45 V, F_{MAX} now becomes 670 MHz. Hence, any drop or reduction in the power supply voltage causes the microprocessor to operate at a lower frequency. In other words, the PDN causing the variations on the power supply terminals of the IC results in the slowdown of the microprocessor. Similarly, a voltage rise across the power supply of the IC, if it exceeds the maximum voltage allowable, causes the IC to malfunction.

In the design of PDNs, the focus is always on minimizing the voltage droop on the power supply terminals of the transistor circuits within an IC and also on ensuring that the voltage maximum does not cause reliability problems [4].

1.1.4 Power Delivery Network

A PDN consists of a power supply, DC–DC converters (also called voltage regulator modules, or VRMs), lots of decoupling capacitors, and interconnections

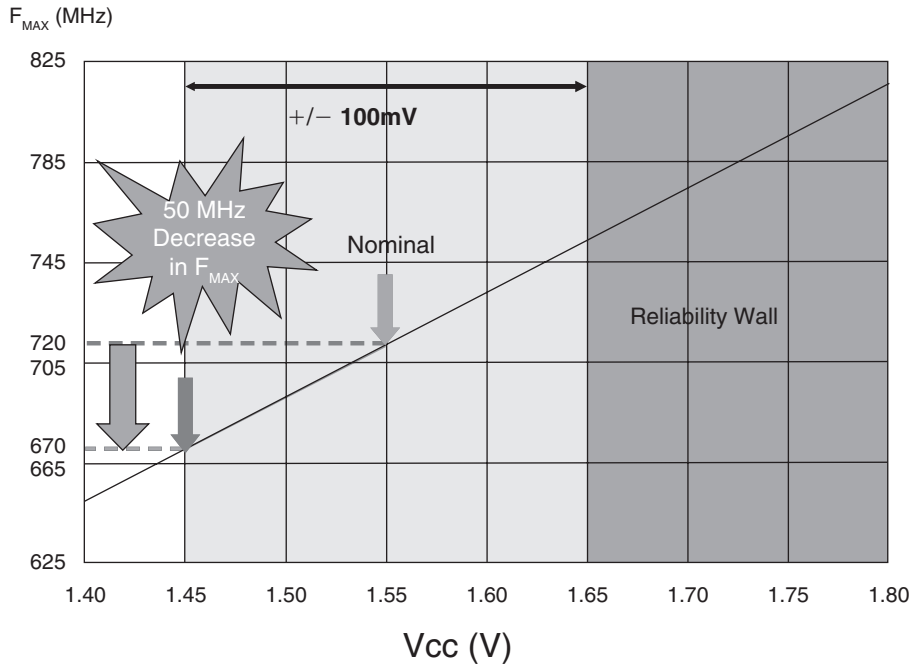


Figure 1-5 Relationship between voltage fluctuation and performance for a microprocessor [4].

that act as conduits for the supply and removal of charge to and from the switching circuits. In a typical computer system, the IC is packaged and placed on a motherboard (with or without a socket) with a power supply on the motherboard. The power supply provides high voltage and current to the motherboard. The voltage is reduced through a DC–DC converter and supplied to the IC through the interconnections in the motherboard and package. The decoupling capacitors are distributed on the motherboard, package, and IC; they act as reservoirs where charge can be stored. The charge is supplied as needed to the transistors from the decoupling capacitors. The proximity of the capacitors to the switching circuits determines the time required to supply the charge. The required time is controlled by the speed of light in the medium, which is the minimum time required to transfer the charge from the capacitor to the transistors. As an example, the minimum time required to supply charge from a capacitor placed on the motherboard 6 inches away from a transistor circuit is 1 ns, since the speed of light in typical printed circuit boards (PCBs) is 166 ps/inch.

A typical PDN for a semiconductor is shown in Figure 1-6 [5]. Since the inverse of time delay is frequency, the proximity of the capacitors to the transistors

determines if the capacitor supplies charge at high frequencies, middle frequencies, or low frequencies. The high-, mid-, and low-frequency capacitors are shown in the figure, where a capacitor farther away from the IC is always large and bulky, thereby operating at a lower frequency. The charge storage capacity of the large capacitors is of the order of thousands of microfarads, much higher than either the high- or mid-frequency capacitors, which are in the nanofarad range.

1.1.5 Transients on the Power Supply

Although the operating frequency of a microprocessor can be high (1 GHz and higher), power supply fluctuations can be caused over a range of frequencies, because a computer is a broadband system in which transistors switch at multiple frequencies. For example, a 1 GHz microprocessor in a system may be executing instruction at 1 GHz, causing voltage fluctuations at the 1 GHz frequency. At the same time, the microprocessor may be writing data to the cache on the PCB at 400 MHz and operating the Joint Test Access Group (JTAG) line for testing the hardware at 1 MHz. Such a switching activity can cause voltage fluctuations over a range of frequencies, which makes the design of the PDN very difficult. Voltage variation

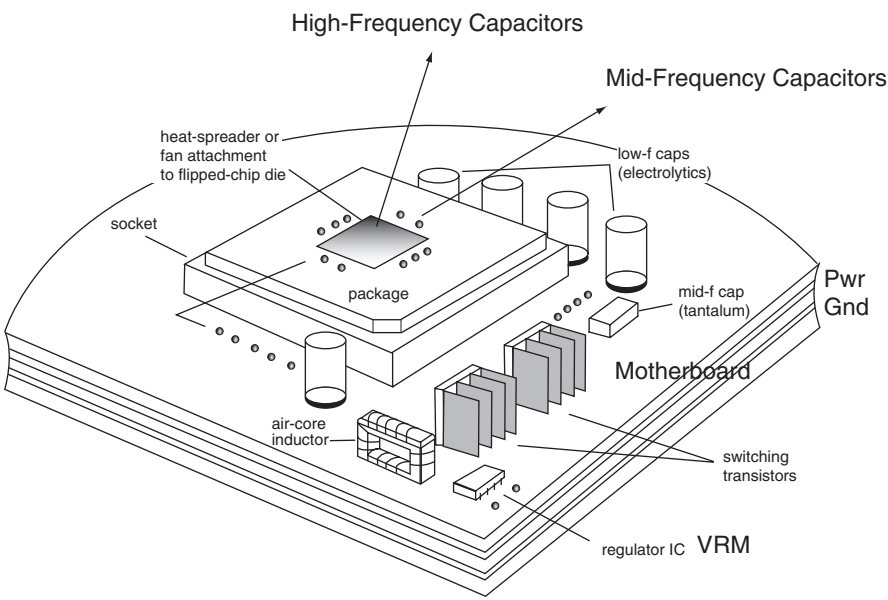


Figure 1-6 Power delivery network. By permission from D. Herrell and B. Beker, "Modeling of power distribution systems in PCs," in *Proceedings of the EPEP '98 Conference,* pp. 159–162, © 1998 IEEE.

on the power supply at multiple frequencies is shown in Figure 1-7 for a microprocessor [5]. For an IC, the transient current flowing through an inductor, L in Figure 1-3, causes voltage drop, V_L, across it, given by

$$V_L = L\frac{dI}{dt} \tag{1.1}$$

where dI/dt is the rate of change of current in the circuit. The inductor L can be equal to L_V or L_G, or a combination of the two depending on the current path. A positive dI/dt through the inductor causes a voltage drop across it, resulting in a reduction in the supply voltage across the IC terminals and causing a performance problem due to a negative spike in the IC supply voltage. Similarly, a negative dI/dt through the inductors increases the supply voltage across the IC terminals, resulting in a positive spike, which causes reliability problems. The power supply noise has four components (1) ultra-high-frequency noise in the 10 to 100 GHz range, (2) high-frequency noise in the 100 to 1000 MHz range, (3) mid-frequency noise in the 1 to 10 MHz range, and (4) low-frequency noise in the 1 to 100 KHz range. The inductance on-chip affects both the ultra-high- and high-frequency noise (>1 GHz), while the package has a large effect on the high-frequency and mid-frequency noise

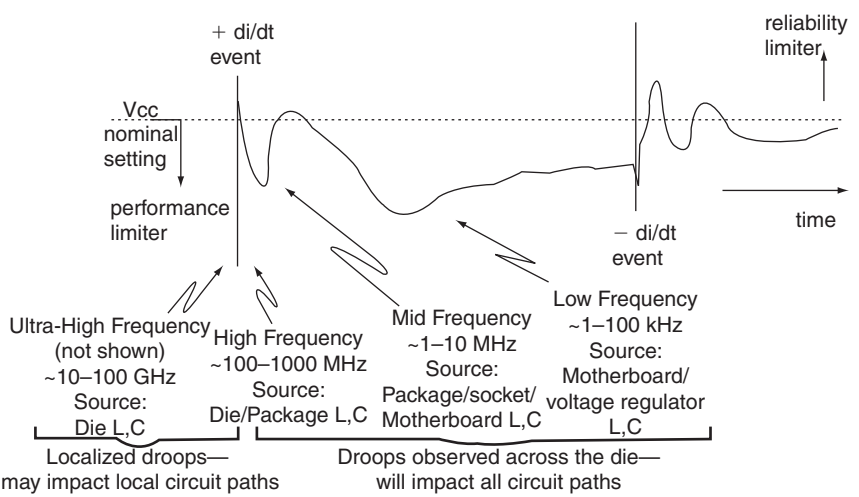


Figure 1-7 Noise signature. By permission from A. Muhtaroglu, G. Taylor, and T. Rahal-Arabi, "On-die droop detector for analog sensing of power supply noise," *IEEE Journal of Solid-State Circuit,* vol. 39, no. 4, pp. 651–660, Apr. 2004, © 2004 IEEE.

components (10 MHz−1 GHz). The inductance of the motherboard and the voltage regulator module affect the mid-frequency and low-frequency noise components (<1 MHz), as shown in Figure 1-7.

The ultra-high-, high-, mid-, and low-frequency noise are also called, respectively, the first, second, third, and fourth droops or spikes on the power supply.

1.2 Simple Relationships for Power Delivery

In any IC, two kinds of circuits need to be powered: the core and I/O. The core consists of transistors that are contained within an IC and that communicate with each other. The I/O, on the other hand, has to communicate with other ICs through the package and motherboard. Because the wires connected to I/O circuits exit the IC, they are very noisy and often are isolated from the core circuits using a separate PDN, as shown in Figure 1-8, where both the core and I/O circuits during switching create voltage fluctuations across the power supply. In this section, simple relationships are derived for the voltage fluctuations on a power supply for both the core and I/O circuits.

1.2.1 Core Circuits

A very simple circuit is shown for the core circuits in Figure 1-9(a), where the driver and receiver circuits are shown as 2 and 1, respectively. The PDN contains some resistance and inductance due to the parasitics of the interconnections in the network. The resistance is assumed to be negligible here. A simple equivalent

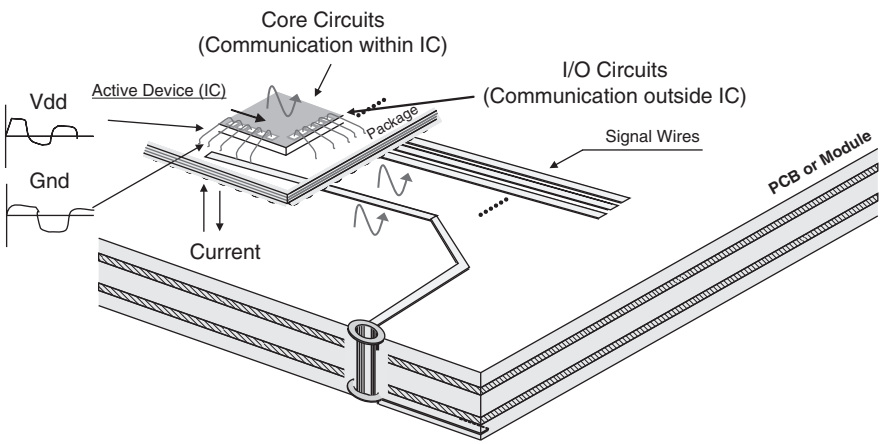


Figure 1-8 Core and I/O circuits. (Courtesy of Professor Joungho Kim KAIST, South Korea.)

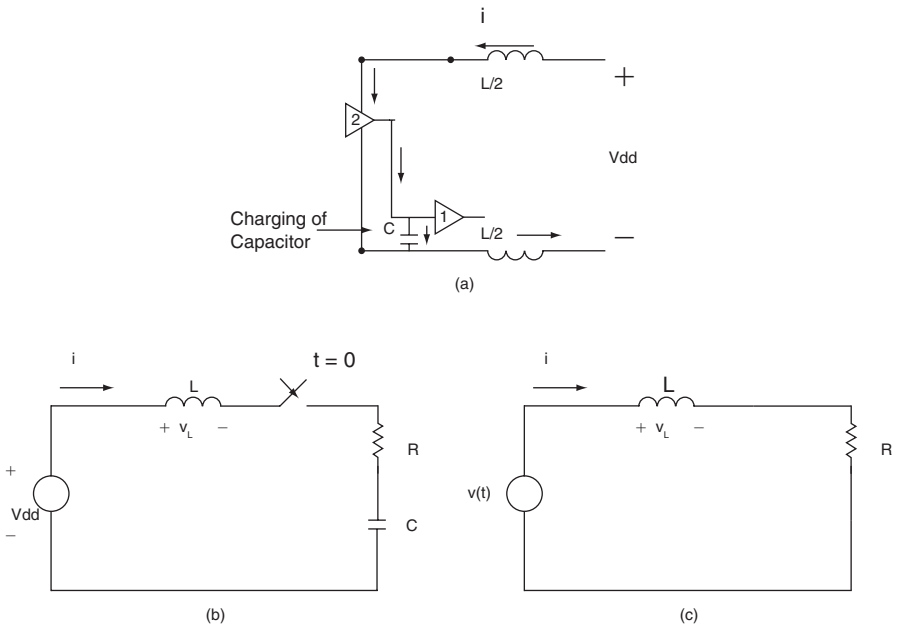


Figure 1-9 (a) Core circuits switching. (b) Equivalent circuit. (c) Simplified equivalent circuit.

circuit for Figure 1-9(a) is shown in Figure 1-9(b). In the simplified equivalent circuit, the switch represents the PMOS transistor that closes at time $t = 0$. The resistance R is the on-resistance of the transistor, and C is the input capacitance of receiver circuit 1 that needs to be charged. The total inductance of the voltage and ground paths is represented by a single inductance L.

The goal of the core PDN is to ensure that sufficient charge is supplied to the switching circuit so that the capacitance can be charged to the required voltage. To minimize delay, the charge has to be supplied within a short time. The circuit in Figure 1-9(b) has two time constants: L/R and RC. The delay of the transistor circuit is defined by the RC delay. Since the L/R time constant should have minimum impact on the RC delay of the transistor, it is desired that [2]

$$\frac{L}{R} \ll RC \tag{1.2}$$

Under this assumption, the simplified equivalent circuit in Figure 1-9(c) can be used, where the voltage drop across the inductor can be obtained by solving equation (1.3),

$$v_L(t) = L\frac{di(t)}{dt} \tag{1.3}$$

where the current is obtained by solving the differential equation:

$$L\frac{di(t)}{dt} + Ri(t) = v(t) \tag{1.4}$$

In equation (1.4), $v(t)$ is an equivalent source voltage with rise time t_r (that combines the switch and Vdd) given by

$$v(t) = \begin{cases} \dfrac{\text{Vdd} \times t}{t_r} & 0 \le t \le t_r \\ \text{Vdd} & t \ge t_r \end{cases} \tag{1.5}$$

The rise time is dictated by the speed of the switch. The maximum voltage across the inductor occurs at time $t = t_r$ and is given by

$$v_{L\max} = \Delta v = \frac{L \times \text{Vdd}}{Rt_r}(1 - e^{-t_r/(L/R)}) \tag{1.6}$$

Example

Assume $t_r = 0.1$ ns, $L = 0.1$ nH, $R = 1\ \Omega$, $C = 1$ nF, and Vdd = 1 V. The ratio $L/R = 0.1\ RC$, and therefore the condition in equation (1.2) is met. The maximum voltage drop across the inductor can be obtained as 632 mV from equation (1.6).

By changing the inductance value to $L = 0.01$ nH, the maximum voltage drop across the inductor can be obtained as 100 mV.

The voltage waveform across the inductor is shown in Figure 1-10. As the inductance is changed from 0.1 nH to 0.01 nH, the voltage drop across the inductor looks more like a rectangular pulse. Therefore, when t_r is much greater than L/R, equation (1.6) can be simplified to

$$\Delta v \approx \frac{L \times \mathrm{Vdd}}{R t_r} \tag{1.7}$$

with a pulse width of t_r.

Example

Consider an IC in which the total capacitance to be charged is 10 nF and the on-resistance of all the transistors in parallel is 0.1 Ω. Let the required maximum voltage drop across the inductor be 10% of Vdd for a rise time t_r = 1 ns. The inductance required in the PDN to meet the voltage drop can be calculated from equation (1.6) by solving:

$$\frac{\Delta v}{\mathrm{Vdd}} = 0.1 = \frac{L}{(0.1*1)}(1 - e^{-1/(L/0.1)}) \tag{1.8}$$

where L is measured in nanohenry. By iterating, the inductance can be calculated as L = 0.01 nH. The inductance satisfies the condition in equation (1.2). It can also be obtained from equation (1.7), since t_r is much greater than L/R.

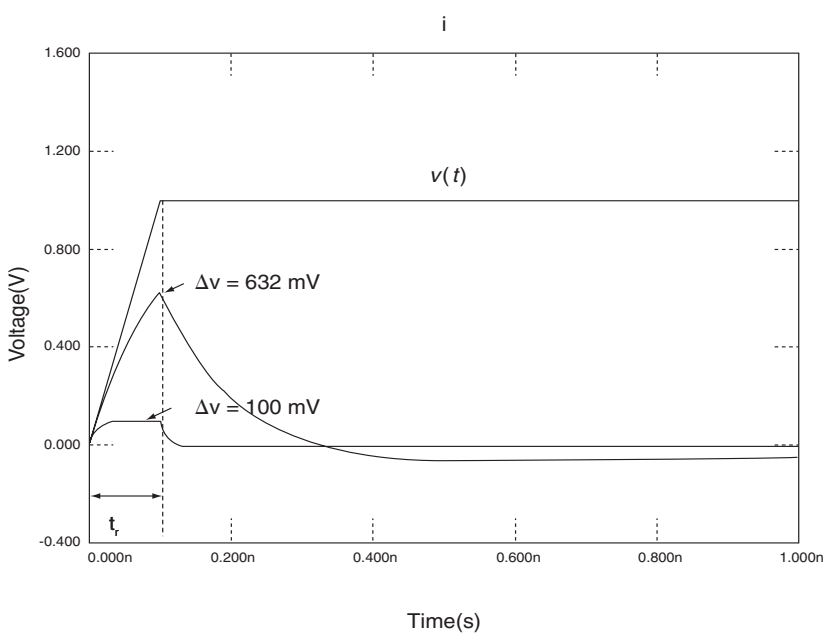


Figure 1-10 Voltage drop across inductor.

1.2.2 I/O Circuits

I/O circuits, unlike core circuits, drive off-chip interconnections. With increase in frequency, the interconnections behave as transmission lines where the delay becomes important. The PDN used to drive an I/O circuit is shown in Figure 1-11(a): the transmission line has a characteristic impedance of Z_0 and delay T. The far end of the transmission line is terminated with a resistor $R = Z_0$. The inductance L represents the PDN loop inductance from the power supply to the chip terminals. As before, the transistor is represented using a switch with an on-resistance R, where R is much less than Z_0 to allow for the maximum voltage to be launched on the transmission line, as shown in Figure 1-11(b).

When the switch closes, the power supply inductance L acts as an open circuit and behaves as a short circuit at time t = infinity. As in the previous section, the voltage source and the switch can be combined and represented as a pulse with rise time t_r. Since the far end of the transmission line is terminated in the characteristic impedance of the transmission line, there are no reflections. The maximum voltage drop across the inductor occurs at time $t = t_r$ and can be calculated as in the previous section by replacing R with Z_0:

$$v_{L\max} = \Delta v = \frac{L \times \mathrm{Vdd}}{Z_0 t_r}(1 - e^{-t_r/(L/Z_0)}) \tag{1.9}$$

Based on equation (1.9), a signal line with low Z_0 (highly capacitive) will always result in a larger voltage drop across the inductor, assuming the inductance is fixed, as described later in this chapter. When t_r is much greater than L/Z_0, the maximum voltage drop across the inductor simplifies to

$$\Delta v \approx \frac{L \times \mathrm{Vdd}}{Z_0 t_r} \tag{1.10}$$

When N parallel transmission lines of characteristic impedance Z_0 are switched simultaneously, it is equivalent to switching a single transmission line of impedance Z_0/N. Hence, the maximum voltage drop across the inductor can be obtained by replacing Z_0 by Z_0/N in equations (1.9) and (1.10).

Example

Assuming L = 1 nH, Z_0 = 50 Ω, Vdd = 1 V, and t_r = 0.1 ns, the maximum voltage drop across the inductor can be calculated as 200 mV from

equation (1.9). Assuming the inductance is reduced to $L = 0.1$ nH, the voltage drop becomes 20 mV. Figure 1-11(c) shows the voltage waveforms across the inductor: the shape of the voltage drop looks more like a rectangular pulse when t_r is much greater than L/Z_0.

Example

Consider a 32-bit bus with $Z_0 = 50\ \Omega$. The driver is switching with a $t_r = 0.1$ ns. Assuming $V = 10\%$ of Vdd is desired as the voltage drop across the inductor, the maximum power supply inductance that must be supported can be obtained by solving the following equation iteratively:

$$\frac{\Delta v}{\mathrm{Vdd}} = 0.1 = \frac{32 \times L}{50 \times 0.1}\left(1 - e^{-0.1/(L/(50/32))}\right) \tag{1.11}$$

The result is an inductance of 16 pH. Since t_r is much greater than $L/(Z_0/N)$, the same result can also be obtained from equation (1.10).

1.2.3 Delay Due to SSN

The presence of the inductor increases the delay of the I/O circuit. The voltage at the input end of the transmission line for a pulse with rise time t_r can be computed as

$$v(t) = \frac{Z_0 \times \mathrm{Vdd}}{L t_r}\left(\frac{L^2}{Z_0^2}\left[e^{\frac{-t}{(L/Z_0)}} - 1\right] + \frac{L}{Z_0}t\right) \qquad t \le t_r \tag{1.12}$$

and

$$v(t) = A + B\left(1 - e^{\frac{-t}{(L/Z_0)}}\right) \qquad t > t_r \tag{1.13}$$

where

$$A = \mathrm{Vdd} - \left[\mathrm{Vdd} - v(t_r)\right]e^{\frac{t_r}{(L/Z_0)}}$$
$$B = \left[\mathrm{Vdd} - v(t_r)\right]e^{\frac{t_r}{(L/Z_0)}} \tag{1.14}$$

and $v(t_r) = v(t = t_r)$ from equation (1.12). A transistor circuit at the receiver requires a minimum voltage at its input to switch states. Let's assume that the minimum voltage required for this to happen at the driver output V_{chip} (input end of the transmission line) is 0.5 × Vdd. Equations (1.12) and (1.13) can be used to calculate the time required to reach 0.5 × Vdd and hence represent the delay incurred because of the power supply inductance. Equation (1.12) can be used when t_r is greater than L/Z_0, and equation (1.13) can be used when t_r is less than L/Z_0 to calculate a 50% delay. This delay does not include the transmission line delay and is valid for a matched load, as in Figure1-11(b).

Example

Consider the previous example in which L = 0.1 nH, Z_0 = 50 Ω, Vdd = 1 V, and t_r = 0.1 ns. Since t_r is greater than L/Z_0, equation (1.12) can be used to calculate the 50% delay. Using an iterative process, the 50% delay at the input of the transmission line is 0.052 ns.

Let's now assume that a 100-bit bus, each of impedance Z_0, is switched simultaneously. This translates to an equivalent impedance Z_0 = 50/100 = 0.5 Ω (transmission lines in parallel). Since t_r is less than L/Z_0, equation (1.13) can be used to calculate the 50% delay. Through iteration, this delay can be calculated as 0.191 ns. Hence, the delay increases because of an increased voltage drop across the inductor caused by an increase in current. The voltage at the input end of the transmission line is shown in Figure 1-11(d) along with the 50% delay.

1.2.4 Timing and Voltage Margin Due to SSN

Timing and voltage margins are affected by crosstalk, process variation, SSN, reflection, and other effects. In this section, we address only the effect of SSN. SSN can affect the voltage margin because power supply noise can corrupt the voltage levels of the signal waveform. In the previous section, a relationship was derived between the SSN and delay: as the SSN increased for a larger number of switching drivers, the 50% delay increased as well. This delay manifests itself as jitter that affects the signal integrity of the waveform and therefore increases the timing error; see Figure 1-11(e). As an example, consider an 8-bit-wide bus. If all the bits transition simultaneously from 0 to 1 (00000000 to 11111111 for the bus), the maximum transient current from the power supply is drawn, resulting in maximum noise and hence maximum delay. If only the alternate bits transition

(00000000 to 10101010), fewer drivers switch and therefore the noise (and delay) is lower than in the previous case. For a pseudorandom bit stream (PRBS), the number of switching drivers changes at random, resulting in random SSN. Therefore, the 50% delay associated with the rising edge changes with the bit pattern, resulting in an uncertainty in the position of the rising edge. This effect is called jitter, shown in Figure 1-11(e). Jitter results in a timing uncertainty whereby a longer time interval may be required to latch the data for all the bit patterns if the jitter is large. Hence, the goal in I/O signaling is to ensure the smallest timing error by controlling jitter, which is possible by reducing SSN in addition to other parameters. This ensures a suitable timing margin. In Chapter 5, this effect is described in more detail through an example.

1.2.5 Relationship between Capacitor and Current

As mentioned earlier, decoupling capacitors serve as charge reservoirs and provide current to the switching circuits. Let's assume that the power supply inductance is small such that equation (1.10) is valid. Consider a single 50-Ω driver, which requires a current of 0.1 A assuming Vdd = 5 V ($\triangle I$ = 5/50). Let's assume that a 100-nF capacitor is available to provide charge to the switching circuits during a time interval of 10 ns (t_r) that keeps the power supply fluctuations to within 10% of Vdd. The current that can be supplied by the capacitor that maintains $\triangle v$ to be 10% of Vdd is given by [6]

$$\Delta I = \frac{C\Delta v}{t_r} = \frac{100\times10^{-9}\times0.5}{10\times10^{-9}} = 5 \text{ A} \tag{1.15}$$

Since a single driver requires 0.1 A to charge the interconnections, the 100-nF capacitor can provide the current to 50 I/O circuits over a period of 10 ns.

1.3 Design of PDNs

Since a computer system supports multiple frequencies, a PDN is best designed in the frequency domain. The response of the PDN to switching circuits can then be viewed in the time domain to evaluate the transient noise voltages generated on the power supply terminals of the IC or between any other nodes in the system. The response of the PDN in the frequency domain enables a designer to understand all the resonances and antiresonances in the system produced by the interaction of inductances and capacitances in the network. An antiresonance, when excited by a

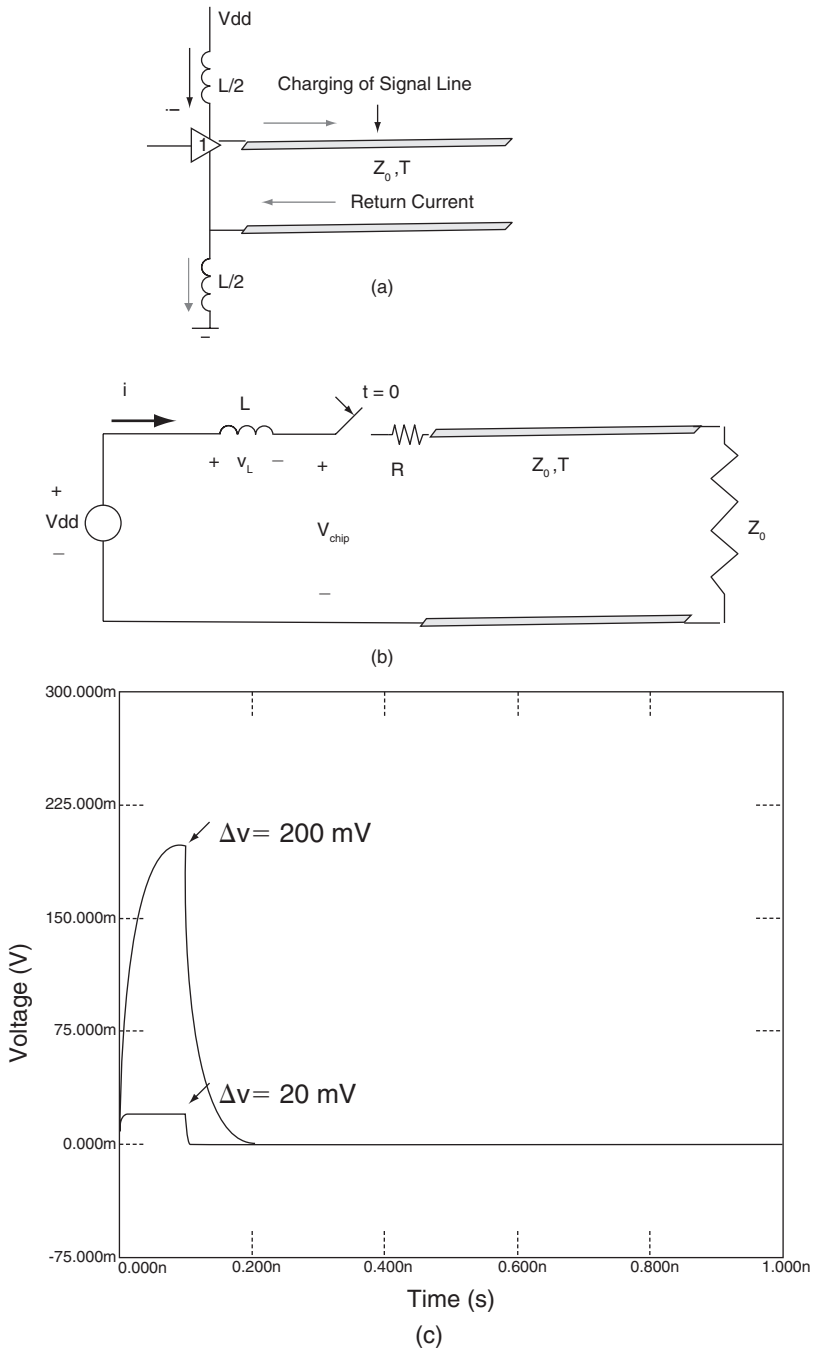
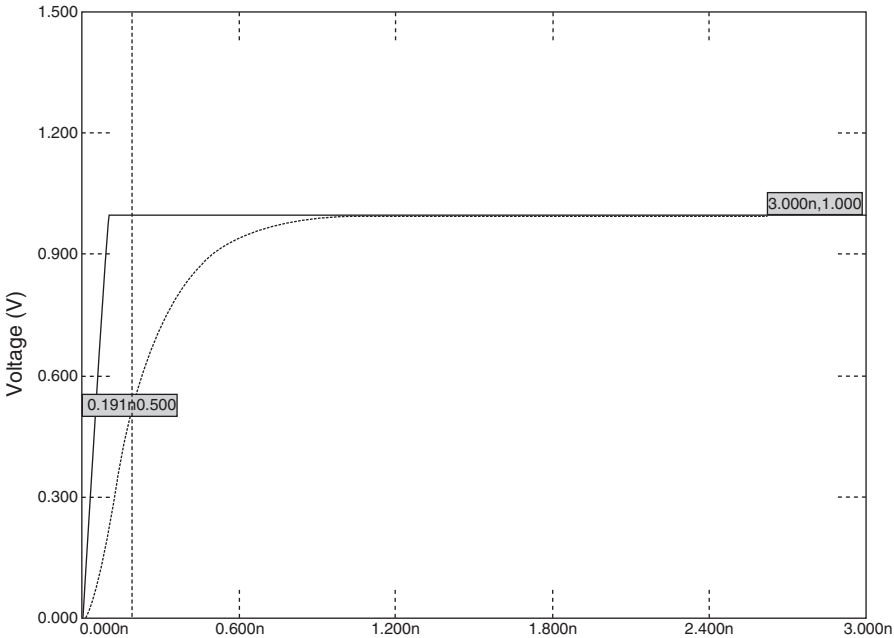
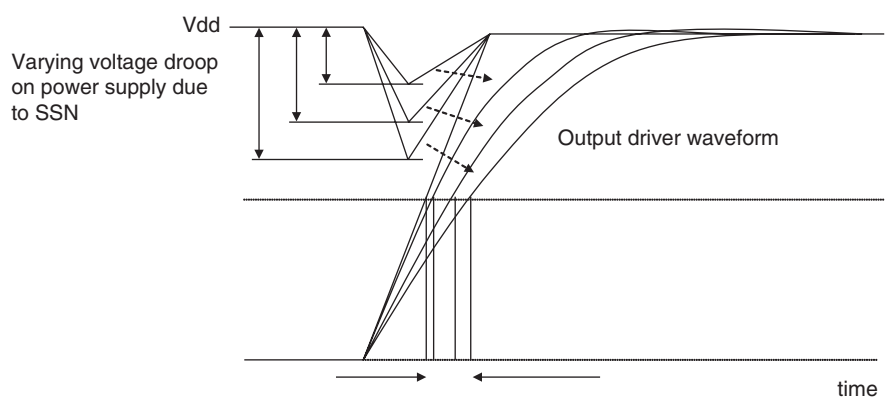


Figure 1-11 (a) I/O circuit switching. (b) Simple equivalent circuit. (c) Voltage drop across inductor.



(d)



(e)

Figure 1-11 (d) Delay due to 1 driver and 100 drivers switching. (e) Jitter caused by simultaneous switching noise. (Courtesy of Sony.)

source, always generates the maximum noise voltage across the power supply terminals of the IC. Based on the frequency response of the PDN, the designer can evaluate the importance of the antiresonances in the system and decide if the source (switching circuits) will ever excite these antiresonances. Hence, the signature of the source along with the frequency response of the PDN decides the noise voltages on the power supply in the time domain. In this section, the concept of target impedance is introduced. The use of target impedance as a design parameter is discussed by evaluating a simple circuit in the frequency and time domain.

1.3.1 Target Impedance

The target impedance is based on Ohm's law, which states that the ratio of voltage to current has to equal the impedance in the network. For a PDN, the voltage is the allowed ripple ($\triangle v$) on the power supply. The target impedance Z_T (in ohms) of a PDN can then be calculated as [7]

$$Z_T = \frac{\text{Vdd} \times \text{ripple}}{50\% \times I_{\max}} \ (\Omega) \tag{1.16}$$

where the average current drawn by the switching circuits is assumed to be 50% of the maximum current and Vdd is the power supply voltage. Assuming a voltage of 5 V with a ripple of 5% and a maximum current of 1 A, the target impedance can be calculated as

$$Z_T = \frac{5 \times 5\%}{50\% \times 1} = 0.5\Omega \tag{1.17}$$

The maximum current drawn by an IC can always be calculated by using the relationship $P = VI_{max}$, since both power P and voltage V for an IC are known. The target impedance Z_T establishes an upper limit for the maximum impedance for the PDN across the power supply terminals of the IC in the frequency domain. An impedance below Z_T ensures that any current transients will always generate noise voltages of less than 5% of 5 V. Hence, Z_T is a very useful parameter for designing PDNs in which the noise voltages have to be controlled within, say, 5% of the supply voltage.

A plot of Z_T versus frequency is shown in Figure 1-12. The frequency axis represents the frequency components associated with the source excitation. According to the figure, if the impedance exceeds the target impedance at any frequency where the current transients can excite the network, then the resulting

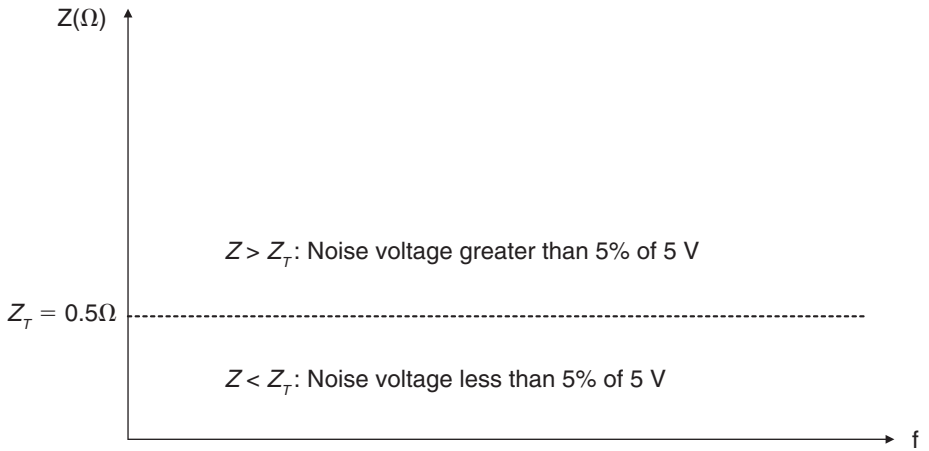


Figure 1-12 *Z* versus frequency.

power supply noise will exceed 5% of 5 V = 250 mV. The figure assumes that the magnitude of the current transients is 50% of the maximum current.

The target impedance calculations for five microprocessors introduced between 1990 and 2002 are shown in Table 1-1. As can be seen, the target impedance has decreased 500-fold over a decade because of the lowering of the supply voltage and increase in power. Since the impedance of the PDN is also given by $Z = \sqrt{L/C}$, where *L* and *C* are the inductances and capacitances in the network, a low target impedance always implies large capacitance and low inductance in the network. In Table 1-1, the frequency of the microprocessor has increased from 16 MHz to 1.2 GHz over a decade, which implies that the target impedance has to be maintained at least up to the fundamental frequency of the clock. However, this

Table 1-1 Target Impedance Trends

Year	Voltage (Volts)	Power dissipated (Watts)	Current (Amps)	Ztarget (mΩ)	Frequency (MHz)
1990	5.0	5	1	250	16
1993	3.3	10	3	54	66
1996	2.5	30	12	10	200
1999	1.8	90	50	1.8	600
2002	1.2	180	150	0.4	1200

Information from Smith [7].

overly restrictive condition may not be satisfied at all frequencies and can often increase the cost of the system. Hence, care should be taken to correlate the frequency response with the current transients in the system to better understand the frequencies at which the PDN will be excited. The target impedance should be maintained at these excitation frequencies.

1.3.2 Impedance and Noise Voltage

Consider the circuit shown in Figure 1-13(a). The circuit has a supply voltage of 2.0 V. The 3-mΩ resistance and 320-pH inductance are the spreading resistance and inductance from the power supply to the capacitor. Spreading resistance and inductance produce resistive and inductive drops when the current travels from the power supply to the capacitors (through the interconnects) for charging them. The capacitor parameters are equivalent series resistance (*ESR*) = 10 mΩ, equivalent series inductance (*ESL*) = 1 nH, and *C* = 100 μF, resulting in a resonant frequency of 0.5 MHz, which are explained in detail later. The on-chip capacitance is 800 nF in the circuit. The current source is 1 A between the voltage and ground terminals of the IC, and through an AC analysis, the voltage (or impedance in ohms) can be obtained as shown in Figure 1-13(b). In Figure 1-13(a), a 1-A current source is used to represent the current, and hence the voltage across it is the impedance in ohms ($Z = V/I$). In the frequency response, the resonant frequency of the decoupling capacitor can be seen, and the large impedance at approximately 13 MHz is caused by the antiresonance between the chip capacitance and ESL of the decoupling capacitor, which is explained later. The null in the impedance profile is called a resonance; the peak in the impedance profile is called the antiresonance. For a 2-V supply, 5% tolerance, and a 10-A average current, the target impedance is 10 mΩ. Therefore, the maximum impedance allowed across the current source (which represents the switching circuit) is 10 mΩ. Clearly, the target impedance is met up to a frequency of 5 MHz in Figure 1-13(b). In the frequency range from 5 MHz to 100 MHz, the target impedance has been exceeded.

Let's now look at the response of this network to two current signatures. The circuit used to compute the time-domain response is shown in Figure 1-14(a). The switching circuit is represented using a time-dependent resistor, the resistance of which changes from 97 mΩ to 197 mΩ, which corresponds to a 10-A change in current in the circuit, assuming only 3 mΩ of resistive impedance (no inductance) is present in the PDN. The current changes from 20 A (2/100 mΩ) to 10 A (2/200 mΩ) in the circuit. The voltage across the time-dependent resistor is shown in Figure 1-14(b) for a current transient with rise time of 10 ns and period of 1 μs. As explained earlier, the transient voltage across the IC power supply contains

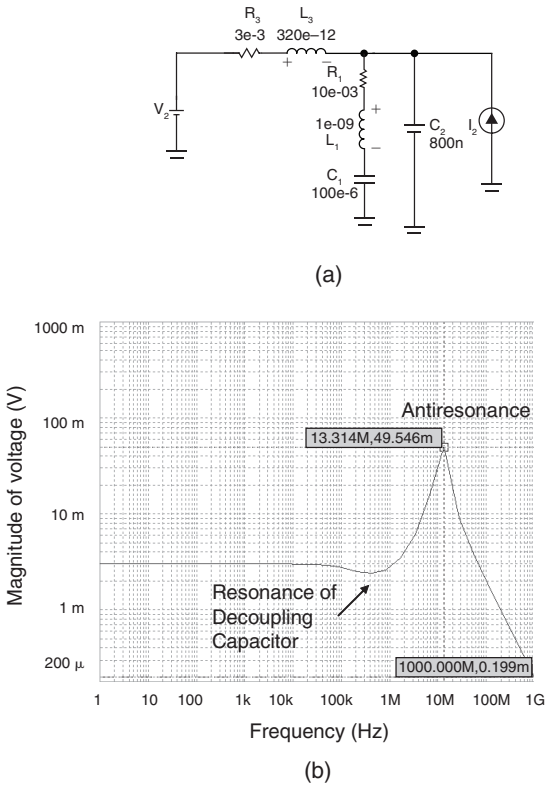


Figure 1-13 (a) Circuit of PDN. (b) Frequency response.

transients with both positive and negative peaks. The noise voltage settles to within 5% of 2 V after 50 ns following the switching activity. Hence, during a large part of the 1-µs period, the noise is below the 5% tolerance value. The 10-ns rise time has enough frequency components that exceed the target impedance initially, causing the first negative glitch to exceed the 100 mV tolerance value. If this negative glitch is a problem, then the impedance at frequencies corresponding to the rise time must be reduced.

Let's now consider the noise voltage when the current transient has a rise time of 10 ns and period of 80 ns, corresponding to a frequency of about 13 MHz, which coincides with the antiresonant frequency. The noise voltage is shown in Figure 1-14(c), which is 200 mV for the entire period of 1 µs and hence exceeds the noise budget of 100 mV. This example shows the importance of managing the impedance of the PDN in the frequency domain to manage excessive noise caused by the current transients.

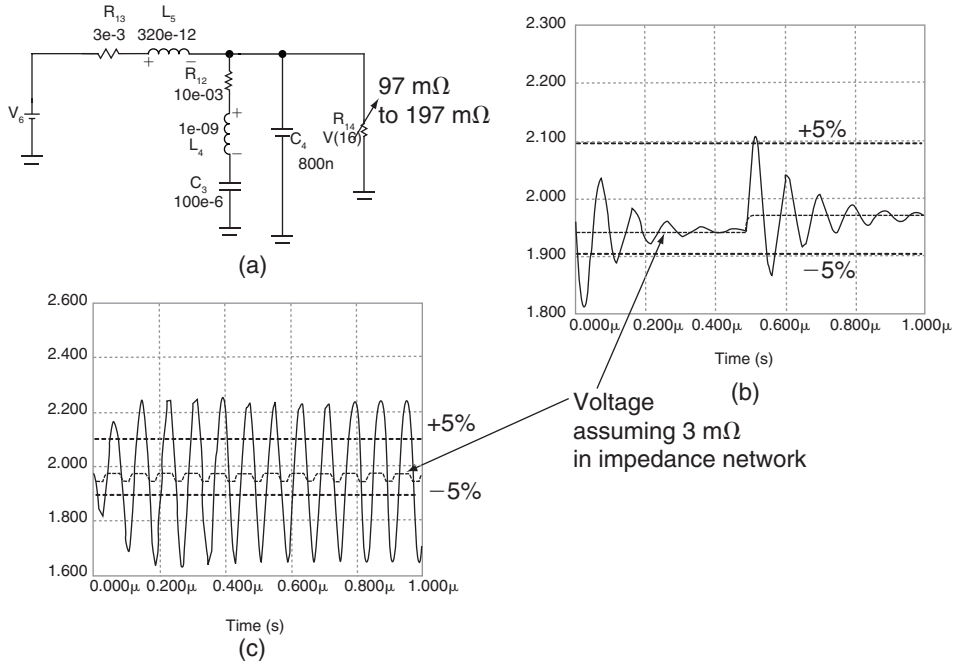


Figure 1-14 (a) Time-domain circuit; (b) 1-µs current period; (c) 80-ns current period.

1.4 Components of a PDN

The elements of the PDN are shown in Figure 1-15 [8]; they include the chip-level power distribution with thin-oxide decoupling capacitors; the package-level power distribution with planes and mid-frequency decoupling capacitors; and the board-level power distribution with planes, low-frequency decoupling capacitors, and VRM. The frequency ranges covered by these elements are also shown in the figure: the power distribution operates at a higher frequency as the proximity to the active devices decreases because of the parasitic inductance and resistance of the interconnections between the active circuitry and the various elements of the PDN. These parasitic effects are explained in the next section.

1.4.1 Voltage Regulator

Computer systems require multiple DC voltages to operate. These voltages have to be well regulated and should be able to supply the required current over a range of frequencies. The trend of increasing power and lowering supply voltage requires designers to move AC–DC and DC–DC converters closer to the electronics

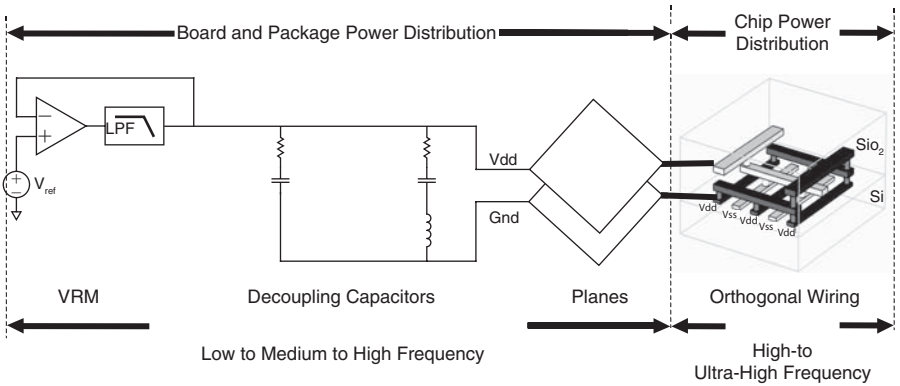


Figure 1-15 Components of a PDN. By permission from M. Swaminathan, J. Kim, I. Novak, and J. P. Libous, "Power distribution networks for system on package: status and challenges," *IEEE Transactions on Advanced Packaging,* vol. 27, no. 2, pp. 286–300, May 2004, © 2004 IEEE.

they feed [8]. A representative class of low-voltage high-current application is the core supply of central processing units (CPUs), digital signal processors (DSPs), and large switching chips. The voltage required may be in the 0.8 V to 2.5 V range, with the current in excess of 100 A for the largest devices. Since the core voltage is often unique and may be required only by the particular device, the DC–DC converters usually feed only one load and hence are also called point-of-load (POL) converters.

Besides the large current requirements, modern electronic circuits contain elements with several different supply voltages. Legacy 5-V and 3.3-V logic devices are still common, but newer devices often require 2.5 V, 1.8 V, 1.5 V, or even lower supply voltage. The pressing need for optimizing device speed while minimizing current consumption leaves little room to combine supply rails with similar but not exactly the same nominal voltage. The solution is therefore to place several DC–DC converters on the board to create the different supply voltages. The topology of these DC–DC converters is determined by two major system constraints: (1) most of the supply voltages are lower than the voltage of the primary source to the board (output of AC–DC converter or battery), so these converters usually have to step down the voltage; and (2) isolation is very seldom required in these converters. In AC-powered systems, the isolation can be easily provided in the AC–DC converters.

Because of these constraints, the single-phase, nonisolated buck converter is the most widely used DC–DC converter topology today, though for high-current

applications, multiphase converters are also becoming popular. In a few applications, step-up boost converters and polarity-reversing buck-boost converters are also used.

1.4.1.1 Operating Principle

The VRM converts one DC voltage to another [9]. It has a reference voltage and a feedback loop. It senses the voltage near the load and adjusts the output current to regulate the voltage at the load. The bandwidth of the regulation loop is usually between one and several hundred kilohertz. At frequencies above the loop bandwidth, the VRM becomes high impedance, and therefore the voltage is no longer well regulated.

1.4.1.2 Four-Element Model

A four-element linear model for a VRM as described in [9] is explained in this section. Figure 1-16(a) is a simplified block diagram for a buck-switching regulator, commonly found in VRMs. At the left of the figure is an input voltage, assumed to be relatively constant. The function of inductor L_1 is to store up energy when switch S_1 is closed, and deliver current to the load. If L_1 has more current than the load is demanding, S_1 opens and S_2 closes. Current continues to flow to the load, but in an ever-diminishing amount until S_2 opens and S_1 closes again. There is an amplifier A with frequency compensation that senses the load voltage with respect to a reference voltage. When the load voltage is too low, it causes the switches and inductor to ramp up the current. When the load voltage is too high, it causes the switches and inductor to ramp down the current. The inductor current is integrated in C_1, which smooths the voltage. C_1 has an ESR. The buck regulator is nonlinear because switches open and close as a function of time. Figure 1-16(b) shows the linearized model of the VRM consisting of an ideal voltage source and four passive elements. In the linear model, R_0 is the value of the resistor between the VRM sense point and the actual load and is usually only a few milliohms. $L_{\_out}$ represents the output inductance of the VRM. It may be the inductance of cables that connect the VRM to a system board or it may be the inductance of pins that connect a VRM to a module (about 200 and 4 nH, respectively). The maximum effective frequency for the VRM is determined by $L_{\_out}$. $R_{\_flat}$ represents the ESR (explained in the next section) of the capacitor associated with the VRM. Generally, the capacitor determines the output impedance of the VRM at frequencies beyond the response time of the loop. The ideal voltage source has the value of the power supply voltage. The value of $L_{\_slew}$ is chosen so that current will be ramped up in the linear model in about the same

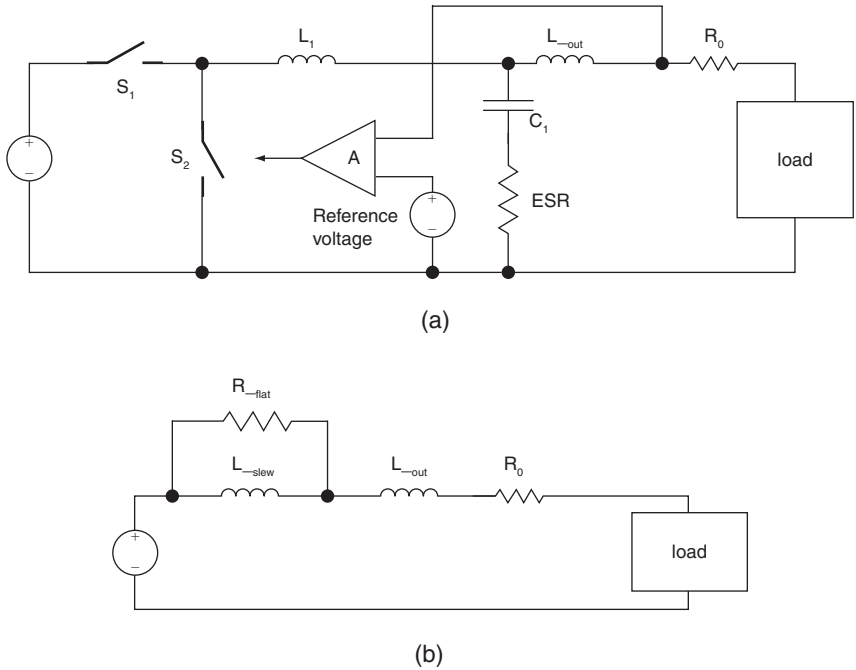


Figure 1-16 (a) Diagram of buck-switching regulator. (b) Four-element model. By permission from L. D. Smith, R. E. Anderson, D. W. Forehand, T. J. Pelc, and T. Roy, "Power distribution system design methodology and capacitor selection for modern CMOS technology," *IEEE Transactions on Advanced Packaging*, vol. 22, no. 3, pp. 284–291, Aug. 1999, © 1999 IEEE.

time that it is ramped up in a real VRM. It is calculated from the equation V = *LdI*/*dt*. In the equation, *V* is the amount of voltage droop or spike that can be accepted (say 5% of 1.8 *V*). The maximum transient current is used for *dI*. The total amount of time for the VRM to ramp this transient current either up or down is used for *dt*. As described in [7], typical model values for a VRM are R_0 = 1 mΩ, $L_{\_out}$ = 4 nH, $R_{\_flat}$ = 30 mΩ, and $L_{\_slew}$ = 67.5 nH.

1.4.1.3 Design Challenges

The challenges for the DC–DC converters are multifold [8]. As a first challenge, the converters have to feed the low-voltage load with reasonable efficiency over a widely varying load-current range, which often requires synchronous rectification to keep losses low. Since the POL converters have to be placed close to the load, which will eventually dissipate the full output power, increasing the efficiency of the POL converter barely reduces the total power dissipation. However, higher converter efficiency can result in a smaller converter volume, which is usually the