# Latchup

**Steven H. Voldman,** ***IEEE Fellow***
*Vermont, USA*

John Wiley & Sons, Ltd

# LATCHUP

# Latchup

**Steven H. Voldman,** ***IEEE Fellow***
*Vermont, USA*

John Wiley & Sons, Ltd

Copyright © 2007 John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England

Telephone (+44) 1243 779777

Email (for orders and customer service enquiries): cs-books@wiley.co.uk
Visit our Home Page on www.wileyeurope.com or www.wiley.com

All Rights Reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except under the terms of the Copyright, Designs and Patents Act 1988 or under the terms of a licence issued by the Copyright Licensing Agency Ltd, 90 Tottenham Court Road, London W1T 4LP, UK, without the permission in writing of the Publisher. Requests to the Publisher should be addressed to the Permissions Department, John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England, or emailed to permreq@wiley.co.uk, or faxed to (+44) 1243 770620.

This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold on the understanding that the Publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

***Other Wiley Editorial Offices***

John Wiley & Sons Inc., 111 River Street, Hoboken, NJ 07030, USA

Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA

Wiley-VCH Verlag GmbH, Boschstr. 12, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 42 McDougall Street, Milton, Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01, Jin Xing Distripark, Singapore 129809

John Wiley & Sons Canada Ltd, 6045 Freemont Blvd, Mississauga, ONT, L5R 4J3

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Anniversary Logo Design: Richard J. Pacifico

***British Library Cataloguing in Publication Data***

A catalogue record for this book is available from the British Library

ISBN 978-0470-01642-8

Typeset in 9/11 pt Times by Thomson Digital, India.
Printed and bound in Great Britain by Antony Rowe Ltd, Chippenham, Wiltshire.
This book is printed on acid-free paper responsibly manufactured from sustainable forestry
in which at least two trees are planted for each one used for paper production.

*To My Wife*
*Annie Brown Voldman*

# Contents


**About the Author** **xvii**

**Preface** **xix**

**Acknowledgements** **xxi**

**Chapter 1 CMOS Latchup** **1**

1.1 CMOS Latchup 1
1.1.1 CMOS Latchup–What is Latchup? 1
1.1.2 CMOS Latchup–Why is Latchup Still an Issue? 5
1.1.3 Early CMOS Latchup History 6
1.2 Fundamental Concepts of Latchup Design Practice 10
1.3 Building a CMOS Latchup Strategy 14
1.3.1 Building a CMOS Business Strategy – 18 Steps in Building a CMOS Latchup Business Strategy 15
1.3.2 Building a CMOS Latchup Technology Strategy – 18 Steps in Building a CMOS Latchup Technology Strategy 15
1.4 CMOS Latchup Technology Migration Strategy 16
1.5 Key Metrics of Latchup Design Practice 18
1.6 CMOS Latchup Technology Trends and Scaling 19
1.7 Key Developments 21
1.7.1 Key Innovations 21
1.7.2 Key Contributions 21
1.7.3 Key Patents 25
1.8 Latchup Failure Mechanisms 25
1.9 CMOS Latchup Events 28
1.9.1 Power-Up Sequence Initiated Latchup 29
1.9.2 Input Pin Overshoot and Power-Up Sequence Initiated Latchup 29
1.9.3 Input Pin Undershoot and Power-Up Sequence Initiated Latchup 30
1.9.4 Multiple Power Supply Power-Up Sequence Initiated Latchup 30
1.9.5 Power Supply Overshoot Initiated Latchup 31
1.9.6 Power Supply Undershoot Initiated Latchup 32
1.9.7 Power Supply (Ground Rail) Undershoot Initiated Latchup 32
1.10 Electrostatic Discharge Sources 33
1.10.1 Human Body Model ESD Event 33

1.10.2 Machine Model ESD Event 34
1.10.3 Cable Discharge Event Source 34
1.11 Single Event Latchup 36
1.11.1 High-Energy Photon Emissions 36
1.11.2 Alpha Particle Ionizing Source 37
1.11.3 Cosmic Ray Source 39
1.11.4 Heavy Ion Source 42
1.12 Summary and Closing Comments 44
Problems 44
References 46

**Chapter 2 Bipolar Transistors** **55**
2.1 The Bipolar Transistor and CMOS Latchup 55
2.1.1 Fundamental Equations of Semiconductors and the Drift–Diffusion Current Constitutive Relationships 55
2.1.2 Diode Forward Bias Conditions 56
2.1.3 Diode Forward Bias Conditions and High-level Injection 57
2.2 Bipolar Transistor 60
2.2.1 Bipolar Current Gain 60
2.2.2 Bipolar Collector-to-Emitter Transport Factor 60
2.2.3 Bipolar Current Characteristics 62
2.2.4 Bipolar Model Gummel Plot 64
2.2.5 Bipolar Current Model–Ebers–Moll Model 64
2.2.6 Bipolar Transistor Base Defect 67
2.2.7 Bipolar Transistor Emitter Defect 69
2.2.8 Bipolar Base Current – Base Defect and Emitter Defect Relation to Bipolar Current Gain 71
2.3 Recombination Mechanisms 72
2.3.1 Shockley–Read–Hall (SRH) Generation–Recombination Model 73
2.3.2 Auger Recombination Model 74
2.3.3 Surface Recombination Mechanisms 76
2.3.4 Surface Recombination Velocity 76
2.3.5 Recombination Mechanisms and Neutron Irradiation 76
2.3.6 Recombination Mechanisms and Gold Recombination Centers 78
2.4 Photon Currents in Metallurgical Junctions 79
2.5 Avalanche Breakdown 81
2.5.1 Bipolar Transistor Breakdown 81
2.5.2 MOSFET Avalanche Breakdown 81
2.6 Vertical Bipolar Transistor Model 83
2.7 Lateral Bipolar Transistor Models 85
2.7.1 Lindmayer–Schneider Model 86
2.7.2 Bipolar Current Gain with Lateral and Vertical Contributions 87
2.7.3 Lateral Bipolar Transistor Models – Nonfield-Assisted 89
2.7.4 Lateral Bipolar Transistor Models – Nonfield-Assisted Base Transport Factor 90
2.8 Lateral Bipolar Transistor Models with Electric Field Assist 91
2.8.1 Lateral Bipolar Transistor Models – Base Transport Factor with Electric Field Assisted Base Transport 93

2.8.2 Lateral Bipolar Transistor Models – Bipolar Current Gain with Electric Field Assisted Base Transport 94
2.8.3 Lateral Bipolar Transistor Models – Two-Dimensional Hole Current Analysis with Electric Field Assist 94
2.8.4 Lateral Bipolar Transistor Models – Vertical Hole Current with Lateral Electric Field Assist 96
2.9 Lateral Bipolar Transistor Models−Nonuniform Vertical Profile 97
2.9.1 Lateral Bipolar Transistor Models−Gunn, Dutton, and Whittier n−/n+ Step Junction Model 98
2.9.2 Lateral Bipolar Transistor Models – Bipolar Current Gain with Nonuniform Vertical Profile 100
2.9.3 Lateral Bipolar Transistor Models – Base Effective Diffusion Length 101
2.10 Triple-Well Bipolar Transistor Models – Lateral and Vertical Contributions 106
2.11 Merged Triple-Well Bipolar Models 108
2.11.1 Merged Triple-Well Models – Lateral and Vertical Contributions with Buried Layer 108
2.11.2 Merged Triple-Well Models – Lateral and Vertical Contributions 110
2.11.3 The Lateral Bipolar Transistor – Substrate Spreading Resistance Representation (Estreich) 111
2.11.4 The Lateral Bipolar Transistor – Substrate Transmission Line Representation (Troutman and Hargrove) 112
2.11.5 The Lateral Bipolar Transistor – Substrate Lossy Transmission Line Representation (Troutman and Hargrove) 115
2.11.6 The Bipolar Transistor – Substrate Transfer Resistance Representation 117
2.12 Summary and Closing Comments 120
Problems 121
References 122

**Chapter 3 Latchup Theory 125**
3.1 Regenerative Feedback 125
3.1.1 Regenerative Feedback without Shunt Resistors and Alpha Representation 126
3.1.2 Regenerative Feedback without Shunt Resistors and Beta Representation 128
3.1.3 Regenerative Feedback with Shunt Resistors and Alpha Representation 128
3.1.4 Regenerative Feedback with Shunt Resistors and Beta Representation 129
3.1.5 Sensitivity Factors – Beta Product 131
3.2 Latchup Criterion with Emitter Resistance 132
3.3 Holding Point Conditions 135
3.3.1 Holding Voltage 135
3.3.2 Holding Current 135
3.3.3 Holding Current Contours 136
3.3.4 Sensitivity Factors – Holding Current 137
3.4 Resistance Space 138
3.4.1 Resistance Space – Constant Holding Condition Contours 139
3.4.2 Resistance Space Plots 141
3.5 Beta Space 142
3.6 CMOS Latchup Differential Tetrode Condition 142
3.6.1 CMOS Latchup Differential Tetrode Condition – First Representation 144

3.6.2 CMOS Latchup Differential Tetrode Condition – Second Representation 145
3.6.3 CMOS Latchup Differential Tetrode Condition – Third Representation 147
3.7 CMOS Latchup Differential Holding Current Relationship 151
3.8 CMOS Latchup Differential Holding Voltage Relationship 152
3.9 CMOS Latchup Differential Resistance Relationship 153
3.10 Differential Generalized Alpha Space Relationship 154
3.11 High-Level Injection 156
3.11.1 High-Level Injection – Base Width Modulation 156
3.11.2 High-Level Injection – Knee and Base Width Modulation Correction 157
3.11.3 High-Level Injection – CMOS Latchup Differential Tetrode Condition 157
3.12 Transient Latchup 160
3.12.1 Transient Latchup – The Four States 160
3.12.2 Transient Latchup – Generalized Excitation Differential Equation and Analysis of the Four States 161
3.12.3 Transient Latchup – Power-Up Excitation 164
3.13 External Latchup 168
3.13.1 External Latchup Diode Injection Source Analysis 168
3.13.2 External Latchup Criterion 170
3.13.3 External Latchup Propagation 173
3.14 Alpha Particle Induced Latchup 178
3.14.1 Charge Collection Process and Analysis 179
3.14.2 Circuit Analysis and Monte Carlo Simulation 181
3.14.3 Maximum Collection Evaluation in a Parallelpiped Region 182
3.15 Summary and Closing Comments 183
Problems 183
References 184

**Chapter 4 Latchup Structures, Characterization and Test 189**
4.1 Guard Rings 189
4.1.1 Intradevice Integrated Guard Ring 191
4.1.2 Intracircuit Guard Ring 191
4.1.3 Intercircuit Guard Ring 192
4.1.4 Interchip Sector Guard Ring 192
4.1.5 ESD Guard Rings – Usage of Guard Rings in ESD Protection 192
4.2 Latchup Characterization Structures – Single- and Dual-Well CMOS PNPN Test Structures 192
4.2.1 Basic pnpn Structure with Single Substrate Contact Spacing 193
4.2.2 Basic pnpn Structure with Internal Well Guard Ring Structure 194
4.2.3 Basic pnpn Structure with Internal Substrate Guard Rings 196
4.2.4 Basic pnpn Structure with Internal Well Guard Ring and Substrate Guard Ring 197
4.2.5 Basic pnpn Structure with Multiple p+/n+ Spacings 199
4.2.6 Basic pnpn Structure with Multiple Well and Substrate Contact Spacings 199
4.2.7 Basic pnpn with Large Substrate Resistance Measurements 200
4.2.8 Basic pnpn with External Injection Source 201
4.3 Latchup Characterization – Basic Triple-Well pnpn Latchup Test Structures 205
4.4 Latchup Characterization Techniques – pnpn Structures with Deep Trench 206

4.5 Latchup Characterization and Testing – Nonautomated Test Systems and Methodology 209
4.5.1 Latchup Testing – dc Test Methods 210
4.5.2 Latchup Testing – Pulsed ac Latchup Test Systems and Methodology 211
4.6 Latchup Characterization and Testing – Automatic Test Systems 213
4.6.1 Noncommercial Wafer-level Automatic Test Systems 213
4.6.2 Commercial Automated Product Test System 214
4.7 Latchup Characterization – Wafer-Level Test Procedures 217
4.7.1 Latchup Characterization Test – Ramped Power Supply Voltage Study 217
4.7.2 Latchup Characterization Test – p+ Anode Power Supply Overshoot Study 218
4.7.3 Latchup Characterization Test – Ramped Negative Power Supply Voltage Study 219
4.7.4 Latchup Characterization Test – n+ Cathode Ground Power Supply Undershoot Study 220
4.8 Latchup Characterization Techniques – Wafer-level Transmission Line Pulse Methodology 221
4.9 Latchup Characterization – Transient Latchup 223
4.9.1 Latchup Characterization – Wafer-level Transient Latchup 224
4.9.2 Latchup Characterization – Transient Latchup (TLU) Amplifier Product-level Test 226
4.10 Guard Ring Characterization 230
4.10.1 Guard Ring Efficiency 231
4.10.2 Guard Ring Theory – A Generalized Bipolar Transistor Perspective 232
4.10.3 Guard Ring Theory – A Probability of Escape Perspective 233
4.10.4 Guard Ring Characterization – Electrical Measurements 234
4.11 Latchup Failure Analysis Techniques 236
4.11.1 Failure Analysis – Optical Microscope 237
4.11.2 Failure Analysis – Emission Microscope (EMMI) Tool 238
4.11.3 Failure Analysis Technique – EMMI Tool Stellari Animation Methodology 240
4.11.4 Failure Analysis Technique – Scanning Superconducting Quantum Interference Device (SQUID) 242
4.11.5 Failure Analysis Techniques – Modified Picosecond Imaging Circuit Analysis Tool – the TLP-PICA Methodology 243
4.11.6 Failure Analysis – Thermal Imaging Techniques 250
4.12 Summary and Closing Comments 250
Problems 251
References 253

**Chapter 5 CMOS Latchup Process Features and Solutions – Dual-Well and Triple-Well CMOS 257**
5.1 CMOS Semiconductor Process Solutions and CMOS Latchup 257
5.2 Substrates 257
5.2.1 Epitaxial Thickness 259

5.3 n-Wells 261
5.3.1 Diffused n-Well Design 261
5.3.2 Retrograde n-Well Design 263
5.3.3 Retrograde n-Well Design and Design Point Optimization 266
5.3.4 Retrograde n-Well Design for Latchup and ESD Design Synthesis 267
5.3.5 n-Wells: n-Well to Substrate Modulation 267
5.3.6 n-Well Depth Scaling 269
5.4 p-Well 269
5.4.1 p-Well Design and Design Point Synthesis 269
5.4.2 p-Well and p++ Substrates 269
5.4.3 p-Well Connecting Implants (Epitaxial Buried Implants) 270
5.5 p+/n+ Scaling 273
5.6 Isolation and Latchup 273
5.6.1 Recess Oxide (ROX) and LOCOS Isolation 274
5.6.2 Shallow Trench Isolation (STI) 275
5.7 Silicide 277
5.8 Triple Well 277
5.8.1 Triple Well–Separate Wells 278
5.8.2 Decoupling of pnp and npn by Spatial Separation 278
5.8.3 Merged Triple-Well CMOS 279
5.8.4 Merged Triple Well with Blanket Implant 288
5.9 High-Dose Buried Layer 289
5.9.1 Buried Implanted Layer for Lateral Isolation (BILLI) Structure 289
5.9.2 Continuous High-Dose Buried Layer 290
5.9.3 HDBL Characterization 292
5.9.4 HDBL Recombination Time and Resistance Measurements 292
5.9.5 HDBL and Reachthrough Implant 293
5.10 Future Concepts 295
5.11 Summary and Closing Comments 295
Problems 296
References 296

**Chapter 6 CMOS Latchup Process Features and Solutions – Bipolar and BiCMOS Technology 301**
6.1 CMOS Latchup in Bipolar and RF BiCMOS Technology 301
6.2 Substrates – High-Resistance Substrates 301
6.2.1 Fifty Ohm Centimeter Substrate Resistance 306
6.2.2 Ultrahigh Substrate Resistance 310
6.3 Subcollectors 312
6.3.1 Subcollector – npn and pnp Bipolar Current Gain 313
6.3.2 Subcollector – $\beta$ Product $\beta_{pnp}\beta_{npn}$ 315
6.3.3 Subcollector – Overshoot and Undershoot Currents 315
6.4 Alternative Isolation Concepts 317
6.5 Trench Isolation (TI) 317
6.5.1 Trench Isolation and Latchup 318
6.5.2 Trench Isolation and Subcollector 323
6.6 Deep Trench 328
6.6.1 Deep Trench as a Guard Ring 329

6.6.2 Deep Trench Guard Ring Characterization – Guard Ring Efficiency Methodology 329
6.6.3 Deep Trench Guard Ring Characterization – Bipolar Transistor Methodology 333
6.6.4 Deep Trench Structure Within the pnpn Structure 334
6.6.5 Deep Trench Structure and Subcollector 340
6.6.6 Electrically Connected Deep Trench Structure 343
6.7 Triple-Well and BiCMOS Processes Integration 346
6.7.1 Triple Well – Deep Trench and Triple Well Integration 347
6.7.2 Triple Well – Deep Trench and Subcollector 348
6.8 Heavily Doped Buried Layer Implant and BiCMOS Technology 351
6.9 Summary and Closing Comments 352
Problems 352
References 353

**Chapter 7 CMOS Latchup – Circuits 357**
7.1 Table of Circuit Interactions 357
7.2 Intrabook Latchup Mechanisms 359
7.2.1 CMOS Latchup Within an ESD Input Network 359
7.2.2 Intrabook Latchup Mechanisms – ESD Input Circuit and I/O Circuit Latchup 361
7.3 Interbook Latchup Mechanisms 362
7.3.1 Interbook Latchup Mechanisms – ESD Input Circuit and Power Clamp Latchup 362
7.3.2 Interbook Latchup Mechanisms – ESD Input Circuit and Power Clamp Latchup (Hsu *et al.*) 362
7.3.3 Interbook Latchup Mechanisms – I/O to ESD Power Clamps (Huh *et al.*) 365
7.3.4 Interbook Latchup Mechanisms – Power Supply to Power Supply (Huh) 366
7.3.5 Interbook Latchup Mechanisms – I/O to I/O Networks (Salcedo-Suner *et al.*) 367
7.3.6 System-Level Latchup 368
7.4 Circuit Solutions – Input Circuit 369
7.4.1 Latchup Prevention Circuit – Dual-Well ESD Networks 369
7.4.2 Latchup Prevention Circuit – Triple-Well ESD Networks 371
7.4.3 Sequence-Independent Input Networks 373
7.4.4 Sequence-Independent Input Networks: Off-Chip Drivers 373
7.4.5 Sequence-Independent Input Networks: ESD Networks 374
7.5 Power Supply Concepts 375
7.5.1 Power Supply Current Limit – Series Resistor 376
7.5.2 Power Supply Current Limit – Current Source 376
7.5.3 Power Supply Solutions – Voltage Regulator 378
7.5.4 Latchup Circuit Solutions – Power Supply Decoupling 378
7.6 Latchup Circuit Solutions – Power Supply to Power Supply Sequencing Circuitry 381
7.6.1 Power Supply to Power Supply Sequence-Independent Networks 382
7.6.2 Power Supply to Power Supply Control Circuits (Lin *et al.*) 383

7.7 Overshoot and Undershoot Clamp Networks 385
7.7.1 Passive Clamp Networks 385
7.7.2 Active Clamp Networks 387
7.7.3 Dynamic Threshold Triple-Well Passive and Active Clamp Networks 390
7.8 Passive and Active Guard Rings 391
7.8.1 Passive Guard Ring Circuits and Structures 392
7.8.2 Active Guard Ring Circuits and Structures 392
7.9 Triple-Well Noise and Latchup Suppression Structures 400
7.10 System-Level Issues 400
7.11 Summary and Closing Comments 401
Problems 401
References 402

**Chapter 8 Latchup Computer Aided Design (CAD) Methods 407**
8.1 Latchup CAD Rules 407
8.1.1 Fundamental Latchup Design Rules 408
8.1.2 Local Latchup Rules 408
8.1.3 Voltage Condition Rules 409
8.1.4 Off-Chip Driver Rules 409
8.1.5 Gate Array Design Rules 410
8.1.6 Mixed Voltage Applications and Special Circuits Rules 410
8.1.7 Global Chip Level Rules 411
8.1.8 External Latchup Design Rules 412
8.2 Design Rule Checking 413
8.2.1 Identification of Guard Rings 413
8.3 Computer-Aided Design Extraction Methodologies – Searching for the pnpn 413
8.3.1 Extraction of the Parasitic pnpn Using Model Graphs (Li) 415
8.3.2 CAD Extraction Methodology (Zhan–Feng–Wu–Chen–Guan–Wang Method) 418
8.3.3 CAD Extraction Methodology – Extraction of the Parasitic pnpn (Habitz–Galland–Washburn Method) 419
8.4 CAD Extraction Methods – Searching for the Guard Rings 420
8.4.1 Virtual Design Level Methodology (Voldman–Sullivan–Nickel–Bass) 421
8.4.2 Built-in Guard Rings 421
8.4.3 Guard Ring Parameterized Cells (Pcell) (Voldman-Perez) 422
8.4.4 CAD Automation of Guard Ring Resistance 423
8.4.5 CAD Automation of Guard Ring Modification Methodology (Ker–Jiang–Peng–Shieh) 425
8.5 Latchup Extraction Methods and Tools 426
8.5.1 Latchup Extraction Tool (Ramaswamy–Sinha–Kadamati–Gharpurey Method) 426
8.5.2 CAD Verification – Design Error Detector Methodology (Venugopal–Sinha–Ramaswamy–Duvvury–Prasad–Raghu–Kadamati) 427
8.5.3 CAD Verification Methodology (Kimura–Tsujikawa) 428

8.5.4 CAD Design Methodology – Modification of Well and Substrate Contact Placement Based on Injection Source 429
8.5.5 Method of Well and Contact Spacing from External Source (Chatty-Brennan) 431
8.5.6 Method of Well and Contact Spacing for Array I/O 431
8.5.7 Transmission Probability Methodology (Voldman-Watson) 433
8.5.8 Global Placement Methodology (Voldman) 435
8.5.9 Method of Latchup, Placement and Routing Cosynthesis (Cohn-Basaran) 437
8.6 Latchup CAD Simulation 439
8.6.1 Latchup CAD Semiconductor Device Simulation 439
8.6.2 Latchup CAD Circuit Models 440
8.7 Summary and Closing Comments 440
Problems 441
References 441

**Index** **445**

# About the Author

Steven H. Voldman is an IEEE Fellow for 'Contributions in ESD Protection in CMOS, Silicon on Insulator and Silicon Germanium Technology'. He has a B.S. in engineering science from University of Buffalo (1979), a first M.S. EE (1981) from Massachusetts Institute of Technology (MIT), a second EE degree (engineering degree) from MIT, a M.S. in engineering physics (1986) and a Ph.D. EE (1991) from University of Vermont under IBM's Resident Study Fellow program.

Since 1984, Voldman has provided experimental research, invention, chip design integration, circuit design, customer support and strategic planning for ESD and latchup. His latchup and ESD work consist of pioneering work on advanced CMOS and BiCMOS semiconductor processing, and presently he is working on RF CMOS, RF BiCMOS silicon germanium (SiGe) technology, image processing and high-voltage smart power technology.

Dr Voldman has written over 150 technical papers between 1982 and 2007. He is a recipient of over 160 issued US patents and 80 US patents are pending, in the area of ESD and CMOS latchup. Dr Voldman is an author of the John Wiley & Sons ESD book series – the first book, *ESD: Physics and Devices*; the second book, *ESD: Circuits and Devices*; and the third book, *ESD: RF Technology and Circuits* – as well as a contributor to the book, *Silicon Germanium: Technology, Modeling and Design*.

Dr Voldman was chairman of the SEMATECH ESD Working Group from 1995 to 2000, to establish a national strategy for ESD in the United States; this group initiated ESD technology benchmarking strategy, test structures and commercial test system strategy. Dr Voldman was also part of the SEMATECH vertical modulated well PTAB in 1992 that focused on MeV implantation for latchup. He is a member of the ESD Association Board of Directors, ESDA Education Committee, as well ESD Standards Chairman for Transmission Line Pulse (TLP) and Very Fast TLP (VF-TLP) testing committee. He has served on various symposia internationally from technical program committee to tutorials on ESD and latchup – EOS/ESD Symposium, International Reliability Physics (IRPS), Taiwan ESD Symposium (T-ESDC), International Conference on Electromagnetic Compatibility (ICEMAC), International Physical and Failure Analysis (IPFA) Symposium and Bipolar/BiCMOS Circuit Technology Meeting (BCTM).

Steve Voldman initiated the 'ESD on Campus' program to bring ESD lectures and interaction to university faculty and students internationally and has provided lectures in the United States, Europe, Taiwan, Singapore, Malaysia, Philippines, China and Thailand. Dr. Voldman received the ESD Association Outstanding Contribution Award in 2007.

# Preface

The book is targeted for the semiconductor process and device engineer, the circuit designer, ESD engineers and latchup specialists. In this book, a balance is established between the theory, semiconductor technology, circuits, testing and computer-aided design (CAD).

The first goal of this book is to teach the fundamentals of the latchup process and design discipline, to teach the latchup methodology, latchup design practices and CMOS latchup discipline.

The second goal is to teach a general methodology of latchup design. This involves understanding of the interrelationship between layout design and device physics, characterization, testing, and process and circuit solutions. With the growing interest in semiconductor management and engineering management, it is also important to understand the business side of CMOS latchup and how this integrates into the business, technology methodology and strategy.

The third goal is to present the material that is relevant to many of today's semiconductor technologies. This book will provide a wide spectrum of technologies including CMOS, RF CMOS, BiCMOS technology, BiCMOS silicon germanium (SiGe) technology, BiCMOS silicon germanium carbon (SiGeC) technology, complementary Bipolar technology, and smart power and image processing technology; it is valuable to see how the technologies influence the latchup results and solutions.

The fourth goal is to expose the reader to the literature in the field of CMOS latchup. An objective of this book is to bridge the gap between the prior work of the 1970s and 1980s and the advanced technologies that are being used in the 1990s, 2000 and beyond. In the early materials, the focus was on semiconductor device physics, device simulation and analytical models. In recent years, the focus has been on new semiconductor process features, CMOS scaling, commercial testing methods and standard development, CAD methodology and tools, and circuit solutions.

The fifth goal is to demonstrate new failure analysis techniques and tools. In the last 20 years, a significant number of failure analysis techniques and tools were developed that have been utilized to understand and quantify CMOS latchup; these tools include emission microscopes, CCD camera techniques, the pico-current analysis (PICA) tool, quantum tool (e.g. SQUID) and other techniques. With these advancements in the field of failure analysis, new issues in latchup can be visualized both spatially and temporally.

The sixth goal is to expose the reader to the growing number of latchup CAD methodologies. With CMOS latchup, a number of new methods for extraction, checking and verification are being used.

The seventh goal is to expose the reader to the patent art in the CMOS field. A significant amount of activity in the CMOS field is associated with process solutions, structures, circuit solutions and CAD methodologies. A majority of the CAD checking and verification methodologies can be found by reading the patent art.

This book also contains the following:

Chapter 1 introduces the reader to the fundamentals and concepts of CMOS latchup design. In this chapter, the discussion addresses the uniqueness of the CMOS latchup design discipline and

methodology. In this chapter, we discuss concepts of the parasitics, parasitic coupling, guard rings, resistive shunting, recombination methods, power supply decoupling methods and others. The chapter reviews key contributions, innovations and patents; this will serve as a reference and a resource. The chapter covers event sources that initiate latchup from noise, to ionizing particles.

Chapter 2 reviews the basics of bipolar semiconductor device physics, with a unique twist. The chapter will discuss the language of the bipolar transistor and physics relevant to the understanding of lateral and vertical bipolar transistors. Although a standard semiconductor text will focus on vertical transistors, this book will focus on lateral transistors, and primarily lateral parasitic bipolar transistors. In Chapter 2, the early models will be extended to address nonuniform vertical profiles, lateral electric fields, active guard rings and triple-well technology.

Chapter 3 focuses on CMOS latchup physics. This chapter will build on the basic concepts discussed in Chapter 2, extending the understanding of the transistor and addressing the cross-coupled pnp and npn. Latchup stability criteria will be discussed. These will be discussed in relevance to the structural features in the CMOS technology. In this chapter, both dc transient latchup will be discussed, connecting early work to present-day analysis.

Characterization, test structures, failure analysis tools and test equipment are looked at in Chapter 4. The chapter addresses latchup test structures followed by failure analysis tools such as photoemission tools, PICA tools and SQUID. For product qualification, CMOS latchup standards will be discussed. Examples of commercial test equipment used today are also included.

Chapter 5 focuses on the influence of CMOS semiconductor process features on CMOS latchup sensitivity. This chapter covers substrates, n-wells, p-wells, isolation and junctions in dual-well and triple-well CMOS technology. In addition, Chapter 5 will discuss high-dose buried layer (HDBL) concepts.

Chapter 6 covers the influence of semiconductor process features on CMOS derivative technologies, such as bipolar to BiCMOS. This chapter will build on the basic learning discussed in Chapter 5 and addresses the addition of high-resistance substrates, subcollectors, multiple subcollectors, trench isolation, deep trench (DT) isolation, biased trench, subcollector isolated epitaxy and triple-well processes. Chapter 6 will discuss the integration of many of these structural features and its influence on latchup.

Chapter 7 examines the latchup circuit solution discipline. First, we will address intradevice, intracircuit, intercircuit and interfunction latchup in integrated chip design. As part of the CMOS latchup design discipline, circuit solutions exist to address power-up, to circuits that determine the onset of latchup. A key area of interest is sequence-independent circuits. The chapter will show examples of active clamp networks and active guard ring networks. A connection is established between the active guard ring concepts and the developments of Chapter 2 on lateral bipolar transistors.

Chapter 8 presents CMOS latchup CAD methodologies. The chapter looks at latchup design checking rules and verification methods. Many concepts, creative techniques and different paths are being taken to address latchup, and it is important to note the range of methods being used today. Some methods address the extraction of pnpn structures, other methods focus on the design rule checking and additional methods are focused on verification. The methods address all things from local to global interactions in a semiconductor chip design.

In this book, the future trends and directions of CMOS latchup discipline are addressed connecting the early work in the 1970s and 1980s to today's technologies. New relevant test systems and tools are provided.

Enjoy the text and enjoy the subject of CMOS Latchup. Although it was thought to be cured, it is not. CMOS Latchup is back.

Steven H. Voldman
IEEE Fellow

# Acknowledgements

I would like to thank the individuals who have helped me on the Right Path in my academic and professional career. The academic foundation gave me the ability to cross from semiconductor physics to circuits, allowing the ability to address the field of CMOS latchup. Faculty from the University of Buffalo, Massachusetts Institute of Technology and University of Vermont had significant impact on my direction and interests. And, of course, many of my teachers were authors of technical textbooks.

As MIT graduate student teaching assistant in 1979, I was fortunate to be able to observe and participate in the teaching of semiconductor devices and circuits to the undergraduate semiconductor course EE 6.012, which at that time had a large focus on bipolar transistors – not MOSFETs. As a graduate student teaching assistant, I was able to observe the MIT faculty discuss the philosophy, method and construction of how each lecture was to be taught, and why; this analysis was fruitful in my concept for the construction of this book. After leaving MIT in 1982, I arrived at IBM Burlington Vermont and began to teach an IBM internal short course on bipolar semiconductor device physics from my lecture notes to IBM engineers; this course did get the interest of many engineers, in particular Ronald R. Troutman. Although I was not really focused on the latchup topic, Ron and I began to cross paths and take a mutual interest in each other's work. He also began serving as my part-time mentor, giving me advice, guidance and direction. By 1984, I was completing most of my bipolar SER simulation and experimental work, and completed my CMOS SER simulation; by coincidence, in 1984, Ron Troutman left on IBM Sabbatical program to MIT to write a book titled *CMOS Latchup – The Problem and the Cure*. In 1984, I was moved to Ron's department of CMOS device design and was to start working on CMOS latchup. Even after he went on IBM Sabbatical, Ron served as my latchup mentor and sent me to IEEE Standards meetings to address installing some of the first CMOS latchup characterization structures into the IEEE standards development. I began my CMOS latchup work through many more of IBM's CMOS technology generations for the next 20 years – giving me the opportunity to be the first to explore many CMOS innovations that influenced CMOS latchup – from p++ substrates, MeV implanters, retrograde n-wells, retrograde p-wells, shallow trench isolation (STI), titanium to cobalt silicide, 'connecting' implants, to p++ buried layers, and more recently BiCMOS, BiCMOS SiGe, RF CMOS, image sensors and power technologies, which gave me the opportunity to explore deep trench, subcollectors, implanted subcollectors, trench isolation, triple-well technology and LDMOS technology. So, I am indebted to Ron Troutman for setting the stage and sending me on this 20-year latchup journey.

In the latchup discipline, I would like to thank for the years of support and the opportunity to provide lectures, invited talks and tutorials on latchup: the 1993 SEMATECH PTAB on Vertical Modulated Wells, the Electrical Overstress/Electrostatic Discharge (EOS/ESD) Symposium, the International Reliability Physics Symposium (IRPS), the Taiwan Electrostatic Discharge Conference (T-ESDC), the International

Physical and Failure Analysis (IPFA), as well as ESD Association Education Committee and the ESD Association Device Testing Standards Committees. In the latchup discipline, I would like to thank the different semiconductor corporations, institutions, high-energy implant corporations and latchup tester corporations for interest in my early latchup work – Wes Morris (SEMATECH), John Lowell (AMD), Leonard Rubin (Eaton Corporation/Axcelis Corporation), John Borland (GENUS), David Bennett (Thermo KeyTek) and Evan Grund (Thermo Fisher Scientific). Wes Morris had a large influence on this book and pushed the need to add materials on single event latchup (SEL), high-dose buried layers, CMOS scaling and simulation. In ESD/latchup community, I would like to thank members of the Association involved in latchup discussions and the Transient Latchup (TLU) Standards Development Committee: Gary Weiss (AT&T), Robert Ashton (White Mountain Labs), David Bennett (Thermo KeyTek), Evan Grund (Oryx Instruments), Mark Kelly (Delphi Auto), Chris O'Connor (UTI), Tom Meuse, Mike Hopkins (Thermo KeyTek), Leo G. Henry, Jon Barth (Barth Electronics), Brenda McCaffrey (White Mountain Labs), Ian Morgan (AMD), Charvaka Duvvury (TI), Professor Ming-Dou Ker (NCTU), Gianluca Boselli (TI), Tong Li (Intrel) and Yoon Huh (GTL). I would also like to thank technical authors who contributed their work for this book – from IBM, I would like to thank Michael Hargrove, Jim Slinkman, William Craig, Jeff Brown, Al Weger, Pia N. Sanda, Franco Stellari, Myra McManus, Jeanne Mechler, Ciaran Brennan, Kiran Chatty, Min Woo, Mujahid Muhammad, Dimitri Kontos and David S. Collins. I would like to thank my friends Tong Li (Intrel), Yoon Huh (GTL), Jin Min (GTL), Peter Bendix (GTL), C.T. Hsu (Winbond), Professor Ming-Dou Ker of NCTU, Albert Wang of Illinois Institute of Technology and Rou Ying Zhan (Freescale) for assistance with the materials for this book. In the area of BiCMOS and smart power technology, I would like to thank Peter Moens and Marie Denison for their suggestions in the area of guard rings and smart power technology issues.

I would like to thank my summer students who worked on CMOS latchup with me – Tong Li of University of Illinois Urbana Champaign, Anne Watson of Penn State University and Ephrem G. Gebreselasie of IBM – for hard work, collaboration and participating in latchup experimentation and discovery.

I would also like to thank the publisher and staff of John Wiley & Sons for taking on this book as part of the ESD book series.

And most important, I would like to thank my children, Aaron Samuel Voldman and Rachel Pesha Voldman, and my wife Annie Brown Voldman – for support and for keeping our lives going forward and on the Right Path.

And of course, I would like to thank my parents, Carl and Blossom Voldman, for being a source for perseverance and overcoming any obstacle placed in front of us. Blessed is the Master of the Universe, who has given us strength.

Steven H. Voldman
IEEE Fellow

# 1 CMOS Latchup

## 1.1 CMOS LATCHUP

Latchup!
In this chapter, a brief overview of latchup is provided. We will provide a first quick look on what latchup is. As a starting point, this discussion will be followed by a summary of evolution, history, key innovations and patents. This chapter discusses the key innovations, contributions and patents associated with the process of understanding how to address the latchup issue in semiconductor technology. In addition, this chapter will provide the reader what are the sources of latchup, from pulses to particles; latchup issues associated with ionizing radiation events, as well as current and voltage excursions outside of the native current and voltage conditions of a technology, will be shown. In this discussion, the issue of technology scaling and how scaling leads to latchup concerns will be reviewed. When this chapter is completed, the following chapters will spiral backward into more depth, on each individual area from models, testing and tools.

### 1.1.1 CMOS Latchup–What is Latchup?

Latchup is a state where a semiconductor device undergoes a high-current state as a result of interaction between a pnp and an npn bipolar transistor. The pnp and npn transistors can be natural to the technology, or parasitic devices. In CMOS technology, these are typically parasitic devices. For each p-channel MOSFET (metal oxide semiconductor field effect transistor) device, there is a corresponding parasitic pnp element formed between the p-channel diffusion, the n-well and the substrate. For each n-channel MOSFET (NMOS) device, there is a corresponding parasitic npn element formed between the n-channel diffusion, the p− substrate and the n-well of the p-channel MOSFET. For each inverter gate, there are corresponding pnp and npn parasitic bipolar elements. Figure 1.1 shows an example of a cross section of a CMOS inverter circuit.

When interaction occurs between a pnp and an npn bipolar transistor, regenerative feedback between the two transistors can lead to electrical instability. This interaction between a three-region pnp and a three-region npn that share base and collector regions can be viewed as a four-region pnpn device [1–3]. As a result of the feedback between the two transistors, there exist stable and unstable

*Latchup* Steven H. Voldman
© 2007 John Wiley & Sons, Ltd

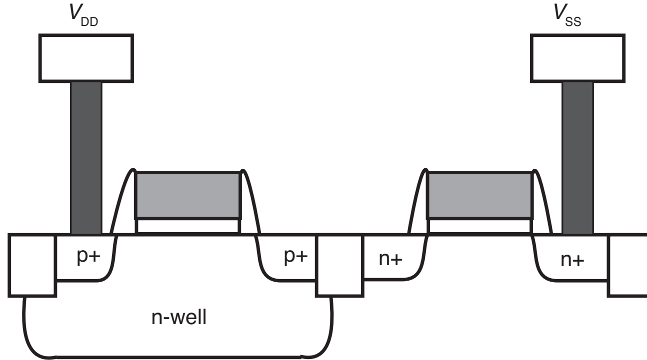



**Figure 1.1** A cross section of a CMOS inverter structure in a CMOS technology highlighting the pnp and npn parasitic bipolar transistors.

regions in the *I–V* characteristic. The *I–V* characteristic is an S-type *I–V* characteristic with both a low-current/high-voltage state and a high-current/low-voltage state (Figure 1.2). In an S-type *I–V* characteristic, there are multiple current states for a given voltage level; the state it chooses is a function of the circuit load line. It is 'off' in normal operation and can be triggered 'on' in a high-current state. In this state, it establishes a high current at a low-voltage, allowing a low impedance shunt. When the two transistors are coupled, the combined device acts as a four-region device of alternating p- and n-doped regions with three physical p–n metallurgical junctions, forming a pnpn structure.

Why are we concerned about latchup? When these parasitic pnpn elements undergo a high-current state, latchup can initiate thermal runaway and can be destructive [4]. Latchup events can lead to destruction of a semiconductor chip, package or system. The current magnitude is such that typically the semiconductor silicon, aluminum and copper metallization fails, and sometimes the package

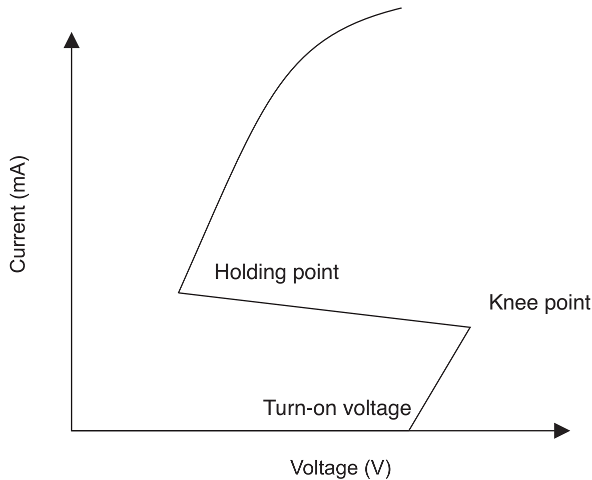



**Figure 1.2** The latchup *I–V* characteristic.

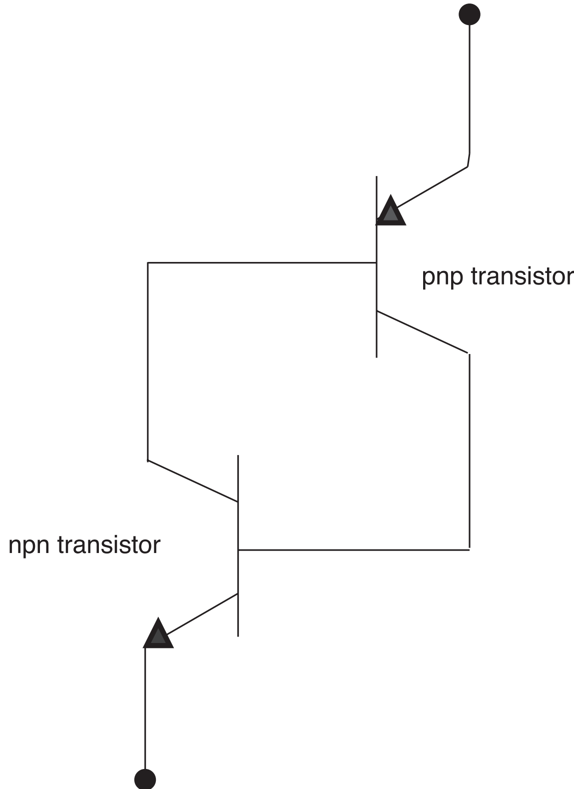



**Figure 1.3** Two-transistor representation of the cross-coupled pnp and npn transistors.

materials melt. Note that another indicator of latchup is the package cracking, melting, delamination, separation and outgassing. Another clear indicator, on a system level, is smoke. In these cases, it is difficult to provide chip-level failure analysis due to the magnitude of the package and system damage. When the card smoke is evident, the module package is melted and the silicon chip is molten, this is a good indicator that latchup has occurred in your semiconductor chip.

Back to the semiconductor device level, conceptually the two transistors can be understood as a cross-coupled pnp and npn bipolar junction transistor (BJT) device, where the base of the pnp BJT device is the collector of the npn BJT device and the base of the npn is the collector of the pnp BJT device. The two cross-coupled devices can be represented as a four-region pnpn (Figure 1.3). This pnp–npn BJT coupling establishes regenerative feedback leading to the S-type *I*–*V* characteristic and causing the electrical instability that is observed as a negative resistance state ($dI/dV < 0$). It is this feature that makes this interaction a danger and an enabler of latchup.

Application of a positive bias on the emitter of the pnp element and a ground potential on the emitter of the npn element establishes a voltage across the pnpn. The positive voltage provides forward biasing of the emitter–base junctions of the pnp and npn transistors. The base–collector junction of the pnp (which is also the base–collector junction of the npn) is in a reverse-biased state. This prevents current flow from the anode of the pnpn to the cathode. As the voltage is increased, the voltage across the base–collector junction increases. This mode of operation is called the forward blocking state. In order for current to flow efficiently from the pnpn anode to the cathode, the base–collector junction

must allow current to flow. For current continuity at the cross-coupled nodes, the collector current of the pnp transistor must equal the base current of the npn transistor, as well as the collector current of the npn transistor must equal the base current of the pnp transistor. Mathematically, the coupling is established through solving Kirchoff's current law at the base–collector nodes. In this form, the standard equations of bipolar transistors can be used to quantify the interaction and current in the pnpn structure. Hence, the two nodal equations can be expressed as

$$I_{\mathrm{cp}} = I_{\mathrm{bn}},$$

$$I_{\mathrm{cn}} = I_{\mathrm{bp},}$$

where $I_{\mathrm{cp}}$ and $I_{\mathrm{cn}}$ are the collector currents of the pnp and npn bipolar junction transistors, respectively, and likewise, $I_{\mathrm{bn}}$ and $I_{\mathrm{bp}}$ are the base currents of the npn and pnp bipolar junction transistors, respectively. The total current through the pnpn structure is equal to the emitter current of the pnp or npn bipolar transistor, $I_{\mathrm{ep}}$ and $I_{\mathrm{en}}$, respectively. From Kirchoff's current law in the transistor, the emitter current must equal the sum of the base and collector currents:

$$I = I_{\mathrm{ep}} = I_{\mathrm{cp}} + I_{\mathrm{bp}}.$$

From the coupling relationships, the current can be expressed as

$$I = I_{\mathrm{ep}} = I_{\mathrm{cp}} + I_{\mathrm{cn}} = I_{\mathrm{bp}} + I_{\mathrm{bn}} = I_{\mathrm{cn}} + I_{\mathrm{bn}} = I_{\mathrm{en}} = I.$$

Solving for the current as a function of the two collector current relationships, the collector current can be represented as a function of the emitter current:

$$I_{\mathrm{cp}} = \alpha I_{\mathrm{ep}} + I_{\mathrm{cp0}},$$
$$I_{\mathrm{cn}} = \alpha I_{\mathrm{en}} + I_{\mathrm{cn0}},$$

where the collector current is equal to the product of the collector-to-emitter transport factor $\alpha$ and the emitter current summed with the base–collector leakage. Solving for the current through the pnpn structure,

$$I = \frac{I_{\mathrm{cp0}} + I_{\mathrm{cn0}}}{1 - (\alpha n + \alpha p)}.$$

From this analysis, it is clear that the current is infinite, when the denominator is equal to zero. This condition can be expressed as

$$\alpha_{\mathrm{p}} + \alpha_{\mathrm{n}} = 1.$$

This expression shows that when the sum of the collector-to-emitter gains is equal to unity, the current goes to infinity. This can also be expressed as a function of the bipolar transistor current gain $\beta$, substituting in for the collector–emitter transport,

$$I = \frac{I_{\mathrm{cp0}} + I_{\mathrm{cn0}}}{1 - \left(\dfrac{\beta_{\mathrm{n}}}{\beta_{\mathrm{n}} + 1} + \dfrac{\beta_{\mathrm{p}}}{\beta_{\mathrm{p}} + 1}\right)},$$

where

$$\alpha_{\mathrm{n}} = \frac{\beta_{\mathrm{n}}}{\beta_{\mathrm{n}} + 1},$$

$$\alpha_{\mathrm{p}} = \frac{\beta_{\mathrm{p}}}{\beta_{\mathrm{p}} + 1}.$$

In this form, when the denominator is equal to zero, the product of the bipolar current gains is equal to unity:

$$\beta_{n}\beta_{p} = 1.$$

The pnpn current magnitude can be large given that the numerator is large, or if the denominator approaches zero.

Hence, from a simple derivation, it can be observed that the regenerative feedback between the parasitic pnp and the npn has a condition where a large current can be established between the power supply and the ground rail in an inverter circuit through the parasitics associated with the p- and the n-channel MOSFET transistor. The transport properties of the parasitic pnp and the npn are also involved, which are influenced by the semiconductor process and the chip design layout. As a result, latchup involves the chip architecture, the circuit layout, the circuit design and the semiconductor process.

What is the frequency of these parasitics in a modern semiconductor design? Today, in an advanced CMOS microprocessor with 200 million transistors, there are 100 million p-channel MOSFETs and 100 million n-channel MOSFETs. As a result, there are 100 million pnp and 100 million npn transistors. Hence, there are on the order of 100 million pnpn elements in a CMOS logic chip that contains 100 million inverter circuits.

*Question*: How many circuits does it take to initiate CMOS latchup? *Answer*: One.

As the number of circuits increases, the number of peripheral circuits and I/O circuits increases. In a CMOS chip with 200 million transistors, there may be on the order of 1000–10 000 interface circuits. It is likely that voltage and current events on the interface circuits can lead to latchup.

*Question*: How many peripheral circuits does it take to initiate CMOS latchup? *Answer*: One.

Yet, today, the number of products exhibiting latchup is a small minority of designs and applications. Hence, there must be a method or methods to diagnose and discriminate which parasitic pnpn structures are important, and which are not. Additionally, there are semiconductor process solutions, circuit solutions and system solutions.

### 1.1.2 CMOS Latchup–Why is Latchup Still an Issue ?

A key question on CMOS latchup is the following–Why is latchup still a concern? There are many reasons why latchup is an issue in today's semiconductor chips. The reasons why is is a concern in some corporations differ based on the choices made in the semiconductor technology, latchup design strategy as well as the latchup methodology. The following is a list of why this issue reoccurs in semiconductor chips:

- lack of characterization of parasitic devices;
- lack of semiconductor process control of parasitic devices;
- lack of dc and ac models of parasitic devices;
- lack of parasitic devices in the circuit simulation;
- lack of extraction, checking and verification of parasitic devices;
- lack of tools addressing parasitic devices between devices, circuits and subfunctions;
- lack of ground rules that sufficiently provide 100 % coverage of design environment;

- low business priority of addressing CMOS latchup, until it is a concern;
- lack of educational training on CMOS latchup in university and college course work for circuit and layout design engineers and technicians;
- education focus on the understanding of bipolar transistors and parasitic transistors;
- lack of focus on latchup during semiconductor process and device design point definition;
- lack of preservation of dimensional and electrical similitude of the lateral and vertical dimensions in the scaling of CMOS technology;
- lack of awareness of the CMOS latchup technology scaling and making CMOS latchup technology part of the technology roadmap;
- lack of semiconductor device and CMOS latchup design point cosynthesis;
- isolation, well and substrate scaling;
- new design methodologies in digital, mixed signal and RF designs;
- nonnative voltages integrated into products well above the native-voltage CMOS latchup capability;
- introduction of high-voltage CMOS (HVCMOS) integrated with low-voltage CMOS (LVCMOS) circuitry;
- new issues associated with subsystem and system integration;
- reversal of battery installation;
- negative polarity on pins;
- wire-bond mismatching in multichip power systems.

These are just some of the reasons why today CMOS latchup is not 'cured' and remains an issue in today's semiconductor chips. In this book, we will address many of these issues in the following chapters.

### 1.1.3 Early CMOS Latchup History

The first transistor, invented at Bell Labs in 1947, was not a CMOS transistor, but was a bipolar junction transistor device. At that time, an advantage of the bipolar junction transistor was the ability to manufacture it, as well as the speed of the transistor. One of the early advantages of the bipolar junction transistor was that it was not a surface device, but a bulk device. In 1955, Bell Labs manufactured and demonstrated the first metal oxide semiconductor field effect transistor. The MOSFET structure was proposed in 1930 by Lilienfeld [5] and Heil [6], but was not manufactured successfully due to fixed and mobile positive charge issues. Fixed charge and mobile charge problems also remained a large issue in p-channel transistors, leading to the early implementation of n-channel MOSFET technology. CMOS had to wait for the p-channel MOSFET device to emerge while the development and research community continued to look for a solution to the oxide charge problem.

By the late 1960s, many manufacturers offered integrated circuits based on either a p-channel or an n-channel MOSFET device. CMOS was invented by RCA and was first demonstrated by Al Medwin at RCA's technology center in Somerville, New Jersey. The RCA Corporation called this technology

COS/MOS, which stood for complementary symmetry metal oxide silicon. The first circuits using COS/MOS included a circuit that contained 13 transistors with a 15-V power supply.

Early interest in CMOS was also due to the potential for military and space applications. Publications in the 1960s by Kinoshita *et al.* [7], Poll and Leavy [8, 9] and Dennehy *et al.* [10] focused on the initiation of CMOS latchup in radiation environments.

CMOS latchup became of growing interest in the early 1970s as interest increased in the usage of CMOS for mainstream technology applications. Interest in CMOS continued in Sandia Laboratories and RCA [11–16]. Gregory and Shafer [12–14], Gallace and Pujol [15] and Barnes *et al.* [16] of Sandia Laboratories focused on latchup as an impediment to the mainstream integration of CMOS. During the first development of RCA's COS/MOS, CMOS latchup was discovered by Gallace and Pujol [15]. Ironically, the parasitic bipolar transistors unintentionally formed by the CMOS inverter switch impacted the introduction of CMOS. With the early problems of CMOS technology, most corporations began to focus on NMOS technology.

In this time frame, prior to mainstream introduction of CMOS, model development, guidelines and radiation implications were analyzed. Alexander *et al.* evaluated MOS model implications [17]. Brucker evaluated the implications on CMOS and silicon-on-sapphire (SOS) memory [18]. Coppage and Evans evaluated the characteristics of the destruction induced by CMOS latchup [19]. Total dose characterization and other nuclear radiation effects were evaluated by London and Wang [20], Simons [21] and Ricketts [22].

It was during the late 1970s that the focus on how to improve CMOS latchup using design techniques and process solutions was first addressed. From a practical perspective, application guidelines were released by Stephenson to address CMOS latchup guidelines [23]. The focus shifted toward addressing CMOS latchup by screening, testing and selection processes. Sivo *et al.* focused on methods of latchup screening [24]. Crowley *et al.* used radiation as a method of selection [25, 26]. In this time frame, the use of neutron irradiation for the prevention of CMOS latchup was proposed by Adams and Sokel [27]. In this time frame, the use of recombination centers (e.g. using gold dopants) was demonstrated by Dawes and Derbenwick [28].

In many corporations, the first attempts to establish a mainstream CMOS technology for memory and logic development discovered the problem of latchup. Early developers of CMOS were hindered by the latchup problem, impeding the mainstream introduction of CMOS in the 1970s. In those time frames, MOSFET-based static read access memory (SRAM) introduced resistor-load NMOS cells and four-device NMOS cells, and avoided integration of p-channel MOSFETs into the network. As a result, CMOS logic was delayed until solutions to latchup were resolved.

Early CMOS development generations addressed CMOS latchup with the introduction of epitaxial wafers, guard rings and ground rules. In the late 1970s to early 1980s, early CMOS technologies had low-doped substrate wafers, recessed oxide (ROX) or LOCOS isolation, and diffused n-wells. These early CMOS technologies had strong lateral bipolar and vertical bipolar transistors and were latchup prone. The early work of Estreich *et al.* [29–33] addressed the physics and modeling of CMOS latchup to a level suitable for mainstream semiconductor manufacturing. Estreich studied the use of a heavily doped masked well region, similar to the bipolar subcollector, to improve latchup tolerance.

In the early 1980s, the focus in the CMOS latchup discipline shifted toward the practical integration of CMOS in manufacturing facilities and mainstream CMOS development. Many of the proposed concepts prior to this were not acceptable for mainstream CMOS technology (e.g. neutron irradiation and gold doping). For mainstream implementation of CMOS, it was necessary to understand guard ring structures, guard ring efficiency, the role of the substrate, semiconductor device models, transient latchup physics and model development, as well as new process features that could be integrated with the evolution of CMOS technology.

The works by Dressendorfer and Armendariz [34], Schroder *et al.* [35], Payne *et al.* [36] and Raburn [37] focused on the elimination of CMOS latchup in CMOS technology. Rung *et al.* [38] and Combs

[40] discussed the use of p-wells and scaleable p-wells for CMOS integration and for CMOS latchup improvement. Iizuka and Moll [41] initiated new figures of merit to quantify CMOS latchup. By 1981 and 1982, Dressendorfer and Ochoa [42, 43], Wieder *et al.* [44], Hu *et al.* [45], Rung and Momose [46] and Huang *et al.* [48] focused on the model development, distribution effects, device simulation and characterization of CMOS latchup. Troutman and Zappe [49, 50] and Rung and Momose [51] focused on transient CMOS latchup associated with power-up conditions.

In 1983, the focus was on achieving two goals. Many developers focused on n-well and p-well optimization for integration with mainstream CMOS. The second goal was to achieve CMOS latchup 'immunity'. Schwabe *et al.* [52], Wollesen *et al.* [53], Goto *et al.* [54], Hu *et al.* [55] and the team of Takacs *et al.* [56] focused on CMOS n-well and p-well engineering for CMOS technology that was being cosynthesized with the CMOS latchup objectives for the future 1-μm CMOS technology era.

By 1983, the goal of achieving 'CMOS latchup immunity' was being advertised and emphasized for alleviation of fears of both technology management and customers. During this time frame, latchup was still prevalent, and the quality of the technologies and the products did not have many of the proposed solutions that were being proposed in the publications and literature. For example, Manoliu *et al.* [57] and Wakeman [58], as well as many other authors, discussed the elimination of CMOS latchup and CMOS latchup 'immunity'.

In 1983, the focus continued on the modeling of CMOS latchup. At that time, the focus was on the improved understanding of guard rings, transient latchup and distribution effects. Troutman focused on an improved understanding of the guard ring and guard ring efficiency [59]. Niitsu *et al.* [60] and Hamdy and Mohsen [61], and G. Hu [63] focused on transient latchup simulation and model development. With the focus on substrate modeling, distribution effects and better latchup models, Troutman and Hargrove developed a transmission line model for better quantification of the substrate on a p− epitaxial/p++ substrate wafer [62]. Research continued in the understanding of CMOS latchup current paths [64], characterization and test structure development [65] and temperature effects [66].

By 1984, while the research and development publications were being printed, CMOS latchup technology solutions were being integrated into future semiconductor technologies. Troutman was responsible for the manufacturing implementation in a 0.8-μm CMOS technology; in this technology generation, the first high-energy MeV implanter was used for a MeV retrograde well. (*Note*: it was integrated into the technology flow prior to 1984, but qualification of the technology did not occur until 1988.) By 1984, the research and development continued toward improved model development [67–78] as well as a shift in focus on the scaling implications of CMOS latchup [73–75]. Lewis [73], Schwabe *et al.* [74] and Odanaka *et al.* [75] focused on the reduced dimensions and the scaling of the p+/n+ spacings and scaled CMOS.

In this time frame, semiconductor process solutions, characterization and cosynthesis were integrated into commercial CMOS technologies with significant success. It was believed at this time that the problem of CMOS latchup was understood and 'cured' [79].

It was at this junction that the beginning of the issues of process integration and cosynthesis into CMOS memory and logic began. In the 1980s, CMOS logic was derived from a DRAM technology base. In this time frame, the issues of cointegration of CMOS DRAM and logic in a common base led to CMOS latchup optimization issues that needed to be addressed. In the CMOS IV 4-MB DRAM technology, Cottrell and Voldman addressed the semiconductor device cointegration of trench DRAM issues [80]. In this work, it was found that the integration issues associated with DRAM trench parasitic devices, DRAM retention time and CMOS latchup were all influenced by the n-well design point. The retrograde well provided a process optimization tool to enhance CMOS latchup immunity. n-well compensation effects caused by p−/p+ substrate wafers in a LOCOS technology were first addressed by Voldman in 1987 [80]. In 1987, Voldman and Fitzgerald evaluated transient latchup as a function of retrograde well dose and energy to improve CMOS latchup. In 1991, Voldman evaluated the integration of retrograde n-well, shallow trench isolation (STI) and epitaxial p−/p++ wafer optimization in

mainstream 0.5-μm CMOS technology and 16-MB DRAM technology [82]. By 1993, the first 0.25-μm 2.5-V CMOS technology integrated solutions for STI and CMOS latchup. After this time frame, CMOS DRAM development separated from CMOS logic development, leading to the separation of the two different design points [81–83].

By 1994, the need to achieve good CMOS electrostatic discharge (ESD) protection in CMOS logic and CMOS microprocessors took precedence over the CMOS latchup design point [82–85]. During this time frame, the solutions for latchup provided significant CMOS latchup margins. It was found that with the retrograde well implant process step both excellent ESD protection and latchup robustness could be achieved as the microprocessors were scaled through successive CMOS technologies [82–86]. In this period, the mainstream technologies also introduced a p-well implant for the n-channel MOSFET, which improved latchup technology margin.

Between 1995 and 2000, the latchup design point adequately prevented latchup in the majority of semiconductor applications in memory and logic applications. But, during this time frame, CMOS technology scaling of the fundamental parameters reduced the latchup margin [87]. Although the technology latchup robustness prevented latchup from occurring, new computer-aided design (CAD) methods were needed to integrate I/O design, CMOS latchup and ESD. A shift occurred toward better design methods, anticipating integration issues in future designs. For example, Li addressed extraction techniques, peripheral I/O integration and generalized substrate transfer resistance representations [88, 89]. In this transition period, solutions to address latchup such as heavily doped buried layers (HDBLs) were proposed, but no evidence of CMOS latchup concerns was evident [91, 92].

Since 2000, with the scaling of technologies from 0.25 μm to 35 nm, mixed signal integration and the growth of the system-on-chip (SOC), the focus on CMOS latchup has shifted again. During this time, some of the focus remained on CMOS scaling issues [92, 93]. Morris quantified the scaling implications of key latchup metrics and process variables for the prediction of future CMOS latchup concerns [92]. With the increased focus on cost, density, foundry compatibility, mixed signal products, RF technology and high-level integration (i.e. system-on-a-chip), the environment has significantly shifted in CMOS integration. Today, advanced CMOS is primarily manufactured on lightly doped p− substrate wafers. Multiple-well technologies require the optimization of isolation structures to provide the maximum CMOS latchup tolerance. With multiple-well implants, lateral scattering of implanted ions, multiple-circuit voltages and digital/analog mixed signal chips, the ability to provide a robust latchup-free technology has once again become a challenge.

During this time, the focus shifted toward CMOS latchup propagation in semiconductor chips, bipolar CMOS (BiCMOS) process solutions, triple-well CMOS, SOC integration latchup issues, smart power, high-voltage CMOS and new CAD representations. Latchup propagation, also known as ‘latchup domino effect’ occurred in high-density CMOS chips [94–97]. Weger *et al.* demonstrated this using animation techniques of the interaction between ESD networks and CMOS latchup, providing both spatial and temporal understanding of the CMOS latchup propagation [94]. This work was extended using CCD camera methods by Stellari [95]. Voldman developed the theory for CMOS latchup domino effect–addressing both the primary trigger source and the secondary interactions [97]. In the BiCMOS technology, the integrated dense CMOS logic addressed new CMOS latchup process solutions for wired and wireless applications. Watson and Voldman evaluated the utilization of deep trench (DT) structures for CMOS latchup improvements as guard rings to bordered n-well structures [98–102]. The first experimental work on the CMOS latchup robustness of a new low-cost isolation, known as trench isolation (TI), was also explored for its CMOS latchup advantages [102, 103]. In this period, triple-well CMOS became important in the technology. Analysis of CMOS ‘merged triple-well’ latchup, where the n-well was interconnected with the buried layer, was also addressed for the first time [104, 105]. At this time, with the direction of higher substrate resistance, the understanding of the substrate resistance issue on future technologies is a concern. Voldman demonstrated CMOS latchup in ultrahigh resistive wafers of 1–5 kΩ cm CMOS substrates [106]. With integration, Salcedo-Suner *et al.*

addressed latchup issues within new ESD circuit concepts [107]. With the SOC integration, Huh *et al*. addressed new CMOS latchup occurrences between power supplies and between circuits [108]. In addition, more sophisticated methods of CAD extraction and verification using model graph (MG) representation were pursued by Zhan and Wang [109, 110].

In the background, the smart power technology engineers with bipolar, CMOS and DMOS (BCD) technology were addressing unique issues in the smart power integration [172–185]. Smart power has introduced HVCMOS and LVCMOS technologies on a common substrate. HVCMOS technology has introduced drain-extended NMOS and lateral diffused MOS (LDMOS) technologies at voltage levels significantly above the CMOS native-voltage power supply. In each successive technology generation node, the 25-V, 40-V, 100-V and 120-V applications do not scale, whereas LVCMOS continues to lower its power supply voltage; this leads to a larger voltage difference between HVCMOS and CMOS in each generation. With smart power and HVCMOS, a larger focus is on injection interaction between the HVCMOS circuits and LVCMOS; HVCMOS LDMOS injection can lead to disturbing the CMOS circuitry. As a result, new interdomain passive and active guard ring strategies have been implemented.

Additionally, new niche fields are established that must address CMOS latchup issues. Image processing semiconductor chips have introduced core-limited designs with the introduction of CMOS triple-well technology and reintroduction of the p−/p++ substrates and p−/n++ substrates. With the growth of image chips in cell phones, this area will have growing interest.

It was once believed that latchup would cease to be an issue for advanced technologies as a result of process solutions, model development, simulation and circuit solutions. Unfortunately, this has turned out not to be the case. As a result, new latchup solutions from semiconductor process structures, implants, design tools and circuit solutions still continue to be pursued. In the following chapters, this will be discussed in more depth.

## 1.2 FUNDAMENTAL CONCEPTS OF LATCHUP DESIGN PRACTICE

In semiconductor technology, a ‘latchup design practice’ is developed for the analysis of latchup from semiconductor process to product design. For latchup, there are fundamental latchup design practices that include semiconductor process, layout design and testing practices. The fundamental concepts and objectives for the latchup discipline are as follows:

- Decoupling of parasitics – spatial separation;
- Decoupling of parasitics – physical isolation (triple well);
- Decoupling of parasitics – physical isolation (insulating regions);
- Recombination and trap states;
- Beta spoiling – placement of defects for parasitic bipolar current gain reduction;
- Semiconductor doping;
- Silicide design;
- Isolation design;
- Isolation beta spoiling;
- Decoupling of parasitics via virtual collector regions;

- Decoupling of parasitics via integrated guard rings;
- Separated guard rings;
- Shunt well resistors;
- Shunt substrate resistors;
- Circuit design;
- Current-limiting circuit elements;
- Power application issues – sequence-independent circuits;
- Power application issues – power sequencing with sequence-dependent circuitry;
- Current-limiting circuit solutions – external current-limiting circuit elements;
- Use of on-chip ESD power clamp networks;
- Use of on-chip ESD rail-to-rail networks;
- Use of on-chip active clamp networks;
- Active guard rings – local substrate potential modulation;
- Active guard rings – introduction of substrate lateral electric field;
- Active guard rings – inversion of injection;
- Active and passive guard ring synthesis;
- Failure analysis – photoemission failure analysis techniques;
- Failure analysis – particle beam methodology;
- Computer-aided design – identification of parasitic elements;
- Computer-aided design – inclusion of parasitics in circuit analysis;
- Computer-aided design – guard ring implementation;
- Computer-aided design – well and contact spacing optimization;
- Computer-aided design – latchup contact placement cosynthesis using cost functionals;
- Chip architecture – floor planning based on parasitic pnpn.

*Decoupling of parasitics – spatial separation*: To avoid the interaction between the pnp and the npn transistor, spatial separation of the parasitic transistors inherently increases the latchup robustness.

*Decoupling of parasitics – physical isolation* (*triple well*): To avoid interaction between the pnp and the npn parasitic transistors, full decoupling can be achieved by physical separation using other structures between the transistors. Using a triple-well region, one can fully isolate the npn and the pnp transistor parasitic elements.

*Decoupling of parasitics – physical isolation* (*insulating regions*): To avoid interaction between the pnp and the npn parasitic transistors, isolation can be placed in between. In this case, the isolation can provide decoupling when the depth of the isolation is significant, and/or fully isolating on the edges as well as below the structure.

*Recombination centers and trap states*: To avoid interaction between the pnp and the npn parasitic transistors, recombination centers and trap states can be added to the base region of the pnp and/or npn transistor. Recombination centers and trap states can be generated in the bulk region (e.g. bulk recombination centers) or on surfaces (e.g. surface recombination centers). Recombination centers can be generated by implantation damage, implanted impurities or irradiation. Recombination centers can be neutral or charged centers.

*Beta spoiling – placement of defects for parasitic bipolar current gain reduction*: To avoid interaction between the pnp and the npn parasitic transistors, defects and trap states can be added to the base region of the pnp and/or npn transistor. Defects can be generated in the bulk region (e.g. bulk recombination centers) or on surfaces (e.g. surface recombination centers). Recombination centers can be generated by implantation damage, implanted impurities or irradiation. Recombination centers can be neutral or charged centers.

*Semiconductor doping – high doping concentration*: The semiconductor region can be doped in order to reduce the bipolar current gain through reduction in the minority carrier diffusion length. This can be achieved by reduction in both recombination time and minority carrier mobility.

*Silicide design*: The silicide design is established to introduce resistance in the emitter regions. Emitter resistance leads to a reduction in the regenerative feedback and improves the latchup robustness. In the latchup design practice, the introduction of emitter resistance is a technique to increase the latchup robustness.

*Isolation design*: The semiconductor isolation design can be optimized to improve the latchup robustness of the semiconductor technology. Isolation structures can include STI, dual-depth (DD) STI, TI and DT isolation structures.

*Isolation beta spoiling*: It is a latchup design practice for the degradation of the parasitic bipolar transistors' current gain by influencing the isolation surfaces. These can include surface state generation, defect generation or channel stop implants. Additional methods under isolation are provided to lower the bipolar current gain transport along the isolation edges.

*Decoupling of parasitics via virtual collector regions*: It is a latchup design practice to add a virtual collector region. By the addition of a virtual collector region, minority carriers can be collected lowering the regenerative feedback between the sources of the parasitic pnpn. With the introduction of a virtual collector, the bipolar current gain of either the pnp or the npn is lowered, which improves the latchup robustness. These virtual collector structures can be vertical or lateral elements.

*Decoupling of parasitics via integrated guard rings*: It is a latchup design practice to add an integrated guard ring to serve as a virtual collector. Minority carriers can be collected lowering the regenerative feedback between the sources of the parasitic pnpn.

*Decoupling of parasitics via separated guard rings*: It is a latchup design practice to add a separated guard ring to serve as a virtual collector. By a separate region, minority carriers can be collected lowering the regenerative feedback between the sources of the parasitic pnpn. The separate guard ring structures can be placed lateral or vertical and lead to a loss of the current lowering the bipolar current gain.

*Shunt well resistors*: It is a latchup design practice to provide low-resistance well shunts in parallel with the emitter–base metallurgical junction of the parasitic element contained within the well structure. This prevents the forward bias of the parasitic transistor element contained within the well.

*Shunt substrate resistors*: It is a latchup design practice to provide low-resistance substrate shunts in parallel with the emitter–base metallurgical junction of the parasitic element contained within the well structure. This prevents the forward bias of the parasitic transistor element contained within the substrate.

*Circuit design*: It is a latchup design practice for the usage of circuit techniques to prevent latchup. The methodology includes (a) current-limiting solutions, (b) decoupling of parasitics from substrate and power connections and (c) sequence-independent power supplies and circuits.

*Use of on-chip ESD power clamp networks*: ESD power clamp networks can be used to prevent latchup. This can be achieved by limiting voltage overshoot and undershoot, and by providing alternative current paths. Given these occur prior to a latchup triggering condition, semiconductor chips are less vulnerable to latchup.

*Use of on-chip ESD rail-to-rail networks*: ESD rail-to-rail networks can be used to prevent latchup between two independent power supply rails and between two ground rails. This can be achieved by limiting voltage overshoot and undershoot, and by providing alternative current paths. Given these occur prior to a latchup triggering condition, semiconductor chips are less vulnerable to interrail power supply latchup.

*Use of on-chip active clamp networks*: Active clamp networks are used in high-performance semiconductor chips that are designed to limit overshoot and undershoot associated with chip-to-chip signal transfer in multichip environments. The objective is to design the circuits to avoid reflections. These 'active clamp' circuit networks provide clamping at the power supply value, preventing overshoot and undershoot excursions.

*Active guard rings – local substrate potential modulation*: Active guard rings collect minority carriers and electrically reconnect to a soft grounded substrate contact to modulate the electrical potential of the substrate for latchup advantages.

*Active guard rings – introduction of substrate lateral electric field*: Active guard rings collect minority carriers and electrically reconnect to a soft grounded substrate contact to modulate the lateral electric field in the substrate to reduce lateral carrier transport.

*Active guard rings – inversion of injection*: Active guard rings can sense the potential of the substrate and using an inverting amplifier invert the potential response for latchup advantages.

*Active and passive guard ring synthesis*: Active and passive guard rings can be integrated and synthesized to utilize the advantage of active guard rings and still provide a passive guard ring that allows collection to the power grid. Active guard rings can be integrated with compensation networks using multiplier–divider logic.

*Current-limiting circuit elements*: It is a latchup design practice to provide current-limiting elements that can prevent CMOS latchup. For example, current-limiting elements can be placed in series with the power or ground to prevent a high-current runaway state. In addition, this can be integrated with power shutdown and restart circuit functions.

*Power application issues – sequence-independent circuits*: It is a latchup design practice to introduce circuits that are independent of the power-up and power-down sequencing. These circuits may be peripheral receivers, off-chip drivers (OCDs) or ESD networks that are not a function of the power sequencing. This will lead to the prevention of forward biasing of junctions that can lead to latchup.

*Power application issues – power sequencing with sequence-dependent circuitry*: It is a design practice to introduce circuits that are dependent on the power-up and power-down sequencing, by adding additional elements to avoid forward biasing. This will lead to the prevention of forward biasing of junctions that can lead to latchup.

*Current-limiting circuit solutions – external current-limiting circuit elements*: It is a latchup design practice to provide current-limiting elements that can prevent latchup. For example, current-limiting elements external to the semiconductor chip can be placed in series with the power or ground to prevent thermal overload and semiconductor chip, board or system failure. Without the current-limiting element, latchup can introduce chip, board or system failure.

*Use of off-chip clamping elements*: It is a common latchup practice to have off-chip solutions to avoid overshoot and undershoot phenomena from the board. Use of diode clamping elements or off-chip ESD elements can minimize board- or card-level phenomenon that can trigger latchup.

*Failure analysis – photoemission failure analysis techniques:* Photoemission techniques are a latchup design practice to determine the photon emissions associated with avalanche breakdown, recombination and latchup. Photoemissions provide both a temporal and a spatial dependence of the time sequence, as well as the spatial location. With frequency filters, it is possible to determine whether the phenomenon is avalanche breakdown or electron–hole pair (EHP) recombination.

*Failure analysis – particle beam methodology*: In latchup analysis, a design practice is the use of particle beams as a diagnostic to identify potential locations of latchup. This can be used for the evaluation of single event latchup (SEL) sensitivity or 'weak spots' in a given design.

*Computer-aided design – identification of parasitic elements*: A CMOS latchup design practice in computer-aided design is the identification of parasitic pnpn that are potential latchup concerns. Using latchup CAD tools, the parasitic pnp and npn elements are identified and classified according to their likelihood to induce CMOS latchup.

*Computer-aided design – inclusion of parasitics in circuit analysis*: A CMOS latchup design practice in CAD is the identification and inclusion of the parasitic pnpn in the circuit analysis. The latchup CAD tool extracts the pnpn element, identifies its importance and includes the SCR circuit schematic into the standard circuit. This postextraction process updates the circuit model with the additional element.

*Computer-aided design – guard ring implementation*: A CMOS latchup design practice is computer-aided design methods that (a) autogenerate guard rings where required, (b) evaluate the guard ring effectiveness and (c) modify the guard rings to improve the latchup robustness. CAD practices vary from shape-based tools to parameterized guard ring cells. In addition, checking and verification CAD methods exist that address the compliance to the CMOS latchup requirements.

*Computer-aided design – well and contact placement and spacing optimization*: A latchup design practice is the placement of well and contacts for CMOS latchup optimization of the shunt resistances and satisfying the checking and verification design rules. For external latchup injection function, tools exist that modify the well and contact spacing based on the magnitude and distance from an injection source.

*Computer-aided design – latchup contact placement cosynthesis using cost functionals*: Latchup contact optimization can be achieved using cost functional optimization with placement and routing using greedy algorithms. The latchup contact placement optimization can be integrated into the area, density and routing calculations by minimizing analysis of the 'zone of protection' and 'unprotected zones'. A constraint of optimization algorithms is that 'unprotected zone' terms are driven to zero.

*Chip Architecture – floor planning based on parasitic pnpn*: A latchup design practice is the placement of functional blocks based on the density and presence of parasitic pnpn elements. For example, given a subfunction where there is no pnpn, elements can be placed next to injection sources (e.g. fill shapes and decoupling capacitors can be placed near I/O circuitry).

## 1.3 BUILDING A CMOS LATCHUP STRATEGY

In the construction of a semiconductor business, there are many aspects to provide reliable semiconductor components that are verified and qualified to not undergo latchup. Hence, there is a

business strategy, as well as a technology strategy. In the following sections, these two issues are separated as distinctly different list of considerations.

### 1.3.1 Building a CMOS Business Strategy – 18 Steps in Building a CMOS Latchup Business Strategy

In the practical implementation of a quality and reliability engineering latchup program, there are a number of steps to be taken in delivering semiconductor parts successfully. It is a key to develop a CMOS latchup management program for building a latchup prevention strategy from the product to the customer. The focus of this latchup program is the management of a facility and corporation in managing its staff, tooling and establishing corporate objectives. These are the 18 steps to build a latchup business strategy:

- a full-time latchup coordinator;
- a full-time latchup technologist;
- semiconductor process technology definition for latchup prevention;
- latchup roadmap strategy for technology migration and evolution of technology;
- technology latchup design test structure and design methodology;
- technology design manual latchup rules for circuit designers;
- semiconductor chip designer latchup verification release process;
- latchup quality and reliability design tools for design release verification;
- latchup manufacturing in-line test and characterization strategy;
- latchup technology characterization strategy;
- latchup quality and reliability chip product test facility and equipment;
- latchup human factor engineering;
- customer application notes on power states and sequencing of chips and systems;
- customer strategy of integration from chip to system environment;
- latchup system test facility;
- system-level latchup prevention strategy;
- an awareness and avoidance procedures for latchup sensitivity of products;
- latchup system-level quality and reliability audit capability.

### 1.3.2 Building a CMOS Latchup Technology Strategy – 18 Steps in Building a CMOS Latchup Technology Strategy

To address the latchup semiconductor technology design phase, there are a number of critical steps that need to be taken. In order to avoid latchup in semiconductor design, these steps must be taken for

ensuring consistent and acceptable latchup-resistant results. The following are the 18 steps needed to ensure the establishment of a latchup prevention strategy:

- semiconductor process latchup robust process control and solutions;
- semiconductor process/device design latchup simulation design tools;
- latchup and semiconductor device design synthesis and optimization strategy;
- latchup test structure methodology and strategy;
- latchup characterization test equipment;
- latchup characterization test strategy;
- latchup design manual strategy;
- semiconductor device-to-circuit design synthesis integration strategy;
- semiconductor circuit-to-function design synthesis integration strategy;
- semiconductor function-to-function design synthesis integration strategy;
- latchup device and circuit simulation strategy;
- latchup CAD design implementation strategy;
- latchup CAD design rule checking (DRC) tools;
- latchup CAD parasitic extraction tools;
- latchup CAD chip-level verification tools;
- latchup design engagement, review and release process;
- latchup technology benchmarking strategy;
- latchup technology scaling strategy.

## 1.4 CMOS LATCHUP TECHNOLOGY MIGRATION STRATEGY

In the migration of semiconductors, it is important to have a technology strategy that prevents latchup in successive generations. In each corporation, the integration of the semiconductor solutions for latchup may differ depending on the semiconductor tooling, cost and other design point integration issues. Table 1.1 is an example of a typical evolution of a corporate semiconductor technology latchup strategy. In this table, in each successive technology generation, a clearly proactive path was chosen to develop semiconductor process solutions that allowed for continuous scaling of CMOS technology. In the earliest days of CMOS integration, prior to 1975, a CMOS technology was developed where CMOS latchup was not addressed as part of the semiconductor design point process. No products were developed with this technology. In the first CMOS technology generation, the CMOS latchup sensitivity evaluation was regarded as part of the device design ‘CMOS design point’. To address the latchup concerns, benchmark test structures were established, as well as CMOS latchup ground rules (e.g. guard rings, n-well contact and p− substrate contact rules). In the next generation, device

**Table 1.1** Example of the history of a technology evolution with incorporated latchup semiconductor process solutions.

| Technology node | Latchup semiconductor process solutions |
|---|---|
| 2.5 μm | None |
| 2.0 μm | Guard rings |
| 1.2 μm | Diffused well optimization p− epitaxial integration |
| 0.8 μm | (a) p++ substrate, p− epi |
| | (b) Retrograde n-well implant |
| | (c) Ground rule development |
| | (d) Guard ring development |
| 0.5 μm | (a) Shallow trench isolation |
| | (b) Retrograde well optimization |
| 0.5 μm | (a) Epi thickness increase |
| | (b) n-well dose control |
| 0.25 μm | (a) Increase of n-well dose |
| | (b) Increase of n-well energy |
| 0.22 μm | (a) p-well integration |
| | (b) Increase n-well dose |
| 0.22 μm | (a) Epitaxial buried layer implant |
| | (b) STI scaling |

design simulation and optimization of a diffused well technology were initiated. In this next technology generation, a p− epitaxy was integrated on a p++ wafer – more aggressive latchup solutions were pursued. Logic and DRAM were developed off a common base technology generation. For example, trench DRAM capacitors as the base technology led the way to the integration of a scaled p− epitaxial on a p++ heavily doped substrate [80]. Additionally, the epitaxial region was scaled significantly compared to the prior technology generation. This 0.8-μm LOCOS isolation-defined CMOS technology introduced the first retrograde n-well using a commercial MeV implanter in a commercial mainstream technology for the purpose of latchup improvements [80]. With the combination of p++ substrate, scaled epitaxy, retrograde n-well and titanium salicide, a significant boost in the latchup robustness was observed. In the next technology generation, STI was integrated into the technology for the purpose of density and elimination of LOCOS isolation bird's beak effects. In the 0.5-μm technology, the combination of a p++ substrate, a scaled p− epitaxial region, a retrograde n-well, silicide and finally shallow trench isolation provided a significant CMOS latchup robustness improvement compared to the prior generation [81–85]. In the first commercial 2.5-V 0.25-μm channel length technology, to address the p+/n+ scaling of the physical dimensions, the n-well sheet resistance was lowered to provide ESD robustness improvements and at the same time to reduce the vertical and lateral bipolar transistors. In the next generation, in a 0.22-μm technology, dual-well CMOS optimized the p-well and n-well regions to higher doping concentrations for latchup and ESD improvements. This technology was followed by another one, with a new implant called a 'buried epitaxial implant' between the p-well and the p++ substrate for CMOS latchup.

From this discussion on the semiconductor process technology evolution, a clear message is evident; to improve or maintain latchup robustness, semiconductor process solutions are needed to be added to avoid latchup. When the process solutions are not added, latchup sensitivity increases with technology scaling; these can be compensated by increased spacing (e.g. descaling), more guard rings and more restrictive ground rules.

## 1.5 KEY METRICS OF LATCHUP DESIGN PRACTICE

For CMOS latchup, there are fundamental latchup design metrics that include semiconductor process, layout design and testing practices:

- p+/n+ space;
- beta product;
- alpha sum;
- overshoot voltage and current;
- undershoot voltage and current;
- latchup specification;
- margin from latchup specification;
- radiation dose.

*p+/n+ space*: The spacing between the p+ anode and the n+ cathode is referred to as the 'p+/n+ space'. This metric is a key measure of the scaling of the technology associated with latchup.

*Beta product*: The beta product metric is a measure of the likelihood of the parasitic or network to undergo latchup. The beta product refers to the product of the pnp parasitic bipolar current gain and the npn parasitic bipolar current gain, $\beta_{pnp}\beta_{npn}$. As the beta product decreases, the likelihood of latchup decreases. (*Note*: This metric serves as a simplified or special case of the stability criterion.)

*Alpha sum*: The 'alpha sum' metric is a measure of the likelihood of the parasitic or network to undergo latchup. The alpha sum refers to the sum of the pnp and the npn parasitic bipolar collector-to-emitter transport factor, $\alpha_{pnp} + \alpha_{npn}$. As the alpha sum decreases, the likelihood of CMOS latchup decreases. (*Note*: This metric serves as a simplified or special case of the stability criterion.)

*Overshoot voltage and current*: This metric determines the amount of voltage or current required to initiate latchup in a circuit or chip from an overshoot event.

*Undershoot voltage and current*: This metric determines the amount of voltage or current required to initiate latchup in a circuit or chip from an undershoot event.

*Latchup specification*: Latchup specifications satisfy the requirement of a given magnitude of current required to 'pass the latchup specification'. Given it 'passes' the latchup specification is a measure of the integrity of the semiconductor chip to CMOS latchup events. Note that the specification magnitude is an arbitrary value established in the industry and does not guarantee latchup does not occur.

*Margin from latchup specification*: Latchup specifications are based on satisfying the requirement of a given current without inducing CMOS latchup. The margin of current or voltage above this specification is a metric for assurance that the semiconductor chip will survive a current or voltage above the latchup specification.

*Radiation dose*: The magnitude of the radiation level for a given radiation type required to initiate latchup can serve as a metric of the integrity of a semiconductor chip. The radiation dose level is a function of the radiation type (e.g. photon, heavy ion and cosmic ray).

## 1.6 CMOS LATCHUP TECHNOLOGY TRENDS AND SCALING

Latchup robustness and the latchup metrics are strongly influenced by technology scaling. With MOSFET constant electric field scaling theory, the dimensions of semiconductors continue to be decreased. With CMOS technology scaling, the following key CMOS latchup metrics and technology variables have a strong influence on the trend of CMOS robustness:

- p+/n+ scaling;
- substrate resistance;
- isolation scaling;
- n-well and p-well design;
- power supply scaling;
- holding voltage.

As the lateral physical dimensions of semiconductor components are reduced from micron to nanometer dimensions, a key latchup metric is the p+/n+ space. Figure 1.4 shows a plot of the p+/n+ space as a function of technology generation. To achieve higher density of circuits in each technology generation, the minimum dimension is scaled. From a latchup perspective, this requires a smaller isolation space between the p+ diffusion in the n-well and the n+ diffusion in the p− substrate, leading to a decreased p+/n+ space. Figure 1.4 shows the change in the p+/n+ spacing [87, 92]. Additionally, since there are semiconductor

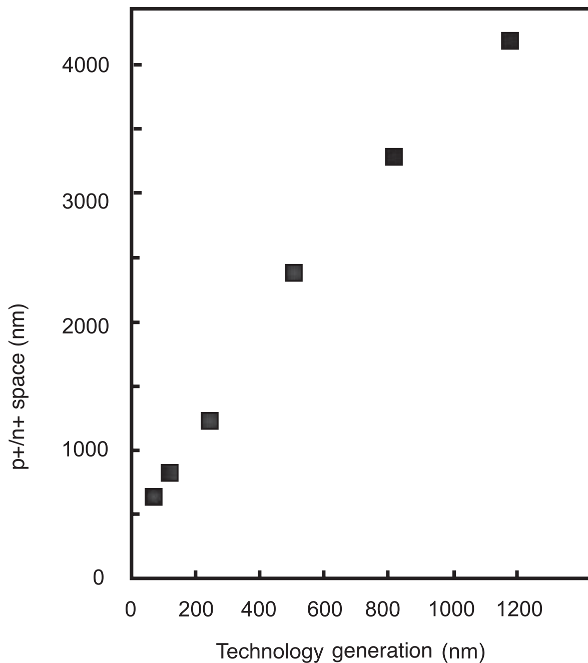



**Figure 1.4** CMOS latchup metric p+/n+ space as a function of technology generation.

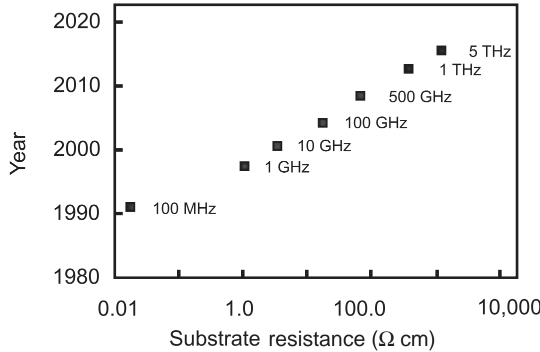



**Figure 1.5** Scaling of substrate wafer resistivity as a function of time and application frequency.

manufacturing limits on the isolation, the reduction of the isolation width leads to a decrease in the isolation depth. This also impacts CMOS latchup. Finally, the reduction of the dimensions also leads to a more stringent restriction of acceptable implantation dose and energies for the n-well and p-well due to scattering phenomenon, MOSFET threshold modulation, MOSFET back-bias, leakage and physical alignment.

With future technology generations and the power supply scaling, noise reduction between digital, analog and RF circuits becomes more critical. To reduce noise and junction capacitance, the substrate wafer resistance is scaled to higher substrate doping concentrations. Figure 1.5 shows the scaling of the substrate resistance with time, as a function of application frequency. Figure 1.5 shows that with scaling the direction is toward higher substrate resistance. Substrate resistance will strongly influence the CMOS latchup robustness. First, the substrate resistance is critical in the shunt substrate resistance (which will be discussed in the following sections). Second, low doping concentration leads to long minority carrier lifetimes. Finally, the ability to dissipate local and global self-heating becomes an increasing concern.

One of the key issues in CMOS latchup scaling is that the lateral scaling occurs at a much faster rate than the vertical scaling. As a result, the scalings of the isolation, n-well, p-well and epitaxial regions are not synchronous with lateral scaling issues. Hence, the geometric similitude is violated in the two-dimensional cross section, leading to difficulty in the production of a scalable 'model'. Hence, one of the issues associated with CMOS latchup is the n-well and p-well design, which will be discussed in detail in the following sections.

To abide by MOSFET constant electric field scaling theory, the power supply voltage is reduced with technology scaling. The scaling of the power supply voltage allows key CMOS latchup metrics to have CMOS latchup conditions above the power supply voltage (e.g. holding voltage). In the case where the latchup injection and voltage sources are also scaling, there is no concern. But, in the case where the latchup source event is not scaling (e.g. cable discharge events (CDEs) and ionizing radiation), this can be an issue.

One of the key issues is the CMOS latchup holding voltage $V_H$. The holding voltage is the voltage at the low-voltage/high-current state that occurs during latchup. A concern is the relative value of the holding voltage versus the power supply voltage. As a technology is scaled, the holding voltage decreases. Table 1.2 show an example of the trend of $V_H$ and p+/n+ space as a function of technology

**Table 1.2** Holding voltage and p+/n+ space as a function of technology generation.

| Technology generation(μm) | p+/n+ space(μm) | Holding voltage (V) |
|---|---|---|
| 0.5 | 2.0 | 6 |
| 0.35 | 1.6 | 5.5 |
| 0.22 | 1.3 | 3.7 |
| 0.18 | 1.0 | 2.7 |
| 0.13 | 0.8 | 2.7 |

generation. In each generation, the power supply voltage is also lowered. As a result, $V_H$ and $V_{DD}$ are scaling with semiconductor device size reduction.

## 1.7 KEY DEVELOPMENTS

There are key developments that influenced the directions and solutions of semiconductor technology and the art of the latchup discipline. In this section, a brief list of key innovations, contributions, technical works and patents that influenced the direction of latchup is shown.

### 1.7.1 Key Innovations

Key innovations that significantly influenced latchup are primarily those associated with semiconductor evolution and layout design directions. The semiconductor technology features that highly influenced CMOS and BiCMOS latchup robustness are as follows:

- introduction of the implanted MeV retrograde well;
- introduction of STI;
- introduction of p++ substrates;
- introduction of p++ buried layers;
- introduction of DT isolation;
- introduction of subcollector implants.

The innovations that have had the greatest influence on the semiconductor design are the following:

- introduction of single and dual guard ring structures;
- introduction of n-well and substrate contact resistance design rules and verification systems;
- introduction of the JEDEC latchup specification and commercial qualification of semiconductor components.

### 1.7.2 Key Contributions

In the latchup field, accomplishments to advance the field are in the form of development of experimental discovery, analytical models, introduction of new semiconductor devices and circuits, test equipment and test methods. Below is a short chronological list of key events that influenced the field of latchup:

- 1967: Poll and Leavy demonstrated 'transient radiation induced CMOS latchup' [8, 9]. This work demonstrated that 'single event latchup' and other radiation environments can induce CMOS latchup in space applications.
- 1973: Shafer and Gregory addressed commercial application usage of CMOS and CMOS latchup implications [12–14].
- 1976: Dawes and Derbenwick demonstrated CMOS latchup improvements with gold doping [28].

- 1978: Estreich *et al.* demonstrated latchup improvements using epitaxial buried layer process [29–33].
- 1979: Adams and Sokel demonstrated usage of neutron irradiation to control CMOS latchup [27]. This work showed that nuclear irradiation modifies the parasitic bipolar current gains.
- 1980: Estreich's thesis 'The physics and modeling of latch-up in CMOS integrated circuits' provided significant advancement on latchup theory [33]. This work significantly advanced analytical models, characterization and latchup characterization structures of CMOS.
- 1982: Parillo demonstrated 'dual-well CMOS' advantages for CMOS latchup. The significance of this work is the demonstration of the CMOS latchup advantages of an n-well and a p-well region [39].
- 1983: Takacs *et al.* demonstrated CMOS latchup in CMOS structures with and without epitaxy in a field oxide (FOX) technology [56]. This work demonstrated that a two order of magnitude improvement is achievable in the CMOS latchup critical current with epitaxial layers as well as demonstrated that CMOS latchup critical voltage is surface induced due to field oxide scaling.
- 1982: Rung *et al.* demonstrated usage of deep trench isolation for CMOS technology. This work demonstrated the advantage of deep trench structures for isolation of CMOS p-channel and n-channel MOSFET devices [47]. Today, this work is significant for BiCMOS and smart power applications.
- 1983: Troutman demonstrated the improvement in guard ring efficiency utilizing p− epitaxial/p++ substrate wafers [59].
- 1983: Troutman and Zappe developed theory for transient latchup [49, 50]. From Zappe's thesis, significant progress was made in the understanding of transient latchup analysis.
- 1982: Troutman developed an automated latchup test system that varies substrate and well resistances for the quantification of CMOS latchup design space as a function of well and substrate resistances. This work established a design methodology for CMOS latchup design point analysis.
- 1983: Troutman developed the generalized differential tetrode criterion for CMOS latchup. This work provided a better analytical relationship for the 'beta product' and 'alpha sum' latchup criteria [79].
- 1984: Troutman integrated the retrograde well into IBM's CMOS IV technology for a 0.8-μm CMOS DRAM and logic technology generation. The significance of this work was the first commercial integration of an MeV retrograde well implanter into a commercial CMOS technology.
- 1984: Terrill and Hu developed a substrate resistance model for CMOS latchup modeling. The significance of the work is that the model addresses spreading resistance in the substrate [71].
- 1985: Troutman wrote the first textbook on CMOS latchup, 'CMOS Latchup: The Problem and the Cure' [79].
- 1985: Hargrove and Troutman developed a substrate transmission line model for CMOS latchup [62]. This work was the initial step in providing a better substrate model for the understanding of CMOS latchup in a p− epitaxial/p++ wafer.
- 1986: Voldman demonstrated retrograde well–substrate modulation leading to variations in the CMOS latchup sensitivity in a 4-MB DRAM 0.8-μm technology [80]. This work demonstrated the need for cosynthesis of epitaxy thickness control, retrograde well energy and dose module in a p− epitaxial/p++ heavily doped wafer [80–82].

- 1986: Chern and Wu developed a new analytical three-dimensional model for substrate resistance in CMOS latchup structures. The significance of this work is that it addresses the substrate spreading resistance associated with majority carrier current during latchup events.
- 1991: Johnston *et al.* addressed the effect of temperature on single-particle latchup (e.g. SEL). The significance is experimental verification of the temperature dependence relationship of the latchup capture cross section and the linear energy transfer (LET) threshold [155]. This is critical to the evaluation of SEL in space and terrestrial applications.
- 1992: Voldman demonstrated the issue of retrograde well–substrate modulation on both ESD and CMOS latchup in a 0.5-μm technology. The first significance of the work was the cooptimization of CMOS latchup and ESD networks. The second significant issue was the impact of the MeV retrograde well implant energy and dose on CMOS latchup scaling and ESD scaling for scaled CMOS technology. Finally, this work demonstrated how to develop semiconductor processes that can achieve ESD robust diode networks and improved CMOS latchup in each technology generation; this work influenced the commercial usage of MeV high-energy implanters for advanced CMOS technology and future CMOS scaling [81–86].
- 1992: Borland proposed the 'buried implant layer and lateral isolation' (BILLI) semiconductor process for the utilization of a blanket implant and LOCOS isolation technology. The significance of the work is the utilization of buried layers for latchup improvement. This concept did not integrate into the mainstream CMOS technology, instead chose the path of heavily doped retrograde n-wells and p-wells, as well as triple well.
- 1993: SEMATECH established a 'vertical modulated well (VMW)' research effort on the integration of retrograde well with scaled CMOS. During this time frame, only one corporation was using retrograde well in standard offerings, and all other US corporations used diffused well CMOS. This accelerated the integration of retrograde wells in many corporations today.
- 1993: Morris (at SEMATECH) focused on the utilization of heavily doped buried layers for future scaled CMOS. This work demonstrated the value of HDBL for latchup in scaled CMOS technology.
- 1995: Bafluer *et al.* addressed latchup concerns in smart power technology [172]. The significance was highlighting the reverse-current problem and LDMOS devices' latchup-related issues.
- 1996: Barak *et al.* addressed the relationship of single event upset (SEU) and SEL. This work highlighted the interaction between circuit soft error rate, system disturbances and single event latchup, as well as addressed global chip interactions [171].
- 1998: Li's thesis 'Design automation for reliable CMOS chip I/O circuits' demonstrated advancement in CAD concepts related to latchup [88–90]. This work (a) implemented Troutman's transfer resistance representation of the substrate to a full CAD tool implementation and (b) extracted parasitic pnp or npn bipolar elements that are extracted, identified and added to its supported device representation (e.g. associated with the parasitic transistor) [88–90].
- 2000: Winkler and Herzl focused on 'active guard ring' concepts for noise suppression in mixed signal chip using on-chip driven guard rings. These concepts are presently being applied to smart power applications for BCD technology [176].
- 2002: Vashenko and Concannon demonstrated the relationship between injection current magnitude and distance between the injection source and the CMOS circuit. This work highlighted the domains of CMOS latchup failure. This work is unpublished.

- 2002: Rubin and Morris integrated heavily doped buried layers into technologies to demonstrate CMOS latchup advantages versus technology trade-offs [91]. This work showed the CMOS latchup improvement with the lower substrate shunt resistance in a p− substrate wafer.
- 2000: Parthasarathy *et al.* addressed the usage of deep trench for isolation in HVCMOS [181].
- 2003: Morris demonstrated CMOS scaling implications for future CMOS technologies [92]. The significance of the work is the demonstration that CMOS latchup is predicted to be an issue in the future, and new process solutions will be needed in the future.
- 2002: Voldman developed the concept of integration of the 'transmission line pulse' test methodology with the picocurrent analysis (PICA) tool for visualization of CMOS latchup propagation in space and time. This tool concept was first experimentally performed by the IBM T. J. Watson team of Weger and Sanda [94]. This work demonstrated the ability to determine the time delay between the injection phenomenon and the CMOS latchup event.
- 2003: Stellari demonstrated CMOS latchup propagation using CCD methodology [95]. This work similarly to the work of Weger *et al.* showed the propagation of CMOS latchup within a dense CMOS logic region.
- 2003: Boselli *et al.* evaluated through simulation the CMOS scaling in a 65-nm technology node [93].
- 2003: Horn's PhD thesis 'On the reverse-current problem in integrated smart power circuits' analyzed the effectiveness of 'compensating active guard ring' techniques for smart power circuits [185].
- 2003: ESD Association released the Transient Latchup (TLU) standard practice (SP) document ESD DSP5.4-2003 on power supply transient stimulation events. Key contributors included I. Morgan (AMD), G. Weiss (AT&T) and C. O'Connor (UTI). The significance of the work was to address power supply transients for both positive and negative events on power supply rails.
- 2004: Watson and Voldman evaluated the CMOS latchup characteristics of a BiCMOS silicon germanium DT structure. The study extended the work of Rung *et al.* by evaluating deep trench and a subcollector for CMOS latchup improvements. This work demonstrated that the subcollector leads to significant improvements when added to the deep trench; this has significant implications for CMOS latchup robustness improvement for automotive and space applications [98–101].
- 2005: Voldman demonstrated the implications of 'merged triple-well' technology. This work demonstrated that the merged triple-well implementation leads to undershoot CMOS latchup robustness degradation and overshoot CMOS robustness enhancement [104,105].
- 2005: Voldman demonstrated the advantage of a 'trench isolation' technology utilized for a low-cost BiCMOS silicon germanium technology [103]. The significance of this work is the evaluation of a second trench concept – a new low-cost trench technology integrated with implanted subcollectors demonstrating CMOS latchup robustness well above 0.22-, 0.18- and 0.13-μm CMOS technologies on low-doped substrate wafers [103].
- 2005: Huh *et al.* demonstrated CMOS latchup issues associated with SOC integration (power supply to power supply and I/O to I/O latchup events). This work has significant implications on CMOS design integration and CAD methodologies [108].
- 2005: Voldman further developed analytical development of CMOS latchup domino theory. A significant result is the extension of the theory of the differential generalized latchup criterion

(developed by Troutman) including the external injection source. This criterion can be applied for external latchup processes [97].

- 2005: Voldman demonstrated CMOS latchup in ultrahigh resistive wafers of 1–5 kΩ cm CMOS substrates. This work demonstrated that the p-well region has a significant influence on the CMOS latchup; this work demonstrated that ultrahigh resistivity wafers will be suitable in the future for RF CMOS, RF BiCMOS, RF BiCMOS silicon germanium, and other SOC and NOC applications that require high performance and low noise [106].
- 2005: Zhan and Wang provided a 'model graph' representation of the extraction of a parasitic pnpn. The significance of the work was utilization of model graph representation applied to CMOS latchup extraction, checking and verification CAD tool development [109, 110].
- 2006: ESD Association Work Group 14 (WG14) released system-level CDE measurement standard practice document. The significance of this work is associated with the growing problem of CMOS latchup events due to cable discharge in systems, boards and chips in advanced CMOS technologies.

### 1.7.3 Key Patents

CMOS latchup was a concern in CMOS technology, as the interest increased for commercial applications. As a result, there were a significant number of patents to address the semiconductor process, circuit and CAD solutions. In the early 1980s, CMOS latchup focus was on circuit solutions and semiconductor processes that integrated latchup solutions with single-well CMOS technology. Early patents addressed CMOS isolation methods and the integration of guard rings with CMOS technology for CMOS latchup. This was followed by dual-well CMOS technology patents. In this time frame, high-energy implanters were developed, leading to a focus on retrograde well implantation as well as high-energy MeV implantation buried layers. In the 1980s, focus also began on circuit solutions for CMOS latchup. In the 1990s and more recently in 2000, the focus has been on CAD tools for extraction, checking and verification, new process features, novel processes and novel failure analysis techniques. With CMOS scaling, patents have focused on quantifying CMOS latchup and design integration into high-density CMOS technology. Additionally, the cosynthesis of CMOS scaling, noise and latchup for RF applications and SOC environments has led to an interest in BiCMOS solutions from process, circuit and integration perspectives for improved CMOS latchup robustness.

## 1.8 LATCHUP FAILURE MECHANISMS

Latchup failure mechanisms can occur on both a local and a global level in chip, board and system environments. Table 1.3 contains failure mechanisms observed in CMOS latchup events. For latchup events that occur in a signal pad, the failure mechanism is typically observed in CMOS inverter circuits. In CMOS inverter circuits, latchup is observed in the parasitic pnpn formed between the p- and the n-channel MOSFET. The failures typically occur near the insulators and metallurgical junction regions. These failures are evident for contact, silicide and metallurgical junction failures.

At the power pads, typically latchup is evident from a signature of metallization failure of $V_{DD}$ and $V_{SS}$ power rails. Power rail failure of the metallurgy can also occur from HBM (human body model), MM (machine model) and CDM (charged device model) events; in the case of latchup, there is most likely evidence of logic circuitry failure.

**Table 1.3** CMOS latchup failure mechanisms.

| | Structure | Failure location | Failure mechanism | Signature |
|---|---|---|---|---|
| Input pad | CMOS inverter | (a) p-channel MOSFET<br>(b) n-channel MOSFET | | (a) Contact failure<br>(b) Silicide failure<br>(c) Metallurgical junction failure |
| Power pad | | Aluminum interconnects | (a) Aluminum | (a) Molten aluminum<br>(b) Insulator cracking |
| ESD networks | Dual diode | (a) n-well diode<br>(b) Inverter circuit | (a) Negative current injection | (a) Molten silicon<br>(b) Contact damage<br>(c) silicide damage |
| | pnpn | (a) pnpn | (a) Triggering of the ESD pnpn device | (a) Metallurgical junction<br>(b) LOCOS or STI failure |
| | *RC*-triggered power clamp | (a) Interaction between *RC* trigger and the p-channel output device | (a) Parasitic pnpn formed between n-resistor of *RC* trigger network and p-channel ESD clamp element | (a) n-channel resistor diffusion and contact failure |
| | Field oxide device (FOD) | (a) Interaction between a dual diode ESD device and a FOD power clamp | (a) Silicon damage | (a) p+/n-well diode<br>(b) FOD power clamp |
| Chip level | | Power bus | (a) Interconnect failure | (a) Cracked insulators<br>(b) Molten metal |
| Chip and system level | | Wire bond | (a) Wire-bond failure | (a) Wire-bond displacement<br>(b) Bond-pad failure |
| Chip and system level | | Packaging | (a) Plastic explosion<br>(b) Delaminated package<br>(c) Molten package | (a) Package discoloration<br>(b) Package displacement<br>(c) Package material separation<br>(d) Smoke |
| Single event latchup | Analog and digital circuits | Analog and digital functional blocks | Metallurgy failure of aluminum interconnects and insulators | (a) Molten aluminum<br>(b) Insulator cracking<br>(c) Spherical extrusions |

ESD networks can also be a participant in the initiation of latchup. Thyristors, also known as silicon-controlled rectifiers (SCRs), are intentionally utilized for ESD protection networks. One of the issues with the ESD SCR network is that utilization of these networks can lead to CMOS latchup on input pins or on power supply rails. When an ESD SCR network is used on a power rail, overvoltage conditions can lead to 'false triggering' from $CdV/dt$ transient states of the ESD network. This can lead to CMOS latchup. The failure mechanism typically observed is silicon failure of the ESD SCR structure.

The second latchup mechanism observed is in ESD power clamp networks that utilize *RC*-triggered MOSFET networks. Ker demonstrated that latchup can occur in *RC*-triggered p-channel MOSFET power clamps between the p-channel MOSFET and an n-type resistor (used in the *RC* discriminator circuit).

On a global level, CMOS latchup failure mechanisms can be observed in the chip power bus. At high current, the metal power bus can be observed to be either melted or displaced. Insulator cracking of the interlevel dielectric material is also evident lateral to the power bus interconnects.

On a package level, CMOS latchup can lead to failure of wire bonds. Wire bonds that have undergone high current stress can be displaced after a CMOS latchup event. This is evident from pre- and postlatchup stress. In addition, bond-pad integrity may also show evidence of the CMOS latchup failure.

On the physical package level, plastic package has demonstrated evidence from plastic delamination, material discoloration, to plastic package separation and displacement from CMOS latchup events. In the process of the package failure, a signature of latchup is molten plastic, burning smell and smoke. Historically, evidence of plastic separation and decapsulation has occurred in early CMOS development. In the case of package failure, it is hard to distinguish CMOS latchup from overvoltage events without an electrical signature and temporal history.

For SEL accelerated testing, it was found that latchup failures do not occur on the peripheral I/O circuits but occur internal to the chip. In analog–digital converter semiconductor chips, accelerated testing shows that the SEL failure leads to latent mechanisms in the interconnects. The failure mechanism was similar to the failures observed in ESD testing, with a few unique features. Becker *et al.* noted that SEL failure occurred in the digital and the analog functional circuit blocks in the interior but not in the peripheral circuits. A spherical metal extrusion was also noted significantly different from that observed in ESD metal failure mechanisms [154].

Table 1.4 shows a CMOS failure mechanism in complementary bipolar CMOS (C-BiCMOS) technology. In complementary BiCMOS technology, the pnp transistor can interact with an npn transistor leading to latchup. In CMOS technology, latchup typically occurs from parasitic transistors of both the p-channel and n-channel MOSFETs; in this technology, a vertical pnp transistor can interact with a parasitic lateral npn element between the pnp region and a subcollector of an npn bipolar transistor.

Smart power technology today consists of HVCMOS integrated with LVCMOS. Application voltages vary with the product. In the 80–120 V application range, products consists of automotive, telecommunication and power supplies. In the 20–80 V application range, products consists of automotive displays, inkjet printers, dc/dc converters, to controllers. Below 20 V application voltages,

**Table 1.4** Bipolar technology latchup failure mechanisms.

| Bipolar | Structure | Failure location | Failure mechanism | Signature |
|---|---|---|---|---|
| Complementary BiCMOS technology | BiCMOS inverter | Bipolar pnp and bipolar npn | Parasitic pnpn formed between the bipolar pnp and the bipolar npn transistor | Silicon failure between the pnp and the npn transistor |

**Table 1.5** Smart power latchup failure mechanisms.

| Smart power | Structure | Failure location | Failure mechanism | Signature |
|---|---|---|---|---|
| | LDMOS and CMOS logic | Low-voltage CMOS logic | External electron injection from LDMOS triggering low-voltage CMOS circuitry | Silicon failure between the PFET and NFET of CMOS logic circuit |
| | LDMOS to LDMOS | LDMOS | Lateral npn bipolar transistor | Silicon failure |

**Table 1.6** Multichip power system latchup failure mechanisms.

| Power systems | Structure | Failure location | Failure mechanism | Signature |
|---|---|---|---|---|
| Chip and system level | IGBT | Wire-bond packaging | (a) Wire-bond failure | (a) Wire-bond displacement<br>(b) IGBT failure |
| Chip and system level | IGBT | Wire-bond IGBT packaging | (a)Wire-bond inductive differences<br>(b) Nonuniform current conduction | (a) IGBT failure |

products consist of controllers, cell phones, to standard low-voltage electronics. SOC integration can include a mixture of high-voltage CMOS, digital and analog circuitry; these systems can include voltage regulators, interface networks, digital-to-analog converters, analog-to-digital converters, current references, oscillators, current references, filters, to power management supervisors. Crosstalk, noise and latchup are common concerns in these products. Table 1.5 is an example of a smart power latchup event. During inductive load dumps, the LDMOS transistor can discharge electrons into the substrate of an SOC product. Electron injection into the substrate can lead to disruption and triggering of the low-voltage CMOS logic circuitry.

In high-power systems that address power levels in the 1–10 kV range and 1–10 kA range, small variations can lead to nonuniform distribution of current and power. Table 1.6 shows examples of CMOS latchup in power systems. From overcurrent or mechanical stress, bond-wire failure can occur. With a bond-wire failure, a nonuniform current can occur in parallel devices. The nonuniform current can lead to nonideal thermal distribution, leading to both IGBT failure and system thermal destruction. The second issue is regarding nonuniform inductances. In a system, given the inductive lead frames are nonidentical, a differential voltage can establish during a current transient. In that case, the differential voltage can lead to CMOS latchup across the two components, or nonuniform current distribution.

## 1.9 CMOS LATCHUP EVENTS

CMOS latchup events can occur from different types of electrical and current responses on both signal pins and power pins. In the following sections, these different conditions and configurations will be highlighted. These different conditions are important because they influence the test procedure and the understanding of how a semiconductor chip can initiate a latchup condition.

### 1.9.1 Power-Up Sequence Initiated Latchup

Latchup can be initiated from power-up sequence conditions. Missequencing of the power supplies, ground rails and input signals can lead to latchup.
This can occur from the following conditions:

- power-up sequencing of the power rail;
- power-up sequencing of the ground rail;
- power-up sequencing of the input signal.

In addition, these conditions can occur due to undershoot and overshoot phenomena on the power rails and the signal pins.

### 1.9.2 Input Pin Overshoot and Power-Up Sequence Initiated Latchup

Latchup can be initiated at a signal pad from an applied pulse, missequencing or during a transient. In the case where a positive pulse is applied to a signal pad, p-type devices can forward bias and initiate latchup. Figure 1.6 shows a circuit schematic highlighting a positive event on a signal pad. In this figure, the power supply $V_{DD}$ and the ground power supply rail $V_{SS}$ are at a fixed voltage. The input

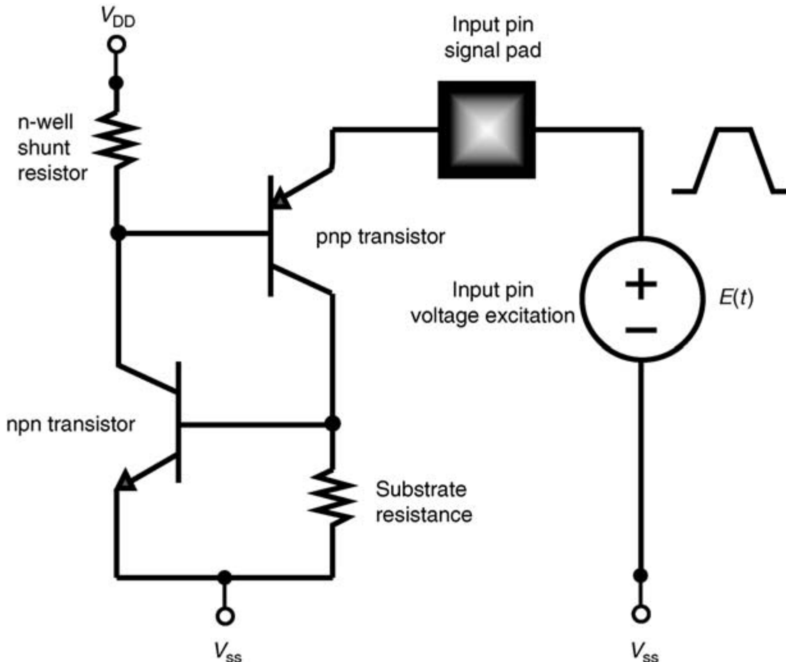



**Figure 1.6** Input signal pin CMOS latchup.

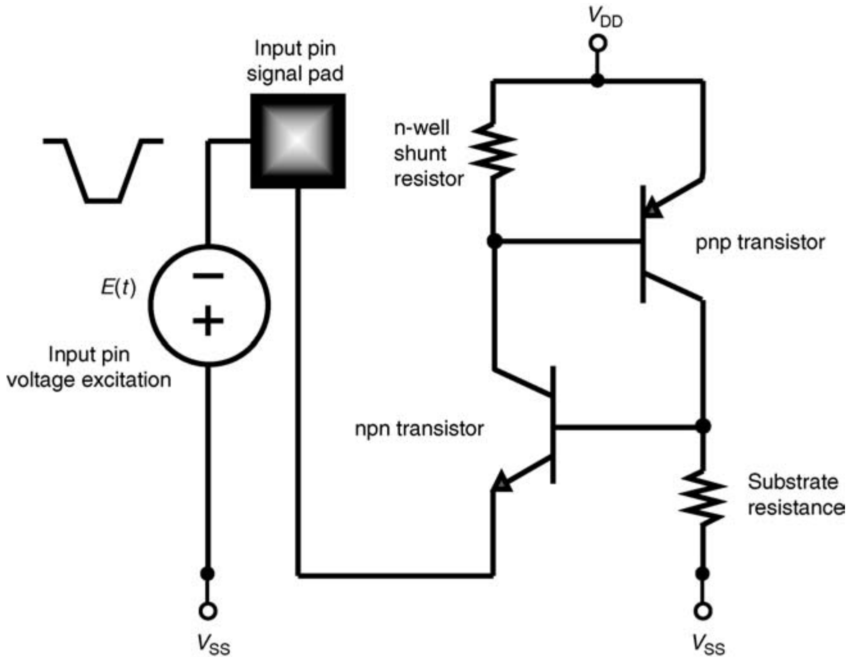



**Figure 1.7** Input signal pin undershoot initiated CMOS latchup.

node is represented as a p+ diffusion that undergoes a positive pulse event. In this event, the p+ diffusion forward biases leading to the parasitic pnp transistor become forward active. Note that this condition is similar to the condition that the signal pad rises faster than the $V_{DD}$ power supply.

### 1.9.3 Input Pin Undershoot and Power-Up Sequence Initiated Latchup

Latchup can be initiated at a signal pad from an applied pulse, missequencing or during a transient. In the case where a negative pulse is applied to a signal pad, n-type devices can forward bias and initiate latchup. Figure 1.7 shows a circuit schematic highlighting a negative pulse event on a signal pad. In this figure, the power supply $V_{DD}$ and the ground power supply rail $V_{SS}$ are at a fixed voltage. The input node is represented as an n+ diffusion that undergoes a negative pulse event. In this event, the n+ diffusion forward biases leading to the parasitic npn transistor become forward active. Note that this condition is similar to the condition that the signal pad rises faster than the $V_{SS}$ power supply.

### 1.9.4 Multiple Power Supply Power-Up Sequence Initiated Latchup

Latchup can occur by the missequencing in a multiple power supply system or a multiple power supply chip. In systems, there exist multiple power supplies that require sequencing to power up or power

down a system. With improper sequencing, networks can forward bias leading to CMOS latchup. In mixed voltage interface semiconductor chips, there can be chip 'sectors' that are at different power supply voltages. For example, peripheral circuits can be placed on a power supply voltage above the native power supply voltage of a semiconductor technology. In this case, the peripheral power rail $V_{DD}$ (I/O) will be separate from the core $V_{DD}$ power supply rail. In the power-up condition (and power-down condition) of a multiple power supply chip, given the power supplies do not rise at the same rate (e.g. which occurs due to the size of the different capacitance loading), semiconductor devices can forward bias leading to latchup. In addition, ESD elements may also exist between the power supply rails that can also forward bias and lead to CMOS latchup.

## 1.9.5 Power Supply Overshoot Initiated Latchup

Latchup can occur when a pulse occurs on the power supply rail. During this event, a positive polarity pulse occurs on the $V_{DD}$ power supply rail. Latchup can be initiated by an applied pulse, missequencing or during a transient. In this case, a positive pulse is applied to the $V_{DD}$ power rail, as well as the n-well and p-type devices. Figure 1.8 shows a circuit schematic highlighting a positive event on a power supply pad. In this figure, the ground power supply rail $V_{SS}$ is at a fixed voltage. Latchup occurs if the p-type device becomes forward active. In this case, it is a function of the n-well shunt resistance.

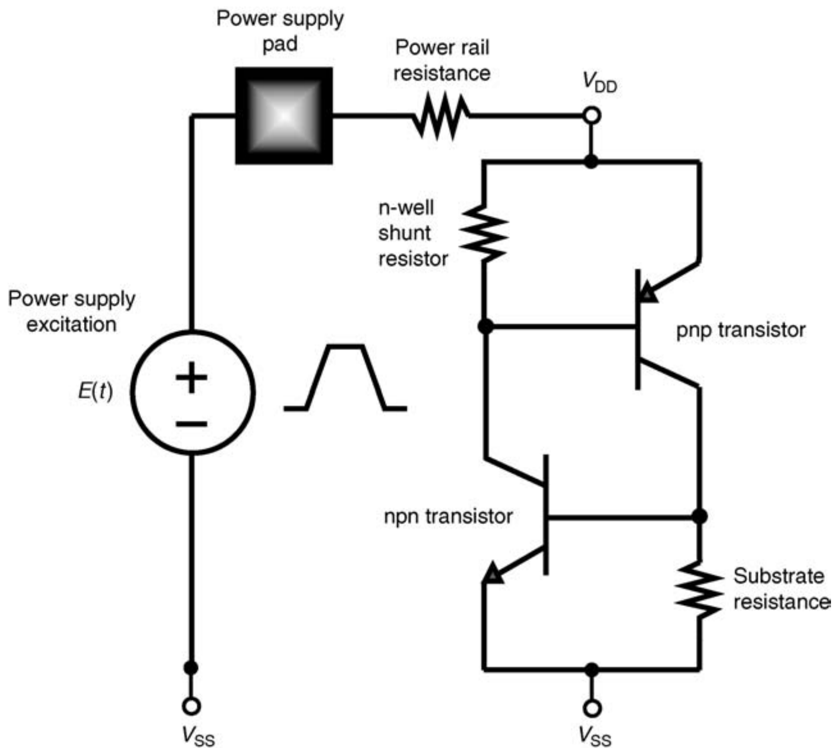



**Figure 1.8** Power supply overshoot initiated CMOS latchup.

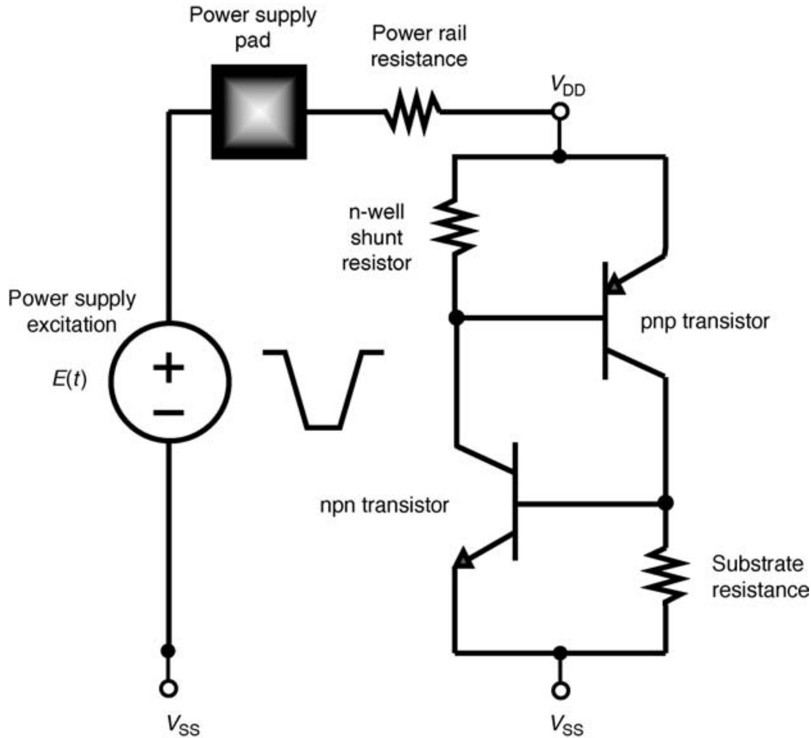



**Figure 1.9** Power supply undershoot initiated CMOS latchup.

## 1.9.6 Power Supply Undershoot Initiated Latchup

Latchup can occur when a negative pulse occurs on the power supply rail. During this event, a negative polarity pulse occurs on the $V_{DD}$ power supply rail. Latchup can be initiated by an applied pulse or an oscillating signal. In the case where a negative pulse is applied to the $V_{DD}$ power rail, when the $V_{DD}$ power rail goes below the $V_{SS}$ power supply rail, the n-well forward biases leading to injection into the substrate. Figure 1.9 shows a circuit schematic highlighting a negative event on a power supply pad. In this figure, the ground power supply rail $V_{SS}$ is at a fixed voltage. Latchup occurs if the negative signal is followed by a positive rising signal leading to the p-type device become forward active.

## 1.9.7 Power Supply (Ground Rail) Undershoot Initiated Latchup

Latchup can occur when a pulse occurs on the ground power supply rail. During this event, a negative polarity pulse occurs on the $V_{SS}$ power supply rail. Latchup can be initiated by an applied pulse, missequencing or during a transient. In this case, a negative pulse is applied to the $V_{SS}$ power rail, as well as the p− substrate and n-type devices. Figure 1.10 shows a circuit schematic highlighting a negative event. In this figure, the power supply rail $V_{DD}$ is at a fixed voltage. Latchup occurs if the n-type device becomes forward active. In this case, it is a function of the p− substrate shunt resistance.

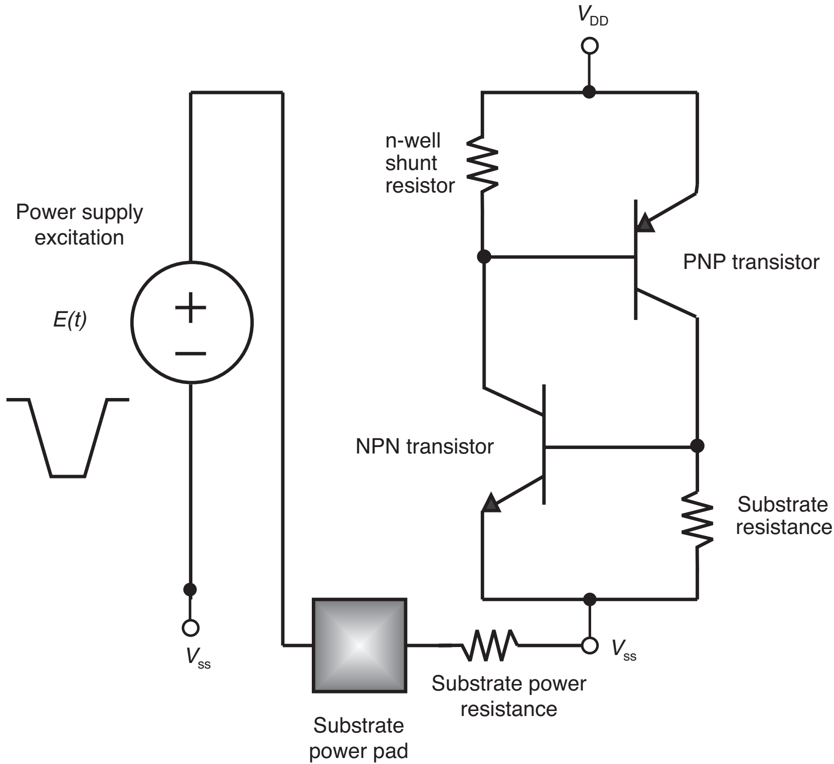



**Figure 1.10** Ground power supply undershoot initiated CMOS latchup.

## 1.10 ELECTROSTATIC DISCHARGE SOURCES

ESD events and latchup have interrelations in the semiconductor physics, circuits, and events. First, latchup and ESD can occur on both signal pins and power pins. Second, CMOS latchup and ESD concerns can occur at the signal pin circuits. Third, the ESD and CMOS latchup damage in the electronics can occur at the same location, or different. Finally, ESD events can initiate CMOS latchup when a chip is powered. For example, ESD events on semiconductor chips can consist of HBM, MM and CDM. On a board, card or system level, other system-level events can also occur. In the following section, the commonality and distinction of some of these events will be briefly discussed.

### 1.10.1 Human Body Model ESD Event

HBM events are high-current events that can initiate latchup. These events typically are single polarity events. HBM events affect semiconductor chips by inducing a impulse on signal pins and power pads. HBM events are represented as a 100-pF charged capacitor and a 1500-Ω resistor [111]. In the discharge process, the HBM pulse has a 17–22-ns rise time and a 150-ns decay time. HBM events can range from low voltage to 35 000 V. The magnitude of the event is a function of the triboelectric

charging, the charging process and the humidity. There are a few reasons why HBM events do not cause CMOS latchup:

- HBM events occur when the semiconductor chip is unpowered.
- HBM events' total charge is small (small capacitance and short time event).

But what if the chip was powered when an HBM was applied?

How does one distinguish damage from an HBM event from CMOS latchup?

- Damage patterns of HBM events are restricted typically to peripheral circuits, ESD networks and interconnect wiring (input wires and bussing).
- HBM time response is a faster event; hence, the material motion (e.g. electromigration) is typically not observed.
- Damage patterns of latchup events can involve semiconductor devices, circuits, bussing and packaging. Typical HBM events do not demonstrate package damage.

### 1.10.2 Machine Model ESD Event

MM events are high-current events that can initiate latchup. Machine models are oscillatory polarity events that are more likely to induce latchup events [112]. MM events are represented as a 200-pF charged capacitor and no resistor. In the discharge process, the rise time is faster than that for an HBM event. There are a few reasons why MM events do not typically cause CMOS latchup:

- MM events occur when the semiconductor chip is unpowered.

Although MM events are not typically observed, they do induce an oscillating transient that can lead to latchup given a system was powered. When the signal pad or power pad undergoes a negative transient, electrons can be discharged into the substrate potentially initiating the npn parasitic devices. With the positive polarity transition, the pnp transistor is also initiated. The primary reason that this is not a concern is these events occur during an unpowered chip state. But, what if the chip or system is powered? How does one distinguish damage from an MM event from latchup?

- Damage patterns of MM events are restricted typically to peripheral circuits, ESD networks and interconnect wiring (input wires and bussing).
- MM events are faster events, and hence do not involve significant material motion (e.g. electromigration).
- Damage patterns of latchup events can involve semiconductor devices, circuits, bussing and packaging. Typical MM events do not demonstrate package damage.

### 1.10.3 Cable Discharge Event Source

CDE is a growing source of latchup events that occur on a card or system level [113–127]. Electrical discharge from charged cables has been a concern in electrical systems. An electrical cable behaves as

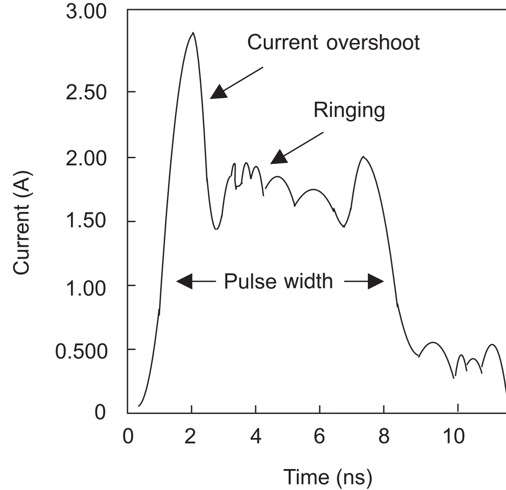



**Figure 1.11** CDE waveform.

a transmission line having both capacitive and inductive characteristics. The charge stored in an electrical cable can be evaluated according to the stored capacitance per unit length. As the length of the cable increases, the amount of stored charge in the cable increases. In an 'unsocketed' cable, the charge is stored between the center and outer conductors across the insulator region. When the voltage across the insulator exceeds the insulator breakdown voltage, an electrical discharge occurs between the outer and inner conductors. In the case where the voltage remains below the breakdown voltage, the stored charge remains in the cable, leaving the cable as a charged cable. When the charged cable center conductor approaches a system, an electrical arc occurs between the center conductor and the electrical system input signals. With the cable discharge event to the system-level inputs, a significant current can discharge to the signal pin leading to latchup in the semiconductor chips integrated into the system components. Given a system that is unpowered, no latchup event can occur. But if the CDE occurs while a system is powered, the current event can lead to CMOS latchup. With the integration of wide area network (WAN) and local area network (LAN), the Ethernet is playing a larger role. When a charged twisted pair cable connects to an Ethernet port with a lower electrical potential, CDE can occur in LAN systems. Figure 1.11 shows an example of a typical CDE waveform. In the past, standards (e.g. IEEE 802.3 Section 14.7.2) have noted the potential for CDE processes in LAN cables. As the number of electronic circuits increases, the number of I/O ports also increases (e.g. Rent's Rule). As the number of I/O ports increases, there is an increase in the number of electrical cable connections. As a result, future systems have an increasing number of electrical cable connections. This increases the likelihood of a charged cable connection leading to CMOS latchup. Hence, with both system and technology evolution, the reasons for the increased concern for this issue are the following:

- WAN and LAN integration;
- category 5 and 6 LAN cabling;
- higher level incidents of disconnection;
- high level incidents of reconnections;

- CMOS technology migration to p− substrates;
- CMOS technology scaling of p+/n+ scaling;
- CMOS technology scaling of shallow trench isolation;
- competitive focus on cost reduction and the elimination of future latchup solutions.

The primary reasons that CDEs are a latchup concern are as follows:

- The systems that are experiencing cable discharge events are powered systems where the cable is inserted.
- The total charge and current magnitude from many cable discharge events exceed the CMOS latchup design specification (e.g. JEDEC specification).
- The CDEs are occurring on high data rate transfer ports that have interface circuits consisting of advanced CMOS technology with low CMOS latchup robustness.

## 1.11 SINGLE EVENT LATCHUP

SEL is an event where a single-particle interaction with a semiconductor chip or system can lead to latchup and destruction of the data transfer, semiconductor chip or system failure [128–171]. SEL can be initiated by a wide range of particle types and events. SEL can be initiated by the following set of particles:

- high-energy photons;
- alpha particles;
- cosmic rays;
- heavy ions.

The primary reasons why CMOS latchup is a concern in space applications are the following:

- wide spectrum of particle types (e.g. neutrons, protons and muons);
- wide energy spectrum and environments (e.g. space and terrestrial);
- single event source does not scale with technology generation (e.g. nonscalable source);
- destructive nature of CMOS latchup events (e.g. device damage and system-level damage);
- inability to repair or introduce redundant functions in space applications;
- cost of the space application.

### 1.11.1 High-Energy Photon Emissions

High-energy photons can lead to photocurrent generation in a metallurgical junction. The two processes of the current collection at a metallurgical junction are associated with photogeneration

within the metallurgical junction and the photogenerated carriers that diffuse to the metallurgical junction. These carriers can lead to latchup given the photogeneration current is significant enough. Additionally, by linear superposition, the photon emissions can contribute photocurrent in addition to already existing currents. A time-dependent volumetric generation rate $g(t)$ is defined as follows:

$$g(t) = g_0 \gamma(t),$$

where $g_0$ is the volumetric generation rate per 'rad' per unit volume, and $\gamma(t)$ is the dose rate. The volumetric generation rate per unit volume can be expressed as the number of electron–hole pairs generated per rad per unit volume. The $\gamma(t)$ dose rate can be expressed as rad/s. (Note: 1 rad equals 100 ergs of energy within 1 g of material.) For silicon this can be expressed as [128]

$$g_0 = 4 \times 10^{13} \text{EHP}/(\text{rad cm}^3).$$

## 1.11.2 Alpha Particle Ionizing Source

Alpha particle emissions occur due to the radioactive decay of atomic species [130]. Alpha particle emissions are a result of the natural radioactive decay process of elements. There are three major series, known as the uranium series, the actinium series and the thorium series. In these processes, a successive cascade of alpha emissions occur in sequence from one isotope to another, until radioactive stability occurs in the element (Figures 1.12 and 1.13). In the uranium series (also known as the '$4n + 2$' series), the nuclide ${}_{92}\text{U}^{238}$ decays to ${}_{90}\text{Th}^{234}$ through an alpha particle emission, as well as other beta emission processes; this process produces a 4.20-MeV alpha emission. The second alpha

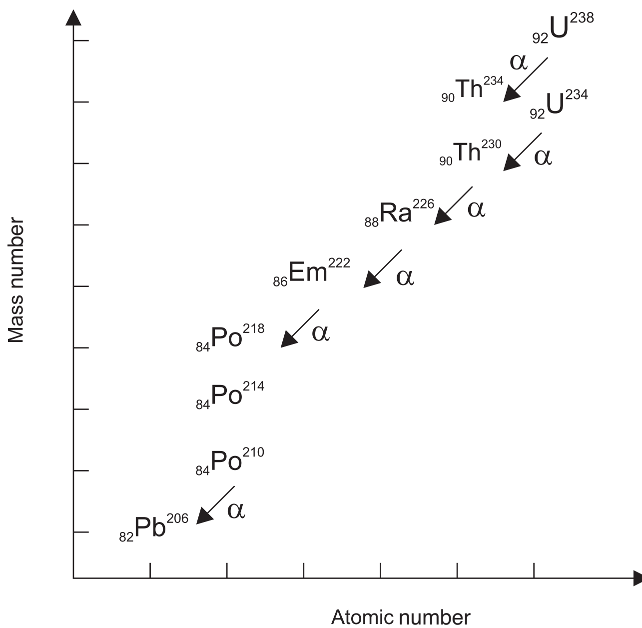



**Figure 1.12** Uranium nuclide series.

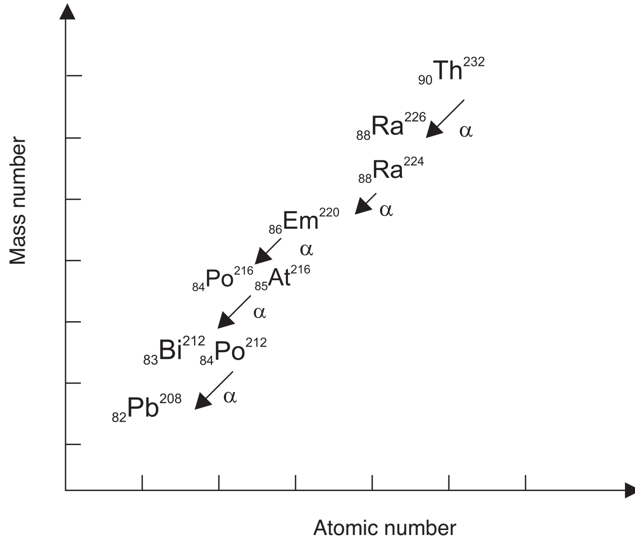
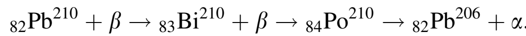



**Figure 1.13** Thorium nuclide series.

emission in the uranium series occurs from ${}_{92}U^{234}$ to ${}_{90}Th^{230}$ with an energy of 4.768 MeV; this is followed by another cascade from ${}_{90}Th^{230}$ to ${}_{88}Ra^{226}$, then ${}_{86}Em^{222}$, then ${}_{84}Po^{218}$ and alternative paths to ${}_{84}Po^{214}$ and ${}_{84}Po^{210}$. This uranium series ends at a stable nuclide of ${}_{82}Pb^{206}$. As a result, a spectrum of discrete alpha particle energies exists with energy levels of 4.20, 4.768, 4.68, 4.77, 5.3 5.486, 5.998 and 7.683 MeV. Hence, the largest energy of an alpha particle of the uranium series is 7.683 MeV [130]. In the thorium series (also known as the 4*n* series), the decay process begins with the decay of nuclide ${}_{90}Th^{232}$ to a mesothorium nuclide ${}_{88}Ra^{228}$ with an alpha particle emission energy of 4.007 MeV. The second alpha emission occurs from ${}_{90}Th^{232}$ to ${}_{88}Ra^{224}$ with a 5.423-MeV emission, followed by a second alpha emission of 6.280 MeV from ${}_{86}Em^{220}$. Additional alpha emission processes continue with the alpha particle emissions from ${}_{84}Po^{216}$, ${}_{85}At^{216}$, ${}_{83}Bi^{212}$ and ${}_{84}Po^{212}$ of 6.774, 7.79, 6.086 and 8.78 MeV, respectively. The thorium series becomes stable at the end of the alpha particle decay process with a final stable nuclide of ${}_{82}Pb^{208}$. In the thorium series, the highest energy alpha particle emission is the 8.78-MeV emission [130].

Lead, used in solder balls, also undergoes natural radioactive decay processes; this lead decay process is a source of alpha particle emissions. Lead isotopes undergo a beta decay process from the isotope ${}_{82}Pb^{210}$ to ${}_{83}Bi^{210}$, to ${}_{84}Po^{210}$. The isotope ${}_{84}Po^{210}$ then decays to ${}_{82}Pb^{206}$ with a 5.3-MeV alpha particle emission [130]:



$$ {}_{82}Pb^{210} + \beta \rightarrow {}_{83}Bi^{210} + \beta \rightarrow {}_{84}Po^{210} \rightarrow {}_{82}Pb^{206} + \alpha. $$

Lead has a high variation of alpha particle fluence rates, which is dependent on the source of the lead materials. Alpha particle flux rate for deposited aluminum can be on the order of 10–30 $\alpha/(\mathrm{mm}^2\,\mathrm{kh})$. All materials contain some small percentage (e.g. ppm) of atoms that generate alpha emissions. Ceramic substrates, used for flip-chip package technology, can have a high alpha particle flux rate. The alpha particle emission occurs in metallurgy and materials used in semiconductor components, such as aluminum wiring. For example, alpha particle flux rate for deposited aluminum can be on the order of

**Table 1.7** Typical alpha particle sources in semiconductor materials.

| Source | Emission type | Flux rate ($\alpha/(\text{cm}^2\,\text{h})$) |
|---|---|---|
| Solder balls | Pb/Sn | 1 |
| Ceramic substrates | Uranium/thorium | 0.1 |
| Plastic package | Uranium/thorium | 0.04 |
| Silicon wafer | Uranium/thorium | Less than 0.004 |

0.02 $\alpha/(\text{mm}^2\,\text{kh})$. Table 1.7 contains a short list of alpha particle sources, emission spectrum type and flux rates.

Figure 1.14 shows an alpha particle penetration into a silicon chip. As the α-particle penetrates through the silicon region, EHPs are generated along the entire track length [131–140]. As the particle energy decreases, the generation rate of EHPs increases. The EHPs diffuse from the initial track, recombine or are collected by the metallurgical junctions. EHPs are generated in all physical regions (e.g. n+ diffusion, p+ diffusion, n-well, p-well and p− substrate). Figure 1.15 shows an α-particle event through a retrograde n-well and a p++ substrate. Figure 1.16 shows the electrical potential of the α-particle event through a retrograde n-well and p++ substrate technology. As the EHPs are generated, they influence the electric potential in the region of the particle. As the particle penetrates the n-well and p− substrate, the metallurgical junctions and the bulk regions are influenced. When the particle extends through the metallurgical junctions, the depletion regions extend into the n-well and substrate regions. EHPs generated within this extended depletion region (e.g. also known as a 'funnel') are swept into the junctions by the electric field. As the metallurgical junctions and the electric field relax, the other minority carriers diffuse to the junctions. One of the primary reasons that α-particles do not introduce latchup is that in today's circuits the total amount of charge collected is not enough to trigger the latchup event.

## 1.11.3 Cosmic Ray Source

Cosmic rays can lead to single event latchup (SEL). There are two types of cosmic rays: galactic and solar cosmic rays. Solar cosmic rays are cosmic rays associated with our Sun, whereas galactic cosmic

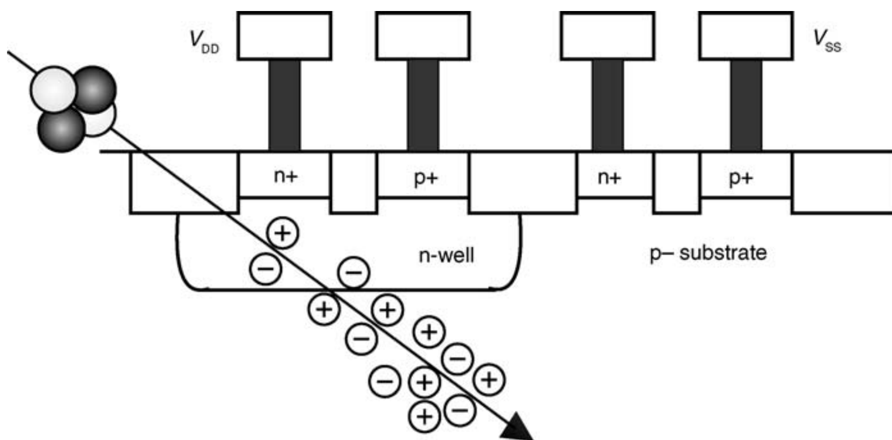



**Figure 1.14** Alpha particle interaction with CMOS latchup structure.

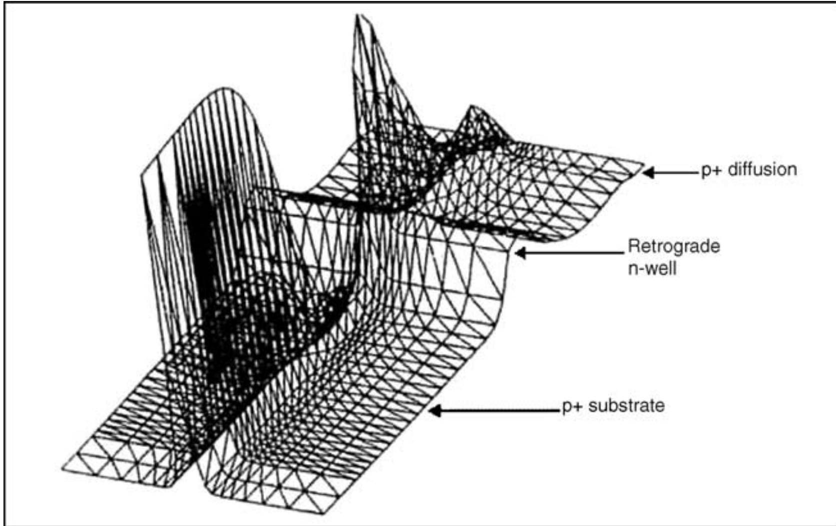



**Figure 1.15** Alpha particle simulation electrostatic potential distribution results through a CMOS structure consisting of a retrograde n-well and p++ substrate.

rays are from a source beyond our solar system. These cosmic rays are classified into three groups – primary cosmic rays, cascade cosmic rays and sea-level cosmic rays. Primary cosmic rays are cosmic rays that exist outside of our Earth's atmosphere, which have not undergone collision processes. The 'cascade cosmic rays' are the cosmic rays that are generated after a primary cosmic ray enters the Earth's atmosphere. When a primary cosmic ray enters the Earth's atmosphere, it undergoes a series of particle collisions leading to a plethora of particle events. Sea-level cosmic rays are the cosmic rays observed on the Earth.

Galactic particles consist of protons, alpha particles and heavy ions. Galactic particles are 92 % protons, 6 % alpha particles and 2 % heavy ions. Heavy ions consist of O, N, C, B, Be, Li and He. The

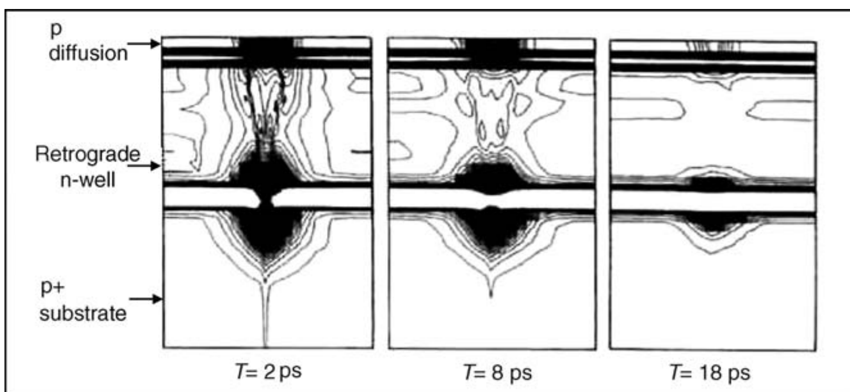



**Figure 1.16** Alpha particle simulation electrostatic potential distribution time sequence through a CMOS structure consisting of a retrograde n-well and p++ substrate.

energy spectrum of the galactic particles extends from 10 to 10 000 MeV. Hence, these particles can induce SEL in satellites and space applications. Very few of these particles can reach the sea level, and hence do not induce SEL in ground-based applications – in today's electronics.

Cascade cosmic rays in the atmosphere are the result of the collisions of the galactic or solar cosmic rays. In the collision of primary incident cosmic rays with the Earth's atmosphere, elementary particles are generated such as pions, muons, neutrons, protons, electrons, positrons and gamma rays. In the case of the charged particles, as they pass through the atmosphere, they lose energy through both electrostatic interaction and nuclear forces. Neutrons and protons have a wide energy distribution and significant fluence rate in space and in the Earth's atmosphere. As protons enter the Earth's atmosphere, the particles lose energy. Hence, the flux density and the energy spectrum change with altitude.

Sea-level cosmic rays consist of elementary particles that are dominated by neutrons and protons. Due to the electric and nuclear force interactions, proton flux rates are significantly lower than neutron flux rates for particle energies below 10–100 MeV. At the very high energy of the spectrum, protons, muons and pions are a significant part of the sea-level cosmic ray population. Neutrons are at high densities at very low energy ranges, sometimes referred to as 'thermal neutrons'. As neutrons enter the silicon, collisions can occur between the neutron and the silicon lattice. Neutrons that collide with the silicon nucleus can lead to fission events and silicon recoil events. In the case of fission events, multiple tracks emanate from the point of the neutron–silicon nucleus. Figure 1.17 shows an example of a fission fragmentation and the multiple tracks formed from the location of the neutron–Si nucleus collision.

A distinction of terrestrial alpha particles and neutron cosmic ray events is that in the range and energy of these particles in the silicon wafer. Alpha particles generated from uranium and thorium series radioactive decay have less than 10-MeV energy levels, where the cosmic ray spectrum is a broad energy spectrum to well over 100–1000 MeV. As a result, the alpha particle events occur only near the surface of the devices (except those generated from the silicon wafer itself). In the case of neutron events, no carriers are generated in the wafer until a collision event occurs with the silicon lattice.

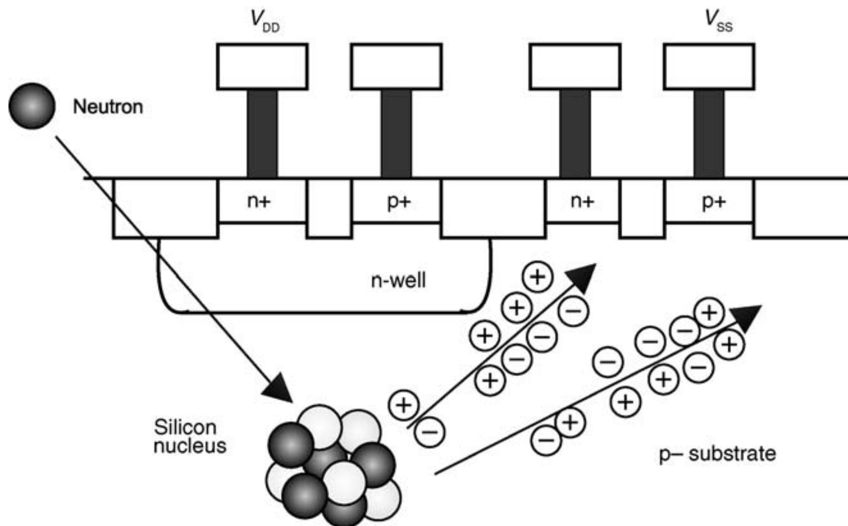



**Figure 1.17** Cosmic ray interaction with CMOS latchup structure.

When the cosmic ray–silicon lattice event occurs, it can occur anywhere in the silicon wafer volume, interlevel dielectrics or back-end-of-line materials. During these collision events, multiple fission fragment tracks can occur. In each collision, the fission fragment type can be different depending on the energy of the event. As a result, the charge generation from cosmic rays can occur at the device surface or deep in the substrate region. When the events occur deep in the lattice or the tracks do not extend toward the device surface, electron–hole pair recombination can lead to the recombination prior to influencing the semiconductor structures. As a result, the wafer doping concentration, thickness and recombination centers can highly influence SEL from cosmic rays for deep events.

For charged fission fragments or other generated particles, the linear energy transfer associated with the generation of the electron–hole pairs will differ. As a result, the range and energy transfer from the particle fragments are significantly different from alpha particle events.

The similarity of these events to alpha particle events is that the generation of electron–hole pairs in the silicon can lead to perturbation of the various nodes of the pnpn structure as the minority carriers are collected at the metallurgical junctions. Hence, the charge transport within the pnpn can be diagnosed similarly.

## 1.11.4 Heavy Ion Source

Heavy ions can be an SEL concern in space applications. First, these heavy ions have a wide variety of potential elements. Heavy ions can consist of O, N, C, B, Li and He. Second, the energy spectrum of these events can be wide leading to the difficulty in preventing SEL latchup through semiconductor process solutions; the energy spectrum of these galactic particles extends from 10 to 10 000 MeV. At these energies, it is difficult to design advanced technology that is immune to such events.

Figure 1.18 shows an example of a heavy ion event. In a heavy ion event, the heavy ion can interact with a silicon lattice inducing silicon nucleus–atom collision, as well as silicon fission events. In an event known as 'silicon nuclear recoil', the energy is transferred to the Si atom through collision. In the silicon nuclear recoil event, the lattice atom is displaced, and electron–hole pairs are generated in the process. In this case, only a single electron–hole pair track is formed. In a silicon fission event, the

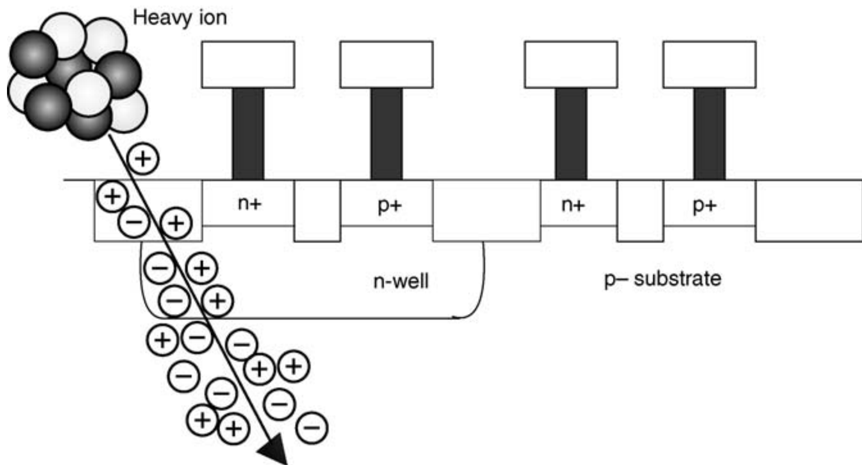



**Figure 1.18** Heavy ion SEL event.

silicon nucleus disintegrates into fission fragments, and each fission fragment introduces a separate track. In the case of Si fission, there are many different cases of fragmentation, and hence a wide variety of events occur in the Si fragment range and energy. As a result, multiple tracks of various lengths and energies occur. Each track generates electron–hole pairs along the track.

Latchup can occur from both tracks contained within the parasitic pnpn (e.g. within the n-well region or triple-well isolated epitaxial region) or from the minority carrier diffusion to the n-well or p-well regions. In these events, it is expected that there is a fast time constant associated with EHP collection in the metallurgical junctions, 'funnel regions' and within the pnpn regions, as well as a slower time constant associated with the time to diffuse to the semiconductor junctions.

SEL from particle event is evaluated from a statistical probability perspective. For the prediction of the single event effect (SEE), calculations evaluate the following:

- distribution of particles;
- critical charge, sensitive area and sensitive volume for the single event effect;
- device cross section.

For the evaluation of the particle distribution, a distribution of flux versus the linear energy transfer is used for rate calculation; the distribution used for galactic cosmic rays is known as the Heinrich curve.

The linear energy transfer is a measure of the energy transferred to the material from the particle. The LET threshold is the minimum LET value that can cause a latchup event. Typically, the definition is for the case of a particle fluence of $10^7$ ions/cm$^2$. Single event effect 'immunity' is when the LET threshold exceeds 100 MeV cm$^2$/mg.

Table 1.8 shows examples of ions that are used in accelerators to simulate single event latchup. Table 1.7 contains the linear energy transfer for a specific ion at a given energy. Note that as the atomic number and energy increase, the LET also increases. As a result, for SEL testing, it is desirable to use ions with high atomic number. In simulation of single event latchup, it is essential to have the effective range of the ions penetrate into the substrate region. As a result, an ion penetration range of above 40 μm is suitable.

A 'latchup capture cross section' can be defined as a function of area. The latchup capture cross section can be plotted as a function of the linear energy transfer. Figure 1.19 shows an example of the relationship of latchup capture cross section and the linear energy transfer as a function of temperature. Johnston showed that as temperature increases, the threshold linear energy transfer decreases and the latchup capture cross section increases [155]. As temperature increases, the pnp and npn bipolar current gains of the parasitic transistors, as well as the well and substrate shunt resistances, increase.

A key distinction between latchup initiation from ESD events (e.g. HBM, MM, CDM and CDE) is that the source of latchup involves the signal pads and peripheral circuits. Simulated testing of mixed signal chips demonstrates that the latchup events occur within the digital and analog core networks, and do not involve the peripheral circuitry.

**Table 1.8** Examples of linear energy transfer for different ions.

| Ion | Energy (MeV) | LET (MeV cm$^2$/mg) |
|---|---|---|
| ${}_{11}B^{+3}$ | 108.2 | 0.89 |
| ${}_{40}Ar^{+11}$ | 400 | 9.88 |
| ${}_{65}Cu^{+18}$ | 659.2 | 21.6 |
| ${}_{136}Xe^{+37}$ | 1330 | 53.7 |

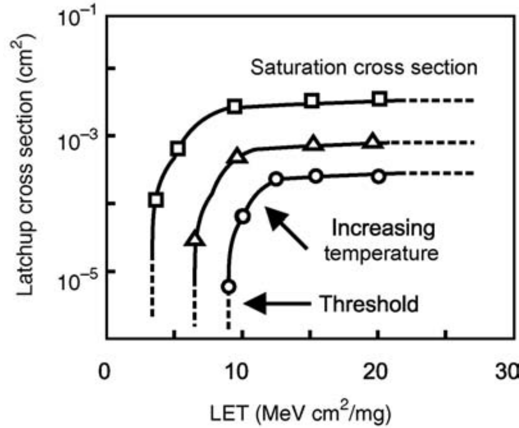



**Figure 1.19** Latchup capture cross section versus linear energy transfer as a function of temperature.

## 1.12 SUMMARY AND CLOSING COMMENTS

This chapter provided an overview of the CMOS latchup discipline. This chapter discussed the history, key innovations, key contributions and patents associated with the process of understanding how to address the latchup issue in semiconductor technology. Both issues associated with ionizing radiation events and current and voltage excursions outside of the native current and voltage conditions of a technology were discussed. The issue of technology scaling and how scaling leads to CMOS latchup concerns should provide insight into future technology trends.

In Chapter 2, we begin from a bottom-up approach from fundamentals of the transistor, and build toward the understanding of CMOS latchup theory in Chapter 3. In Chapter 2, the fundamentals of bipolar transistor models associated with CMOS latchup will be shown, with a focus on recombination physics, transport and the models of vertical and lateral components.

## PROBLEMS

1. Plot the trend of technology scaling dimension as a function of time (year) from 1980 to 2015. Extend the data from 2.0 μm to sub-35 nm. Plot the following semiconductor variables as a function of time:

   a. junction depth;

   b. epitaxial thickness;

   c. n-well depth;

   d. p-well depth.

2. Plot the trend of epitaxial thickness, p-well depth, n-well depth, junction depth and p+/n+ space as a function of the technology generation from 2.0-μm to 35-nm technology. What do you observe?

3. Given a metric of the 'beta product', derive the scaling of the beta product as a function of the scaling variable $\alpha$. Assume an original beta product value of $(\beta_{npn})_0(\beta_{pnp})_0$. Assume MOSFET constant electric field scaling theory. How does it scale?

4. Given that a technology must maintain the 'beta product' equal to unity, how would one provide a 'constant latchup robustness scaling' value as a technology is scaled according to MOSFET constant electric field scaling theory. By constant latchup robustness scaling, the latchup robustness is preserved as a constant as the technology is scaled as scaling parameter $\alpha$. (*Hint*: Invent a second scaling parameter to offset the natural scaling and show the relationship of the new scaling parameter as a function of the MOSFET constant electric field scaling parameter.)

5. Plot the trend of substrate resistance as a function of technology generation from 2.0 μm to 35 nm. Define a relationship of the trend. Assuming that a constant substrate resistance must be maintained between a p+ substrate contact and an n-channel transistor, show the derivation for the spacing between the substrate contact and the transistor.

6. Given a standard industry cable (category 5 or 6 LAN cable), provide an estimate of the range of typical capacitance per unit length values. Plot total capacitance of a cable as a function of the cable length from 1 to 300 ft (in pF). Given these cables are charged through triboelectric charging to 100 V, what is the total stored charge in the cable for 1, 10, 20, 50, 100, 200 and 300 ft lengths.

7. Given a charged cable, the width of a pulse from the cable is proportional to the length of the cable. Calculate the pulse width from a category 5 or 6 cable for cables in the range of 1–300 ft.

8. Given a JEDEC latchup specification of 100 mA current, derive the relationship of a cable that provides a pulse of this magnitude as a function of capacitance per unit length, cable length and predischarge voltage.

9. The range of an alpha particle in silicon can be calculated as a function of the particle:

$$R = 1.53E^{1.67},$$

where $E$ is the energy in MeV and $R$ is the range in microns. Plot the range as a function of the energy from 1 to 10 MeV. Given a uranium series, the maximum energy is 8.8 MeV. Calculate the range of the highest energy alpha particle. Given all the charge along the range is collected at a given junction (e.g. an n-well-to-substrate junction), what is the number of electron–hole pairs generated and what is the total charge?

10. The range of an alpha particle in silicon can be calculated as a function of the particle:

$$R = 1.53E^{1.67},$$

where $E$ is the energy in MeV and $R$ is the range in microns. Plot the range as a function of the energy from 1 to 10 MeV. Given a lead solder ball, the maximum energy is 5.5 MeV. Calculate the maximum range of the alpha particle emission. What is the range of energy emitted? Why? Given all the charge along the range is collected at a given junction (e.g. an n-well-to-substrate junction), what is the number of electron–hole pairs generated and what is the total charge?

11. Given an ionizing radiation source generates electron–hole pairs in the base region of a bipolar transistor. Explain the transport and the collection processes of both the electron and the hole.

12. Show the difference in the interaction from alpha particles, cosmic ray events and heavy ion events. Where does the interaction occur? How does the time response and spatial dependencies differ? How does this affect the time response of the drift and diffusion processes? Which processes have 'funneling' and why?

## REFERENCES


1. J.L. Moll, M. Tannenbaum, J.M. Goldey and N. Holonyak. P-N-P-N transistor switches. *Proceedings of the IRE*, **44**, 1956, 1174–1182.
2. A.K. Jonscher. P-N-P-N switching diodes. *Journal of Electronic and Control*, **3**, 1957, 573–586.
3. F.E. Gentry, F.W. Gutzwiller, N. Holonyak and E.E. von Zastrow. *Semiconductor Controlled Rectifiers: Principles and Applications of p-n-p-n Devices*. Englewood Cliffs, NJ: Prentice-Hall; 1964.
4. H.A. Schafft. Second breakdown – a comprehensive review. *Proceeding of the IEEE*, **55**, 1967, 1272–1288.
5. J.E. Lilienfeld. US Patent No. 1,745,175 (1930).
6. O. Heil. UK Patent No. 439,457.
7. K. Kinoshita, C.T. Kleiner and E.D. Johnson. Radiation induced regeneration through the p–n junction isolation monolithic IC's. *IEEE Transactions on Nuclear Science*, **12**, 1965, 83–90.
8. R.A. Poll and J.F. Leavy. *Study of transient radiation induced latchup*. General Atomic Division Final Report (Contract N0014-66-C-0347), No. GA-7969, May 1967.
9. J.F. Leavy and R.A. Poll. Radiation-induced integrated circuit latchup. *IEEE Transactions on Nuclear Science*, **16**, 1969, 96–103.
10. W.J. Dennehy, A.G. Holmes-Seidle and W.F. Leopold. Transient radiation response of complementary-symmetry MOS integrated circuits. *IEEE Transactions on Nuclear Science*, **16**, 1969, 114–119.
11. S.S. Eaton. *Noise immunity of COS/MOS integrated-circuit logic gates*. RCA Application Note ICA-6176, Harrison, NJ, 1970.
12. B.L. Gregory and B.D. Shafer. Latchup in CMOS integrated circuits. *IEEE Transactions on Nuclear Science*, **20**, 1973, 293–299.
13. B.L. Gregory. *Latchup studies in bulk silicon CMOS integrated circuits*. Sandia Laboratory Report No. 73-3003, Albuquerque, NM, January 20, 1973.
14. B.D. Shafer. *CMOS latchup analysis and prevention*. Sandia Laboratories Report SAND75-0371, Albuquerque, NM, June 1975.
15. L.J. Gallace and H.L. Pujol. *Reliability considerations for COS/MOS devices*. RCA Technical Note ST-6418, RCA Corporation, Somerville, NJ, July 1975.
16. C.E. Barnes, W.R. Dawes Jr., G.F. Debenwick, B.L. Gregory, C.B. Norris Jr., H.H. Sander and B.D. Shafer. *Latchup prevention in CMOS*. Sandia Laboratories Report SAND76-0048, Albuquerque, NM, March 1976.
17. D.R. Alexander, R.M. Turfler and L.D. Ray. *MOS modeling: Part II*. BDM Corp. Report BDM/A-207-75-TR-R2, Albuquerque, NM, September 30, 1976.
18. G.J. Brucker. Characteristics of CMOS/bulk and SOS memories in a transient environment. *IEEE Transactions on Nuclear Science*, **24**, 1977, 2209–2212.
19. F.N. Coppage and D.C. Evans. Characteristics of destruction from latchup in CMOS. *IEEE Transactions on Nuclear Science*, **24**, 1977, 2226–2229.
20. A. London and R.C. Wang. Dose rate and extended total dose characterization of radiation hardened metal gate CMOS integrated circuits. *IEEE Transactions on Nuclear Science*, **25**, 1978, 1172–1175.
21. M. Simons. *Radiation effects in MOS integrated circuits*. Final Technical Report RTI/0991/00-03F, Research Triangle Institute, Research Triangle Park, NC, February 1978.
22. L.W. Ricketts. *Fundamentals of Nuclear Hardening of Electronic Equipment*. New York: Wiley-Interscience; 1972.
23. K. Stephenson. *CMOS: some application guidelines*. Harris Semiconductor Application Note 209, Harris Semiconductor, Melbourne, FL, March 1975.
24. L.L. Sivo, F. Rosen and L.C. Jeffers. Latchup screening of LSI devices. *IEEE Transactions on Nuclear Science*, **25**, 1978, 1534–1537.

25. J.L. Crowley, F.A. Junga and T.J. Stultz. Technique for selection of transient radiation-hard junction-isolated integrated circuits. *IEEE Transactions on Nuclear Science*, **23**, 1976, 1703–1708.
26. T.J. Stultz, F.A. Junga and J.L. Crowley. *Radiation-induced latchup of junction isolated integrated circuits*. Lockheed Palo Alto Research Laboratory Report LMSC-D497933, Palo Alto, CA, March 20, 1976.
27. J.R. Adams and R.J. Sokel. Neutron irradiation for prevention of latchup in CMOS integrated circuits. *IEEE Transactions on Nuclear Science*, **26**, 1979, 5069–5073.
28. W.R. Dawes Jr. and G.F. Derbenwick. Prevention of CMOS latchup by gold doping. *IEEE Transactions on Nuclear Science*, **23**, 1976, 2027–2030.
29. D.B. Estreich, A. Ochoa Jr. and R.W. Dutton. An analysis of latchup prevention in CMOS IC's using an epitaxial-buried layer process. *International Electron Device Meeting (IEDM) Technical Digest*, December 1978. p. 230–234.
30. D.B. Estreich and R.W. Dutton. Modeling latchup in CMOS integrated circuits. *1978 Asilomar Conference on Circuits, Systems, and Computer Digest*, Monterey, CA, November 1978. p. 489–492.
31. D.B. Estreich and R.W. Dutton. Latchup in CMOS integrated circuits. *1978 Government Microcircuit Applications Conference (GOMAC) Digest of Papers*, Monterey, CA, November 1978. p. 110–111.
32. D.B. Estreich. *Latchup and radiation integrated circuit (LURIC): a test chip for CMOS latchup investigation*. Sandia Laboratories Report SAND78-1540, Albuquerque, NM, November 1978.
33. D.B. Estreich. *The physics and modeling of latch-up and CMOS integrated circuits*. Technical Report No. G-201-9, Integrated Circuits Laboratory, Stanford Electronic Laboratories, Stanford University, Stanford, CA, November 1980.
34. P.V. Dressendorfer and M.G. Armendariz. A SEM technique for experimentally locating latchup paths in integrated circuits. *IEEE Transactions on Nuclear Science*, **27**, 1980, 1688–1693.
35. J.E. Schroder, A. Ochoa and P.V. Dressendorfer. Latchup elimination in bulk CMOS LSI circuits. *IEEE Transactions on Nuclear Science*, **27**, 1980, 1735–1738.
36. R.S. Payne, W.R. Grant and W.J. Bertram. Elimination of latchup in bulk CMOS. *International Electron Device Meeting (IEDM) Technical Digest*, December 1980. p. 248–251.
37. W.D. Raburn. A model for parasitic SCR in bulk CMOS. *International Electron Device Meeting (IEDM) Technical Digest*, December 1980. p. 252–255.
38. R.D. Rung, C.J. Delloca and L.G. Walker. A retrograde p-well for higher density CMOS. *IEEE Transactions on Electron Devices*, **28**(10), 1980, 1115–1119.
39. L. Parillo, *et al*. Twin-tub CMOS – a technology for VLSI circuits. *International Electron Device Meeting (IEDM) Technical Digest*, 1980. p. 752–755.
40. S.R. Combs. Scaleable retrograde p-well CMOS technology. *International Electron Device Meeting (IEDM) Technical Digest*, December 1981. p. 346–349.
41. T. Iizuka and J. Moll. A figure of merit for CMOS latchup tolerance. *The 1981 CMOS Workshop*, San Francisco Airport Hilton, San Francisco, CA, May 18, 1981.
42. P.V. Dressendorfer and A. Ochoa. An analysis of modes of operation of parasitic SCR's. *IEEE Transactions on Nuclear Science*, **28**(6), 1981, 4288–4291.
43. A. Ochoa and P.V. Dressendorfer. A discussion of the role of distributed effects in latchup. *IEEE Transactions on Nuclear Science*, **28**(6), 1981, 4292–4294.
44. A.W. Wieder, C. Werner and J. Harter. Design model for bulk CMOS scaling enabling accurate latchup prediction. *International Electron Device Meeting (IEDM) Technical Digest*, 1981. p. 354–357.
45. G.J. Hu, M.R. Pinto and S. Kordic. Two-dimensional simulation of latchup in CMOS structure. *Device Research Conference Proceedings*, VA-5, June 21–23, 1982.
46. R.D. Rung and H. Momose. Improved modeling of CMOS latchup and VLSI implications. *Symposium on VLSI Technology*, September 1982. p. 52–53.
47. R.D. Rung, H. Momose and Y. Nagabuko. Deep trench isolated CMOS devices. *International Electron Device Meeting (IEDM) Technical Digest*, 1982. p. 237–240.
48. C.C. Huang, M.D. Hartanft, N.F. Pu, C. Yue, C. Rahn, J. Schrankler, G.D. Kirchner, F.L. Hampton and T.E. Hendrickson. Characterization of CMOS latchup. *International Electron Device Meeting (IEDM) Technical Digest*, 1982. p. 454–457.
49. R.R. Troutman and H.P. Zappe. Power-up triggering condition for latchup in bulk CMOS. *1982 Symposium on VLSI Technology*, Oiso, Japan, September 1982. p. 52–53.

50. R.R. Troutman and H.P. Zappe. A transient analysis of latchup in bulk CMOS. *IEEE Transactions on Electron Devices*, **30**, 1983, 170–179.
51. R.D. Rung and H. Momose. DC holding and dynamic triggering characteristics of bulk CMOS latchup. *IEEE Transactions on Electron Devices*, **30**, 1983, 1647–1655.
52. U. Schwabe, H. Herbst, E.J. Jacobs and D. Takacs. N- and p-well optimization for high-speed n-epitaxy CMOS circuits. *IEEE Transactions on Electron Devices*, **30**, 1983, 1339–1344.
53. D.L. Wollesen, J. Haskell and J. Yu. N-well and p-well performance comparison. *International Electron Device Meeting (IEDM) Technical Digest*, 1983. p. 31–34.
54. G. Goto, H. Takahashi and T. Nakamura. Latchup immunity against noise pulses in a CMOS double well structure. *International Electron Device Meeting (IEDM) Technical Digest*, 1983. p. 168–171.
55. G.J. Hu, Y. Taur, R.H. Dennard, L.M. Terman and C.Y. Ting. A self-aligned 1-μm CMOS technology for VLSI. *International Electron Device Meeting (IEDM) Technical Digest*, 1983. p. 731–734.
56. D. Takacs, J. Harter, E.P. Jacobs, C. Werner, U. Schwabe, J. Winnerl and E. Lange. Comparison of latchup in n-well and p-well CMOS circuits. *International Electron Device Meeting (IEDM) Technical Digest*, 1983. p. 159–164.
57. J. Manoliu, F.H. Tseng, B.J. Woo and T.J. Meier. High density and reduced latchup susceptibility CMOS technology for VLSI. *IEEE Electron Device Letters*, **4**, 1983, 233–235.
58. L. Wakeman. Silicon gate CMOS chips gain immunity to SCR latchup. *Electronics*, **56**(16), 1983, 136–140.
59. R.R. Troutman. Epitaxial layer enhancement of n-well guard rings for CMOS circuits. *IEEE Electron Device Letters*, **4**, 1983, 438–440.
60. Y. Niitsu, H. Nihara, K. Kanzaki and S. Kohyama. Resistance modulation effect due to current injection and CMOS latchup. *International Electron Device Meeting (IEDM) Technical Digest*, 1983. p. 164–167.
61. E. Hamdy and A. Mohsen. Characterization and modeling of transient latchup in CHMOS technology. *International Electron Device Meeting (IEDM) Technical Digest*, 1983. p. 172–175.
62. R.R. Troutman and M.J. Hargrove. Transmission line model for latchup in CMOS circuits. *Symposium on VLSI Technology*, September 13–15, 1983. p. 56–59.
63. G.J. Hu. A better understanding of CMOS latchup. *IEEE Transactions on Electron Devices*, **31**, 1984, 62–67.
64. R.C. Fang and J.L. Moll. Latchup model for the parasitic path in bulk CMOS. *IEEE Transactions on Electron Devices*, **31**, 1984, 113–120.
65. W. Craig. Latchup test structures and their characterization. *IEEE VLSI Workshop on Test Structures*, San Diego, February 20, 1984.
66. J.G. Dooley and R.C. Jaeger. Temperature dependence of latchup in CMOS circuits. *IEEE Electron Device Letters*, **5**, 1984, 41–43.
67. D. Takacs, C. Werner, J. Harter and U. Schwabe. Surface induced latchup in VLSI CMOS circuits. *IEEE Transactions on Electron Devices*, **31**, 1984, 279–286.
68. R.R. Troutman and H.P. Zappe. Layout and bias considerations for preventing transiently triggered latchup in CMOS. *IEEE Transactions on Electron Devices*, **31**, 1984, 279–286.
69. I.H. Leventhal. *Comparison of DC latchup characterization techniques for CMOS technology.* BS thesis, Massachusetts Institute of Technology, June 1984.
70. G.J. Hu and R.H. Bruce. A CMOS structure with high holding voltage. *IEEE Electron Device Letters*, **5**, 1984, 211–214.
71. K.W. Terrill and C. Hu. Substrate resistance calculation for latchup modeling. *IEEE Transactions on Electron Devices*, **31**, 1984, 1152–1155.
72. K.W. Terrill, P.F. Byrne, H.P. Zappe, N.W. Cheung and C. Hu. A new method for preventing CMOS latchup. *International Electron Device Meeting (IEDM) Technical Digest*, 1984. p. 406–409.
73. A.G. Lewis. Latchup suppression in fine dimension shallow p-well CMOS circuits. *IEEE Transactions on Electron Devices*, **31**, 1984, 1472–1481.
74. U. Schwabe, E.P. Jacobs, D. Takacs, J. Winnerl and E. Lange. Reduced n+/p+ spacing with high latchup hardness in self-aligned double well CMOS technology. *International Electron Device Meeting (IEDM) Technical Digest*, 1984. p. 410–413.
75. S. Odanaka, M. Wakabayashi and T. Ohzone. The dynamics of latchup turn-on behavior in scaled CMOS. *IEEE Transactions on Electron Devices*, **32**, 1985, 1334–1340.

76. M.R. Pinto and R.W. Dutton. Accurate triggering condition analysis for CMOS latchup. *IEEE Electron Device Letters*, **6**(2), 1985, 100–102.
77. K.Y. Fu. Transient latchup in bulk CMOS with a voltage-dependent well–substrate junction capacitance. *IEEE Transactions on Electron Device Letters*, **32**, 1985, 717–720.
78. J.E. Hall, J.A. Seitchik, L.A. Arledge and P. Yang. An improved circuit model for CMOS latchup. *IEEE Electron Device Letters*, **6**, 1985, 320–322.
79. R. Troutman. *Latchup in CMOS Technology: The Problem and the Cure.* New York: Kluwer Academic Publishers; 1985.
80. P. Cottrell, S. Warley, S. Voldman, W. Leipold and C. Long. N-well design for trench DRAM arrays. *International Electron Device Meeting (IEDM) Technical Digest*, December 1988. p. 584–587.
81. S. Voldman, M. Marceau, A. Baker, E. Adler, S. Geissler, J. Slinkman, J.B. Johnson and M. Paggi. Retrograde well and epitaxial thickness optimization for shallow- and deep-trench collar merged isolation and node trench (MINT) SPT cell and CMOS logic technology. *International Electron Device Meeting (IEDM) Technical Digest*, 1992. p. 811–814.
82. S. Voldman. Retrograde well implants and ESD. *SEMATECH Vertical Modulated Well PTAB*, November, Austin, TX, November 1994.
83. S. Voldman. Optimization of MeV retrograde wells for advanced logic and microprocessor/PowerPC and electrostatic discharge (ESD). *Smart and Economic Device and Process Designs for ULSI Using MeV Implant Technology Seminar: SEMICON West, SEMICON West GENUS Seminar*, San Francisco, CA, July 1994.
84. S. Voldman. MeV implants boost device design. *IEEE Circuits and Devices* 1995, **11**(6), 8–16.
85. S. Voldman. The impact of MOSFET technology evolution and scaling on electrostatic discharge protection. *Microelectronics Reliability*, **38**, 1998, 1649–1668.
86. S. Voldman. Electrostatic discharge protection, scaling, and ion implantation in advanced semiconductor technologies. *Process Integration Issues/Technical Trends Session*, *Proceedings of the Ion Implantation Conference ($I^2$CON)*, Napa, CA, 1999.
87. M. Hargrove, S. Voldman, J. Brown, K. Duncan and W. Craig. Latchup in CMOS technology. *Proceedings of the International Reliability Physics Symposium (IRPS)*, 1998. p. 269–278.
88. T. Li and S.M. Kang. Layout extraction and verification methodology for CMOS I/O circuits. *IEEE/ACM Design Automation Conference (DAC)*, 1998. p. 291–296.
89. T. Li. *Design automation for reliable CMOS chip I/O circuits.* PhD thesis, University of Illinois Urbana-Champaign, UILU-ENG-98-2219, August 1998.
90. T. Li, C.H. Tsai, E. Rosenbaum and S.M. Kang. Substrate modeling and lumped substrate resistance extraction for latchup/ESD circuit simulations. *IEEE/ACM Design Automation Conference (DAC)*, 1999. p. 549–554.
91. K.K. Bourdelle, Y. Chen, R. Ashton, L. Rubin, A. Agarwal and W. Morris. Epi-replacement in CMOS Technology by high dose, high energy boron implantation into Cz substrates. *International Electron Device Meeting (IEDM) Technical Digest*, 1999. p. 312–316.
92. W. Morris. Latchup in CMOS. *Proceedings of the International Reliability Physics Symposium (IRPS)*, 2003. p. 76–84.
93. G. Boselli, V. Reddy and C. Duvvury. Latchup in 65 nm CMOS technology: a scaling perspective. *Proceedings of the International Reliability Physics Symposium (IRPS)*, 2004. p. 137–144.
94. A. Weger, S. Voldman, F. Stellari, P. Song, P. Sanda and M. McManus. A transmission line pulse (TLP) picosecond imaging circuit analysis (PICA) methodology for evaluation of ESD and latchup. *Proceedings of the International Reliability Physics Symposium (IRPS)*, 2003. p. 99–104.
95. F. Stellari, P. Song, M.K. McManus, A.J. Weger, K. Chatty, M. Muhammad and P. Sanda. Study of critical factors determining latchup sensitivity in ICs using emission microscopy. *Proceeding of the International Symposium for Testing and Failure Analysis (ITSFA)*, 2003. p. 19–24.
96. K. Chatty, P. Cottrell, M. Muhammad, F. Stellari, A. Weger, P. Song and M. McManus. Model-based guidelines for cable discharge events (CDE) induced latchup in CMOS ICs. *Proceedings of the International Reliability Physics Symposium (IRPS)*, 2004. p. 130–134.
97. S. Voldman. Latchup and the domino effect. *Proceedings of the International Reliability Physics Symposium (IRPS)*, 2005. p. 145–156.

98. A. Watson, S. Voldman and T. Larsen. Deep trench guard ring structures and evaluation of the probability of minority carrier escape for ESD and latchup in advanced BiCMOS SiGe technology. *Proceedings of the Taiwan Electrostatic Discharge Conference (T-ESDC)*, 2003. p. 97–103.
99. S. Voldman and A. Watson. The influence of deep trench and substrate resistance on the latchup robustness in a BiCMOS silicon germanium technology. *Proceedings of the International Reliability Physics Symposium (IRPS)*, 2004. p. 135–142.
100. A. Watson and S. Voldman. The effect of deep trench and sub-collector on the latchup robustness in BiCMOS silicon germanium technology. *Proceedings of the Bipolar Circuit Technology Meeting (BCTM)*, Montreal, Canada, September 12–14, 2004. p. 172–175.
101. S. Voldman and A. Watson. The influence of polysilicon-filled deep trench and sub-collector implants on latchup robustness in RF CMOS and BiCMOS SiGe technology. *Proceedings of the Taiwan Electrostatic Discharge Conference (T-ESDC)*, 2004. p. 15–19.
102. S. Voldman. A review of CMOS latchup and electrostatic discharge (ESD) in bipolar complimentary MOSFET (BiCMOS) silicon germanium technologies: Part II. Latchup. *Microelectronics and Reliability*, **45**, 2005, 437–455.
103. S. Voldman, E.G. Gebreselasie, L.W. Lanzerotti, N.B. Feilchenfeld, S.A. St. Onge, A. Joseph and J. Dunn. The influence of silicon dioxide-filled trench isolation (TI) structure and implanted sub-collector on latchup robustness. *Proceedings of the International Reliability Physics Symposium (IRPS)*, 2005. p. 112–121.
104. S. Voldman, E.G. Gebreselasie, D. Hershberger, D.S. Collins, N.B. Feilchenfeld, S.A. St. Onge, A. Joseph, J. Dunn. Latchup in merged triple well technology. *Proceedings of the International Reliability Physics Symposium (IRPS)*, 2005. p. 129–136.
105. S. Voldman and E. Gebreselasie. The influence and integration of merged triple well, deep trench, and sub-collector on CMOS latchup. *Proceedings of the Taiwan Electrostatic Discharge Conference (T-ESDC)*, 2006. p. 49–52.
106. S. Voldman, E. Gebreselasie, X.F. Liu, D. Coolbaugh and A. Joseph. The influence of high resistivity substrates on CMOS latchup. *Proceedings of the Electrical Overstress/Electrostatic Discharge (EOS/ESD) Symposium*, 2005. p. 90–99.
107. J. Salcedo-Suner, R. Cline and C. Duvvury. A new I/O signal latch-up phenomenon in voltage tolerant ESD protection circuits. *Proceedings of the International Reliability Physics Symposium (IRPS)*, 2003. p. 85–91.
108. Y. Huh, K.J. Min, P. Bendix, V. Axelrad, R. Narayan, J.-W. Chen, L.D. Johnson and S. Voldman. Chip-level layout and bias considerations for preventing neighboring I/O cell interaction induced latchup and inter-power supply latchup in CMOS technologies. *Proceedings of the Electrical Overstress/Electrostatic Discharge (EOS/ESD) Symposium*, 2005. p. 100–107.
109. R.Y. Zhan, H.G. Feng, Q. Wu, G. Chen, X.K. Guan and A.Z. Wang. A technology-independent CAD tool for ESD protection device extraction: ESD Extractor. *Proceedings of the International Conference on Computer-Aided Design*, 2002. p. 510–513.
110. R.Y. Zhan. *ESDCat: a new CAD software package for full-chip ESD protection circuit verification*, October 2005.
111. ESD Association. *ESD STM 5.1-2001: standard test method for electrostatic discharge sensitivity testing-human body model (HBM) component level*, 2001.
112. ESD Association. *ESD STM 5.2-1998: standard test method for electrostatic discharge sensitivity testing-machine model (MM) component level*, 1998.
113. Intel Corporation. *Cable discharge event in local area network environment*. Intel White Paper, Order No. 249812-001, July 2001.
114. R. Brooks. *A simple model for the cable discharge event*. IEEE802.3 Cable-Discharge Ad-Hoc Committee, March 2001.
115. Telecommunications Industry Association (TIA). Category 6 cabling: static discharge between LAN cabling and data terminal equipment. *Category 6 Consortium*, December 2002.
116. J. Deatherage and D. Jones. Multiple factors trigger discharge events in Ethernet LANs, *Electronic Design* 2000; **48**(25) 111–116.
117. H. Geski. *DVI compliant ESD protection to IEC 61000-4-2 level 4 standard*. Conformity, September 2004. p. 12–17.

118. A. Weger, S. Voldman, F. Stellari, P. Song, P. Sanda and M. McManus. A transmission line pulse (TLP) pico-second imaging circuit analysis (PICA) methodology for evaluation of ESD and latchup. *Proceedings of the International Reliability Physics Symposium (IRPS)*, 2003. p. 99–104.
119. F. Stellari, P. Song, M.K. McManus, A.J. Weger, K. Chatty, M. Muhammad and P. Sanda. Study of critical factors determining latchup sensitivity in ICs using emission microscopy. *Proceeding of the International Symposium for Testing and Failure Analysis (ITSFA)*, 2003. p. 19–24.
120. K. Chatty, P. Cottrell, M. Muhammad, F. Stellari, A. Weger, P. Song and M. McManus. Model-based guidelines for cable discharge events (CDE) induced latchup in CMOS ICs. *Proceedings of the International Reliability Physics Symposium (IRPS)*, 2004. p. 130–134.
121. ESD Association. *Electrostatic Discharge Association Draft Standard Practice (SP) for the Protection of Electrostatic Discharge Susceptible Items System Level – Simulation Verification*, ESD DSP14.1-2003, 2003.
122. ESD Association. *Electrostatic Discharge Association Draft Standard Practice (SP) for the Protection of Electrostatic Discharge Susceptible Items System Level – Measurement of Cable Discharge Events*, ESD DSP14.3-2006, 2006.
123. S. Voldman. Latchup and the domino effect. *Proceedings of the International Reliability Physics Symposium (IRPS)*, 2005. p. 145–156.
124. S. Voldman. Cable discharge event and CMOS latchup. *Proceedings of the Taiwan ESD Conference (T-ESDC)*, 2005. p. 63–68.
125. T. Brodbeck, W. Stadler, R. Gartner, H. Gossner, W. Hartung and R. Losehand. ESD-schutznassnahmen fur schnittstellan in der kommunikationselektronik (ESD protection measures for communication IC interfaces). *Proceedings of the 9th ESD Forum*, 2005. p. 115–120.
126. T.X. Lai and M.D. Ker. Method to evaluate cable discharge event (CDE) reliability of integrated circuits in CMOS technology. *Proceedings of the 2006 IEEE International Symposium on Quality Electronic Design (ISQED)*, 2006. p. 597–602.
127. W. Stadler, T. Brodbeck, R. Gartner and H. Gossner. Cable discharges into communication interfaces. *Proceedings of the Electrical Overstress/Electrostatic Discharge (EOS/ESD) Symposium*, 2006. p. 144–151.
128. J.L. Wirth and S.C. Rogers. The transient response of transistors and diodes to ionizing radiation. *IEEE Transactions on Nuclear Science*, **11**, 1964, 24–38.
129. W.H. Dierking, G.E. Katz and E.L. Steele. Transient radiation current generator model for semiconductor devices. *IEEE Transactions on Nuclear Science*, **16**, 1969, 144–152.
130. I. Kaplan. *Nuclear Physics*, 2nd ed. Reading, MA: Addison-Wesley Publishing Company; 1962.
131. T.C. May and M.H. Woods. Alpha particle-induced soft errors in dynamic memories. *IEEE Transactions on Electron Devices*, **26**(1), 1979, 2–9.
132. G.A. Sai-Halasz and M.R. Wordeman. Monte Carlo modeling of the transport of ionizing radiation created carriers in integrated circuits. *IEEE Electron Device Letters*, **1**(10), 1980, 211–213.
133. C.H. Hsieh, P.C. Murley and R.R. O'Brien, A field-funneling effect on the collection of alpha particle generated carriers in silicon devices. *IEEE Electron Device Letters*, **2**(4), 1981, 103–105.
134. C. Hu. Alpha particle-induced field and enhanced collection of carriers. *IEEE Transactions on Electron Letters*, **3**(2), 1982, 31–34.
135. G. Sai-Halasz, M.R. Wordeman and R.H. Dennard. Alpha-particle-induced soft error rate in VLSI circuits. *IEEE Transactions on Electron Devices*, **29**, 1982, 725–731.
136. G. Sai-Halasz and D.D. Tang. Soft errors rates in static bipolar RAMs. *International Electron Device Meeting (IEDM) Technical Digest*, Session 20.7, 1983. p. 344–347.
137. S. Voldman and L. Patrick. Alpha particle induced single event upsets in bipolar static emitter coupled logic (ECL) cells. *Proceedings of the IEEE Nuclear and Space Radiation Conference*, Boulder, CO, 1983.
138. S. Voldman and L. Patrick. Alpha particle induced single event upsets in bipolar static emitter coupled logic (ECL) cells. Special issue on 1984 Annual Conference on Nuclear and Space Radiation Effects. *IEEE Transactions on Nuclear Science*, **31**(6), 1984, 1196–1200.
139. S. Voldman, L. Patrick and D. Wong. Alpha particle effects on bipolar emitter-coupled logic static arrays. *Proceedings of the International Solid State Circuit Conference (ISSCC)*, Vol. XXVII, 1985. p. 262–263.
140. S. Voldman, P. Corson, L. Patrick, K. Nguyen, L. Gilbert, R. Goodwin, T. Maffit and S. Murphy. CMOS SRAM alpha particle modeling and experimental results. *International Electron Device Meeting (IEDM) Technical Digest*, Session 20.7, 1987. p. 518–521.

141. A.H. Johnston, R.E. Plaag and M.P. Baze, The effect of circuit topology on radiation-induced latchup. *IEEE Transactions on Nuclear Science*, **36**(6), 1989, 2229–2238.
142. A.H. Johnston. The influence of VLSI technology evolution on radiation-induced latchup in space systems. *IEEE Transactions on Nuclear Science*, **43**(2), 1996, 505–521.
143. S.C. Moss, S.D. LaLumondiere, J.R. Scarpulla, K.P. MacWilliams, W.R. Crain and R. Koga, Correlation of picosecond laser-induced latchup and energetic particle-induced latchup in CMOS test structures. *IEEE Transactions on Nuclear Science*, **42**(6), 1995, 1948–1956.
144. P.E. Dodd, M.R. Shaneyfelt, J.R. Schwank and G.L. Hash. Neutron-induced latchup in SRAMs at ground level. *Proceedings of the International Reliability Physics Symposium (IRPS)*, 2003. p. 51–55.
145. R.E. Plaag, M.P. Baze and A.H. Johnston. A distributed model for radiation-induced latchup. *IEEE Transactions on Nuclear Science*, **35**(6), 1988, 1563–1569.
146. D.K. Nichols, J.R. Coss, R.K. Watson, H.R. Schwartz and R.L. Pease. An observation of proton-induced latchup in CMOS microprocessor. *IEEE Transactions on Nuclear Science*, **39**(6), 1992, 1654–1656.
147. H. Iwata and T. Aoki. Numerical simulation of single event latchup in the temperature range of 77-450 K. *IEEE Transactions on Nuclear Science*, **42**(3), 1995, 148–154.
148. J.R. Schwank, P.E. Dodd, M.R. Shaneyfelt, J.A. Felix, G.L. Hash, V. Ferlet-Cavrois, P. Paillet, J. Baggio, P. Tangyunyong and E. Blackmore. Issues for single-event proton testing of SRAMs. *IEEE Transactions on Nuclear Science*, **51**(6), 2004, 3692–3700.
149. R.C. Lacoe, S.C. Moss, J.V. Osborn, B.K. Janousek, S.D. La Lumondiere, S. Brown and D.C. Mayer. Neutron and proton irradiation for latchup suppression in a radiation-tolerant commercial submicron CMOS process. *Proceedings of Fifth European Conference on Radiation and Its Effects on Components and Systems (RADECS)*, 1999. p. 340–345.
150. J. Levinson, O. Even, E. Adler, M. Hass, D. Ilberg and Y. Lifshitz. Single Event Latchup (SEL) in IDT 7187 SRAMs-dependence on ion penetration depth. *Proceedings of Fifth European Conference on Radiation and Its Effects on Components and Systems (RADECS)*, 1993. p. 438–440.
151. T. Aoki. Dynamics of heavy-ion-induced latchup in CMOS structures. *IEEE Transactions on Electron Devices*, **35**(11), 1988, 1885–1891.
152. A.H. Johnston. Mechanisms for single-particle latchup in CMOS structures. *Proceedings of the Second European Conference on Radiation and its Effects on Components and Systems (RADECS)*, 1993. p. 433–437.
153. Y. Saitoh, Y. Akiba, H. Konishi, H. Watanabe, K. Nonaka, M. Kamiya, Y. Tsuboyama and T. Miura. Radiation damage of 1.0 μm CMOS VLSI as a function of latchup immunity. *Proceedings of the IEEE Nuclear Science Symposium and Medical Imaging Conference*, Vol. 2, 1992. p. 788–790.
154. H.N. Becker, T.F. Miyahira and A.H. Johnston. Latent damage in CMOS devices from single event latchup. *IEEE Transactions on Nuclear Science*, **49**(6), 2002, 3009–3015.
155. A.H. Johnston, B.W. Hughlock, M.P. Baze and R.E. Plaag. The effect of temperature on single-particle latchup. *Proceedings of the IEEE Transactions on Nuclear Science*, **38**(6), 1991; 1435–1441.
156. A.H. Johnston and B.W. Hughlock. Latchup in CMOS from single particles. *IEEE Transactions on Nuclear Science*, **37**(6), 1990, 1886–1893.
157. G. Bruguier and J.-M. Palau. Single particle-induced latchup. *IEEE Transactions on Nuclear Science*, **43**(2), 1996, 522–532.
158. A.H. Johnson, G.M. Swift and L.D. Edmonds. Latchup in integrated circuits from energetic protons. *IEEE Transactions on Nuclear Science*, **44**(6), 1997, 2367–2377.
159. A.I. Chumakov. A.N. Egorov, O.B. Mavritsky, A.Y. Nikiforov and A.V. Yanenko. Single event latchup threshold estimation based on laser dose rate test results. *IEEE Transactions on Nuclear Science*, **44**(6), 1997, 2034–2039.
160. P.J. Layton, D.R. Czajkowski, J.C. Marshall, H.F.D. Anthony and R.W. Boss. Single event latchup protection of integrated circuits. *Proceedings of the Radiation and Its Effects on Components and Systems (RADECS)*, 1997. p. 327–331.
161. E. Normand, J.L. Wert, P.P. Majewski, D.L. Oberg, W.G. Bartholet, S.K. Davis, M. Shoga, S.A. Wender and A. Gavron. Single event upset and latchup measurements in avionics devices using the WNR neutron beam and a new neutron-induced latchup model. *IEEE Radiation Effects Data Workshop*, *Proceedings of the Nuclear and Space Radiation Effects Conference (NSREC)*, 1995. p. 33–38.

162. F.M. Roche, S.D. Bocus and B. Pistoulet. Influence of low energy electronic radiation on the latchup triggering level in CMOS integrated circuits. *Proceedings of the Radiation and Its Effects on Components and Systems (RADECS)*, 1991. p. 440–444.
163. P.E. Dodd, M.R. Shaneyfelt, E. Fuller, J.C. Pickel, F.W. Sexton and P.S. Winokur. Impact of substrate thickness on single-event effects in integrated circuits. *IEEE Transactions on Nuclear Science*, **48**(6), 2001, 1865–1871.
164. J.P. Spratt, J.C. Pickel, R.E. Leadon, R.C. LaCoe, S.C. Moss and S.D. LaLumondiere. A single event latchup suppression technique for COTS CMOS ICs. *IEEE Transactions on Nuclear Science*, **50**(6), 2003, 2219–2224.
165. J.R. Schwank, M.R. Shaneyfelt, J. Baggio, P.E. Dodd, J.A. Felix, V. Ferlet-Cavrois, P. Paillet, D. Lambert, F.W. Sexton, G.L. Hash and E. Blackmore. Effects of particle energy on proton-induced single-event latchup. *IEEE Transactions on Nuclear Science*, **52**(6), 2005, 2622–2629.
166. H. Shindou, S. Kuboyama, T. Hirao and S. Matsuda. Local and pseudo SELs observed in digital LSIs and their implication to SEL test method. *IEEE Transactions on Nuclear Science*, **52**(6), 2005, 2638–2641.
167. J. Barak, J. Levinson, A. Akkerman, M. Hass, M. Victoria, A. Zentner, D. David, O. Even and Y. Lifishitz. A new approach to the analysis of SEU and SEL data to obtain the sensitive volume thickness. *IEEE Transactions on Nuclear Science*, **43**, 1996, 907–911.
168. B.E. Fischer, M. Schlogl, J. Barak, E. Adler and S. Metzger. Simultaneous imaging of upsets and latchup sensitive regions of a static RAM. *Nuclear Instrumentation and Methods in Physics Research B*, **130**, 1997, 478–485.
169. J. Levinson, A. Akkerman, M. Victoria, M. Hass, D. Ilberg, M. Alluralde, R. Henneck and Y. Lifshitz. New insight into proton-induced latchup: experiment and modeling. *Applied Physics Letters*, **63**, 1993, 2952–2954.
170. P. Foulliat, H. Lapuyade, A. Touboul, J.P. Dom and R. Gaillard. Numerical modeling of mechanisms involved in latchup triggering by laser beam. *IEEE Transactions on Nuclear Science*, **43**, 1996, 944–951.
171. J. Barak, E. Adler, B.E. Fischer, M. Schlogl and S. Metzger, Micro-beam mapping of single event latchups and single event upset in CMOS SRAMs. *IEEE Transactions on Nuclear Science*, **45**(3), 1998, 1595–1602.
172. M. Bafluer, J. Buxo, M.P. Vidal, P. Givelin, V. Macary and G. Sarrabayrouse. Application of a floating well concept to a latchup-free low-cost smart power high-side switch technology. *IEEE Transactions on Electron Devices*, **40**(7), 1993, 1340–1342.
173. R. Peppiette. A new protection technique for ground recirculation parasitics in monolithic power IC's. *Sanken Technical Report*, **26**(1), 1994, 91–97.
174. M. Bafluer, M.P. Vidal, J. Buxo, P. Givelin, V. Macary and G. Sarrabayrouse. Cost-effective smart power CMOS/DMOS technology: design methodology for latchup immunity. *Analog Integrated Circuits and Signal Processing*, **8**(3), 1995, 219–231.
175. W.W.T. Chan, J.K.O. Sin and S.S. Wong. A novel crosstalk isolation structure for bulk CMOS power IC's. *IEEE Transactions on Electron Devices*, **45**(7), 1998, 1580–1586.
176. W. Winkler and F. Herzl, Active substrate noise suppression in mixed-signal circuits using on-chip driven guard rings. *Proceedings of the IEEE 2000 Custom Integrated Circuits Conference*, May 2000. p. 356–360.
177. O. Gonnard and G. Charitat. Substrate current protection in smart power IC's. *Proceedings of the International Symposium on Power Semiconductor Devices (ISPSD)*, 2000. p. 169–172.
178. R. Zhu, V. Parthasarathy, V. Khemka and A. Bose. Implementation of high-side, high-voltage RESURF LDMOS in a sub-half micron smart power technology. *Proceedings of the International Symposium on Power Semiconductor Devices (ISPSD)*, 2001. p. 403–406.
179. O. Gonnard, G. Charitat, P.Lance, M. Susquet, M. Bafluer and J.P. Laine. Multi-ring active analogic protection (MAAP) for minority carrier injection suppression in smart power IC's. *Proceedings of the International Symposium on Power Semiconductor Devices (ISPSD)*, 2001. p. 351–354.
180. M. Schenkel, P. Pfaffli, W. Wilkening, D. Aemmer and W. Fichtner. Transient minority carrier collection from substrate in smart power design. *Proceedings of the European Solid State Device Research Conference (ESSDERC)*, 2001.
181. V. Parthasarathy, R. Zhu, V. Khemka, T. Roggenbauer, A. Bose, P. Hui, P. Rodriguez, J. Nivison, D. Collins, Z. Wu, I. Puchades and M. Butner. A 0.25-μm CMOS based 70V smart power technology with deep trench for high-voltage isolation. *International Electron Device Meeting (IEDM) Technical Digest*, 2002. p. 459–462.
182. V. Parthasarathy, V. Khemka, R. Zhu, I. Puchades, T. Roggenbauer, M. Butner, P. Hui, P. Rodriquez and A. Bose. A multi-trench analog + logic protection (M-TRAP) for substrate crosstalk prevention in a 0.25-μm smart power

platform with 100V high-side capability. *Proceedings of the International Symposium on Power Semiconductor Devices (ISPSD)*, 2004. p. 427–430.
183. J.P. Laine, O. Gonnard, G. Charitat, L. Bertolini and A. Peyre-Lavigne. Active pull-down protection for full substrate current isolation in smart power IC's. *Proceedings of the International Symposium on Power Semiconductor Devices (ISPSD)*, 2003. p. 273–276.
184. V. Khemka, V. Parthasarathy, R. Zhu, A. Bose and T. Roggenbauer. Trade-off between high-side capability and substrate minority carrier injection in deep sub-micron smart power technologies. *Proceedings of the International Symposium on Power Semiconductor Devices (ISPSD)*, 2003. p. 241–244.
185. W. Horn. *On the reverse-current problem in integrated smart power circuits*. PhD thesis, Technical University of Graz, Austria, April 2003.

# 2 Bipolar Transistors

## 2.1 THE BIPOLAR TRANSISTOR AND CMOS LATCHUP

Bipolar junction transistors (BJTs) were invented in the 1940s and continued to have substantial focus in technology through the 1980s for a wide range of applications [1–11]. In this chapter, our interest is the bipolar transistor and its relationship to latchup. Latchup involves the interaction of bipolar transistors that either are native to the technology or are parasitic transistors. Parasitic bipolar transistors are formed naturally with the placement of features or circuits adjacent to each other. These elements can have vertical or lateral components. The chapter will first address bipolar fundamental variables and concepts that are important for the understanding of latchup.

### 2.1.1 Fundamental Equations of Semiconductors and the Drift–Diffusion Current Constitutive Relationships

In latchup, the analysis of the semiconductor drift and diffusion properties of the minority and majority carriers is important to understand the current flow in the parasitic elements. The fundamental equations in the semiconductor formulation include the Poisson equation,

$$\nabla \cdot (\varepsilon \nabla \psi) = -q(p - n + N_{\mathrm{D}} + N_{\mathrm{A}}),$$

where $\varepsilon$ is the dielectric constant, $p$ is the hole concentration, $n$ is the electron concentration, $N_{\mathrm{D}}$ is the donor concentration and $N_{\mathrm{A}}$ is the acceptor concentration. The current continuity equations for the electron and the hole concentration can be expressed as

$$\frac{\partial n}{\partial t} - \frac{1}{q} \nabla \cdot J_{\mathrm{n}} = G - R,$$

$$\frac{\partial p}{\partial t} + \frac{1}{q} \nabla \cdot J_{\mathrm{p}} = G - R,$$

*Latchup* Steven H. Voldman
© 2007 John Wiley & Sons, Ltd

where $J_n$ is the electron current density, $J_p$ is the hole current density, $G$ is the generation rate and $R$ is the recombination rate. In any given volume, the first term in the equations is associated with the local change in the concentration in time within that physical volume and the second term is associated with the flow of carriers in and out of the physical volume. The right-hand side (RHS) is associated with the net rate of increase of the carriers. In latchup, considerable focus is placed on the recombination rate of excess minority carriers. The drift–diffusion constitutive relationship for electrons and holes, with temperature variation, can be expressed as

$$J_n = -q\mu_n nE + qD_n \nabla n + qnD_n^T \nabla T,$$
$$J_p = -q\mu_p pE - qD_p \nabla p + qpD_p^T \nabla T.$$

For example, the drift–diffusion constitutive relationship for electrons contains an electron drift term and an electron diffusion term, and an additional current drive term is present when there is a gradient in the lattice temperature. The drift term is most significant inside the metallurgical junction, and the diffusion terms are typically dominant in the emitter, base and collector volumes outside of the metallurgical junctions. In a CMOS process with a p− substrate and an n-well region, the hole is transported between the p+ diffusion and the p− substrate. The minority carrier hole diffusion term is important in the understanding of the transport between the emitter and collector of the parasitic pnp. Also note that electric fields and thermal gradients also can influence the hole carrier transport. During latchup events, the heat flow in the regions of the latchup event can influence the latchup failure. The heat energy flow is

$$J_Q = -K(T)\nabla T,$$

where $K(T)$ is temperature-dependent thermal conductivity. The thermal conduction equation can be expressed as

$$\nabla \cdot (K(T_L))\nabla T_L = J \cdot E + E_g(R - G).$$

## 2.1.2 Diode Forward Bias Conditions

A latchup condition is that the bipolar transistors involved in the pnpn structure are in a forward active bias state. For a BJT to become ‘forward active’, the emitter–base (E–B) metallurgical junction diode must establish an E–B voltage that leads to a forward current flow. Hence, an understanding of the ‘forward active’ condition of diodes is important for understanding latchup. A key condition for latchup is the forward active initiation of both a pnp and an npn transistor.

The second issue of interest in CMOS latchup is when the diode element is an injection source. For example, an n+ diffusion in a p− substrate can inject minority carrier current into the p− substrate region. These minority carriers can be collected at an adjacent n-well region, serving as a perturbation of CMOS logic circuits or as a virtual source of minority carriers. As a result, the understanding of the high-level injection of diode structures is also important in the understanding of ‘external latchup’.

The diode current equation can be obtained by the evaluation of the total current through a p–n metallurgical junction. The current flowing through the diode is expressed as a function of the excess minority electron and hole populations at the p–n metallurgical junction,

$$I = qA\left\{\frac{D_p}{L_p}\Delta p_n + \frac{D_n}{L_n}\Delta n_p\right\},$$

where

$$\Delta p_{\mathrm{n}} = p'_{\mathrm{n}} - p_{\mathrm{n0}},$$
$$\Delta n_{\mathrm{p}} = n'_{\mathrm{p}} - n_{\mathrm{p0}}.$$

The variable $A$ is the area of the metallurgical junction where the minority carrier injection occurs. This expression reduces to the ideal diode case when the excess minority carrier injection is small compared to the doping concentrations. The hole population can be expressed as

$$\Delta p_{\mathrm{n}} = p_{\mathrm{n0}}\mathrm{e}^{V_j/kT} - p_{\mathrm{n0}},$$
$$\Delta p_{\mathrm{n}} = p_{\mathrm{n0}}\{\mathrm{e}^{V_j/kT} - 1\},$$

and the excess electron population can be expressed as

$$\Delta n_{\mathrm{p}} = n_{\mathrm{p0}}\mathrm{e}^{V_j/kT} - n_{\mathrm{p0}},$$
$$\Delta n_{\mathrm{p}} = n_{\mathrm{p0}}\{\mathrm{e}^{V_j/kT} - 1\}.$$

The ideal diode model is expressed as [8–11]

$$I = I_{\mathrm{s}}\left\{\exp\left(\frac{qV_j}{kT}\right) - 1\right\},$$

where $I$ is the diode current and $I_{\mathrm{s}}$ is the reverse saturation current,

$$I_{\mathrm{s}} = qA\left\{\frac{D_{\mathrm{p}}}{L_{\mathrm{p}}}p_{\mathrm{n0}} + \frac{D_{\mathrm{n}}}{L_{\mathrm{n}}}n_{\mathrm{p0}}\right\}.$$

As the voltage across the metallurgical junction increases above the thermal voltage ($V_0 = kT/q$), the injection current increases exponentially. Given this is associated with latchup initiation, the linear change in the electrical voltage associated with forward biasing leads to an exponential increase in the current. In addition, the linear decrease in electrical voltage leads to an exponential decrease in the current (e.g. 'de-biasing').

## 2.1.3 Diode Forward Bias Conditions and High-level Injection

Under high-level injection, the excess minority carrier populations can be expressed as

$$\Delta p_{\mathrm{n}} = \frac{p_{\mathrm{n}}(\mathrm{e}^{V_j/kT} - 1)(1 + (n_{\mathrm{n}}/p_{\mathrm{p}})\mathrm{e}^{q(V_j-\phi_i)/kT})}{1 - \mathrm{e}^{2q(V_j-\phi_i)/kT}},$$
$$\Delta n_{\mathrm{p}} = \frac{n_{\mathrm{p}}(\mathrm{e}^{V_j/kT} - 1)(1 + (p_{\mathrm{p}}/n_{\mathrm{n}})\mathrm{e}^{q(V_j-\phi_i)/kT})}{(1 - \mathrm{e}^{2q(V_j-\phi_i)/kT})}.$$

From this expression, the diode current equation in high-level injection can be expressed as

$$I = I_{\mathrm{S}}\frac{(\mathrm{e}^{V_j/kT} - 1)(1 + \eta\mathrm{e}^{q(V_j-\phi_i)/kT})}{1 - \mathrm{e}^{2q(V_j-\phi_i)/kT}},$$

where

$$I_s = qA\left\{\frac{D_p}{L_p}p_{n0} + \frac{D_n}{L_n}n_{p0}\right\}$$

and for a p+/n diode is

$$\eta = \frac{N_a}{N_d}$$

and for a n+/p diode

$$\eta = \frac{N_d}{N_a}.$$

This expression reduces to the ideal diode case when the excess minority carrier injection is small compared to the doping concentrations. The ideal diode model is expressed as

$$I = I_s\left\{\exp\left(\frac{qV_j}{kT}\right) - 1\right\},$$

where $I$ is the diode current and $I_s$ is the reverse saturation current.

It is also assumed that all the terminal voltage occurs across the metallurgical junction. Resistance in the physical regions of the diode region can also limit the total current through the structure. When the anode is an emitter of a bipolar transistor, the emitter resistance will play a role in the latchup initiation. During the latchup event, the emitter resistance is a function of the contacts resistance, silicide film and the diffusion sheet resistance. On the emitter, the total resistance can be expressed as a function of the metal resistance, the salicide and the diffusion,

$$R = R_M + (R_{Sal})_{eff} + (R_d)_{eff}.$$

From the diode equation, the voltage across the diode can be expressed as the sum of the voltage drops across the series resistance and the metallurgical junction,

$$I_D = I_s\left\{\exp\left(\frac{qV_D - \sum I_D R_i}{kT}\right) - 1\right\}.$$

The resistance can also be obtained from the derivative of the voltage with respect to current,

$$\frac{dV_D}{dI_D} = \frac{d(V_J + V_R)}{dI_D} = \frac{dV_J}{dI_D} + R_d.$$

Solving for the resistance,

$$R_d = \frac{dV_D}{dI_D} - \frac{dV_J}{dI_D}.$$

From the high-level injection relationship,

$$I_D = I_s \exp\left(\frac{qV_j}{2kT}\right)$$

and

$$\frac{dV_J}{dI_D} = \frac{2kT}{q}\frac{1}{I_D}.$$

From this expression, diode series resistance term can be expressed as

$$R_d = \frac{dV_D}{dI_D} - \frac{2kT}{q}\frac{1}{I_D}.$$

During latchup events, the diode structures can undergo high-level injection and self-heating. The diode current equation in high-level injection can be expressed as

$$I = I_s \frac{(e^{V_j/kT} - 1)(1 + \eta e^{q(V_j - \phi_i)/kT})}{1 - e^{2q(V_j - \phi_i)/kT}},$$

where

$$I_s = qA\left\{\frac{D_p}{L_p}p_{n0} + \frac{D_n}{L_n}n_{p0}\right\}$$

and for an abrupt p+/n diode is

$$\eta = \frac{N_a}{N_d}$$

and for an abrupt n+/p diode

$$\eta = \frac{N_d}{N_a}.$$

In addition, the self-heating influences the forward bias state of the metallurgical junction. Let the rise in the junction temperature be equal to the product of the thermal impedance and the input power,

$$T_J = \Theta_{TH} P_D.$$

The power consumption can be estimated as the power consumption in the diode region,

$$P_D = I_D V_D.$$

Hence,

$$T_J = \Theta_{TH}(I_D V_D)$$

The junction temperature can influence the diode forward bias condition as well as the bipolar current gain, influencing the latchup condition. The junction temperature is proportional to the power consumption in the diode region. Hence, for latchup, the second key issue is that the junction temperature is a function of the thermal impedance. The thermal impedance is a strong function of the substrate wafer material and doping concentration. When the junction temperature reaches the

intrinsic temperature, or melting temperature (e.g. aluminum, copper and silicon), thermal failure can occur.

## 2.2 BIPOLAR TRANSISTOR

In the following sections, the focus will be on key aspects of the bipolar transistor that shed light on latchup analysis. With the understanding of the diode operation as a two-terminal device, the bipolar transistor can be understood as a three-terminal device. The pnp bipolar transistor can be understood as two two-terminal p–n semiconductor devices that are coupled. As we progress to latchup analysis, the pnpn can be understood as two three-terminal devices (e.g. a pnp and an npn).

### 2.2.1 Bipolar Current Gain

In bipolar transistors, the bipolar transistor collector current is a function of the bipolar current gain $\beta$ and the base current $I_B$. The bipolar current gain can be defined as a function of the collector current [1–11],

$$\beta = \left|\frac{I_C}{I_B}\right|.$$

The bipolar current gain is a metric of the efficiency of how much base current is required to transport current from the emitter to the collector. Hence, the higher the value of the bipolar current gain $\beta$, the better the device's ability to transport current. Hence, the complimentary case of using a large base current to transport current from emitter to collector is the case when $\beta$ tends to zero; this case improves the latchup immunity. Hence, the lower the bipolar current gain, the higher the latchup immunity.

### 2.2.2 Bipolar Collector-to-Emitter Transport Factor

In bipolar transistors, the bipolar transistor collector current is a function of the bipolar collector-to-emitter transport factor $\alpha$ and the emitter current $I_E$. The bipolar collector-to-emitter transport factor $\alpha$ can be defined as the ratio of the collector current and the emitter current [1–11]:

$$\alpha = \left|\frac{I_C}{I_E}\right|.$$

The collector-to-emitter transport factor is a metric of the efficiency of transporting the minority carrier from the emitter to the collector. Hence, the highest value of the collector-to-emitter transport factor $\alpha$ is unity. Hence, the complimentary case of the inability to transport current lowers the collector-to-emitter transport factor toward zero; this case improves the latchup immunity.

From this definition, dividing both terms by the base current, the collector-to-emitter transport factor is

$$\alpha = \left|\frac{\{I_C/I_B\}}{\{I_E/I_B\}}\right|.$$

From the Kirchoff current law, the current flowing is the sum of the currents flowing into a the three transistor terminals and can be expressed as

$$\sum_{\mathrm{E,B,C}} I_j = 0$$

or

$$I_{\mathrm{E}} + I_{\mathrm{B}} + I_{\mathrm{C}} = 0.$$

These terminal relationships can be substituted into the collector-to-emitter transport factor expression,

$$I_{\mathrm{E}} = -(I_{\mathrm{B}} + I_{\mathrm{C}}),$$

hence

$$\alpha = \left| \frac{\{I_{\mathrm{C}}/I_{\mathrm{B}}\}}{\{I_{\mathrm{B}} + I_{\mathrm{C}}/I_{\mathrm{B}}\}} \right|,$$

$$\alpha = \frac{|\{I_{\mathrm{C}}/I_{\mathrm{B}}\}|}{1 + |\{I_{\mathrm{C}}/I_{\mathrm{B}}\}|}.$$

Substituting the definition of bipolar current gain,

$$\alpha = \frac{\beta}{\beta + 1}.$$

From a current amplification perspective, the term $\alpha$ is also referred to as the collector-to-emitter amplification factor. The term $\beta$ is also known as the base-to-collector current amplification factor. In latchup, current amplification is fundamentally related to the regenerative feedback and how latchup occurs.

The collector-to-emitter current amplification factor is a product of two terms. The first term is associated with the efficiency at which the emitter injects current, known as the emitter injection efficiency $\gamma$. The second term is associated with, given the current is injected, the likelihood of the current reaching the collector prior to recombination. The emitter injection efficiency for a pnp transistor is the efficiency at which the emitter injects the hole currents toward the collector region. The total emitter current is the sum of the hole current $I_{\mathrm{Ep}}$ and the electron current $I_{\mathrm{En}}$.

$$\gamma = \frac{I_{\mathrm{Ep}}}{I_{\mathrm{E}}} = \frac{I_{\mathrm{Ep}}}{I_{\mathrm{Ep}} + I_{\mathrm{En}}}.$$

The second term is the proportion of the injected carriers that can reach the collector. This is referred to as the base transport factor $B$. For a pnp transistor

$$B = \frac{I_{\mathrm{C}}}{I_{\mathrm{Ep}}}.$$

Hence, it is clear that the emitter-co-collector current amplification factor can be expressed as a product of the two amplification variables,

$$\alpha = \left| \frac{I_{\mathrm{C}}}{I_{\mathrm{E}}} \right| = B\gamma.$$

### 2.2.3 Bipolar Current Characteristics

In bipolar transistors, the bipolar transistor collector current is a function of the bipolar current gain $\beta$ and the base current $I_B$. From the bipolar current gain definition,

$$\beta = \frac{I_C}{I_B},$$

the bipolar collector current $I_C$ can simply be expressed as

$$I_C = \beta I_B.$$

The bipolar current–voltage characteristics are typically expressed as the collector current $I_C$ versus collector-to-emitter voltage $V_{CE}$ ($I_C$ versus $V_{CE}$ characteristics) [8–11]. Figure 2.1 shows the collector current $I_C$ versus collector-to-emitter voltage $V_{CE}$. In this plot, the base current $I_B$ is a parameter that is generated to produce the family of the $I_C$ versus $V_{CE}$ characteristics. Using a bipolar curve tracer, the $I_C$ versus $V_{CE}$ characteristics are generated by utilizing an established base step generator. Experimentally, from the bipolar base current steps on a bipolar transistor curve tracer, the bipolar current gain $\beta$ can be evaluated as a function of the base current by using the above equation for collector current.

Two regions exist in the bipolar current–voltage characteristics prior to the onset of avalanche multiplication. The first is the linear region and the second is the saturation region. At low collector-to-emitter voltages, the current is a function of the $V_{CE}$ voltage. At high collector-to-emitter voltages, the collector current characteristics are a constant with voltage.

In this chapter, the focus will be on parasitic bipolar transistors. Figure 2.2 shows examples of parasitic pnp bipolar $I_C$ versus $V_{CE}$ characteristics for three different semiconductor process cases. Figure 2.2 includes the following cases: (1) nonretrograde n-well, (2) a retrograde n-well and (3) a second retrograde n-well with double the dose of case 2.

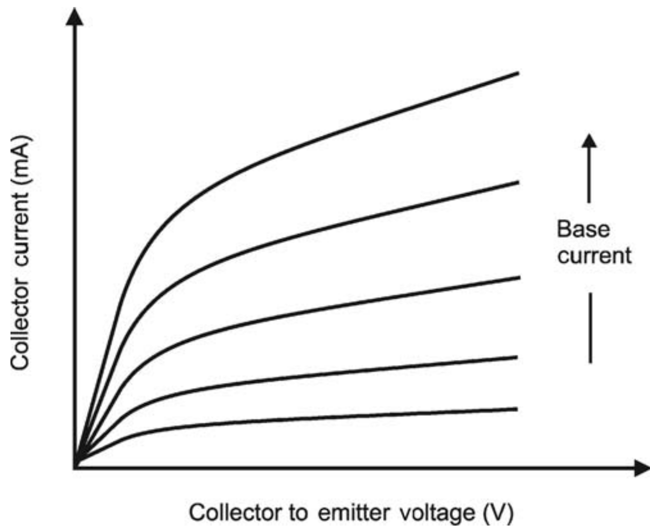



**Figure 2.1** Bipolar collector current versus collector-to-emitter voltage ($I_C$ versus $V_{CE}$) characteristics.

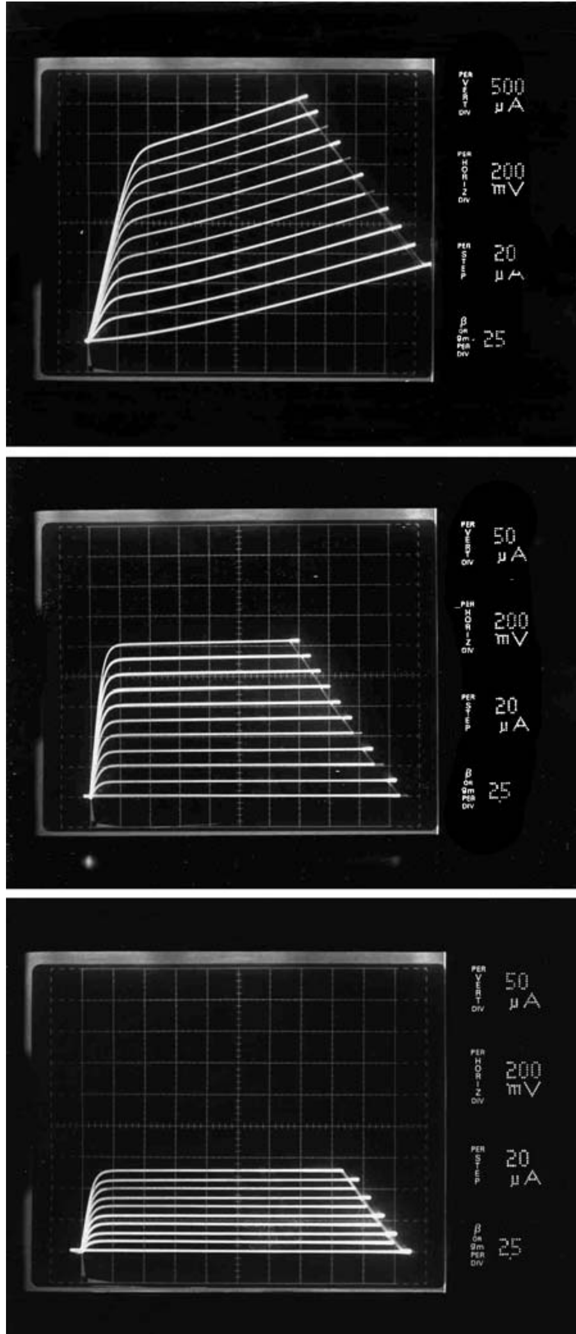



**Figure 2.2** Parasitic pnp bipolar transistor collector current versus collector-to-emitter voltage ($I_C$ versus $V_{CE}$) characteristics for three different well profiles.

## 2.2.4 Bipolar Model Gummel Plot

A means of electrical characterization of the bipolar base current and the bipolar collector current as a function of forward bias is known as the Gummel plot. Figure 2.3 shows an example of a Gummel plot. The Gummel plot shows the logarithm of the base current as a function of the emitter–base voltage. From the slope of the base current characteristics, the ideality of the diode response can be quantified. Additionally, on the Gummel plot, the logarithm of the collector current is a function of the emitter–base voltage. From the ratio of the collector current and bipolar current for any given base–emitter voltage, the dc bipolar current gain can be understood.

Note that in the diode equation, it was assumed that the diode characteristics are ideal. In actuality, a nonideality term can be placed in the exponential expression, and experimentally this can be extracted from the Gummel plot. The presence of this nonideality is indicative of volumetric or surface recombination states in the metallurgical junction. Note that from a latchup perspective, nonideality of the forward bias diode lowers the forward bias current, leading to improvement in the latchup immunity of the technology.

## 2.2.5 Bipolar Current Model–Ebers–Moll Model

The bipolar transistor is a three-region device, where current injection can flow from emitter to collector or from collector to emitter. The flow of current is a function of the bias conditions of the emitter-to-base junction and the base-to-collector junction. As a result, the current transport can be understood as two diodes that are electrically interacting through common physical region. The bipolar transistor can then be expressed as a 'coupled diode model' where the diodes are a function of the

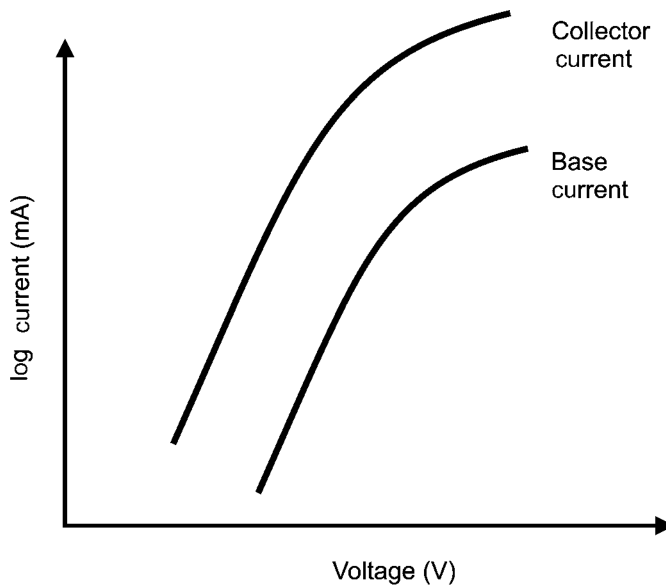



**Figure 2.3** Bipolar transistor Gummel plot.

emitter–base and the collector–base voltages:

$$I_E = I_E(V_{EB}, V_{CB}),$$
$$I_C = I_C(V_{EB}, V_{CB}).$$

Ebers and Moll developed a 'coupled diode representation' of the bipolar transistor, which can be expressed as [6]

$$I_E = I_{ES}\left\{\exp\left(\frac{qV_{EB}}{kT}\right) - 1\right\} - \alpha_I I_{CS}\left\{\exp\left(\frac{qV_{CB}}{kT}\right) - 1\right\},$$
$$I_C = \alpha_N I_{ES}\left\{\exp\left(\frac{qV_{EB}}{kT}\right) - 1\right\} - I_{CS}\left\{\exp\left(\frac{qV_{CB}}{kT}\right) - 1\right\}.$$

The emitter saturation current $I_{ES}$ is the emitter saturation current (in normal mode of operation) with the collector junction short circuited ($V_{CB} = 0$). The collector saturation current $I_{CS}$ is the collector saturation current (in the inverse mode of operation) with the emitter junction short circuited ($V_{EB} = 0$). Ebers and Moll were able to apply linear superposition by having the boundary conditions such that this could be applied.

From this form, it is clear that emitter is represented as a diode current representation as a function of the emitter–base voltage. In addition, a secondary term is represented for the inverse injection from the collector–base region. From this form, it is also clear that the collector is represented as a diode current as a function of the collector–base voltage. In addition, a secondary term is represented for the forward injection from the emitter–base region. These terms can be placed in other forms. Multiplying the first emitter equation by the collector-to-emitter transport factor [11],

$$\alpha_N I_E = \alpha_N I_{ES}\left\{\exp\left(\frac{qV_{EB}}{kT}\right) - 1\right\} - \alpha_N \alpha_I I_{CS}\left\{\exp\left(\frac{qV_{CB}}{kT}\right) - 1\right\}.$$

This expression can be substituted into the collector current expression,

$$I_C = \alpha_N I_{ES}\left\{\exp\left(\frac{qV_{EB}}{kT}\right) - 1\right\} - I_{CS}\left\{\exp\left(\frac{qV_{CB}}{kT}\right) - 1\right\},$$

resulting in the form of the collector current relationship as a function of the emitter current,

$$I_C = \alpha_N I_E - (1 - \alpha_N \alpha_I) I_{CS}\left\{\exp\left(\frac{qV_{CB}}{kT}\right) - 1\right\}.$$

For the other emitter current relationship, the above expression for the collector current is multiplied by the inverse collector-to-emitter transport factor,

$$\alpha_I I_C = \alpha_I \alpha_N I_{ES}\left\{\exp\left(\frac{qV_{EB}}{kT}\right) - 1\right\} - \alpha_I I_{CS}\left\{\exp\left(\frac{qV_{CB}}{kT}\right) - 1\right\},$$

where this is substituted into the emitter current expression below,

$$I_E = I_{ES}\left\{\exp\left(\frac{qV_{EB}}{kT}\right) - 1\right\} - \alpha_I I_{CS}\left\{\exp\left(\frac{qV_{CB}}{kT}\right) - 1\right\},$$

where the collector current is expressed as a function of the emitter current,

$$I_C = \alpha_N I_E - (1 - \alpha_N\alpha_I) I_{CS}\left\{\exp\left(\frac{qV_{CB}}{kT}\right) - 1\right\}.$$

Hence, the emitter and collector currents can be represented as transport factors,

$$I_E = \alpha_I I_C + (1 - \alpha_N\alpha_I) I_{ES}\left\{\exp\left(\frac{qV_{EB}}{kT}\right) - 1\right\},$$
$$I_C = \alpha_N I_E - (1 - \alpha_N\alpha_I) I_{CS}\left\{\exp\left(\frac{qV_{CB}}{kT}\right) - 1\right\},$$

and the set of the equations can also be shown in the form of

$$I_E = \alpha_I I_C + I_{EO}\left\{\exp\left(\frac{qV_{EB}}{kT}\right) - 1\right\},$$
$$I_C = \alpha_N I_E - I_{CO}\left\{\exp\left(\frac{qV_{CB}}{kT}\right) - 1\right\},$$

where

$$I_{EO} = (1 - \alpha_N\alpha_I) I_{ES},$$
$$I_{CO} = (1 - \alpha_N\alpha_I) I_{CS}.$$

In this expression, a reciprocity relationship can be derived, where it can be shown that [6,11]

$$\alpha_N I_{ES} = \alpha_I I_{CS}.$$

Figure 2.4 shows an example of the circuit representation for the bipolar model in the Ebers–Moll representation.

Latchup is both a dc and a transient phenomenon. The Ebers–Moll bipolar current model is relevant to latchup since it addresses large signal variations that can occur during CMOS latchup event. In the

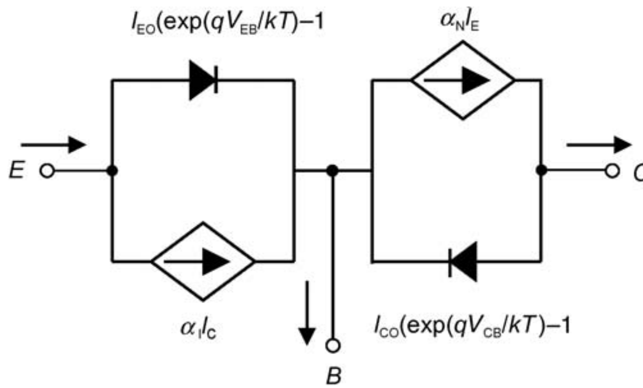



**Figure 2.4** Equivalent circuit associated with the Ebers–Moll model representation.

next section, the focus will be on investigating from a more physical perspective the source of the bipolar current gain and collector-to-emitter transport factors.

### 2.2.6 Bipolar Transistor Base Defect

In the base region of a bipolar transistor for forward injection, minority carriers are transported from the emitter to the collector. In this process, two base current terms exist in the bipolar transistor. The first process that is of interest is the EHP recombination in the base region. As the minority carriers flow from the emitter to the collector, majority carriers recombine with the minority carriers. This base current term is proportional to the total number of excess carriers in the base region. The total charge in the base region from forward injection can be expressed as [10]

$$q_{\mathrm{f}} = qA\left\{\frac{1}{2}p'_{\mathrm{b}}(0)W_{\mathrm{b}}\right\},$$

where $q$ is the charge of an electron, $A$ is the area of the emitter–base metallurgical junction, $p'_{\mathrm{b}}(0)$ is the minority carrier hole magnitude at the emitter–base junction edge and $W_{\mathrm{b}}$ is the base width. This expression assumes a triangular distribution within the base region.

The base current component is the rate at which these carriers recombine within the base region. Hence, the current is equal to the total excess charge divided by the recombination time in the base region,

$$I_{\mathrm{B}_1} = -\frac{q_{\mathrm{f}}}{\tau},$$

where base current component can be expressed as

$$I_{\mathrm{B}_1} = -\left\{\frac{qAW_{\mathrm{b}}}{2\tau}\right\}p'_{\mathrm{b}}(0),$$

where the injection condition at the junction can be expressed as

$$p'_{\mathrm{b}}(0) = p_{\mathrm{b0}}\left\{\exp\left(\frac{qV_{\mathrm{EB}}}{kT}\right) - 1\right\}.$$

The collector current is approximately the current injected at the emitter–base metallurgical junction. At the edge of the junction, this term is dominated by the carrier diffusion term. Hence, the collector current can also be expressed as a function of the slope of the carrier injected into the base region,

$$I_{\mathrm{C}} = qAD_{\mathrm{b}}\left.\frac{\mathrm{d}p'(x)}{\mathrm{d}x}\right|_{x=0};$$

as above, this can be approximated as

$$I_{\mathrm{C}} = -qAD_{\mathrm{b}}\frac{p'_{\mathrm{b}}(0)}{W_{\mathrm{b}}},$$

$$p'_{\mathrm{b}}(0) = p_{\mathrm{b0}}\left\{\exp\left(\frac{qV_{\mathrm{EB}}}{kT}\right) - 1\right\}.$$

From this analysis, a term known as the 'base defect' can be defined as the ratio of the base recombination current and the collector current [10].

$$\delta_{\mathrm{B}} = \frac{I_{\mathrm{B}_1}}{I_{\mathrm{C}}}.$$

Substituting the relationships, for $I_{\mathrm{B}_1}$ and $I_{\mathrm{C}}$, where

$$I_{\mathrm{B}_1} = -\left\{\frac{qAW_{\mathrm{b}}}{2\tau}\right\} p'_{\mathrm{b}}(0)$$

and

$$I_{\mathrm{C}} = -qAD_{\mathrm{b}}\frac{p'_{\mathrm{b}}(0)}{W_{\mathrm{b}}},$$

$$\delta_{\mathrm{B}} = \frac{I_{\mathrm{B}_1}}{I_{\mathrm{C}}} = \frac{-\left\{\dfrac{qAW_{\mathrm{b}}}{2\tau}\right\} p'_{\mathrm{b}}(0)}{-qAD_{\mathrm{b}}\dfrac{p'_{\mathrm{b}}(0)}{W_{\mathrm{b}}}} = \frac{W_{\mathrm{b}}^2}{2\{D_{\mathrm{b}}\tau\}}.$$

From the relationship of the diffusion length and the recombination time,

$$L_{\mathrm{b}} = \{D_{\mathrm{b}}\tau\}^{1/2},$$

$$\delta_{\mathrm{B}} = \frac{1}{2}\left\{\frac{W_{\mathrm{b}}}{L_{\mathrm{b}}}\right\}^2,$$

where $W_{\mathrm{b}}$ is the base width and $L_{\mathrm{b}}$ is the diffusion length of the minority carrier in the base region. In this case, the base defect is defined as a function of two physical scale lengths. As a result, one can view the base defect metric as a measure of the physical competition of diffusion transport versus recombination physics.

From a length-scale perspective, as the ratio of the base width and diffusion length increases, the base defect increases. For a fixed base width, as the diffusion length decreases, the base defect increases. Hence, as the recombination increases within the base region, the base defect increases and the net collector current decreases.

For a fixed diffusion length, as the base width decreases (e.g. dimensional scaling), the base defect decreases. Hence, as a bipolar transistor base width is scaled, the parameter of the base defect will decrease for a fixed diffusion length.

A diffusion transit time can be defined as the amount of time needed for a carrier to diffuse across a base region of width $W_{\mathrm{b}}$. Hence, we can define a diffusion transit time of

$$\{\tau_{\mathrm{t}}\}_{\mathrm{diff}} = \frac{W_{\mathrm{b}}^2}{2D_{\mathrm{b}}}.$$

From the definition of the diffusion transit time to cross the base region, this can be expressed as

$$\delta_{\mathrm{B}} = \frac{\{\tau_{\mathrm{t}}\}_{\mathrm{diff}}}{\tau}.$$

In this representation, as the ratio of two time constants, it is clear that the base defect is a measure of the competition of the minority carrier diffusion transport time versus the recombination time.

From a latchup perspective, the base defect term is an important metric associated with latchup. The base defect term influences the regenerative feedback response of the pnpn network.

As a latchup design practice, the following can be stated:

- The base defect term is a measure of the recombination current divided by the collected current.
- The base defect term is a measure of the physical competition between diffusion transport and recombination within the base region of a transistor.
- The base defect term of a bipolar transistor can be used as a metric to evaluate the latchup sensitivity of a semiconductor parasitic transistor.

## 2.2.7 Bipolar Transistor Emitter Defect

In the evaluation of the bipolar transistor, there is also a second base current term associated with the reverse carrier injection from the base to the emitter [8–10]. In the base region of a bipolar transistor for forward injection, minority carriers are transported from the emitter to the collector. In this process, two base current terms exist in the bipolar transistor. The first process, associated with the 'base defect', is recombination. The second process is associated with the back injection from the base to the emitter. The population of carrier at the edge of the emitter–base junction associated with the injection of carriers to the emitter can be expressed as

$$n_e'(0) = n_{e0}\left\{\exp\left(\frac{qV_{EB}}{kT}\right) - 1\right\}.$$

At the edge of the junction, this term is dominated by the carrier diffusion term. Hence, the base current can also be expressed as a function of the slope of the carrier injected into the emitter region,

$$I_{B_2} = qAD_e \left.\frac{dn'(x)}{dx}\right|_{x=0};$$

as above, this can be approximated as

$$I_{B_2} = -qAD_e \frac{n_e'(0)}{W_e},$$

$$n_e'(0) = n_{e0}\left\{\exp\left(\frac{qV_{EB}}{kT}\right) - 1\right\}$$

or

$$I_{B_2} = -qAD_e \frac{n_{e0}}{W_e}\left\{\exp\left(\frac{qV_{EB}}{kT}\right) - 1\right\}.$$

From the relationship of the collector current, it is known that

$$I_C = -qAD_b \frac{p_b'(0)}{W_b},$$

where

$$p_b'(0) = p_{b0}\left\{\exp\left(\frac{qV_{EB}}{kT}\right) - 1\right\}$$

or

$$I_C = -qAD_b \frac{p_{b0}}{W_b}\left\{\exp\left(\frac{qV_{EB}}{kT}\right) - 1\right\}.$$

From these two expressions of $I_{B_2}$ and $I_C$, the following can be expressed:

$$\frac{I_{B_2}}{\{D_e n_{e0}/W_e\}} = \frac{I_C}{\{D_b p_{b0}/W_b\}}.$$

From this analysis, a term known as the 'emitter defect' can be defined as the ratio of the base current injected from the base to the emitter and the collector current:

$$\delta_E = \frac{I_{B_2}}{I_C},$$

where

$$\delta_E = \frac{D_e W_b n_{e0}}{D_b W_e p_{b0}}.$$

From the relationship

$$np = n_i^2,$$

the carrier concentrations can be approximated at low-level injection as a function of the majority doping concentration. Applying the above relationship twice, it can be shown that

$$\frac{n_{e0}}{p_{b0}} = \frac{N_D}{N_A},$$

and hence, the emitter defect term can be expressed as [10]

$$\delta_E = \frac{D_e W_b N_D}{D_b W_e N_A}.$$

From this expression, it can be observed that the emitter defect term is a function of the doping concentration and the diffusion coefficient on both sides of the emitter–base junction. In addition, it is a function of the ratio of the base and the emitter width. As the ratio of the base width and the emitter width increases, the emitter defect increases (leading to higher base current). As the base width scales to smaller dimensions for a given emitter width, the emitter defect decreases. Hence, the geometric ratio of the base width and the emitter width influences the base current injection characteristics. From a latchup perspective, the base defect term is an important metric associated with latchup. The larger the emitter defect metric, the lower the current transported to the collector.

As a latchup design practice, the following can be stated:

- The emitter defect term is a measure of the base current divided by the collected current associated with back injection into the emitter.
- The emitter defect term is a function of the ratio of the doping concentrations, diffusion coefficients and emitter and base widths within a transistor.

- The emitter defect term of a bipolar transistor can be used as a metric to evaluate the latchup sensitivity of a semiconductor parasitic transistor.

### 2.2.8 Bipolar Base Current – Base Defect and Emitter Defect Relation to Bipolar Current Gain

The bipolar current gain can be expressed as a function of base defect term and the emitter defect term. For the base defect term, where the base current is the recombination current term,

$$\delta_B = \frac{I_{B_1}}{I_C},$$

where

$$\delta_B = \frac{1}{2}\left\{\frac{W_b}{L_b}\right\}^2,$$

and from the emitter defect definition,

$$\delta_E = \frac{I_{B_2}}{I_C},$$

where

$$\delta_E = \frac{D_e W_b n_{e0}}{D_b W_e p_{b0}},$$

or the emitter defect term can be expressed as

$$\delta_E = \frac{D_e W_b N_D}{D_b W_e N_A}.$$

Restating these expressions, the base current can be expressed as the sum of the two terms,

$$I_B = I_{B_1} + I_{B_2}.$$

Substituting the terms as a function of the base and emitter defect terms,

$$I_B = \delta_B I_C + \delta_E I_C = \{\delta_B + \delta_E\} I_C.$$

Defining the bipolar forward current gain as

$$\beta_f = \frac{I_C}{I_B},$$

it can be shown that

$$\beta_f = \frac{1}{\delta_B + \delta_E},$$

or another form of the forward bipolar current gain can be expressed as

$$\beta_f = \frac{1}{\frac{1}{2}\{W_b/L_b\}^2 + \{(D_e W_b N_D)/(D_b W_e N_A)\}}.$$

For the discussion of bipolar transistors, in advanced technologies, the base width is very narrow and the diffusion lengths are significant. As technologies are scaled, the recombination in the base region is very insignificant. In the limit that the base defect is much smaller than the emitter defect, the bipolar current gain can be approximated as follows:

$$\beta_f = \left.\frac{1}{\delta_B + \delta_E}\right|_{\delta_E \gg \delta_B} \approx \frac{1}{\delta_E}.$$

For latchup, the parasitic bipolar transistors can be both lateral and vertical bipolar elements. In both the cases of the parallel and the vertical elements, the scale of the base widths can be significantly higher where the base defect term is important. In these cases, both terms may be important. In the case where the base defect is significantly higher than the emitter defect term, where the bipolar current gain is recombination dominated,

$$\beta_f = \left.\frac{1}{\delta_B + \delta_E}\right|_{\delta_B \gg \delta_E} \approx \frac{1}{\delta_B}.$$

From the collector-to-emitter transport factor, this can also be expressed in terms of the base and emitter defects. From the above definition,

$$\beta_f = \frac{1}{\delta_B + \delta_E} = \frac{\alpha_f}{1 - \alpha_f},$$

the forward collector-to-emitter transport factor can be expressed as

$$\alpha_f = \frac{1}{1 + \{\delta_B + \delta_E\}}.$$

Substituting the terms,

$$\alpha_f = \frac{1}{1 + \left\{\frac{1}{2}(W_b/L_b)^2 + ((D_e W_b N_D)/(D_b W_e N_A))\right\}}.$$

From this form, it can be observed that as the base width is significantly smaller than the diffusion length and the back injection is significantly less than the forward injection, the collector-to-emitter transport factor approaches unity.

For latchup, where these terms are associated with parasitic bipolar transistors, the larger the base defect and emitter defect terms, the lower the current transport away from unity. Hence, it is an advantage for latchup to maximize the base and emitter defect terms to minimize the collector-to-emitter transport.

## 2.3 RECOMBINATION MECHANISMS

A key latchup semiconductor design discipline is the utilization of EHP recombination for latchup improvements. As a result, the understanding of EHP generation–recombination is very important. EHP recombination rate is the rate associated with the probability that a minority carrier will recombine. With each incremental motion of the carrier in the lattice, there is a probability that it recombines. Hence, the higher the probability of recombination, the shorter the time the minority

carrier survives, as well as the shorter the distance it will travel in the lattice. The recombination rate is also proportional to the number of carriers available. In latchup, the more likely that the carriers will recombine spatially and temporally, the lower the probability that the carriers can contribute to latchup regenerative feedback. Additionally, as the recombination rate increases, the distance a carrier traverses the region also decreases; this decreases the spatial extent to which latchup can also propagate through the semiconductor chip.

### 2.3.1 Shockley–Read–Hall (SRH) Generation–Recombination Model

The first recombination model is the SRH model [12,13]. In its simplest form, the SRH model assumes a single trap state energy in the silicon band gap. These trap states are spatially localized, forming independent trap states for carrier capture and are of uniform density. For conduction band-to-trap transitions, there exist two processes that are inverse processes. The first process is electron capture, and its inverse process is electron emission. For the valence band-to-trap state transitions, the same is true for holes. The capture rate for electrons is

$$r_1 = n[N_{\mathrm{t}}(1 - f(E_{\mathrm{t}}))]v_{\mathrm{th}}\sigma_{\mathrm{n}},$$

where $n$ is the number of electrons, $N_{\mathrm{t}}$ is the trap density, $f(E_{\mathrm{t}})$ is the Fermi–Dirac (F–D) distribution, $v_{\mathrm{th}}$ is the thermal velocity and $\sigma$ is the electron capture cross section. The emission rate is

$$r_2 = n[N_{\mathrm{t}}(f(E_{\mathrm{t}}))]e_{\mathrm{n}}.$$

For the holes, the capture rate is

$$r_3 = p[N_{\mathrm{t}}(f(E_{\mathrm{t}}))]v_{\mathrm{th}}\sigma_{\mathrm{p}},$$

and the hole emission rate is

$$r_4 = [N_{\mathrm{t}}(1 - f(E_{\mathrm{t}}))]e_{\mathrm{p}}.$$

These four processes are inherently coupled through the carrier density in the conduction and valence bands. When the system is in equilibrium, the electron capture process is equal to the electron emission process under the thermodynamic law of detailed balancing. When the system is not in equilibrium, the capture and emission rates do not balance, leading to a net generation rate.

Let the net generation rate be equal to the difference between the emission rate and the capture rate of the electrons. This is also equal to the difference between the hole emission and the capture rate. In order to solve for the generation rate, we must eliminate the F–D probability distribution function and substitute it into the generation rate formula. The SRH generation rate for a single trap population can be expressed as [12,13]

$$U = \frac{(pn - n_i^2)}{\tau_{\mathrm{n}}(p + p_1) + \tau_{\mathrm{p}}(n + n_1)},$$

where

$$\tau_n = \frac{1}{N_t v_{th} \sigma_n},$$
$$\tau_p = \frac{1}{N_t v_{th} \sigma_p},$$
$$n_1 = n_i \exp\left(\frac{E_t - E_i}{kT}\right),$$
$$p_1 = n_i \exp\left(\frac{E_i - E_t}{kT}\right).$$

When the generation rate $U$ is negative, it is the recombination rate. The diffusion length can be expressed as a function of the SRH recombination time and the diffusion coefficient. This can be expressed as

$$L = \sqrt{D\tau_{SRH}} \propto \sqrt{\frac{D}{\alpha_{SRH} N_D}}.$$

A key latchup consideration is that the recombination time for minority carriers is inversely proportional to the density of the minority carriers. A key latchup practice is as follows:

- Increase in the doping concentration of the semiconductor lowers the recombination time of the minority carrier.
- The recombination time of the minority carrier is inversely proportional to the doping concentration (at low doping concentration).

## 2.3.2 Auger Recombination Model

The second key latchup design practice is to utilize heavily doped regions to increase the recombination rate. Heavily doped regions are used in devices for both low resistance and high recombination rate. Heavily doped regions for latchup reduction are used in n-wells, p-wells, subcollectors, buried layers and buried grid structures. The doping concentration in silicon increases with, the recombination time. Recombination and generation physics is a function of the probability of interaction with a carrier population. A general expansion can be shown on interaction of a minority carrier with a single majority carrier (e.g. two-body interaction), a minority carrier with two majority carriers (e.g. a three-body interaction) and two minority carriers and a single majority carrier (e.g. a three-body interaction). This process can be expanded to highlight all the existing interactions. The probability of the interactions is a function of the carrier population and interaction. It can be shown that there is a second interaction known as Auger recombination rate that dominates over the SRH recombination rate when the doping concentration is high [14–16,46].

An expansion of excess carrier can be expressed as a series of time constants based on the interactions. Higher order nonlinear terms exist, which can be truncated based on a low-level injection assumption. At low-level injection, the minority carrier expansion can be expressed as follows:

$$\frac{dp'}{dt} = -\left[\frac{1}{\tau_1} + \frac{1}{\tau_2}\right] p'.$$

The first time constant $\tau_1$ is the time constant associated with SRH recombination. In this form, it can be seen that the SRH time constant is linearly proportional to the electron and hole population. The second time constant $\tau_2$ is the Auger recombination time constant. This second time constant is part of the linear response of a semiconductor, which is proportional to the square of the doping concentration. At high doping concentrations, the second time constant $\tau_2$ dominates over the first time constant $\tau_1$. Hence, the linear response of the system (e.g. the low-level injection assumption) is

$$\frac{\mathrm{d}p'}{\mathrm{d}t} = -\left[\frac{1}{\tau_1} + \frac{1}{\tau_2}\right]p',$$

where it can be expressed as a single effective time constant,

$$\frac{1}{\tau} = \left[\frac{1}{\tau_1} + \frac{1}{\tau_2}\right].$$

The first time constant is associated with the SRH model. The second time constant is the Auger recombination effect. The Auger recombination time constant is inversely proportional to the square of the doping concentrations. As a result, as the doping concentration increases, this term dominates the linear response of the excess minority carrier distribution. A key impact of Auger recombination is the significant shortening of the minority carrier diffusion length. At high doping concentrations, the diffusion coefficient is reduced due to the dopants. The Auger recombination also leads to a short lifetime. The Auger recombination time can be represented as

$$\frac{1}{\tau} = \alpha_{\mathrm{Au}} N^2,$$

where $\alpha_{\mathrm{Au}}$ is the Auger recombination constant and $N$ is the majority carrier doping concentration. The Auger recombination coefficient is $(\alpha_{\mathrm{Au}})_{\mathrm{n}} = 1.8 \times 10^{-31}\,\mathrm{cm}^6/\mathrm{s}$ and $(\alpha_{\mathrm{Au}})_{\mathrm{p}} = 8.3 \times 10^{-32}\,\mathrm{cm}^6/\mathrm{s}$ for electrons and holes, respectively [14,15]. Selberherr determined the temperature dependence of the Auger recombination process as $(\alpha_{\mathrm{Au}})_{\mathrm{n}} = 1.8 \times 10^{-31}(T/300)^{0.14}\,\mathrm{cm}^6/\mathrm{s}$ and $(\alpha_{\mathrm{Au}})_{\mathrm{p}} = 8.3 \times 10^{-32}$ $(T/300)^{0.14}\,\mathrm{cm}^6/\mathrm{s}$. Hence, the diffusion length in the heavily doped n-type region can be written in the form

$$L = \sqrt{D\tau_{\mathrm{Au}}} \propto \sqrt{\frac{D}{\alpha_{\mathrm{Au}} N_{\mathrm{D}^2}}}.$$

A key latchup design practice is to utilize large doping concentrations to minimize the recombination time:

- Doping concentrations in regions are chosen to minimize minority carrier recombination time using the Auger recombination.
- The widths of the heavily doped regions are chosen on the order of the minority carrier diffusion length (e.g. subcollectors and, buried layers).
- Substrate doping concentrations and epitaxial thicknesses are chosen to minimize the minority carrier diffusion length.

## 2.3.3 Surface Recombination Mechanisms

Surface recombination plays a role in latchup by influencing the carrier transport near silicon–dielectric interfaces. These structures can have surface recombination on the sidewalls that can influence bipolar parasitic transport and bipolar current gains. The surface recombination rate can be expressed as a function of the number of surface states per unit area, $N_{st}$, where

$$U_s = \frac{(pn - n_i^2)}{\tau_n(p + p_1) + \tau_p(n + n_1)},$$

where

$$\tau_n = \frac{1}{N_{st} v_{th} \sigma_n}$$

and

$$\tau_p = \frac{1}{N_{st} v_{th} \sigma_p}$$

In the latchup discipline, the intentional generation of trap states on silicon–dielectric surfaces in the area of current transport can reduce the parasitic bipolar current gain.

## 2.3.4 Surface Recombination Velocity

The surface recombination rate can also be expressed as surface recombination velocity $S_p$ for holes (or $S_n$ for electrons). The surface recombination velocity is related to the excess minority carrier population as

$$J_p = qp'_n S_p,$$

where the surface recombination can be expressed as a function of the surface state density, the capture cross section and the thermal velocity,

$$S_p = N_{st} \sigma_p v_{th}.$$

## 2.3.5 Recombination Mechanisms and Neutron Irradiation

A semiconductor practice unique to the latchup discipline is the usage of methods to reduce the minority carrier lifetime. It is a latchup design practice to provide a method for reduction of the minority carrier lifetime by changing the recombination mechanisms and the physical semiconductor device through irradiation [17–26]. Neutron irradiation can be used to modify the minority carrier lifetime. When a semiconductor is irradiated with a neutron source, the neutrons undergo collisions with the lattice. Neutron-induced lattice damage leads to changes in the minority carrier lifetime. The minority carrier lifetime is inversely proportional to the neutron fluence. The minority carrier lifetime can be expressed as

$$\frac{1}{\tau} = \frac{1}{\tau_0} + K\Phi,$$

where $\Phi$ is the neutron fluence (neutrons/cm$^2$) and $K$ is the lifetime damage constant (e.g. units of cm$^2$/s). From capture cross section theory, it is clear that the constant $K$ is associated with the capture cross section. From capture cross section theory,

$$\frac{1}{\tau} = N_t v_{th} \sigma.$$

Hence, it is clear that the lifetime damage constant $K$ is associated with the capture cross section. Since the collision process is a function of the energy, the rate must be evaluated by integration over the energy spectrum of the neutrons. Averaging over the energy spectrum of the neutrons, $N(E)$ (e.g. for all energies above the threshold energy $E_0$ for the collision process), we can express the lifetime damage constant as [27]

$$K = \langle \sigma \rangle = \frac{\int_{E=E_0}^{E=\infty} \sigma(E) N(E) \mathrm{d}E}{\int_{E=E_0}^{E=\infty} N(E) \mathrm{d}E}.$$

In silicon, the integration is completed for energies above 10 keV (e.g. $E_0 = 10\,\mathrm{keV}$).

The lifetime damage constant $K$ in a semiconductor is a function of the resistivity and the dopant type. The lifetime damage constant can be expressed as a function of the minority carrier concentration to the majority carrier concentration (for an n-type silicon, where n is the majority carrier) [27].

$$K = 4 \times 10^{-7} + \frac{1}{\left\{4 \times 10^4 + 5.7 \times 10^6 \left(\frac{p}{n}\right)^{0.534}\right\}} (\mathrm{cm}^2/\mathrm{s}).$$

As the recombination time decreases, the bipolar current gain in the parasitic transistors decreases; this can lead to higher latchup robustness. As the product of the bipolar current gain decreases, initiation of latchup becomes more difficult.

Table 2.1 shows the product of the bipolar current gains (e.g. $\beta_{npn}\beta_{pnp}$) as a function of a neutron fluence (cm$^{-2}$) [27]. These experimental results show that the product of the bipolar current gains can be significantly reduced with neutron irradiation of the devices. In the first case, $\beta_{npn}\beta_{pnp}$ is significantly above unity. As the neutron fluence is increased by 5×, $\beta_{npn}\beta_{pnp}$ approaches unity with a 5× reduction. At very high neutron fluence, the technology $\beta_{pnp}\beta_{npn}$ is reduced well below unity, as well as reduced by 100× of the original value.

Hence, these experimental results demonstrate that the recombination of the bulk region can be significantly altered with neutron irradiation, leading to a robust CMOS technology without a change in the design or the semiconductor process.

**Table 2.1** Neutron fluence as a function of the product of the bipolar current gains.

| Splits | Neutron fluence (cm$^{-2}$) | $\beta_{pnp}\beta_{npn}$ |
|---|---|---|
| A | $1 \times 10^{13}$ | 10 |
| B | $5 \times 10^{13}$ | 2.3 |
| C | $2 \times 10^{14}$ | 0.11 |

### 2.3.6 Recombination Mechanisms and Gold Recombination Centers

A semiconductor practice unique to latchup is the usage of dopants that serve as recombination centers [28–44]. It is a latchup design practice to introduce trap states for the reduction of the minority carrier lifetime. Trap states can be introduced with metals such as platinum and gold [27–44]. Gold was introduced into the semiconductor lattice to provide recombination centers. A primary reason for the usage of gold was the introduction of two trap states near the mid-band energy state in silicon. The first acceptor level is at $E = E_C - 0.54\,\text{eV}$. The second donor level is $E = E + 0.35\,\text{eV}$. The existing two deep trap states help the recombination in both p-type and n-type silicon. One of the primary interest in gold as a dopant was for the improvement in the diode reverse recovery time. Early experimental studies on the relationship of gold in silicon were performed by Bankowski and Forster [33], Fairfield and Gokhale [34, 35], and Clement and Voos [36]. The experimental work showed that for a logarithmic increase in the gold concentration, there was a logarithmic decrease in the recombination time (Figure 2.5).

To evaluate the effectiveness of the gold in improving latchup, the bipolar current gain can be evaluated with and without gold dopants. Estreich showed that a metric can be established for the ratio of the average bipolar current gain without gold doping ($\beta$) and the average bipolar current gain with gold doping ($\beta_{\text{Au}}$), referred to as the bipolar current gain reduction factor [27],

$$\frac{\beta}{\beta_{\text{Au}}}.$$

Estreich's experimental work showed that the reduction factor for a lateral pnp was approximately 15, whereas the reduction factor for a vertical npn was approximately 3–5 (experimental work was performed on a n− substrate wafer) [27]. A latchup semiconductor design practice is as follows:

- Dopants that have deep trap states near the mid-band of a semiconductor can serve to decrease the recombination time.
- Dopants that introduce both acceptor and donor trap states can lead to reduction in the recombination time in both the npn and the pnp transistors.

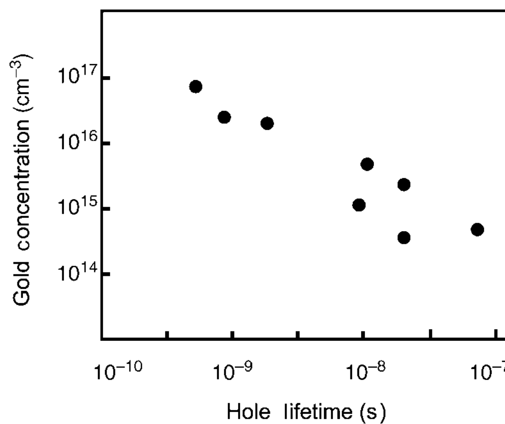



**Figure 2.5** Relationship of the gold concentration versus the hole lifetime.

- The bipolar gain reduction factor is a function of the dopant concentration, process variables and the bipolar current gain of the device.

## 2.4 PHOTON CURRENTS IN METALLURGICAL JUNCTIONS

Photons can lead to the EHP generation, which can initiate latchup [44,45]. Define a time-dependent volumetric generation rate, $g(t)$,

$$g(t) = g_0 \gamma(t),$$

where $g_0$ is the volumetric generation rate per 'rad' per unit volume and $\gamma(t)$ is the dose rate. The volumetric generation rate per unit volume can be expressed as the number of electron–hole pairs generated per rad per unit volume. The $\gamma(t)$ dose rate can be expressed as rad/s. (*Note*: 1 rad equals 100 ergs of energy within 1 g of material.) For silicon, this can be expressed as [44]

$$g_0 = 4 \times 10^{13} \mathrm{EHP}(\mathrm{rad\,cm}^3)$$

For an abrupt p–n metallurgical junction, Wirth and Rogers derived the photocurrent response as [44]

$$I(t) = qA\left[x_\mathrm{d} g(t) + \int_0^t g(t-\lambda)\left\{\frac{\sqrt{D_\mathrm{n}}\exp(-\lambda_\mathrm{n}/\tau_\mathrm{n}) + \sqrt{D_\mathrm{p}}\exp(-\lambda_\mathrm{p}/\tau_\mathrm{p})}{\sqrt{\pi\lambda}}\right\}\mathrm{d}\lambda\right],$$

where $x_\mathrm{d}$ is the width of the depletion region and $A$ is the junction area. The first term is the total generation within the metallurgical junction depletion region and the second term is associated with the diffusion of the minority carriers to the depletion region. In this analysis of Wirth and Rogers, it is assumed that the depletion region is not distorted from the EHP generation itself. (*Note*: the depletion region width is a constant.) Note that in the case of particle tracks, this is not always valid due to the 'funneling' effect. The first term is associated with the 'prompt photocurrent' and the second term, the convolution integral, is the 'primary photocurrent' term associated with the slower diffusion process. The nature of the second term, the convolution integral, is a function of the pulse width of the generation source and the diffusion times in the medium. For the case of a square pulse, the generation source can be defined as a constant magnitude $G$ for a time of $t = 0$ to $t = t_\mathrm{p}$,

$$g(t) = G \quad 0 \le t \le t_\mathrm{p},$$

where the solution of the integral can be expressed as the error function erf ($\chi$). The current during the time of the pulse ($t = 0$ to $t = t_\mathrm{p}$) can be expressed as

$$I(t) = qAG\left[x_\mathrm{d} + L_\mathrm{p}\mathrm{erf}(\sqrt{t/\tau_\mathrm{p}}) + L_\mathrm{n}\mathrm{erf}(\sqrt{t/\tau_\mathrm{n}})\right],$$

and for times longer than the pulse width, the junction current is

$$I(t) = qAG[x_\mathrm{d} + L_\mathrm{p}\mathrm{erf}(\sqrt{t/\tau_\mathrm{p}}) - \mathrm{erf}(\sqrt{(t-t_\mathrm{p})/\tau_\mathrm{p}}) + L_\mathrm{n}\mathrm{erf}(\sqrt{t/\tau_\mathrm{n}}) - \mathrm{erf}(\sqrt{(t-t_\mathrm{p})/\tau_\mathrm{n}})].$$

In a p-well to substrate metallurgical junction region on an n− substrate region, the dominant generation occurs in the n− substrate. In this case, the hole diffusion length is significantly much larger

than the electron diffusion length. The above equation can be simplified by addressing the collection of holes in the n− substrate,

$$I(t) \cong qAG\left[x_{\mathrm{d}} + L_{\mathrm{p}}\mathrm{erf}(\sqrt{t/\tau_{\mathrm{p}}})\right],$$

and for times longer than the pulse width, the junction current is

$$I(t) = qAG\left[x_{\mathrm{d}} + L_{\mathrm{p}}\mathrm{erf}(\sqrt{t/\tau_{\mathrm{p}}}) - \mathrm{erf}(\sqrt{(t-t_{\mathrm{p}})/\tau_{\mathrm{p}}})\right].$$

In an n-well-to-substrate junction on a p− substrate region, the dominant generation occurs in the p− substrate. In this case, the electron diffusion length is significantly much larger than the hole diffusion length. The above equation can be simplified by addressing the collection of holes in the p− substrate. The current during the time of the pulse ($t = 0$ to $t = t_{\mathrm{p}}$) can be expressed as

$$I(t) \cong qAG\left[x_{\mathrm{d}} + L_{\mathrm{n}}\mathrm{erf}(\sqrt{t/\tau_{\mathrm{n}}})\right],$$

and for times longer than the pulse width, the junction current is

$$I(t) \cong qAG\left[x_{\mathrm{d}} + L_{\mathrm{n}}\mathrm{erf}(\sqrt{t/\tau_{\mathrm{n}}}) - \mathrm{erf}(\sqrt{(t-t_{\mathrm{p}})/\tau_{\mathrm{n}}})\right].$$

For the case of a p− substrate, when the pulse width is significantly smaller than the electron recombination time, the error function can be approximated, leading to

$$I(t) \cong qAG\left[x_{\mathrm{d}} + \frac{2}{\sqrt{\pi}}L_{\mathrm{n}}\sqrt{\frac{t}{\tau_{\mathrm{n}}}}\right],$$

for the time of $0 < t < t_{\mathrm{p}}$, and for times longer than the pulse width $t_{\mathrm{p}}(t > t_{\mathrm{p}})$, the current is

$$I(t) \cong qAG\left[\frac{2}{\sqrt{\pi}}L_{\mathrm{n}}\left\{\frac{t_{\mathrm{p}}}{\sqrt{t}\sqrt{\tau_{\mathrm{n}}}}\right\}\exp\left\{-\frac{(t-t_{\mathrm{p}})}{\tau_{\mathrm{n}}}\right\}\right].$$

From the current equation, a diffusion distance and volume of interest can be defined. A time-dependent region can be defined as

$$x(t) \cong \left[x_{\mathrm{d}} + \frac{2}{\sqrt{\pi}}L_{\mathrm{n}}\sqrt{\frac{t}{\tau_{\mathrm{n}}}}\right]$$

for time less than the pulse width and as

$$x(t) \cong \left[\frac{2}{\sqrt{\pi}}L_{\mathrm{n}}\left\{\sqrt{\frac{t_{\mathrm{p}}}{\sqrt{t}\sqrt{\tau_{\mathrm{n}}}}}\right\}\exp\left\{-\frac{(t-t_{\mathrm{p}})}{\tau_{\mathrm{n}}}\right\}\right]$$

for times longer than the pulse width. For a constant generation rate, it is clear that a region can be defined where there is a collection distance (and volume) that influences the total current flowing to the junction.

## 2.5 AVALANCHE BREAKDOWN

Avalanche breakdown is important for latchup, since breakdown mechanisms can serve as trigger conditions for the initiation of latchup. Avalanche breakdown can occur in both parasitic bipolar transistors formed by diffusions and its corresponding wells and substrates, or MOSFET transistors themselves.

### 2.5.1 Bipolar Transistor Breakdown

In bipolar transistors, there are different breakdown criteria that are established because of multiple metallurgical junctions and the potential state of the additional electrode. Additionally, analytical relationships exist between the different breakdown states.

*Emitter–base breakdown voltage with open collector* ($BV_{EBO}$): The emitter–base breakdown voltage is defined as the breakdown voltage when the collector is unbiased or floating; this is referred to as $BV_{EBO}$.

*Collector–base breakdown voltage with open emitter* ($BV_{CBO}$): The collector-to-base breakdown voltage is defined as the voltage when the base is unbiased or floating; this is referred to as $BV_{CBO}$. To obtain $BV_{CBO}$, the methodology is to extract the value from a plot of $M - 1$ versus $V_{CB}$ (collector–base voltage). Using the Miller equation,

$$M = \frac{1}{1 - (V_{CB}/BV_{CBO})^n},$$

the value of $BV_{CBO}$ is extracted.

*Collector–emitter breakdown voltage with open base* ($BV_{CEO}$): The collector-to-emitter breakdown voltage is defined as the voltage when the base is unbiased or floating; this is referred to as $BV_{CEO}$. The value of breakdown value $BV_{CEO}$ can be extracted experimentally or from $BV_{CBO}$. The breakdown voltage $BV_{CEO}$ can be obtained from $BV_{CBO}$ by

$$BV_{CEO} = BV_{CBO}(1 - \alpha)^{1/n}.$$

*Collector–emitter breakdown voltage with base resistor* ($BV_{CER}$): The collector-to-emitter breakdown voltage is defined as the breakdown voltage when the base is unbiased or floating; this is referred to as $BV_{CEO}$. Given a base resistor $R_b$, the $BV_{CER}$ value can be derived from the $BV_{CBO}$ value as

$$BV_{CER} = BV_{CBO} \sqrt[n]{1 - \frac{\alpha}{1 + (kT/q)(1/R_b)(1/I_E)}}.$$

From the latchup perspective, avalanche breakdown of metallurgical junctions associated with inducing or establishing regenerative feedback within the pnpn structure is the highest concern. In the case of latchup, the avalanche breakdown of the base–collector junction of the pnp and the npn transistor can lead to triggering of latchup, as will be discussed in the following sections.

### 2.5.2 MOSFET Avalanche Breakdown

In MOSFET devices, avalanche breakdown occurs in the metallurgical junction formed between the MOSFET drain and its supporting structure [47]. For an n-channel MOSFET, avalanche breakdown

occurs in the n+ source/drain implant to p− epitaxy (or p− substrate) metallurgical junction. From the Townsend criteria,

$$\int \alpha \, dx = 1,$$

where the impact ionization is integrated over the physical space where a nonzero electric field exists. From this form, the avalanche multiplication factor $M$, which can be related to the impact ionization coefficient integrated over the depletion width, can be expressed as

$$M = \frac{1}{1 - \int \alpha \, dx}.$$

In a depletion region, the peak electric field is maximum at the center of the dipole. In the analysis of the MOSFET, the model is simplified to express it as a function of the depletion width. Hence, it can be integrated over the integral and stated in the form

$$M = \frac{1}{1 - \alpha x_d}.$$

From semiconductor physics, it is known that the depletion width can be expressed as a power of the applied voltage

$$x_d \propto (V_D)^n.$$

The power of this relationship is a function of the doping profile at the metallurgical junction. In this form, it is also clear that when the voltage is greater than the avalanche breakdown voltage, the multiplication factor should increase rapidly. Hence, heuristically, it is clear that the multiplication expression should satisfy the form

$$M = \frac{1}{1 - (V_D/V_{av})^n}.$$

In this form, as the drain voltage approaches the avalanche breakdown voltage, the multiplication factor approaches infinity. As the gate electrode voltage in a MOSFET is increased, the electric field in the drain region increases. From the expression,

$$M \approx \frac{1}{1 - \alpha(E)x_d} = \frac{1}{1 - \alpha_0 x_d \exp\left\{-\frac{B}{E}\right\}}.$$

Substituting a voltage condition where the electric field is the voltage over a physical distance, the voltage across the depletion region is the drain voltage minus the drain saturation voltage and the distance is the depletion region,

$$M \approx \frac{1}{1 - (\alpha_0 x_d) \exp\{-[Bx_d/\{V_D - V_{d_{sat}}\}]\}},$$

where the drain saturation velocity is expressed as

$$V_{d_{sat}} = \frac{V_G - V_t}{a + b(V_G - V_t)},$$

where the saturation velocity is the voltage drive divided by a two-parameter expression in the denominator and the voltage drive is the gate voltage minus the MOSFET threshold voltage.

The avalanche generation current can then be calculated from multiplication factor and the current flowing through the drain junction. The current flowing through the drain structure is the MOSFET current flowing through the surface region (e.g. MOSFET source-to-drain current) as well as the current flowing from the parasitic bipolar transistor formed from the MOSFET source, epitaxy region and the MOSFET drain forming a lateral npn transistor. In this case, the MOSFET drain and source serve as the bipolar junction transistor collector and emitter, and the epitaxial region serves as a base. The avalanche current can be expressed as

$$I_{\mathrm{av}} = (M-1)I = (M-1)\{I_{\mathrm{DS}} + I_{\mathrm{C}}\}.$$

The avalanche generation current flows to the p− substrate region of the n-channel MOSFET structure. A portion of this current flows to the base of the lateral parasitic npn transistor serving as the emitter–base current, while the rest of the current flows to the substrate region as substrate current. Hence, the avalanche current can be defined as the sum of the substrate current (flowing to the substrate contact) and the base current of the lateral npn transistor (serving as base drive current),

$$I_{\mathrm{av}} = I_{\mathrm{s}x} + I_{\mathrm{B}}.$$

Then the substrate current can be estimated as

$$I_{\mathrm{s}x} = I_{\mathrm{av}} - I_{\mathrm{B}} = (M-1)\{I_{\mathrm{DS}} + I_{\mathrm{C}}\} - I_{\mathrm{B}};$$

hence we can express the avalanche current and the substrate current as a function of the impact ionization, the depletion width, drain voltage, drain saturation voltage and MOSFET n-channel and parasitic npn current [47]:

$$I_{\mathrm{av}} \approx \left(\frac{1}{1-(\alpha_0 x_{\mathrm{d}})\exp\{-[Bx_{\mathrm{d}}/\{V_{\mathrm{D}} - V_{\mathrm{d_{sat}}}\}]\}} - 1\right)\{I_{\mathrm{DS}} + I_{\mathrm{C}}\}$$

and

$$I_{\mathrm{s}x} \approx \left(\frac{1}{1-(\alpha_0 x_{\mathrm{d}})\exp\{-[Bx_{\mathrm{d}}/\{V_{\mathrm{D}} - V_{\mathrm{d_{sat}}}\}]\}} - 1\right)\{I_{\mathrm{DS}} + I_{\mathrm{C}}\} - I_{\mathrm{B}}.$$

## 2.6 VERTICAL BIPOLAR TRANSISTOR MODEL

In latchup, there exist both vertical and lateral bipolar transistors. Evaluation of bipolar transistor current density from the emitter to the collector for an npn transistor can be represented as [47]

$$J = \frac{qn_{\mathrm{i}}^2\left[\exp\left(\frac{qV_{\mathrm{BC}}}{kT}\right) - \exp\left(\frac{qV_{\mathrm{BE}}}{kT}\right)\right]}{\int_0^{y_{\mathrm{B}}} (p(y)/D_{\mathrm{n}}(y))\mathrm{d}y},$$

where this can be expressed as the current in the form

$$I = I_{\mathrm{s}}\left[\exp\left(\frac{qV_{\mathrm{BC}}}{kT}\right) - \exp\left(\frac{qV_{\mathrm{BE}}}{kT}\right)\right],$$

where $I_{\mathrm{s}}$ is the saturation current, $V_{\mathrm{BC}}$ is the base–collector voltage and $V_{\mathrm{BE}}$ is the emitter–base voltage. The saturation current is a function of the electron charge ($q$), the area of the metallurgical junction ($A$), the intrinsic carrier concentration ($n_{\mathrm{i}}$), the effective diffusion coefficient ($\langle D_{\mathrm{n}}\rangle$) and the base charge ($Q_{\mathrm{B}}$),

$$I_{\mathrm{s}} = \frac{q^2A^2n_{\mathrm{i}}^2\langle D_{\mathrm{n}}\rangle}{Q_{\mathrm{B}}},$$

where the effective diffusion coefficient ($\langle D_{\mathrm{n}}\rangle$) is expressed as

$$\langle D_{\mathrm{n}}\rangle = \frac{qA\int_0^{y_{\mathrm{B}}} p(y)\mathrm{d}y}{\int_0^{y_{\mathrm{B}}} (p(y)/D_{\mathrm{n}})\mathrm{d}y},$$

and stored base charge is expressed as the integral of the hole carrier population integrated over the base width ($y_{\mathrm{B}}$),

$$Q_{\mathrm{B}} = qA\int_0^{y_{\mathrm{B}}} p(y)\mathrm{d}y.$$

The base current can be expressed in a similar form where current is a function of the excess charge in the base divided by the recombination time,

$$Q'_{\mathrm{B}} = qA\int_0^{y_{\mathrm{B}}} [n(y) - n_0]\mathrm{d}y,$$

and the base recombination current can be expressed as

$$I_{\mathrm{r_B}} = \frac{Q'_{\mathrm{B}}}{\tau_{\mathrm{n}}},$$

where $\tau_{\mathrm{n}}$ is the base recombination time.

Latchup will be a function of the geometrical parameters, doping concentration and the bias conditions in the metallurgical junctions. In this section, the expression of the bipolar transistor is represented as a one-dimensional structure. In practice, latchup comprises both vertical and lateral components. Modeling of the latchup parasitic transistors as a vertical or lateral element is valid under the following cases:

- A model can be represented as a vertical bipolar parasitic transistor when the transistor spacing is a significant distance from the n-well to p-well edge (e.g. low lateral current density) compared to the vertical base width.
- A model can be represented as a vertical bipolar transistor when the structure is bordered by isolation structure significantly deeper than the emitter region (e.g. shallow trench isolation (STI) or trench isolation (TI)).

- A model can be represented as a vertical bipolar transistor when there is no lateral transport due to isolation structure that extends beyond the well regions (e.g. deep trench).

In the case of the vertical transistor, the model of the area of the transistor is the diffusion area. In the case of the lateral device, the model of the area may be the edge area of the emitter structure. In both cases, there is an idealization of the effective area.

## 2.7 LATERAL BIPOLAR TRANSISTOR MODELS

Latchup involves the parasitic bipolar transistors formed from the n-channel and p-channel MOSFET devices and the well structures. As a result, the interaction involves typically at least one lateral bipolar parasitic element. The nature of the lateral transistor model is a function of the isolation and the well structure [27].

In technology generations that utilized LOCOS isolation, the junction depth of the n-channel and p-channel transistors extended below the LOCOS isolation depth within the silicon region. For example, when the LOCOS isolation thickness was typically 0.33 μm, half of the LOCOS isolation (e.g. 0.17 μm) extended below the silicon surface. In a 0.8-μm technology, the MOSFET device source and drain region extended below the isolation depth (e.g. 0.5 μm deep). Hence, a p-channel MOSFET diffusion extended below the isolation region, allowing lateral current transport to flow from the edge of the p-diffusion to the n-well metallurgical junction. Additionally, the n-channel MOSFET source and drain region also extended below the LOCOS isolation region, also allowing for lateral electron current flow along the device surface under the isolation region.

Lateral transistor models are important for CMOS latchup. The lateral transistor models are inherently two dimensional in nature [27]. Early models to quantify the physics of the lateral transistor focused on the separation of the vertical and lateral components of current flow [27]. Figure 2.6 shows a simple model that separates the vertical flow from the lateral current flow.

A simple model can be developed where the components of the vertical and lateral currents are separated into first and second components for a p-channel MOSFET within an n-well region (contained within the p− substrate). Assuming a p+ diffusion is emitter, the hole current flowing

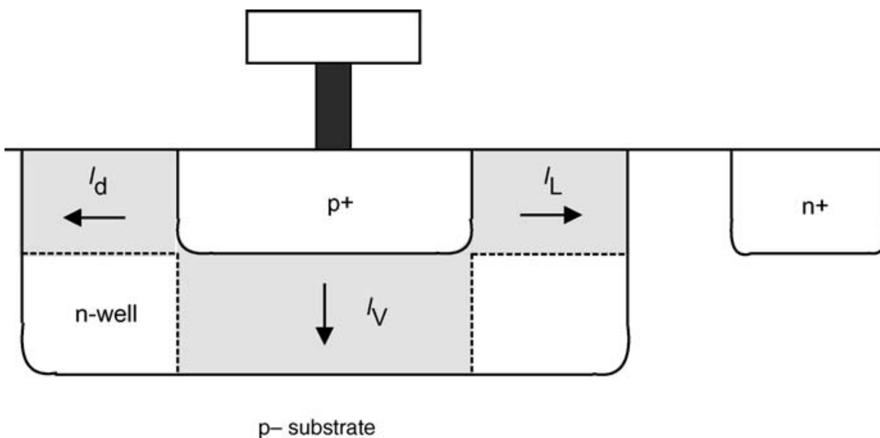



**Figure 2.6** A model for vertical and lateral transport in a lateral transistor structure.

laterally from the p-diffusion edge can be regarded as the lateral current flow $I_L$. The hole current flowing from the bottom of the p+ diffusion region can be regarded as the vertical hole current flow $I_V$.

A first model by Tsang and Busen addressed the issue of lateral injection and the vertical injection in the lateral pnp transistor element [48]. It was noted by Tsang and Busen that the ratio of the lateral and vertical components has an influence on the bipolar current gain of the lateral transistor structure. Semiconductor device models for lateral transistors were also developed by Lindmayer and Schneider [49], Chou [50,51], Long [52], Kidron [53], Seltz and Kidron [54], Fulkerson [55], Parameswaran and Tyagi [56], Demassa and Rispin [57], Berger and Dreckmann [58], and Estreich [27].

A majority of the early lateral pnp device models focused on a narrow base width geometry where the hole diffusion length was significantly greater than the effective base width. For example, in latchup analysis in a dual-well p− substrate CMOS technology, this is the case where the p-channel MOSFET to n-well edge (e.g. n-well to p− substrate junction) spacing is small compared to the hole diffusion length in the n-well. With a narrow base width assumption, these models were able to assume that n+ buried regions and electric fields in the base region had negligible impact on the transport of the hole carriers near the surface.

## 2.7.1 Lindmayer–Schneider Model

The Lindmayer–Schneider model utilized the concept of the separation of the vertical and lateral current components as well as evaluated the unity bipolar current gain cutoff frequency dependence of a lateral pnp transistor [49]. Berger and Dreckmann focused on the quantification of the base current components [58]. Fulkerson utilized a full two-dimensional analysis of the current flow in the lateral transistor [55]. The significance of the work was the demonstration that the simplification of model into lateral and vertical components has validity in the geometry that was evaluated.

Lindmayer and Schneider developed a lateral transistor model that can be applied to the parasitic BJT, as shown in Figure 2.7 [49]. Let us first assume that the vertical pnp base width (e.g. distance between the p+ diffusion metallurgical junction and the n-well to substrate metallurgical junction) is significantly greater than the hole diffusion length, as well as the hole diffusion length is significantly greater than the lateral base width spacing. Given a narrow well region, one can assume that the current

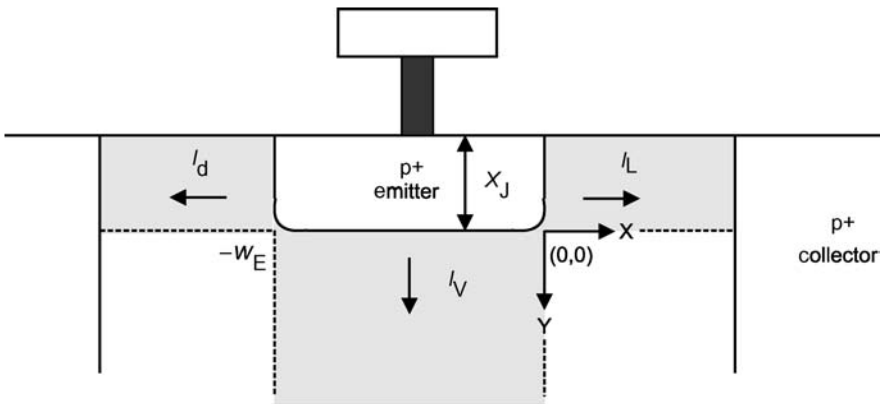



**Figure 2.7** Lateral pnp transistor cross section.

from both edges flows to the well edge. Assuming the hole current transport is primarily from hole diffusion, the lateral and vertical current flows can be estimated as

$$I_L = -2qx_J LD_p \frac{P_0}{W_B},$$
$$I_V = -qW_E LD_p \frac{P_0}{L_p},$$

where $q$ is the electron charge, $x_J$ is the junction depth, $L$ is the lateral width of the diffusion, $W_B$ is the lateral pnp base width, $W_E$ is the width of the p+ diffusion, $L_p$ is the hole diffusion length and $P_0$ is the minority hole carrier injection at the p+ diffusion metallurgical junction. In most cases, the p-diffusion is only adjacent to the n-well edge. Given that the p− diffusion is only adjacent to one edge of the n-well structure, we can assume that the current flowing away from the lateral interior edge also contributes to the total vertical current flow. For latchup, the Lindmayer–Schneider model can be modified as shown:

$$I_L = -qx_J LD_p \frac{P_0}{W_B}$$
$$I_V \approx -q(W_E + x_J)LD_p \frac{P_0}{L_p}.$$

A second correction can also be applied to address the isolation depth. Given an isolation depth, $X_{ISO}$, the model can be expressed as [59]

$$I_L = -q(x_J - x_{ISO})LD_p \frac{P_0}{W_B}$$

and

$$I_V \approx -q(W_E + \{x_J - x_{ISO}\})LD_p \frac{P_0}{L_p}.$$

## 2.7.2 Bipolar Current Gain with Lateral and Vertical Contributions

Assuming a lateral pnp transistor, the lateral bipolar current gain can be solved from the base transport factor (emitter-to-collector transport factor). In the case of lateral and vertical contributions, the definition for the bipolar current gain is modified. A general heuristic definition of the bipolar current gain is the ratio of the collected current at the collector junction and the current lost in the base region. The current that reaches the collector is the product of the collector-to-emitter transport term and the current lateral current, $I_L$,

$$I_C = \alpha_l I_L.$$

In the base region, there are three terms. The first term is associated with the carriers that recombine in the lateral region. This term can be expressed as

$$I_{BL} = (1 - \alpha_l)I_L.$$

The second base current term is associated with the lateral current flowing in the opposite direction. The third base current term is associated with the vertical current that flows away from the collector region. Then the definition can be expressed as [27]

$$\beta_{\mathrm{pnp}} = \frac{I_{\mathrm{C}}}{I_{\mathrm{B}}} = \frac{\alpha_{\mathrm{l}} I_{\mathrm{L}}}{(1-\alpha_{\mathrm{l}})I_{\mathrm{L}} + I_{\mathrm{D}} + I_{\mathrm{V}}}.$$

Assuming the term of the current flowing away laterally in the opposite direction is significantly less than the vertical current, this can be simplified as follows:

$$\beta_{\mathrm{pnp}} = \frac{\alpha_{\mathrm{l}} I_{\mathrm{L}}}{(1-\alpha_{\mathrm{l}})I_{\mathrm{L}} + I_{\mathrm{V}}}.$$

An alternate form of the bipolar current gain with both the extra terms is

$$\beta_{\mathrm{pnp}} = \frac{\alpha_{\mathrm{l}}}{(1-\alpha_{\mathrm{l}}) + (I_{\mathrm{D}} + I_{\mathrm{V}}/I_{\mathrm{L}})},$$

or ignoring the lateral term,

$$\beta_{\mathrm{pnp}} = \frac{\alpha_{\mathrm{l}}}{(1-\alpha_{\mathrm{l}}) + (I_{\mathrm{V}}/I_{\mathrm{L}})}.$$

In the first form, the vertical current flow degrades the lateral bipolar current. In the second form, it is clear that the second term in the denominator is the ratio of the vertical and lateral currents. If the vertical current is significantly less than the lateral current, then the second term in the denominator tends to zero and the definition reduces to

$$\beta_{\mathrm{l}} = \frac{\alpha_{\mathrm{l}}}{1-\alpha_{\mathrm{l}}}.$$

In the above relationship, the vertical current does not contribute to the latchup current, but actually improves the latchup robustness by injecting into the substrate, providing a decrease in the total regenerative feedback current. In the above development, it is assumed that the vertical current recombines in the substrate and does not contribute to the latchup lateral bipolar current. Given that the pnp element is within an n-well structure where there are both lateral and vertical current contributions to the pnp, the definition of the bipolar current gain is different. It is assumed that the vertical and lateral currents both contribute to the total pnp bipolar current gain. The current that reaches the collector is the product of the collector-to-emitter transport term and the lateral current. $I_{\mathrm{L}}$ and the vertical current

$$I_{\mathrm{C}} = \alpha_{\mathrm{l}} I_{\mathrm{L}} + \alpha_{\mathrm{V}} I_{\mathrm{V}}.$$

In the base region, the first term is associated with the carriers that recombine in the lateral region. This term can be expressed as

$$I_{\mathrm{BL}} = (1-\alpha_{\mathrm{l}})I_{\mathrm{L}}.$$

The second term is the current that recombines in the vertical region,

$$I_{\mathrm{BV}} = (1-\alpha_{\mathrm{V}})I_{\mathrm{V}}.$$

The third term, $I_D$, is the current that flows in the opposite direction and does not contribute to the bipolar pnp current gain. This last term is evaluated as a loss term. The generalized form of the bipolar current gain in this representation can be defined as shown:

$$\beta_{pnp} = \frac{I_C}{I_B} = \frac{\alpha_l I_L + \alpha_V I_V}{(1 - \alpha_l)I_L + I_D + (1 - \alpha_V)I_V}.$$

In this definition, the bipolar current gains of both the lateral and vertical components contribute to the total bipolar gain. This definition may be a more appropriate model in some physical structures (e.g. merged triple well). In merged triple well, the npn transistor can have both lateral and vertical components that are in separable in the electrical measurement due to the merging of the lateral n-well structure and the n-type buried layer region; additionally, both the components add to the regenerative feedback.

## 2.7.3 Lateral Bipolar Transistor Models – Nonfield-Assisted

In the early pnp lateral bipolar models, the issue of buried layers and field-assisted phenomena was assumed negligible due to a narrow base width assumption. For latchup, this cannot always be assumed for regions where the base width is significantly greater than the diffusion length. Assume that the current components of the lateral and vertical components are separated into three terms, $I_L$, $I_D$ and $I_V$, respectively. Assuming an n− substrate, the lateral pnp structure is a p+ diffusion and a p-well region, where the base is the n− substrate. The lateral current $I_V$ is the current that flows from the emitter to the collector region. The lateral current $I_D$ is the current that flows from the emitter to the opposite direction of the collector. It is assumed that this current flows away from the collector and recombines. The vertical current $I_V$ is the current that flows from the emitter to the 'substrate'. In the lateral device, the emitter area is the product of the emitter depth $X_J$ and the width of the emitter, $W_e$, on the edge that faces the collector,

$$A_V = W_e X_J.$$

Assuming no lateral electric field, the one-dimensional analysis of the hole excess carrier population in the base region can be represented as follows:

$$D_p \frac{d^2 p'(x)}{dx^2} - \frac{1}{\tau_p} p'(x) = 0.$$

Expressing as a function of the diffusion length $L_p$ and thermal voltage $(kT/q)$, the minority carrier concentration can be expressed as

$$\frac{d^2 p'(x)}{dx^2} - \frac{1}{L_p^2} p'(x) = 0.$$

The solution to the differential equation can be expressed as

$$p'(x) = A e^{S_1 x} + B e^{S_2 x},$$

where the roots of the equation can be expressed as

$$S_{1,2} = \pm \frac{1}{L_p}$$

or alternatively

$$S_{1,2} = \lambda_1 \pm \lambda_2.$$

From the boundary conditions at $x = 0$ (at the emitter) and $x = W_B$ (at the collector), the hole carrier populations are

$$p'(0) = p_{n0}\left\{\exp\left(\frac{V}{V_T}\right) - 1\right\}$$

and

$$p'(W_B) = -p_{n0}.$$

The solution for the excess hole carrier population can be expressed as follows [27]:

$$p'(x) = p_{n0}\,\theta\left\{\exp\left(\frac{x}{L_p}\right) - \exp\left(\frac{W_B}{L_p}\right)\frac{\sinh(x/L_p)}{\cosh(x/L_p)}\right\} - p_{n0}\left\{\frac{\sinh(x/L_p)}{\sinh(W_B/L_p)}\right\},$$

where

$$\theta = \exp\left(\frac{V}{V_T}\right) - 1.$$

### 2.7.4 Lateral Bipolar Transistor Models – Nonfield-Assisted Base Transport Factor

Assuming a lateral pnp transistor, the base transport factor is key to the understanding of the minority carrier diffusion. Figure 2.7 shows a lateral pnp transistor in a CMOS semiconductor process. At the boundary conditions of the base region, $x = 0$ (at the emitter) and $x = W_B$ (at the collector), the hole carrier populations are

$$p'(0) = p_{n0}\left\{\exp\left(\frac{V}{V_T}\right) - 1\right\}$$

and

$$p'(W_B) = -p_{n0}.$$

The solution for the excess hole carrier population can be expressed as follows [27]:

$$p'(x) = p_{n0}\left\{\exp\left(\frac{x}{L_p}\right) - \exp\left(\frac{W_B}{L_p}\right)\frac{\sinh(x/L_p)}{\cosh(x/L_p)}\right\} - p_{n0}\left\{\frac{\sinh(x/L_p)}{\sinh(W_B/L_p)}\right\},$$

where

$$\theta = \exp\left(\frac{V}{V_T}\right) - 1.$$

The diffusion current of hole carriers can be expressed as

$$J_p(x) = -qD_p\frac{dp'(x)}{dx}.$$

The base transport factor (also known as the collector-to-emitter transport factor) is expressed as the two edges of the base region. Hence, the base transport factor is evaluated by the current density arriving at the collector, normalized to the current injected at the emitter junction. For a lateral pnp transistor, the base transport factor can be expressed as

$$\alpha_1 = \frac{J_p(x = W_B)}{J_p(x = 0)}.$$

Solving for the diffusion lengths, the base transport factor is expressed as

$$\alpha_1 = \operatorname{sech}(W_B/L_p).$$

## 2.8 LATERAL BIPOLAR TRANSISTOR MODELS WITH ELECTRIC FIELD ASSIST

In the early pnp lateral bipolar models, the issue of buried layers and field-assisted phenomena was assumed negligible due to a narrow base width assumption. For latchup, this cannot always be assumed for regions where the base width is significantly greater than the diffusion length.

Semiconductor device models that included the 'buried layer' and electric field enhanced transport for lateral transistors were developed by Long [52], Demassa and Rispin [57], and Parameswaran and Tyagi [56], and Estreich [27]. In CMOS latchup, when a lateral electric field exists in the base region, the ratio of the minority carrier lateral and vertical transport is influenced. For example, if there is a lateral electric field in the direction of the hole transport, the hole current will increase; this will influence the ratio of vertical and lateral transport.

Today, there are a few issues why the lateral electric field assist is important. First, in a modern CMOS technology, this can be evident in a dual-well or triple-well technology due to physical designs. Additionally, a new concept is the utilization of the influence of the lateral electric field assist in lateral bipolar transport as a means to provide 'active guard rings' where the electric field hampers the lateral bipolar current gain to reduce the minority carrier transport between injection sources and latchup sensitive circuitry.

In the 1980s, Estreich focused on the issue of lateral electric field enhancement on the transport of carrier associated with latchup in single-well CMOS technology [27]. The Estreich model addressed the lateral pnp model incorporating four features that were not present in the other models. The Estreich model first included both drift and diffusion currents in the base region where the drift component was associated with a lateral electric field that exists in the substrate region (e.g. base region). The second feature of the model is the lateral debiasing along the bottom of the emitter region. The third feature of the model addressed the nonuniform doped base region – the presence of a 'buried layer'. The fourth feature of the model addressed the surface recombination and 'bias-sensitive' depletion layer recombination currents [27].

A model of a lateral electric field can be established by utilizing the solution of two cylindrical electrodes. Estreich first assumed that the structure can be represented as a lateral pnp device with first and second contacts on both sides of the lateral pnp separated by distance $L$. Estreich assumed the two surface contacts in silicon can be represented as having a radius $R$, which is associated with the contact width (e.g. the length of the diffusion can be assumed to be equivalent to the diameter of the half

cylinder). Hence, assuming that the voltage across the two physical contact regions is $V$, the electric field can be estimated as [27]

$$E = \frac{2}{\ln((L + \sqrt{L^2 - 4R^2})/2R)} \frac{V}{\sqrt{L^2 - 4R^2}}.$$

In the model of Estreich, the current components of the lateral and vertical components are divided into three terms, $I_L$, $I_D$ and $I_V$, respectively. In the model, assuming an n− substrate, the lateral pnp is a p+ diffusion and a p-well region, where the base is the n− substrate. The lateral current $I_V$ is the current that flows from the emitter to the collector region, and the lateral current $I_D$ is the current that flows from the emitter to the opposite direction of the collector. It is assumed that this current flows away from the collector and recombines. The vertical current $I_V$ is the current that flows from the emitter to the 'substrate'. In the lateral device, the emitter area is the product of the emitter depth $X_J$ and emitter width $W_e$ on the edge that faces the collector,

$$A_V = W_e X_J.$$

Assuming a lateral electric field, the one-dimensional analysis of the hole excess carrier population in the base region can be represented as (including the lateral electric field term)

$$D_p \frac{d^2 p'(x)}{dx^2} - \mu_p E \frac{dp'(x)}{dx} - \frac{1}{\tau_p} p'(x) = 0.$$

Expressing as a function of the diffusion length $L_p$ and thermal voltage $(kT/q)$, the minority carrier concentration can be expressed as

$$\frac{d^2 p'(x)}{dx^2} - \frac{E}{V_T} \frac{dp'(x)}{dx} - \frac{1}{L_p^2} p'(x) = 0.$$

The solution to the differential equation can be expressed as

$$p'(x) = A e^{S_1 x} + B e^{S_2 x},$$

where the roots of the equation can be expressed as

$$S_{1,2} = \frac{E}{2V_T} \pm \sqrt{\left(\frac{E}{2V_T}\right)^2 + \frac{1}{L_p^2}},$$

or alternatively

$$S_{1,2} = \lambda_1 \pm \lambda_2.$$

From the boundary conditions at $x = 0$ (at the emitter) and $x = W_B$ (at the collector), the hole carrier populations are

$$p'(0) = p_{n0} \left\{ \exp\left(\frac{V}{V_T}\right) - 1 \right\},$$
$$p'(W_B) = -p_{n0}.$$

Estreich showed that the solution for the excess hole carrier population with the electric field assisted transport can be expressed as follows [27]:

$$p'(x) = p_{n0}\left\{\theta - \frac{\exp(-\lambda_1 W_B) + \theta\exp(\lambda_2 W_B)}{2\sinh(\lambda_2 W_B)}\right\}\exp((\lambda_1 + \lambda_2)x)$$
$$+ p_{n0}\left\{\frac{\exp(-\lambda_1 W_B) + \theta\exp(\lambda_2 W_B)}{2\sinh(\lambda_2 W_B)}\right\}\exp((\lambda_1 + \lambda_2)x),$$

where

$$\theta = \exp\left(\frac{V}{V_T}\right) - 1.$$

## 2.8.1 Lateral Bipolar Transistor Models – Base Transport Factor with Electric Field Assisted Base Transport

Assuming a lateral pnp transistor, the base transport factor is key to the understanding of the minority carrier diffusion. At the boundary conditions of the base region, $x = 0$ (at the emitter) and $x = W_B$ (at the collector), the hole carrier populations are

$$p'(0) = p_{n0}\left\{\exp\left(\frac{V}{V_T}\right) - 1\right\},$$
$$p'(W_B) = -p_{n0}.$$

Estreich showed the solution for the excess hole carrier population with the electric field assisted transport can be expressed as follows [27]:

$$p'(x) = p_{n0}\left\{\theta - \frac{\exp(-\lambda_1 W_B) + \theta\exp(\lambda_2 W_B)}{2\sinh(\lambda_2 W_B)}\right\}\exp((\lambda_1 + \lambda_2)x)$$
$$+ p_{n0}\left\{\frac{\exp(-\lambda_1 W_B) + \theta\exp(\lambda_2 W_B)}{2\sinh(\lambda_2 W_B)}\right\}\exp((\lambda_1 - \lambda_2)x),$$

where

$$\theta = \exp\left(\frac{V}{V_T}\right) - 1,$$
$$S_{1,2} = \lambda_1 \pm \lambda_2.$$

The solution for the excess hole carrier population can be expressed as

$$S_{1,2} = \frac{E}{2V_T} \pm \sqrt{\left(\frac{E}{2V_T}\right)^2 + \frac{1}{L_p^2}}.$$

The diffusion of hole carriers can be expressed as

$$J_p(x) = -qD_p\frac{dp'(x)}{dx}.$$

For a lateral pnp transistor, the base transport factor can be expressed as

$$\alpha_l = \frac{J_p(x = W_B)}{J_p(x = 0)}.$$

Solving for the diffusion lengths, the base transport factor with field assisted transport in the base region is

$$\alpha_l = \frac{\lambda_2 \exp(W_B \lambda_1)}{\lambda_1 \sinh(W_B \lambda_2) + \lambda_2 \cosh(W_B \lambda_2)}.$$

### 2.8.2 Lateral Bipolar Transistor Models – Bipolar Current Gain with Electric Field Assisted Base Transport

Assuming a lateral pnp transistor, the lateral bipolar current gain can be solved from the base transport factor (emitter-to-collector transport factor).

$$\beta_l = \frac{\alpha_l}{1 - \alpha_l}.$$

The base transport factor with electric field assisted transport in the base region is

$$\alpha_l = \frac{\lambda_2 \exp(W_B \lambda_1)}{\lambda_1 \sinh(W_B \lambda_2) + \lambda_2 \cosh(W_B \lambda_2)}.$$

From the base transport factor, the solution for the lateral bipolar current gain can be obtained by finding the common denominator and algebraically solving for the lateral bipolar current gain,

$$\beta_l = \frac{\{[\lambda_2 \exp(W_B \lambda_1)]/[\lambda_1 \sinh(W_B \lambda_2) + \lambda_2 \cosh(W_B \lambda_2)]\}}{1 - \{[\lambda_2 \exp(W_B \lambda_1)]/[\lambda_1 \sinh(W_B \lambda_2) + \lambda_2 \cosh(W_B \lambda_2)]\}},$$

then the lateral bipolar current gain with lateral electric field can be expressed as follows [27]:

$$\beta_l = \frac{\lambda_2 \exp(W_B \lambda_1)}{\lambda_1 \sinh(W_B \lambda_2) + \lambda_2 \cosh(W_B \lambda_2) - \lambda_2 \exp(W_B \lambda_1)}.$$

### 2.8.3 Lateral Bipolar Transistor Models – Two-Dimensional Hole Current Analysis with Electric Field Assist

In the evaluation of latchup, the evaluation of the parasitic bipolar transistor must address both the lateral and the vertical components. Figure 2.8 shows the vertical and lateral components and the axis of reference.

A simple model can be developed where the components of the vertical and lateral currents are separated into first and second components for a p-channel MOSFET within an n-well region (contained

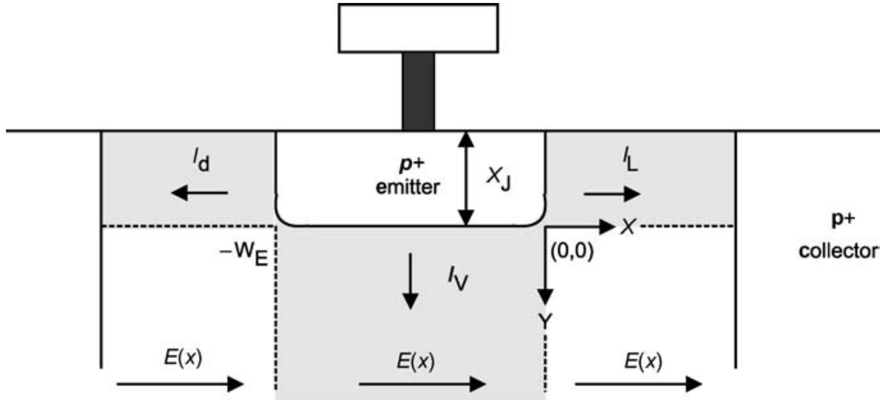



**Figure 2.8** A model for vertical and lateral transport in a lateral transistor structure.

within the p− substrate). The hole current flowing laterally from the p-diffusion edge can be regarded as the lateral current flow $I_L$; this lateral current flow is influenced by the lateral electric field under the p+ diffusion. The vertical hole current flow $I_V$ is a function of the vertical profile. As a simplified model, assume that the vertical profile is semi-infinite and uniform in doping concentration (e.g. the doping concentration of the substrate does not vary in the $y$ dimension). Establishing boundary conditions, the excess hole carrier population in the $x$ direction can be expressed as

$$p'(x,0) = p'(x=0, y=0)\exp\left(\frac{xE}{V_T}\right).$$

where the corner of the p-diffusion is represented as the coordinate reference point $(x=0, y=0)$

$$p'(x=0, y=0) = p_{n0}\left\{\exp\left(\frac{V}{V_T}\right) - 1\right\}.$$

For the case of a semi-infinite substrate (or the limiting case where the junction depth is significantly greater than the hole diffusion length), the excess minority carrier injection can be expressed as

$$p'(x,y) = p'(x=0, y=0)\exp\left(\frac{xE}{V_T}\right)\exp\left(-\frac{y}{L_p}\right).$$

From the hole current constitutive relationship, the hole current in the $x$ direction can be expressed as a function of the drift and the diffusion terms,

$$J_{px}(x,y) = q\mu_p p'(x,y)E_x - qD_p\frac{dp'(x,y)}{dx}.$$

Expressing the hole current as a function of the thermal voltage,

$$J_{px}(x,y) = qD_p\left\{\frac{p'(x,y)}{V_T}E - \frac{dp'(x,y)}{dx}\right\}.$$

From this expression, it is evident that the lateral current transport is dependent on the magnitude of the lateral electric field and the gradient of the excess minority carrier hole population [56].

### 2.8.4 Lateral Bipolar Transistor Models – Vertical Hole Current with Lateral Electric Field Assist

Analysis of the vertical hole current in the lateral pnp transistor can be evaluated at the injection boundary conditions of the emitter. The vertical hole current flow $I_V$ is a function of the vertical profile. From the hole drift–diffusion constitutive relationship,

$$J_{py}(x,y) = q\mu_p p'(x,y)E_y - qD_p \frac{dp'(x,y)}{dy}.$$

Assuming the vertical profile is a constant doping concentration, the drift term is zero. The vertical hole current can then be obtained as

$$I_V = qD_p A_V \frac{dp'(y)}{dy}\bigg|_{y=0},$$

where the area term $A_V$ is the geometrical area of the emitter from the vertical dimension. In this model development, it is assumed that the lateral current flow is from the sidewall of the diffusion and the vertical area is the diffusion bottom area. Estreich noted that in the presence of a lateral electric field, the area term $A_V$ is not the geometrical area of the emitter ($A_V = W_E W_Z$), but in fact only the case when the lateral electric field is zero [27]. In this case, the area of the vertical component is expressed as an effective area term $A_V^*$. This is obtained by integrating the term over the area of the emitter. The reason this is true is because the excess hole population over the complete emitter area is not constant. From the early development, the excess hole carrier population in the $x$ direction can be expressed as

$$p'(x,0) = p'(x=0, y=0)\exp\left(\frac{xE}{V_T}\right),$$

where the corner of the p-diffusion is represented as the coordinate reference point ($x = 0, \quad y = 0$)

$$p'(x=0, \quad y=0) = p_{n0}\left\{\exp\left(\frac{V}{V_T}\right) - 1\right\}.$$

Hence, the effective vertical area of the emitter, $A_V^*$, is obtained by integrating it over the emitter area,

$$A_V^* = \int_0^{W_Z} \int_{-W_E}^{0} \exp\left\{\frac{xE_x}{V_T}\right\} dx\, dz.$$

In the general case, the vertical effective area term can be a function of three dimensions. In the two-dimensional analysis, the effective vertical area term is independent of the $z$ dimension and can be expressed as

$$A_V^* = W_Z \int_{-W_E}^{0} \exp\left\{\frac{xE}{V_T}\right\} dx.$$

Integration over the $x$ dimension, the effective vertical area term can be expressed as follows:

$$A_{\mathrm{V}}^{*} = W_Z \int_{-W_{\mathrm{E}}}^{0} \exp\left\{\frac{xE}{V_{\mathrm{T}}}\right\} \mathrm{d}x = W_Z \left\{\frac{V_{\mathrm{T}}}{E}\right\} \int_{\frac{-W_{\mathrm{E}}E}{V_{\mathrm{T}}}}^{0} \mathrm{d}\mu \exp \mu.$$

Solving for the effective area term,

$$A_{\mathrm{V}}^{*} = W_{\mathrm{E}} W_Z \left\{ \frac{1 - \exp\left(\frac{W_{\mathrm{E}}E}{V_{\mathrm{T}}}\right)}{\left(\frac{W_{\mathrm{E}}E}{V_{\mathrm{T}}}\right)} \right\}.$$

In the case where the $W_{\mathrm{E}}E$ is significantly small compared to the thermal voltage $V_{\mathrm{T}}$, using L'hopital's rule, in the limit, reduces the geometrical area,

$$\lim_{u\to 0} A_V^{*} = \lim_{u\to 0} W_{\mathrm{E}} W_Z \left\{\frac{1 - \exp(-u)}{(u)}\right\} = W_{\mathrm{E}} W_Z \lim_{u\to 0} \frac{\frac{\mathrm{d}}{\mathrm{d}u}(1 - \exp(-u))}{\frac{\mathrm{d}}{\mathrm{d}u}(u)} = W_{\mathrm{E}} W_Z = A_{\mathrm{V}}.$$

Solving for the vertical hole current,

$$I_{\mathrm{V}} = qD_{\mathrm{p}} A_{\mathrm{V}}^{*} \left.\frac{\mathrm{d}p'(y)}{\mathrm{d}y}\right|_{y=0},$$

$$p'(y) = p_{\mathrm{n}0} \exp\left(-\frac{y}{L_{\mathrm{p}}}\right) \left\{\exp\left(\frac{V}{V_{\mathrm{T}}}\right) - 1\right\},$$

which can be expressed as

$$p'(y) = p_{\mathrm{n}0} \exp\left(-\frac{y}{L_{\mathrm{p}}}\right) \theta,$$

$$\theta = \left\{\exp\left(\frac{V}{V_{\mathrm{T}}}\right) - 1\right\}.$$

From the effective area term, the vertical current can be expressed as

$$I_{\mathrm{V}} = \frac{qD_{\mathrm{p}} W_{\mathrm{E}} W_Z}{L_{\mathrm{p}}} \left\{ \frac{1 - \exp\left(-\frac{W_{\mathrm{E}}E}{V_{\mathrm{T}}}\right)}{\left(\frac{W_{\mathrm{E}}E}{V_{\mathrm{T}}}\right)} \right\} p_{\mathrm{n}0}\theta.$$

## 2.9 LATERAL BIPOLAR TRANSISTOR MODELS–NONUNIFORM VERTICAL PROFILE

In the development of lateral and vertical bipolar gains, the early models of Tseng and Busen [48] and Lindmayer and Schneider [49] assumed a uniform doping concentration in the vertical direction. The presence of a doping transition in the vertical direction influences the bipolar current in both the vertical and lateral bipolar transistors.

### 2.9.1 Lateral Bipolar Transistor Models–Gunn, Dutton, and Whittier n−/n+ Step Junction Model

The analysis of the step junction of a n−/n+ transition was first addressed by Low [60], Arthur and coworkers [61,62], Lade and Jordan [63], Scharffetter [64], Dutton and Whittier [65], and Venkateswaran and Roulston [66], and Hauser and Dunbar [67]. In the analysis, the focus was on the evaluation of the influence of the n−/n+ 'space–charge' layer, and the influence on the current transport, as well as recombination time in a p+/ n−/n+. Dutton and Whittier [65] assumed that under a low-level injection assumption, in quasi-equilibrium, the minority carrier concentration on both sides of the n−/n+ high–low junction transition can be evaluated. From the relationship of the intrinsic carrier population and the electron and hole population,

$$np = n_{\mathrm{i}}^2$$
$$(n_0 + n')(p_0 + p') = n_{\mathrm{i}}^2$$

Under a low-level injection assumption, the doping concentration of the n-type region is significantly larger than the excess carrier population. Additionally, the excess hole population is larger than the minority carrier hole equilibrium concentration. Given an n−/n+ high–low (H-L) transition occurs at the epitaxy–substrate junction, Dutton and Whittier represented the junction [65]

$$p'(x = W_{\mathrm{epi}})N_{\mathrm{D_{epi}}} = n_{\mathrm{i}}^2 = p'(x = W_{\mathrm{epi}} + W_{\mathrm{sc_2}})N_{\mathrm{D_{sub}}},$$
$$p'(x = W_{\mathrm{epi}}) = p'(x = W_{\mathrm{epi}} + W_{\mathrm{sc_2}})\frac{N_{\mathrm{D_{sub}}}}{N_{\mathrm{D_{epi}}}},$$

where $W_{\mathrm{epi}}$ is the epitaxial thickness and $W_{\mathrm{sc_2}}$ is a metallurgical junction width of the n−/n+ high–low transition. To evaluate the electrical current flowing to the n−/n+ junction the model assumed conservation of current; the current flowing from the p+/n− junction equals the recombination current within the n−/n+ step junction and the current that is injected into the substrate (Figure 2.9). The current flowing to the substrate (e.g. n+ region) is,

$$I_{\mathrm{sub}} = q\frac{D_{\mathrm{p_{sub}}}}{L_{\mathrm{p_{sub}}}}p'(W + W_{\mathrm{sc_2}}).$$

The recombination current can be evaluated from the change in the excess carrier current divided by the recombination time within the n−/n+ region,

$$U_{\mathrm{sc_2}} = \frac{p - p_{\mathrm{n}}}{\tau_{\mathrm{sc_2}}},$$

where

$$I_{\mathrm{sc_2}} = A\int_{W}^{W+W_{\mathrm{sc_2}}} U_{\mathrm{sc_2}}\,\mathrm{d}x.$$

Dutton and Whittier showed that the carrier population can be expressed as a function of the quasi-Fermi level and the potential as a function of position [65]. To simplify the expression, the potential

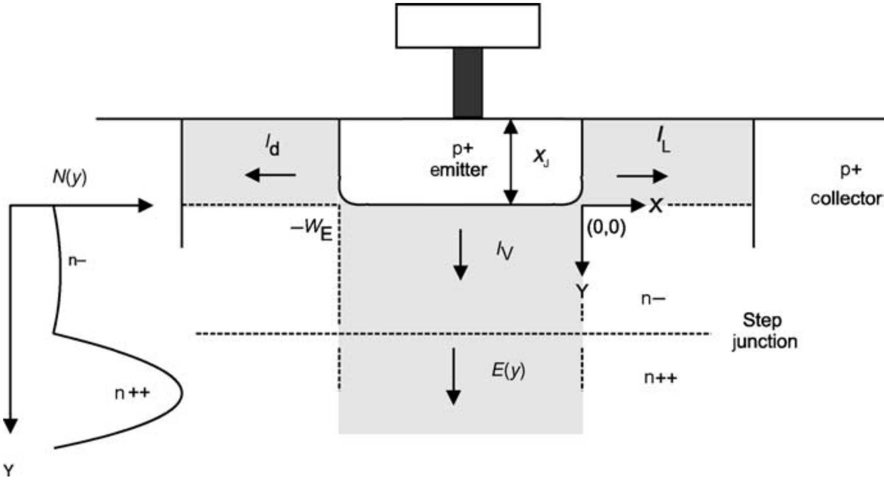



**Figure 2.9** A model for vertical and lateral transports with a vertical step transition.

was defined as a linear change in the potential across the n−/n+ junction region. It can be shown that this can be expressed as

$$J_{sc_2} = q\frac{W_{sc_2}}{\tau_{sc_2}}\left(\frac{kT}{q\Delta}\right)p'(x),$$

where the built-in potential can be expressed as the

$$\Delta = \frac{kT}{q}\ln\left[\frac{N_{D_{sub}}}{N_{D_{epi}}}\right].$$

Gunn noted that the step doping transition can be represented as a recombination velocity [62]. Dutton and Whittier represented this term as a normalized recombination velocity [65]. The boundary condition at the n−/n+ high–low transition can be expressed as a mixed boundary condition

$$qD_p\frac{\partial p'}{\partial x}\bigg|_{x=W} = -(J_{sc_2} + J_{sub}),$$
$$\left\{qD_p\frac{\partial p'(x)}{\partial x} + \xi\frac{p'(x)}{L_p}\right\}\bigg|_{x=W} = 0.$$

It was noted by Gunn [62] and Dutton and Whittier [65] that this can be represented in the form

$$\left\{qD_p\frac{\partial p'(x)}{\partial x}\right\}\bigg|_{x=W} = -\frac{\xi}{L_p}p'(x)\bigg|_{x=W}.$$

The solution for the excess carrier population was shown to be expressed as

$$p'(x) = p'(0)\left\{\frac{\cosh([W-x]/L_p) + \xi\sinh([W-x]/L_p)}{\cosh(W/L_p) + \xi\sinh(W/L_p)}\right\},$$

and the forward current relationship for the p+/n−/n+ structure is

$$J_F = \frac{qD_p p'(0)}{L_p}\left\{\frac{\sinh(W/L_p) + \xi\cosh(W/L_p)}{\cosh(W/L_p) + \xi\sinh(W/L_p)}\right\},$$

where the normalized recombination factor can be expressed as

$$\xi = \frac{(W_{sc_2}/\tau_{sc_2})(kT/q\Delta) + (D_{p_{sub}}/L_{p_{sub}})(N_{D_{epi}}/N_{D_{sub}})}{\frac{D_p}{L_p}}.$$

If the recombination in the n−/n+ region is small, then the term is reduced to

$$\xi \approx \frac{D_{p_{sub}}}{D_{p_{epi}}}\frac{L_{D_{epi}}}{L_{D_{sub}}}\frac{N_{D_{epi}}}{N_{D_{sub}}}.$$

Estreich defined a term expressed as the fractional reduction in vertical current as [27]

$$F_V = \frac{\sinh\gamma + \xi\cosh\gamma}{\cosh\gamma + \xi\sinh\gamma},$$

where

$$\gamma = \frac{W_{epi}}{L_{D_{epi}}}$$

and

$$\xi = \frac{D_{p_{sub}}}{D_{p_{epi}}}\frac{L_{D_{epi}}}{L_{D_{sub}}}\frac{N_{D_{epi}}}{N_{D_{sub}}}.$$

In a two-dimensional structure, the vertical and lateral current components are interrelated. A reduction in the vertical current component leads to an enhancement in the lateral current component. Hence, in the evaluation of latchup, the nonuniform vertical profile influences both the vertical and lateral component current distribution.

### 2.9.2 Lateral Bipolar Transistor Models – Bipolar Current Gain with Nonuniform Vertical Profile

Estreich introduced the n−/n+ high–low (H-L) junction transition into the bipolar current gain of the lateral pnp bipolar transistor. Estreich defined a term expressed as the fractional reduction in vertical current as [27]

$$F_V = \frac{\sinh\gamma + \xi\cosh\gamma}{\cosh\gamma + \xi\sinh\gamma},$$

where

$$\gamma = \frac{W_{epi}}{L_{D_{epi}}}$$

and

$$\xi = \frac{D_{p_{sub}}}{D_{p_{epi}}} \frac{L_{D_{epi}}}{L_{D_{sub}}} \frac{N_{D_{epi}}}{N_{D_{sub}}}.$$

Integrating the fractional reduction factor into the pnp bipolar term, Estreich formed a new definition for the bipolar current gain [27],

$$\beta_{pnp} = \frac{\alpha_l I_L}{(1-\alpha_l)I_L + F_V I_V}.$$

Today, there are a number of reasons why the low–high (L–H) and high-low (H–L) transitions are important for latchup. First, this concept influences the lateral and vertical bipolar current gain characteristics. Secondly, this has a role in triple-well technology. Third, it has a role in high-dose buried layer (HDBL) transport physics. Finally, this concept is also used for both 'passive' and 'active' guard rings. Today, image processing semiconductor chips and smart power technology utilize p− epitaxial regions on p++ substrates and can utilize the L–H transition as an advantage for passive and active guard rings.

## 2.9.3 Lateral Bipolar Transistor Models – Base Effective Diffusion Length

In a semiconductor structure, the nature of the current flow in 'wide base width' lateral transistor is two dimensional at large spacing [68]. There are cases of interest where the analysis can be simplified when boundary conditions are reflective or absorbent. Examples of surface conditions perpendicular to the transport are as follows:

- insulator surface with surface recombination;
- step junction p−/p+ reflective boundary with both reflective and recombinant characteristics;
- metallurgical junctions that collect the minority carriers.

In the case of an insulator surface, the isolation surface can be STI surface or wide TI. In the case of the step junction, the boundary of the step junction can serve as the transport through that physical surface. The second case can be a p++ substrate contact in a p− epitaxial region or a p−/p++ step junction in the p− substrate. In the case of a metallurgical junction, the surface can be of another shape within the base region, which can be a guard ring or other physical element. Hence, it is desirable to have an analytical description to simplify these potential cases.

Figure 2.10 shows a region of analysis and a control volume. To evaluate the minority carrier transport, the current at the first surface, $x = x$, and the second surface, $x = x + dx$, can be evaluated within an incremental region of width, $dx$. Within an incremental distance from $x = x$ to $x = x + dx$, the current at $x = x + dx$ can be expressed as

$$I_n(x + dx) = I_n(x) - \Delta I_n(x, x + dx).$$

In semiconductors, this region is typically semi-infinite and hence a first top surface must be defined. Hence, it is advantageous to define a control volume in both the $x$ and $y$ directions and evaluate the current flowing in and out of the control volume surfaces. From the conservation of carriers, the carrier density is a function of the carrier flowing through the surfaces, as well as the generation rate within the

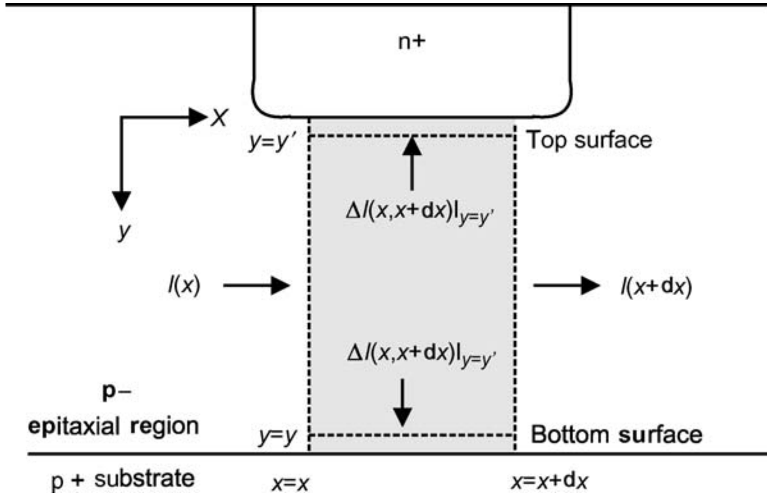



**Figure 2.10** Control volume for current analysis within a base region of a lateral npn transistor.

control volume. In evaluation of the transport, the first top surface and the lower bottom surface can be defined. In this fashion, analysis can address the transport through the top surface, through the bottom surface, and recombination within the incremental volume. Hence, the transport in the direction perpendicular to the $x$ direction can be regarded as a 'loss'.

Within the physical control volume, the recombination within the volume can be defined as

$$\Delta I_{\mathrm{n_r}}(x, x + \Delta x).$$

The loss in the top surface at $y = y'$ can be defined as

$$\Delta I_{\mathrm{n_r}}(x, x + \Delta x)|_{y=y'}.$$

The loss in the bottom surface at $y = y$ can be defined as

$$\Delta I_{\mathrm{n_r}}(x, x + \Delta x)|_{y=y},$$

where we define the loss in the region between $x = x$ and $x = x + \mathrm{d}x$ as

$$\Delta I_{\mathrm{n}}(x, x + \mathrm{d}x) = \Delta I_{\mathrm{n_r}}(x, x + \Delta x) + \Delta I_{\mathrm{n_r}}(x, x + \Delta x)|_{y=y'} + \Delta I_{\mathrm{n_r}}(x, x + \Delta x)|_{y=y}.$$

In this analysis, the distance between the two surfaces $y = y'$ and $y = y$ can be defined as

$$T = |y' - y|.$$

Defining the diffusion current, it can be defined as

$$I_{\mathrm{n}}(x) = qD_{\mathrm{n}}W\frac{\mathrm{d}}{\mathrm{d}x}n(x),$$

where

$$n(x) = \int_{y=0}^{y=T} n(x, y)\mathrm{d}x.$$

The recombination current within the volume can be expressed as the amount of excess charge within the incremental volume $W\mathrm{d}x$,

$$\Delta I_{\mathrm{n_r}}(x, x + \Delta x) = \frac{Q_{\mathrm{n}}(x)W\mathrm{d}x}{\tau_{\mathrm{n}}} = -\frac{qW_{\mathrm{n}}(x)\mathrm{d}x}{\tau_{\mathrm{n}}}.$$

For the upper and lower boundary conditions, it is of interest to evaluate the upper and lower surfaces as if they are a current loss.

From the Dutton–Whittier model [65] and the Estreich derivation [27], the p−/p+ L–H junction transition introduces a term expressed as the fractional reduction in vertical current as

$$F_{\mathrm{V}} = \frac{\sinh\gamma + \xi\cosh\gamma}{\cosh\gamma + \xi\sinh\gamma},$$

where

$$\gamma = \frac{W_{\mathrm{epi}}}{L_{\mathrm{D_{epi}}}}$$

and

$$\xi = \frac{D_{\mathrm{n_{sub}}}}{D_{\mathrm{n_{epi}}}}\frac{L_{\mathrm{D_{epi}}}}{L_{\mathrm{D_{sub}}}}\frac{N_{\mathrm{A_{epi}}}}{N_{\mathrm{A_{sub}}}}.$$

From the model of Gunn, this can also be represented as a recombination velocity [61,62]. To evaluate it as a recombination surface, the evaluation of the electron current through the surface will provide the total current loss through this lower boundary surface.

Given the boundary doping transition, a vertical built-in field can be assumed at the low–high transition region of the magnitude,

$$E_y = \frac{kT}{q}\left\{\frac{1}{p(y)}\frac{\partial p(y)}{\partial y}\right\}.$$

From the drift–diffusion constitutive relationship,

$$J_{\mathrm{n}_y} = q\mu_{\mathrm{n}}nE_y + qD_{\mathrm{n}}\frac{\partial n(y)}{\partial y}.$$

Substituting the low–high electric field condition,

$$J_{\mathrm{n}_y} = qD_{\mathrm{n}}\frac{n}{p}\frac{\partial p}{\partial y} + qD_{\mathrm{n}}\frac{\partial n}{\partial y}.$$

This can be expressed as

$$J_{\mathrm{n}_y}(x, y) = J_{\mathrm{n}_y}(T, x)\exp\left\{-\frac{y - T}{L_{\mathrm{D_{sub}}}}\right\}.$$

Assuming all current flowing into the substrate is loss through recombination, this can be expressed as the integration over the semi-infinite substrate from the low–high transition to the bottom of the substrate (e.g. $y$ at infinity). Integration over the $y$ direction leads to the expression

$$\Delta I_{n_r}(x, x+\Delta x)|_{y=y} = -qWD_n \left\{ \frac{N_{epi}}{N_{sub}} \right\} \frac{n(x,T)}{L_{D_{sub}}} dx.$$

Assuming the second boundary condition is an insulating boundary with no surface recombination velocity, then the second term is zero. Assuming the second boundary is a metallurgical junction, which could be any physical region within the base region or a guard ring structure,

$$\Delta I_{n_r}(x, x+\Delta x)|_{y=y'} = -qD_n W \frac{\partial}{\partial y} n(x, y=y')dx.$$

Hence, the evaluation of the characteristics of the incremental volume can be evaluated as the recombination with the base region and the current flowing out of the upper and lower surfaces, where it is assumed that the current flowing in through the surfaces is 'loss' through carrier collection or recombination. From carrier diffusion theory, it is known that the diffusion length is associated with loss of carriers as a function of distance. Hence, it is clear that this can be represented as a loss in the direction of the flow and in the direction opposite to the direction of the flow. It is anticipated that a generalized term may be possible of the form, based on dimensional analysis,

$$\frac{d^2 Q_n}{dx^2} - \frac{Q_n}{L_{eff}^2} = 0,$$

where the effective diffusion length is associated with three terms,

$$\frac{1}{L_{eff}^2} = \frac{1}{L_{D_n}^2} + \frac{1}{L_1^2} + \frac{1}{L_2^2},$$

where there is an diffusion length in the direction of the flow and two additional equivalent lengths perpendicular to the flow. It was shown by Stella *et al.* [68] that the derivation can be expressed as a second-order differential diffusion equation, where an effective diffusion length is derived, addressing the loss through the two surfaces.

Define an average electron concentration in the region $T$,

$$\langle n(x) \rangle = \frac{Q_n(x)}{qT}.$$

For the special case of a reflection boundary and a metallurgical junction, the differential equation can be expressed as

$$\frac{d^2 Q_n}{dx^2} - Q_n \left\{ \frac{1}{D_n \tau_n} \right\} - Q_n \left\{ \frac{1}{T} \frac{1}{L_{D_{sub}}} \frac{N_{epi}}{N_{sub}} \frac{n(x,T)}{\langle n(x) \rangle} \right\} - Q_n \left\{ \frac{1}{T} ([\partial n(x, y=y')]/[\partial y]) \langle n(x) \rangle) \right\} = 0.$$

From this form, it is apparent that the three last terms are the loss through the incremental volume and can be expressed as an effective diffusion coefficient, given it is a constant value:

$$\frac{d^2 Q_n}{dx^2} - \frac{Q_n}{L_{eff}^2} = 0.$$

Assuming the terms to be constant,

$$r = \frac{n(x,T)}{\langle n(x)\rangle} = \text{constant}$$

and

$$s = \frac{[\partial/\partial y\, n(x,y')]}{\langle n(x)\rangle} = \text{constant}$$

then an effective diffusion length in the base region can be derived. From the analysis of Stella *et al.* [68], the characteristic length can be derived as

$$L_{\text{eff}} = \frac{1}{\sqrt{(1/L_{\mathrm{D_n}})^2 + (1/L_1)^2 + (1/L_2)^2}},$$

where

$$L_{\mathrm{D_n}} = \sqrt{D_{\mathrm{n}}\tau_{\mathrm{n}}}$$

and

$$L_1 = \sqrt{TL_{\mathrm{D_{sub}}}\left(\frac{N_{\text{sub}}}{N_{\text{epi}}}\right)r}$$

and

$$L_2 = \sqrt{\frac{T}{s}},$$

where this can be expressed as

$$L_y = \frac{1}{\sqrt{(1/L_1)^2 + (1/L_2)^2}}.$$

First, note that the expression for the effective length $L_1$ is associated with the arithmetic mean of the thickness of the region and the substrate diffusion length. In addition, as occurs in the Dutton–Whittier expression and Estreich derivation, the ratio of the dopants at the p−/p++ step junction is also present in the effective diffusion length $L_1$. In the case of the second diffusion length, the expression is the ratio of the thickness and the relative gradient scale length, which is contained within the $s$ expression.

The evaluation of the parameter $r$ and $s$ can be obtained by separation of variables, allowing $n(x,y) = n(x)n(y)$. In this fashion, a differential equation as a function of $n(y)$ can be evaluated with $L_y$ as the effective diffusion length, where

$$\frac{\mathrm{d}^2 n}{\mathrm{d}y^2} + \frac{n(y)}{L_y^2} = 0.$$

It has been shown that the solution is

$$\frac{L_y}{T}\tan\left(\frac{T}{L_y}\right) = \frac{L_{D_{sub}}}{T}\left(\frac{N_{sub}}{N_{epi}}\right).$$

In the case where the ratio of substrate doping and epitaxial doping is large, the solution tends to the case that the tangent function is equal to zero. As a result,

$$\frac{T}{L_y} \approx \frac{\pi}{2}.$$

As a result, the effective diffusion length under a junction region, with a L–H step junction can be expressed as

$$L_{eff} = \frac{1}{\sqrt{(1/L_{D_n})^2 + (\pi/2T)^2}}.$$

This analysis is important for the evaluation of the perpendicular boundaries on lateral current transport. In the case of a guard ring structure, this derivation is helpful in the prediction of the guard ring efficiency. A latchup design principle is as follows:

- The lateral diffusion characteristics can be evaluated as a one-dimensional analysis wherein the vertical boundary conditions can be addressed as 'loss' boundary conditions.
- Lossy boundary conditions can be metallurgical junction, insulator with surface recombination or reflective boundary condition (low–high step junctions).
- Special cases exist that allow representation of an effective diffusion length within the base region.

## 2.10 TRIPLE-WELL BIPOLAR TRANSISTOR MODELS – LATERAL AND VERTICAL CONTRIBUTIONS

Triple-well technology was a solution to eliminate latchup by the separation of the p-channel MOSFETs in one region and isolation of the n-channel MOSFETs in the second region. Isolated MOSFETs are practiced today in advanced CMOS, RF CMOS and RF BiCMOS applications with the objective of noise isolation and back-biasing conditions different from the substrate potential. In the case of RF BiCMOS, the isolation of the p− epitaxial region is achieved using subcollector implants on the bottom and diffused isolation, reachthrough implants and trench isolation on the edges. Figure 2.11 shows a cross section of a triple-well technology, as practiced on a p− substrate. In this implementation, the p-channel MOSFET is contained within an n-well region, as practiced in dual-well CMOS. The n-channel MOSFET is contained in a p-well region, but in this process architecture, the distinction is the p-well that is electrically isolated from the p− substrate. The p-well is electrically isolated from the p− substrate by the formation of an n-well and an n-type buried layer. The n-type buried layer (e.g. also known as the n− isolation layer) separates the n-channel MOSFET transistor from the p-channel MOSFET, electrically eliminating the classical pnpn parasitic device and the latchup between the p-channel MOSFET and the n-channel MOSFET. However, it introduces lateral and vertical npn transistors between the MOSFET and the n− buried layer.

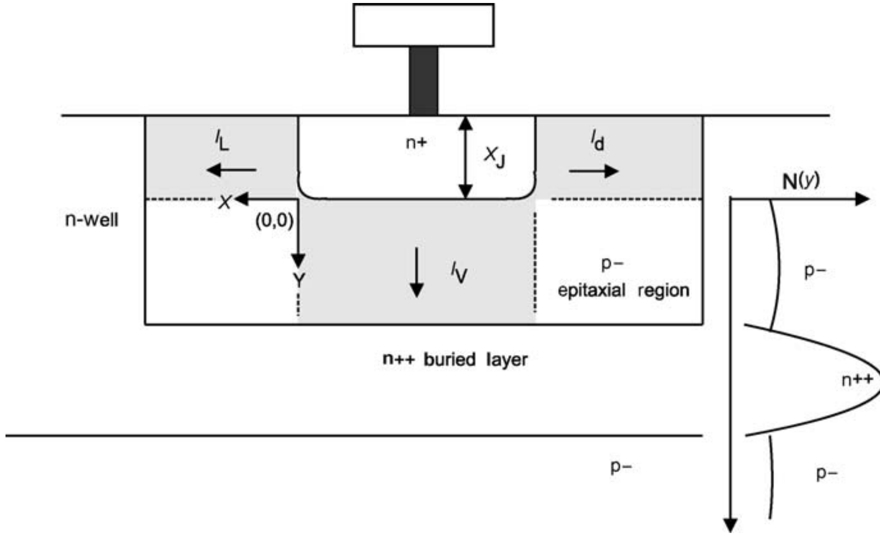



**Figure 2.11** CMOS 'triple-well' technology cross section.

Assuming a lateral npn transistor, the lateral bipolar current gain can be solved from the base transport factor (emitter-to-collector transport factor). In the case of lateral and vertical contributions, the definition for the bipolar current gain is modified.

As shown in the prior section, one can define a collector current associated with a vertical and a lateral current. In the dual-well CMOS, the vertical current is not regarded as contributing to the current gain. The current that reaches the collector is the product of the collector-to-emitter transport term and the current lateral current, $I_L$.

$$I_C = \alpha_l I_L + \alpha_v I_V.$$

In the base region, there are three terms. The first term is associated with the carriers that recombine in the lateral region. This term can be expressed as

$$I_{BL} = (1 - \alpha_l) I_L.$$

The second term is associated with the carriers that recombine in the vertical region.

$$I_{BV} = (1 - \alpha_v) I_V.$$

The third base current term is associated with the lateral current flowing in the opposite direction. Then the definition can be expressed as

$$\beta_{npn} = \frac{I_C}{I_B} = \frac{\alpha_l I_L + \alpha_V I_V}{(1 - \alpha_l) I_L + I_D + (1 - \alpha_v) I_V}.$$

Assuming the term of the current flowing away laterally in the opposite direction is significantly less than the vertical current, this can be simplified as follows:

$$\beta_{npn} = \frac{I_C}{I_B} = \frac{\alpha_l I_L + \alpha_V I_V}{(1 - \alpha_l) I_L + (1 - \alpha_v) I_V}.$$

In this structure, $\beta$ can be expressed as lateral and vertical currents. As the spacing between the n-channel MOSFET and the n-isolation region increases, the lateral bipolar current is dominated by the vertical contribution. In this case, the experimental results will saturate as a function of the npn base width spacings. Hence, the vertical contribution can be evaluated at large spacings.

In RF BiCMOS technology, trench isolation is used on the sidewalls of the isolation region. The isolation structure can be polysilicon-filled deep trench (DT) isolation or an oxide-filled TI structure. In these cases, there is no lateral npn transistor, but only a vertical bipolar transistor. In this case, the lateral current flow can be regarded as a current loss to the vertical bipolar current gain,

$$\beta_{\mathrm{npn}} = \frac{I_{\mathrm{C}}}{I_{\mathrm{B}}} = \frac{\alpha_{\mathrm{V}} I_{\mathrm{V}}}{I_{\mathrm{L}} + (1 - \alpha_{\mathrm{v}}) I_{\mathrm{V}}}.$$

## 2.11 MERGED TRIPLE-WELL BIPOLAR MODELS

Triple-well technology has become important for future CMOS technology after 0.13-μm physical dimensions. In the analysis of the merged triple-well npn bipolar transistor gain, the influence of the vertical transistor is a function of the contribution of the current in the triple-well region and the buried layer. In the following sections, the vertical current contribution and the buried layer are discussed.

### 2.11.1 Merged Triple-Well Models – Lateral and Vertical Contributions with Buried Layer

Figure 2.12 shows cross section of merged triple well structure. Figure 2.13 shows cross section of merged triple well structure highlighting the buried layer resistor. In Figure 2.14, the CMOS 'merged triple-well' technology circuit schematic with two npn transistors is shown (vertical and lateral npn transistors highlighting the buried layer resistance). The buried layer resistance plays a role in the latchup physics.

For the vertical npn bipolar transistor current to contribute to latchup, the current must flow along the n− buried layer (e.g. npn collector) to the n-well region. The resistance of the buried layer

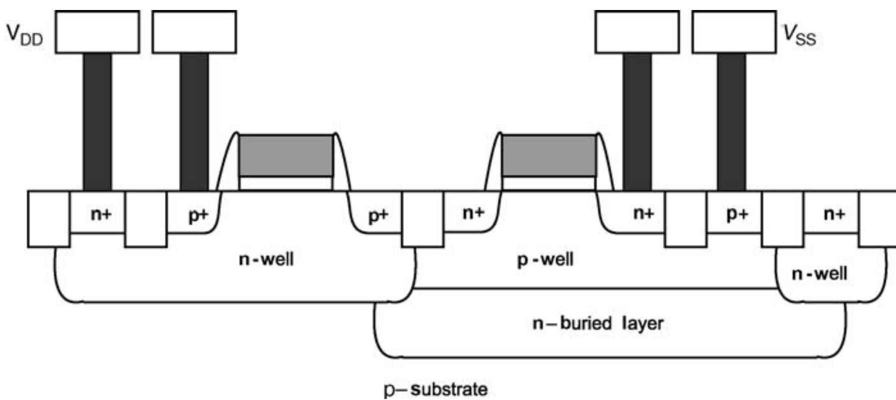



**Figure 2.12** CMOS 'merged triple-well' technology cross section.

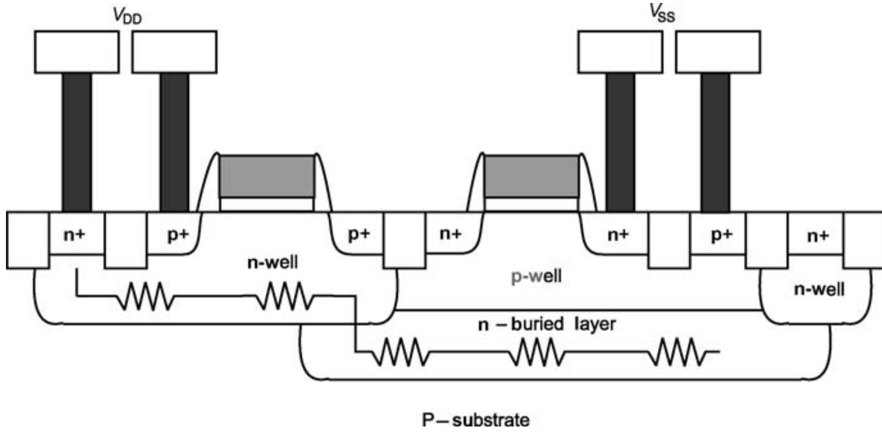



**Figure 2.13** CMOS 'merged triple-well' technology cross section highlighting the buried layer resistance.

influences both the current flow and the generalized tetrode stability criteria. The vertical transistor current must flow to the n-well region in order to contribute to the regenerative feedback. Hence, if the n− buried layer resistance is high, the current will not flow to the n-well.

The model can contain two corrections. First, in the npn bipolar model, a corrective factor can be placed to address the net current that contributes from the vertical transistor. The second

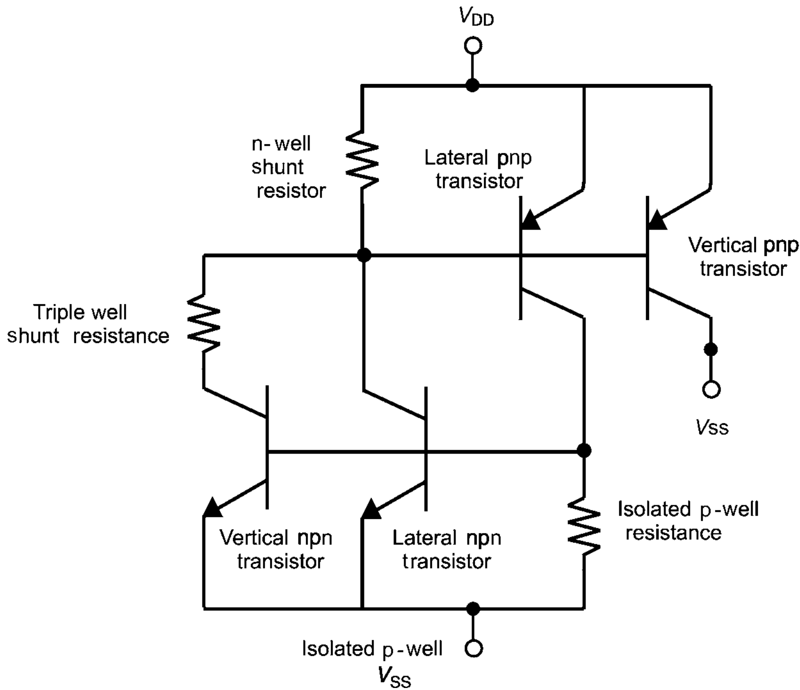



**Figure 2.14** CMOS 'merged triple-well' technology circuit schematic with two npn transistors (vertical and lateral npn transistors highlighting the buried layer resistance).

correction can address the stability by treating as two separate transistors, where the vertical transistor includes the collector resistance in the stability criterion. Figure 2.14 shows the two-transistor representation where the vertical transistor contains the collector resistance. As the spacing between the p+ diffusion and the n+ diffusion (p+/n+ space) increases, the lateral $\beta_{npn}$ decreases. Additionally, the collector resistance of the n− buried layer will play a more significant role in the latchup response.

### 2.11.2 Merged Triple-Well Models – Lateral and Vertical Contributions

The electrical separation of the p− and n-channel MOSFETs using a 'triple-well' technology is a solution to eliminate latchup by the separation of the p-channel MOSFETs in one region and isolation of the n-channel MOSFETs in the second region.

But, in the evolution of CMOS from dual-well CMOS to triple-well CMOS, the following design constraints influenced the implementation of triple-well CMOS:

- a smaller p+/n+ spacing with the migration from a 0.18- to 0.13-μm CMOS technology;
- maintain spacing, pitch and circuit density between the n-channel MOSFET and p-channel MOSFET;
- better noise isolation from the substrate.

As a result, the concept of separation of the p-channel and n-channel by constructing a separate isolated epitaxial p-well region using an n-well ring and buried layer was not established as a design practice in advanced CMOS development.

Three practices of triple well are used in advanced CMOS structures:

- An n− buried layer under the p-well region with a ring electrically isolating the p-well and integrated with the existing n-well region. The n− buried layer is only under the p-well region.
- An n− buried layer under the p-well region with a ring electrically isolating the p-well and integrated with the existing n-well region. The n− buried layer is selectively under the p-well and the n-well region of some circuitry, subfunctions or cores.
- A blanket n− buried layer is placed under the entire chip. The n− buried layer under the p-well region integrates with a partial n-well ring electrically isolating the p-well and integrated with the existing n-well region. This fully isolates the substrate region.

Figure 2.14 shows a cross section of a 'merged triple-well' technology, as practiced on a p− substrate. In this implementation, the p-channel MOSFET is contained within an n-well region, as practiced in dual-well CMOS. The p-well is electrically isolated from the p− substrate by the formation of an n-well and an n-type buried layer. In this structure, the n-type buried layer (e.g. also known as the n− isolation layer) does not separate the n-channel MOSFET transistor from the p-channel MOSFET; but in fact, it introduces lateral and vertical npn transistors between the n-channel MOSFET and the n-buried layer.

As discussed in the previous section, the npn transistor contains both lateral and vertical contributions to the npn bipolar current gain. The current that reaches the collector is the product of

the lateral collector-to-emitter transport term and the lateral current $I_L$ as well as the vertical collector-to-emitter transport term and the vertical current $I_V$,

$$I_C = \alpha_l I_L + \alpha_v I_V.$$

In the base region, the first term is associated with the carriers that recombine in the lateral region,

$$I_{BL} = (1 - \alpha_l) I_L,$$

the second term is associated with the carriers that recombine in the vertical region,

$$I_{BV} = (1 - \alpha_v) I_V,$$

and the third term, which is associated with the current flowing away from the n-well edge current term, is associated with the lateral current flowing in the opposite direction. Then the definition can be expressed as

$$\beta_{npn} = \frac{I_C}{I_B} = \frac{\alpha_l I_L + \alpha_V I_V}{(1 - \alpha_l) I_L + I_D + (1 - \alpha_v) I_V}.$$

Assuming the term of the current flowing away laterally in the opposite direction is significantly less than the vertical current, this can be simplified as follows:

$$\beta_{npn} = \frac{I_C}{I_B} = \frac{\alpha_l I_L + \alpha_V I_V}{(1 - \alpha_l) I_L + (1 - \alpha_v) I_V}.$$

In the merged triple-well structure, the vertical and lateral components cannot be separated using different electrical contacts. But, as the spacing between the n-stripe and the n-well edge increases, the bipolar current is dominated by the vertical contribution. Hence, by varying the p+ to n+ spacing (e.g. the n-well to n-stripe distance), the lateral npn gain decreases. In the case where the resistance of the buried layer is low, the variation above this saturated value is the lateral component. In this fashion, the vertical and lateral contributions can be separated.

For the pnp transistor, the influence on the pnp bipolar transistor gain is dependent on the form of the merged triple-well process.

### 2.11.3 The Lateral Bipolar Transistor – Substrate Spreading Resistance Representation (Estreich)

Substrate resistance and minority carrier diffusion length has a strong influence on latchup for both internal latchup and external latchup. As a result, the model representations of the substrate region, which serves as the base region of lateral bipolar transistors, are very important. For any two independent n-diffusions in a p− substrate, there exists a lateral bipolar transistor. Parasitic lateral BJTs exist within an inverter circuit to interact between any two regions in the substrate. Hence, parasitic lateral transistors exist within a circuit, between circuits and between chip subfunctions.

For example, parasitic npn BJTs are formed between n-channel pull-down OCD transistors and its adjacent n-well guard rings. Parasitic npn transistors exist between n-well diode ESD elements and its adjacent n-well guard rings. Parasitic bipolar transistors exist between LDMOS HVCMOS and low-voltage CMOS in smart power applications. Hence, there are many semiconductor structures that can

be represented as a lateral npn bipolar transistor in a p− substrate (serving as the base region of the npn transistor), where models are needed to quantify the resistance based on simple geometrical regions and doping profiles.

A simple resistance model developed for the evaluation of resistance between a substrate contact and the transistor using the spreading resistance derivation between two points was first utilized by Estreich for latchup [27]. Assuming a two-point spreading resistance model, the resistance can be represented as

$$R = \frac{\rho \ln\left(\frac{2L}{a}\right)}{\pi h},$$

where $\rho$ is the resistivity, $h$ is the thickness of the wafer, $2L$ is the spacing between the two points and $a$ is the geometrical representation associated with the size of the contact structure. The contact structure can be represented as a circular region of radius $a$,

$$a = \sqrt{\frac{W_1 W_2}{\pi}},$$

where $W_1$ and $W_2$ are the geometrical rectangular dimensions of the contact (e.g. product is the arithmetic mean).

The resistance between the contact and the device can be represented as one half of resistance shown above (e.g. at the symmetrical center point between the two contacts at the surface).

### 2.11.4 The Lateral Bipolar Transistor – Substrate Transmission Line Representation (Troutman and Hargrove)

The substrate spans the physical space in a semiconductor chip leading to both coupling effects and distribution effects [69,70]. The modeling of the substrate is key to understanding the response to the lateral parasitic transistor where the active region is influenced by the local potential. Two primary cases are of interest. In the first case, the element is a plurality of active elements, where each spatial region has at least one node that is independent of the other active elements. In this case, a common electrical node may be present, but some of the nodes can be regarded as independent. In the second case, there are spatially separated regions with no active devices between them.

Substrate modeling of a lateral BJT can be addressed by using lumped resistor elements under the semiconductor devices. To address substrate resistance effects, a different lumped resistor element can be placed under each independent active device region. The substrate, the base of the lateral transistor, can be modeled as a lateral resistance to the electrical contact region and vertical resistances under each physical device region.

A better representation is a full transmission line (TL) representation. For high-current phenomenon that occurs near the surface of the wafer, the substrate can be modeled as a lossy transmission line. A lossy TL model is suitable for current flow between two regions near the surface of the wafer and cases where current can flow from backside contacts [69]. From the general form of a lossy transmission line, we can express the current and voltage conditions as

$$V(y) = V_1 \exp\left\{-\frac{y}{\gamma}\right\} + V_2 \exp\left\{-\frac{L-y}{\gamma}\right\}$$

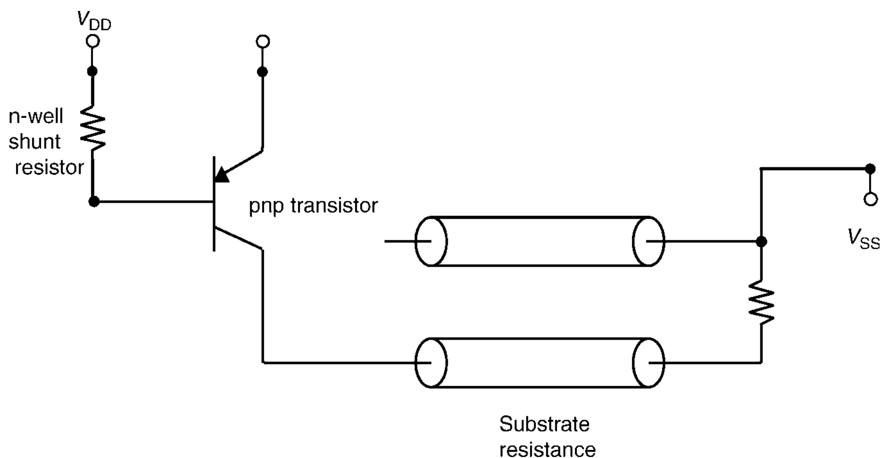



**Figure 2.15** Lossy TL model representation of the substrate.

and

$$I(y) = \frac{1}{Z_0}\left[V_1 \exp\left\{-\frac{y}{\gamma}\right\} - V_2 \exp\left\{-\frac{L-y}{\gamma}\right\}\right].$$

These current and voltage conditions represent a lossy transmission line with the transmission propagating in the $y$ dimension, whose transmission line length is $L$ with the loss factor $\gamma$. This is shown in Figure 2.15.

Assuming a transmission line with a termination resistance $R_1$ at terminal 1 and a termination resistance $R_2$ at terminal 2, we can express the TL system in the form

$$\begin{bmatrix} I_1 \\ I_2 \end{bmatrix} = \frac{1}{Z_0}\begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}\begin{bmatrix} V_1 \\ V_2 \end{bmatrix},$$

where the main diagonal terms are expressed as

$$C_{11} = \left(1 + \left(\frac{Z_0}{R_1}\right)\right),$$
$$C_{22} = \left(1 + \left(\frac{Z_0}{R_2}\right)\right),$$
$$C_{12} = \left(-1 + \left(\frac{Z_0}{R_1}\right)\right)\exp\left\{-\frac{L}{\gamma}\right\},$$
$$C_{21} = \left(-1 + \left(\frac{Z_0}{R_2}\right)\right)\exp\left\{-\frac{L}{\gamma}\right\}.$$

Solving for the voltage equations at terminals 1 and 2, the matrix can be inverted and can be represented as a function of the two current conditions:

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \frac{Z_0}{\Delta}\begin{bmatrix} C_{22} & -C_{12} \\ -C_{21} & C_{11} \end{bmatrix}\begin{bmatrix} I_1 \\ I_2 \end{bmatrix},$$

where the determinant is $\Delta = C_{11}C_{22} - C_{12}C_{21}$.

From this form, the voltage between two points in the base region (substrate) can be evaluated. From this form, the input impedance can be solved as a function of the characteristic impedance, the reflection coefficient and the loss factor. The input impedance can be expressed as

$$Z_{\text{in}} = \frac{V(0)}{I(0)} = Z_0 \frac{1 + \Gamma_{\text{R}} \exp\{-2L/\gamma\}}{1 + \Gamma_{\text{R}} \exp\{-2L/\gamma\}},$$

and the reflection coefficient can be expressed as

$$\Gamma_{\text{R}} = \frac{V_2}{[V_1 \exp\{-L/\gamma\}]}.$$

The transmission line case can be simplified by addressing the case in which the current is not flowing to terminal 2. This can represent the case of the current flowing to a second region such as the backside of the wafer. In this case, the surface potential may be equal the transmission line voltage. This can be solved for from the matrix relationship, as

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \frac{Z_0}{\Delta} \begin{bmatrix} C_{22} & -C_{12} \\ -C_{21} & C_{11} \end{bmatrix} \begin{bmatrix} I_1 \\ 0 \end{bmatrix},$$

where

$$\psi_{\text{s}}(y) = V(y) = V_1 \exp\left\{-\frac{y}{\gamma}\right\} \left[1 + \Gamma_{\text{R}} \exp\left\{-\frac{2(L-y)}{\gamma}\right\}\right],$$
$$I(y) = \frac{1}{Z_0} V_1 \exp\left\{-\frac{y}{\gamma}\right\} \left[1 - \Gamma_{\text{R}} \exp\left\{-\frac{2(L-y)}{\gamma}\right\}\right]$$

and

$$V_1 = \frac{Z_0}{\Delta} C_{22} I_1.$$

For the case where the current flowing from the first terminal equals the current from the second terminal, then $I_2 = -I_1$. The expressions for $V_1$ and $V_2$ coefficients can be expressed as a function of the current $I_1$ and the loss factors in the transmission line,

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \frac{Z_0}{\Delta} \begin{bmatrix} C_{22} & -C_{12} \\ -C_{21} & C_{11} \end{bmatrix} \begin{bmatrix} I_1 \\ -I_1 \end{bmatrix}$$

and surface potential,

$$\psi_{\text{s}}(y) = V(y) - V(L) = \left(V_1 \exp\left\{-\frac{y}{\gamma}\right\} - V_2\right) \left[1 - \exp\left\{-\frac{2(L-y)}{\gamma}\right\}\right].$$

Generalizing this approach, the current in one element is related to that in other, but unequal. This can be represented as a percentage of the first term. Hence, a third derivation can assume that $I_2 = -\delta I_1$, where $\delta$ is a percentage of the current.

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \frac{Z_0}{\Delta} \begin{bmatrix} C_{22} & -C_{12} \\ -C_{21} & C_{11} \end{bmatrix} \begin{bmatrix} I_1 \\ -\delta I_1 \end{bmatrix}.$$

To address the case where surface structures are present along the interface between the points of interest, the substrate can be represented as a plurality of transmission lines in series, where insertion resistors or stubs and other elements can interrupt the transmission line terminations. The voltage and current conditions can use matrix multiplication to introduce the terminations and boundary conditions. For example, along the interface, there are n+ implants and other diffusions that can modify the current and voltage characteristics along the surface. In this fashion, multiple elements can intersect the transmission line to address more complicated circuit environments.

### 2.11.5 The Lateral Bipolar Transistor – Substrate Lossy Transmission Line Representation (Troutman and Hargrove)

The substrate TL model characteristics will be a function of the doping concentration profile in the substrate. The doping concentration profile will have a strong influence on the loss characteristics of the TL model [69]. One case of a lossy TL is that of a p− epitaxy and p++ substrate. The cases of interest primarily are the heavily doped substrate with a thin epitaxial region and the lightly doped substrate. Today, applications that use epitaxial wafers and p++ substrates are trench DRAM, image processing technology and smart power applications. For the p-type heavily doped substrate, the cases of interest are the lightly doped epitaxy without a p-well implant, the p-well case with an epitaxial flat zone and the p-well with a connecting implant. For the lightly doped substrate, the types of substrate models of interest is the p-well with p− wafer, the p− wafer with heavily doped buried layer and the lightly doped wafer with the heavily doped buried layer and a termination implant. Additionally, this can be utilized for models of n− substrates and triple well. In all cases, the transmission line representation must address the transmission line loss factor, the characteristic impedance and the termination resistances.

Troutman showed that a TL representation can be determined if we regard the system as a conductive region near the device surface, a second conductive region below the surface and two termination resistors between the two conductive regions (Figure 2.16). This forms the lossy TL with the termination resistances related to the ‘up’ resistance between the two physical regions. For the

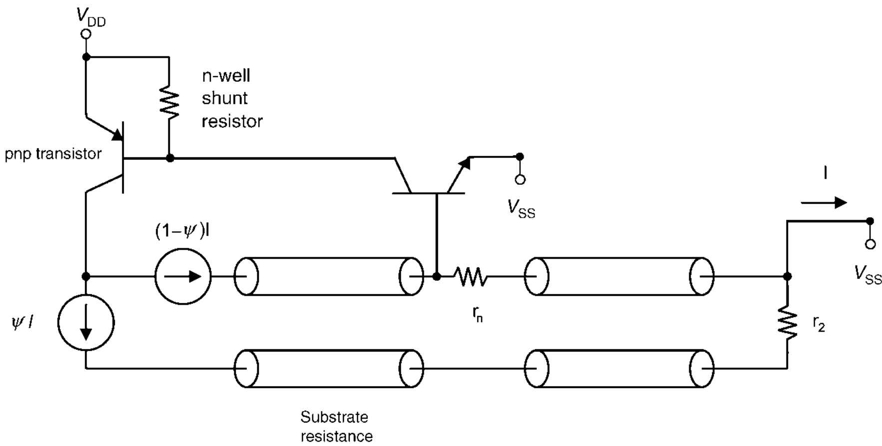



**Figure 2.16** Two-transistor representation of Troutman and Hargrove transmission line model.

lossy TL development, the loss factor and the characteristic impedance is important for the evaluation. These can be expressed as the series resistance and the shunt conductance. The loss factor can be expressed as

$$\gamma = \frac{1}{\sqrt{rg}}$$

and the characteristic impedance,

$$Z_0 = \sqrt{\frac{r}{g}}.$$

The series resistance per unit length can be evaluated as

$$r = \frac{1}{W} \frac{1}{\int_0^{x_1} \sigma(x) \mathrm{d}x},$$

where the integration is completed to an arbitrary depth at which the TL conductive plate is initiated. For the case of the implant near the device surface, with a background doping of $N_{\mathrm{epi}}$, a concentration for the surface region can be defined as

$$N_{\mathrm{A}}(x) = N_{\mathrm{epi}} + N_{\mathrm{A}} \exp\left\{ -\left[ \frac{x - x_0}{\sqrt{2}\sigma} \right]^2 \right\}.$$

In this case, the doping profile may represent a field surface implant or a p-well region. The series resistance per unit length can be represented as

$$r = \frac{\rho_0}{W T_{\mathrm{f}}},$$

where

$$\rho_0 = \frac{1}{q \mu_0 N_{\mathrm{A}}},$$

and $T_{\mathrm{f}}$ is expressed as a function of the number of Gaussian profile standard deviations,

$$T_{\mathrm{f}} = n\sigma.$$

Troutman showed that this can be represented as a dimensionless factor $n$ that is a function of the Gaussian profile, or

$$n = \sqrt{\frac{\pi}{2}} \left[ 1 + \mathrm{erf}\left( \frac{\mu}{\sqrt{2}\sigma} \right) \right] + x_1 \frac{N_{\mathrm{epi}}}{\sigma N_{\mathrm{A}}}.$$

In the case of a p++ substrate or a p++ buried layer, there is a conductive shunt to the TL lower ‘plate’. Let us define a shunt conductance per unit length of

$$g = \frac{W}{\rho_{\mathrm{epi}}(T_{\mathrm{epi}} + \lambda)},$$

where the epitaxial resistivity is

$$\rho_{\text{epi}} = \frac{1}{q\mu_{\text{p}}N_{\text{epi}}},$$

and the effective thickness is determined by the epitaxial flat zone at the end of the surface implant (or p-well) and at the beginning of the p+ substrate (or p++ buried layer). In the p++ substrate model case, Troutman added an effective depth factor to address the extra effective thickness between the epitaxial flat zone and the heavily doped flat zone of the p++ wafer to add an increase in the epitaxial effective resistance.

For the p++ wafer, a concentration profile can be assumed to be equal to

$$N_{\text{sub}}(x) = \frac{N_{\text{sub}}}{2}\operatorname{erfc}\left[\frac{x - x_{\text{e}}}{x_0}\right],$$

where we determine the epitaxial depth position from above expression, assuming $x_2 = x_{\text{e}}$.

From this, we can express the effective depth parameter as

$$\lambda = \frac{x_0}{2}\operatorname{erfc}^{-1}\left\{\frac{2N_{\text{epi}}}{N_{\text{sub}}}\right\}.$$

Substituting these expressions into the loss factor and characteristic impedance relationships,

$$\gamma = \frac{1}{\sqrt{rg}} = T_{\text{epi}}\left[\frac{T_{\text{f}}\rho_{\text{epi}}([1+\lambda]/T_{\text{epi}})}{T_{\text{epi}}\rho_0}\right]^{1/2}$$

and

$$Z_0 = \sqrt{\frac{r}{g}} = \frac{\rho_{\text{epi}}(T_{\text{epi}} + \lambda)}{W\gamma}.$$

For the case of the p++ buried layer implant, the model can be modified by treating the buried layer as a second series resistance term in parallel with the first series resistance term. In this case, the shunt conductance term would not include the effective depth factor. In the buried layer analysis, the second series resistance term determines the series resistance by integration over the buried layer implant. The buried layer implant would be represented by a two-sided Gaussian profile and the integration of the series resistance would integrate over the conductance. The epitaxial thickness boundary point would coincide with the bounds of the integration of the buried layer implant.

The lossy TL terminations are a function of the process implants. In the case of the HDBL implant, a heavily doped implant can be used to connect the layer to the substrate contact, providing a low resistance shunt between the surface and the buried layer. This provides a low-resistance termination of the line for latchup events.

### 2.11.6 The Bipolar Transistor – Substrate Transfer Resistance Representation

In latchup, the lateral BJT elements are coupled to other structures through the substrate. In the case of a lateral npn BJT, the base region is the p− substrate. In a semiconductor chip environment, the lateral

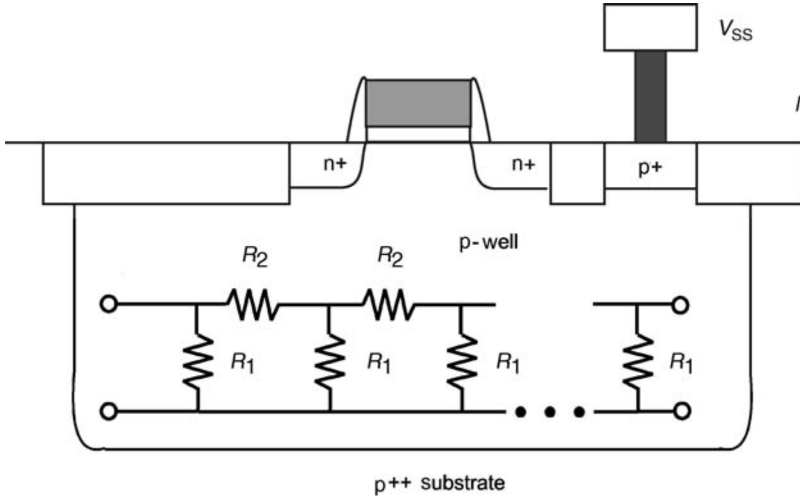



**Figure 2.17** Substrate representation as a ladder network.

npn bipolar transistor is coupled to stimuli of currents and voltages from surrounding elements. Troutman applied the concept of transfer resistance to the substrate to address latchup-related coupling phenomenon [70]. With the usage of a transfer resistance model representation, the analysis of the individual BJT can be analyzed with the substrate (or base transistor) coupling physics contained within the transfer resistance matrix. As will be discussed in Chapter 3, this concept will be valuable for understanding latchup interaction and its propagation through a plurality of circuit elements. In two dimensions, the base region of the lateral BJT network can be represented as a ladder network (Figure 2.17). But in three dimensions, to address a plurality of injection sources within the substrate region, a three-dimensional representation is desired. Figure 2.18 addresses the case where there are 3D and multiple elements in a common substrate. When there are multiple elements contained within a common substrate, the substrate potential for a given element is dependent on the voltage drops or current in the system [69,71,72].

Given $n$ elements contained in a common substrate, we can define a general system with a relationship to $n$ injection currents as

$$\begin{bmatrix} V_1 \\ V_2 \\ \cdot \\ \cdot \\ V_{\mathrm{n}} \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} & \cdot & \cdot & R_{1\mathrm{n}} \\ R_{21} & R_{22} & & & R_{2\mathrm{n}} \\ \cdot & & \cdot & & \\ \cdot & & & \cdot & \\ R_{\mathrm{n}1} & R_{\mathrm{n}2} & \cdot & \cdot & R_{\mathrm{nn}} \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \\ \cdot \\ \cdot \\ I_n \end{bmatrix}.$$

Representing the resistance expression as a second-order tensor and the voltage and current as a first-order vector,

$$V_i = R_{ij} I_j$$

or

$$V_i = \sum_{j=1}^{j=n} R_{ij} I_j = R_{i1} I_1 + R_{i2} I_2 + R_{i3} I_3 + \cdots + R_{in} I_{\mathrm{n}}.$$

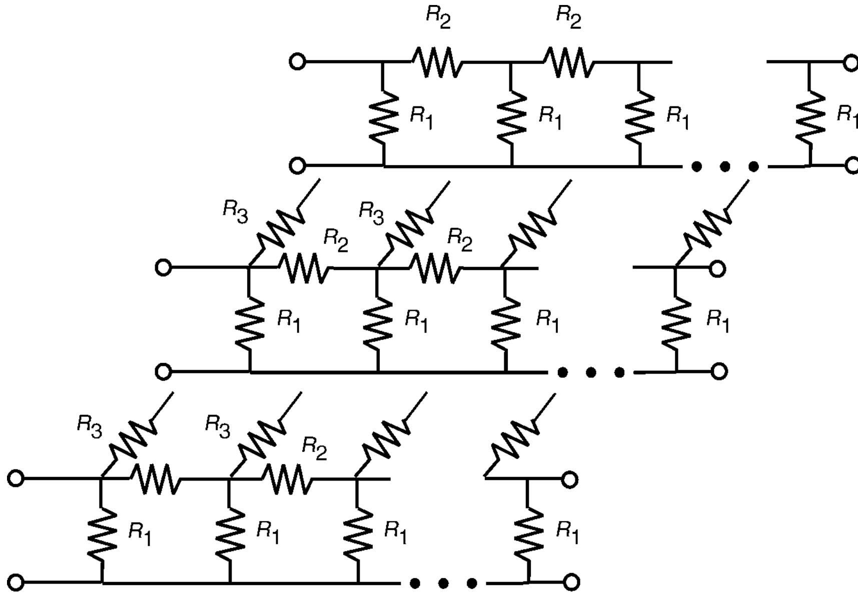



**Figure 2.18** Three-dimensional resistive substrate matrix with a plurality of inputs.

In this formulation, the resistance terms of the main diagonals are 'self-resistance' terms and the off-diagonal elements are 'mutual resistance' terms or transfer resistance terms. In this form, the influence on the $i$th element is a function of the current injected into the substrate from the other elements. In the case where there is a local substrate contact for every element and weak coupling, then we can assume that

$$R_{ij} = 0 \qquad \forall i \neq j,$$

$$\begin{bmatrix} V_1 \\ V_2 \\ \cdot \\ \cdot \\ V_{\mathrm{n}} \end{bmatrix} \approx \begin{bmatrix} R_{11} & 0 & \cdot & \cdot & 0 \\ 0 & R_{22} & & & \\ \cdot & & \cdot & & \\ \cdot & & & \cdot & \\ 0 & \cdot & & \cdot & R_{nn} \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \\ \cdot \\ \cdot \\ I_n \end{bmatrix}.$$

In this case, the individual elements can be treated as independent elements and no interaction through the substrate. In the case where there are $n$ elements, but only $m$ elements are connected to the common substrate contact, and all $n - m$ elements have independent substrate contacts, then the form of the matrix can be described as

$$\begin{bmatrix} V_1 \\ V_2 \\ \cdot \\ V_m \\ \cdot \\ V_{\mathrm{n-1}} \\ V_{\mathrm{n}} \end{bmatrix} = \begin{bmatrix} R_{11} & R_{12} & \cdot & R_{1m} & 0 & 0 & 0 \\ R_{21} & R_{22} & & R2m & 0 & 0 & 0 \\ \cdot & & \cdot & & 0 & 0 & 0 \\ R_{m1} & R_{m2} & & R_{mm} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdot & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & R_{\mathrm{n-1n-1}} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & R_{\mathrm{nn}} \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \\ \cdot \\ \\ \\ I_{\mathrm{n-1}} \\ I_{\mathrm{n}} \end{bmatrix}.$$

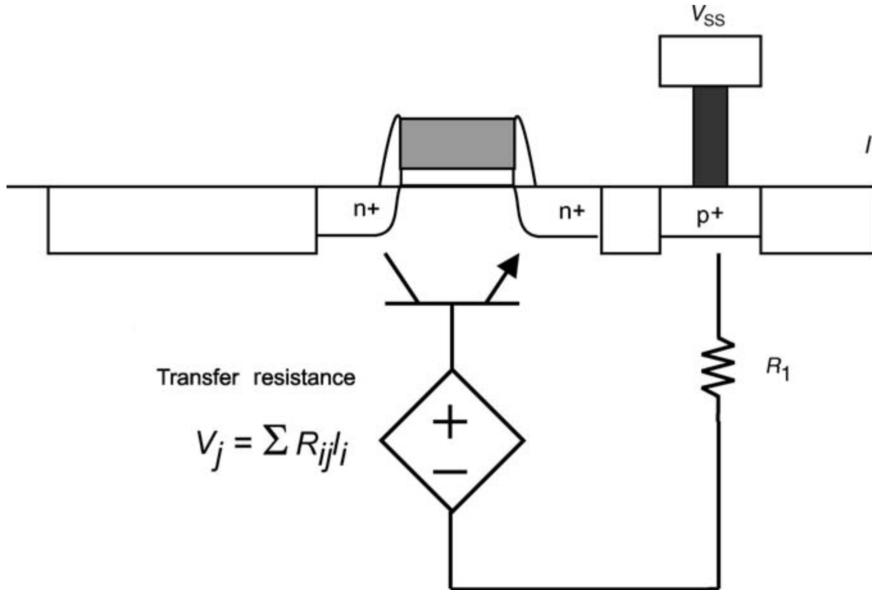



**Figure 2.19** Transfer resistance representation of the substrate.

The second-order resistance tensor main diagonal terms can be estimated from the physical resistance of the device to a common plane in the substrate region. This can be achieved analytically or through device simulation. This resistance may comprise a series of resistance elements in which the injected current flows through. The off-diagonal terms of the second-order resistance tensor can be obtained analytically or through simulation. These are associated with the influence of the substrate potential under the $i$th element as a result of injected current at the $j$th element.

In the solution by Tong Li [71,72], the transfer resistance was expressed in the form of a $m \times n \times p$ matrix of transfer resistances, where there are $m$ injection current sources, $n$ substrate contacts and $p$ locations of interest. In this form, it can be expressed as

$$R_{ji}^{k} = \frac{V_{ji}^{k}}{I_i}$$

where $i$ is the index for the $m$ injection sources, $j$ is the index of the $p$ locations of interest and $k$ is for the $n$ substrate contacts. In this form, the transfer matrix terms can be solved for the $m$ injection locations and $n$ substrate contacts. This can be applied to multiple devices or multifinger elements. Hence, for a single element, the transfer resistance term can be derived in the environment of $n$ surrounding elements to evaluate the base resistance response. This is critical to the understanding of the bipolar transistor response in a coupled substrate environment (Figure 2.19).

## 2.12 SUMMARY AND CLOSING COMMENTS

This chapter addresses the fundamentals of semiconductor bipolar transistor elements for both lateral and vertical devices. In the next chapter, CMOS latchup physics is addressed. In this chapter, the focus

will be on understanding the pnpn structure in the environment of integrated electronics and parasitic bipolar elements that occur in CMOS and BiCMOS technology.

## PROBLEMS

1. Show the relationship between the emitter injection efficiency factor $\gamma$, the base transport factor $B$, and the base defect $\delta_B$ and the emitter defect $\delta_E$ terms. Express the emitter efficiency factor and base transport factors as a function of the base and emitter defect.

2. Plot the Auger recombination time as a function of doping concentration from $10^{15}$ to $10^{21}\,cm^{-3}$ for both p-type and n-type silicon. Plot the recombination time for typical Shockley–Read–Hall recombination as a function of doping concentration. Plot both SRH and Auger recombination times as a function of doping independently, and as an effective recombination time. At what doping concentration does Auger recombination dominate the total SRH recombination? What doping concentration does SRH recombination dominate the Auger recombination? Given typical n-well and p-well implant dose magnitudes in semiconductor processing, what dose levels are needed in order to have Auger recombination doping significant?

3. Given typical subcollector doping concentrations, what doping concentrations and subcollector widths are needed in order to have significant Auger recombination doping? What are the values to prevent injection into the substrate? Given subcollectors are implanted instead of grown, whereas a $10^{16}\,cm^{-2}$ dose achieves standard subcollector doping concentrations, what is the lowest subcollector implant level at which Auger recombination will be dominant?

4. Given an objective of having a highly doped p++ buried layer (HDBL) is placed in a p− substrate, what width and concentration of the HDBL will prevent electrons from surviving penetrations through the HDBL layer?

5. From the Johnson limit relationship for maximum voltage – unity current cutoff frequency product,

$$V_m f_T = E_m \frac{v_s}{2\pi}.$$

   Express the speed of the transistor to switch to the breakdown voltage. Given a SCR rectifier, switching time can be expressed as the sum of the base transit times, express the switching relationship as a function of the ($V_m = BV_{CEO}$) and the Johnson limit relationship.

6. Assuming a MOSFET constant electric field scaling relationship with the scaling parameter $\alpha$, show the scaling relation assuming dimensional similitude scaling in both the lateral and vertical dimensions. Assume the vertical profile scales as a parameter $\beta$, derive the relationship for the transit time scaling as a function of both the lateral and vertical scaling relationships, assuming the pnp is scaled according to vertical scaling and the npn transistor is scaled according to the lateral scaling parameters.

7. Given an n-well CMOS in a p− substrates, show the vertical parasitic npn and pnp Ebers–Moll model superimposed on the CMOS structure (e.g. p diffusion in an n-well, n-well and substrate).

8. Assuming a substrate with a p− substrate doping concentration of $10^{15}\,cm^{-3}$ and a HDBL p++ layer of $10^{20}\,cm^{-3}$, evaluate the reflection factor at the top interface. How does it affect the lateral transport? Assuming a substrate with a p− substrate doping concentration of $10^{15}\,cm^{-3}$ and a

p-well of p+ layer of $10^{17}$ cm$^{-3}$, evaluate the reflection factor at the top interface between the p-well and the p− substrate. How does it affect the lateral transport?

9. Given a triple-well CMOS technology (n− region placed in the p− substrate region), show the vertical and lateral npn Ebers–Moll model superimposed on the triple-well structure.

10. Given a triple-well model, assume a uniform epitaxial region and a physical contact on both sides of the n-channel transistor in the isolated epitaxial region. Assume that the spacing of the p+ contacts in the triple-well region is $L$ and the voltage between the two contacts is $V_{\text{p-well}}$. Assume that the STI trench isolation is $L_{\text{STI}}$ and the depth of the triple-well implant is $L_{\text{TW}}$. Derive an electric field expression for a lateral electric field in the isolated epitaxial region based on the geometrical factors in the triple-well region. Assume a sheet resistance value where the sheet resistance is the value under the STI regions. Derive the vertical npn characteristics of the vertical npn in the presence of the lateral electric field.

## REFERENCES


1. J. Bardeen and W.H. Brittain. The transistor, a semiconductor triode. *Physical Review*, **74**, 1948, 230.
2. W. Shockley. Theories of high values of alpha for collector contacts on germanium. *Physical Review*, **78**, 1950, 294.
3. W. Shockley. The theory of p–n junctions in semiconductors and p–n junction transistors. *Bell Systems Technical Journal*, **28**, 1949, 435–489.
4. W. Van Roosbroeck. Theory of the flow of electrons and holes in germanium and other semicondutors. *Bell Systems Technical Journal*, 29, 1950, 560–607.
5. J.R. Haynes and W. Shockley. The mobility and life of injected holes and electrons in germanium. *Physical Review*, **81**, 1951, 835.
6. J.J. Ebers and J.L. Moll. Large-signal behavior of junction transistors. *Proceedings of IRE*, **42**, 1954, 1761–1772.
7. R.D. Middlebrook. *An Introduction to Junction Transistor Theory.* New York: John Wiley & Sons, Ltd.; 1957.
8. P.E. Gray, D. Dewitt, A.R. Boothroyd and J.F. Gibbons. *Physical Electronics and Circuit Models of Transistors*, Vol. 2. Semiconductor Electronics Education Committee. New York: John Wiley & Sons, Ltd.; 1964.
9. C.L. Searle, A.R. Boothroyd, E.J. Angelo, P.E. Gray, D. Dewitt and D.O. Pederson. *Elementary Circuit Properties of Transistors*, Vol. 3. Semiconductor Electronics Education Committtee. New York: John Wiley & Sons, Ltd.; 1964.
10. P. E. Gray and C.L. Searle. *Electronic Principles: Physics, Models, and Circuits*. New York: John Wiley & Sons, Ltd.; 1969.
11. B. Streetman. *Solid State Electronic Devices*. Englewood Cliffs, NJ: Prentice-Hall; 1972.
12. W. Shockley and Read WT Jr. Statistics of the recombination of holes and electrons. *Physical Review*, **87**(5), 1952, 835–842.
13. R.N. Hall. Electron–hole recombination in germanium. *Physical Review*, **87**(5), 1952, 387.
14. D.J. Fitzgerald and A.S. Grove. Surface recombination in semiconductors. *Surface Science*, **9**, 1968, 347–369.
15. J. Dwiezor and W. Schmid. Auger coefficients for highly doped and highly excited silicon. *Applied Physics Letters*, **31**(5), 1977, 346–347.
16. J.D. Beck and R. Conradt. Auger recombination in silicon. *Solid State Communication*, **13**, 1973, 93.
17. B.L. Gregory. Minority carrier recombination in neutron irradiated silicon. *IEEE Transactions on Nuclear Science*, **16**, 1969, 53–62.
18. B.L. Gregory and C.W. Gwyn. Application of neutron damage models to semiconductor device studies. *IEEE Transactions on Nuclear Science*, **17**, 1970, 325–334.
19. H.J. Stein and R. Gereth. Introduction rates of electrically active defects in n− and p-type silicon by electron and neutron radiation. *Journal of Applied Physics*, **39**, 1968, 2890–2904.
20. H.J. Stein. Defects in silicon: concepts and correlations. *Radiation effects in semiconductors*. New York: Gordon and Breach Science Publishers; 1970.
21. B.L. Gregory. Minority carrier trapping by defect clusters. *Applied Physics Letter*, **16**, 1970, 67–69.

22. M. M. Sanga and W.G. Oldham. Measurements of neutral base lifetime in neutron-irradiated transistors. *IEEE Transactions on Nuclear Science*, **20**, 1973, 266–273.
23. C.A. Goben. A study of neutron-induced base current component in silicon transistors. Sandia Report SC-R-65–912, Albuquerque, NM, July 1965.
24. Pakuti LJ Jr. *Analysis of radiation induced degradation in lateral transistors*. Ph.D. dissertation. Stanford University; 1971.
25. J.R. Adams and R. J. Sokel. Neutron irradiation for prevention of latch-up in MOS integrated circuits. In: *Proceedings of the Nuclear and Space Radiation Effects Conference (NSREC)*, July 19, 1979.
26. Dawes WR Jr., G.F. Derbenwick and B.L. Gregory. Process technology for radiation hardened CMOS integrated circuits. *IEEE Journal of Solid State Circuits*, **11**, 1976, 459–465.
27. D.B. Estreich. *The physics and modeling of latch-up and CMOS integrated circuits.* Technical Report No. G-201-9. Integrated Circuits Laboratory, Stanford University, November 1980.
28. K.P. Lisiak and A.G. Milnes. Platinum as a lifetime control deep impurity in silicon. *Journal of Applied Physics*, **46**, 1975, 5229–5235.
29. K.P. Lisiak and A.G. Milnes. Energy levels and concentrations for platinum in silicon. *Solid State Electronics*, **18**, 1975, 533–540.
30. R.F. Bailey. Semiconductor devices with diffused platinum. US Patent No. 3,640,783 (February 8, 1972).
31. W.M. Bullis. Properties of gold in silicon. *Solid State Electronics*, **9**, 1966, 143–168.
32. W.R. Wilcox and T.J. LaChappelle. Mechanism of gold diffusion into silicon. *Journal of Applied Physics*, **35**, 1964, 240–246.
33. A.E. Bankowski and J.H. Forster. Electrical properties of gold-doped diffused Si computer diodes. S. Chou. *Technical Journal of Bell Systems*, **39**, 1960, 87–104.
34. J.M. Fairfeld and B. V. Gokhale. Gold as a recombination center in silicon. *Solid State Electronics*, **8**, 1965, 685–691.
35. J.M. Fairfeld and B. V. Gokhale. Control of diffused diode recovery time through gold doping. *Solid State Electronics*, **9**, 1966, 905–907.
36. G. Clement and M. Voos. Minority carrier lifetime measurements in gold doped Si MOS structures. *Surface Science*, **11**, 1968, 147–148.
37. A.G. Nassibian. Effect of gold on surface properties and leakage current of MOS transistors. *Solid State Electronics*, **10**, 1967, 891–896.
38. S. F. Cagnina and E. H. Snow. Properties of gold doped MOS structures. *Journal of Electrochemical Society*, **114**, 1967, 1165–1172.
39. P. Richman. The effect of gold doping upon the characteristics of MOSFET field-effect transistors with applied substrate voltage. *Proceedings of the IEEE*, **56**, 1968, 774–775.
40. S.D. Brotherton. Electrical properties of gold at the silicon–dielectric interface. *Journal of Applied Physics*, **42**, 1971, 2085–94.
41. D.R. Lamb, G.R. Moghal and R.J. Hawkins. The effect of gold doping on the threshold voltage, hall mobility, gain, and current noise of MOS transistors. *International Journal of Electronics*, **30**, 1971, 141–147.
42. G. R. Moghal. Determination of gold concentration and the effective impurity doping at the silicon/silicon dioxide interface. *International Journal of Electronics*, **44**, 1978, 205–269.
43. Dawes WR Jr and G.F. Derbenwick. Prevention of CMOS latchup by gold doping. *IEEE Transaction on Nuclear Science*, **23**, 1976, 2027–2030.
44. J. L. Wirth and S.C. Rogers. The transient response of transistors and diodes to ionizing radiation. *IEEE Transactions on Nuclear Science*, **11**, 1964, 24–38.
45. W.H. Dierking, G.E. Katz and E.L. Steele. Transient radiation current generator model for semiconductor devices. *IEEE Transactions on Nuclear Science*, **16**, 1969, 144–152.
46. S. Voldman. *ESD: physics and devices*. Chichester: John Wiley & Sons, Ltd.; 2004.
47. R.S. Muller and T.I. Kamins. *Device Electronics for Integrated Circuits*. New York: John Wiley & Sons, Inc.; 1977.
48. W. K. Tsang and K.M. Busen. Detailed investigations of a lateral pnp transistor. *International Electron Device Meeting (IEDM) Technical Digest*; 1965. p. 32–35.
49. J. Lindmayer and W. Schneider. Theory of the lateral transistor. *Solid State Electronics*, **10**, 1967, 225–234.
50. S. Chou. An investigation of lateral transistors – d.c. characteristics. *Solid State Electronics*, **14**, 1971, 811–826.
51. S. Chou. Small signal characteristics of lateral transistors. *Solid State Electronics*, **15**, 1972, 27–38.

52. E. L. Long. Current-induced field enhancement of lateral transistors. *International Electron Device Meeting (IEDM) Technical Digest*; 1970, p. 72–74.
53. I. Kidron. Integrated circuit model for lateral pnp transistors including isolation junction interactions. *International Journal of Electronics*, **31**, 1971, 421–440.
54. D. Seltz and I. Kidron. A two-dimensional model for the lateral pnp transistor. *IEEE Transactions on Electron Devices*, **21**(9), 1974, 587–592.
55. D. E. Fulkerson. A two-dimensional model for calculation of common-emitter current gains of lateral pnp transistor. *Solid State Electronics*, **11**, 1968, 821–826.
56. N. Parameswaran and S. Tyagi. A two-dimensional analysis of common emitter current gain in a lateral transistor. *International Journal of Electronics*, **40**, 1976, 593–600.
57. T. A. Demassa and L. Rispin. Analysis of modified lateral pnp transistors. *Solid State Electronics*, **18**, 1975, 481–490.
58. H.H. Berger and U. Dreckmann. The lateral pnp transistor – a practical investigation of the d.c. characteristics. *IEEE Transactions on Electron Devices*, **26**, 1979, 1038–1046.
59. E. Yang. *Fundamentals of Semiconductor Devices*. New York: McGraw-Hill; 1978.
60. G.G.E. Low. Carrier concentration disturbances in semiconductors. *Proceedings of the Physical Society B*, **68**, 1955, 310–314.
61. J.B. Arthur, A.F. Gibson and J.B. Gunn. Current gain at L–H junctions in germanium. *Proceedings of the Physical Society B*, **69**, 1956, 705–711.
62. J.B. Gunn. On carrier accumulation, and the properties of certain semiconductor junctions. *Journal of Electronics and Control*, **4**, 1958, 17–50.
63. R.W. Lade and A.G. Jordan. On the static characteristics of high–low junction devices. *Journal of Electronics and Control*, **13**, 1962, 23–31.
64. D.L. Scharfetter. Minority carrier injection and charge storage in epitaxial schottky barrier diodes. *Solid State Electronics*, **8**, 1965, 299.
65. R.W. Dutton and R.J. Whittier. Forward current–voltage and switching characteristics of $p^+$–n–$n^+$ (epitaxial) diodes. *IEEE Transactions on Electronic Devices*, **16**, 1969, 458–467.
66. K. Venkatateswaran and D.J. Roulston. Recombination dependent characteristics of silicon $p^+$–n–$n^+$ epitaxial diodes. *Solid State Electronics*, **15**, 1972, 311–323.
67. J.R. Hauser and P.M. Dunbar. Minority carrier reflecting properties of semiconductor high–low junctions, *Solid State Electronics*, **18**, 1975, 715–716.
68. R. Stella, S. Favilla and G. Croce. Novel achievements in the understanding and suppression of parasitic minority carrier currents in $p-$ epitaxial/$p^{++}$ substrate smart power technologies. In: *Proceedings of the International Symposium on Power Semiconductor Devices* (*ISPSD*); 2004. p. 423–426.
69. R.R. Troutman and M.J. Hargrove. Transmission line model for latchup in CMOS circuits. In: *Symposium on VLSI Technology*; 13–15 September 1983. p. 56–59.
70. R. Troutman. *CMOS Latchup – The Problem and the Cure*. New York: Kluwer, 1985.
71. T. Li. *Design automation for reliable CMOS chip I/O circuits*. Ph.D. thesis. University of Illinois Urbana-Champaign, UILU-ENG-98-2219; August 1998.
72. T. Li, C. H. Tsai, E. Rosenbaum and S.M. Kang. Substrate modeling and lumped substrate resistance extraction for latchup/ESD circuit simulations. In: *IEEE/ACM Design Automation Conference* (*DAC*); 1999. p. 549–554.

# 3 Latchup Theory

In this chapter, the basics of latchup theory and the pnpn will be discussed. The equations for latchup initiation, regenerative feedback as well as sensitivity to all of the physical variables will be discussed. In the first section, basic dc equations are discussed to understand the physical equations and nature of latchup. In the second section, a second look is taken based on the work of differential and ac analysis of latchup. In the differential analysis, the form and conditions of the equations will constantly refer back to the earlier formulations to generate insight into the differences of the two bodies of work. Transient latchup, external latchup injection and guard rings are also discussed.

The early theory for latchup was associated with the understanding of the pnpn structure, also known as the silicon-controlled rectifier (SCR) [1–8]. Shockley, Ebers, Moll and others were involved in the modeling and physics of the pnpn structure [1–8]. In the late 1970s, the emphasis changed in the understanding of parasitic pnpn structures for the purpose of quantifying CMOS latchup in integrated electronics. The primary focus shifted toward understanding of latchup in terrestrial applications for mainstream CMOS technology. In the late 1970s to the mid-1980s, the literature addressed modeling, characterization and semiconductor process solutions for elimination [9–26]. In the mid-1980s, the focus shifted toward understanding of distribution effects, process solutions and transient latchup [26–58]. In the 1990s, the latchup work focused on standardization and standards development for qualification of CMOS and BiCMOS (bipolar CMOS) products [59–70].

## 3.1 REGENERATIVE FEEDBACK

The latchup S-type *I–V* characteristic allows the circuitry to transition from a low-current/high-voltage state to a high-current/low-voltage state. It is 'off' in normal operation and can be triggered 'on' in a high-current state. In this state, it establishes a high current at a low voltage, allowing a low-impedance shunt. At low impedance, the current can lead to thermal failure.

Conceptually, the four-region pnpn can be understood as a cross-coupled pnp and npn bipolar junction transistor (BJT) device, where the base of the pnp BJT device is the collector of the npn BJT device and the base of the npn is the collector of the pnp BJT device. This pnp–npn BJT coupling establishes regenerative feedback leading to the S-type *I–V* characteristic and causing the electrical instability that is observed as a negative resistance state (e.g. $dI/dV < 0$). As discussed in Chapter 1,


© 2007 John Wiley & Sons, Ltd

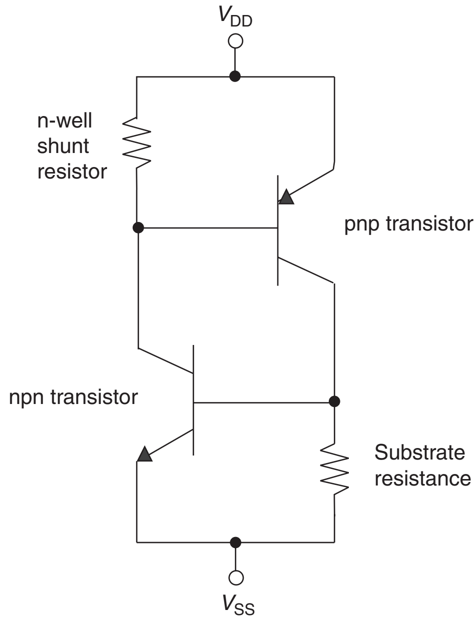



**Figure 3.1** Latchup circuit schematic with shunt resistors (e.g. associated with the well and substrate resistances).

the circuit can be represented as a two-transistor circuit. In the earlier representation, the resistances of the well and the substrate regions were not shown. Figure 3.1 shows the circuit schematic with shunt resistances across the emitter–base junctions of the pnp and the npn transistor.

### 3.1.1 Regenerative Feedback without Shunt Resistors and Alpha Representation

Application of a positive bias on the pnp emitter and a ground potential on the npn emitter establishes a voltage across the pnpn. The positive voltage provides a forward biasing of the emitter–base junctions of both pnp and npn transistors. The pnp base–collector junction (which is also the npn base–collector junction) is reverse biased; this prevents pnp anode-to-cathode current flow. As the anode-to-cathode voltage increases, the base–collector junction voltage increases. This mode of operation is called the forward blocking state [1–7].

For current continuity at the cross-coupled nodes, the pnp transistor collector current must equal the npn transistor base current, as well as the npn collector current must equal the pnp base current. Mathematically, the coupling is established through solving Kirchoff's current law at the base–collector nodes. In this form, the standard equations of bipolar transistors can be used to quantify the interaction and current in the pnpn structure. Hence, the two nodal equations can be expressed as

$$I_{\mathrm{cp}} = I_{\mathrm{bn}},$$
$$I_{\mathrm{cn}} = I_{\mathrm{bp}},$$

where $I_{cp}$ and $I_{cn}$ are the pnp and npn BJT collector currents, respectively, and likewise, $I_{bn}$ and $I_{bp}$ are the npn and pnp BJT base currents, respectively. The pnpn total current is equal to the pnp BJT emitter current or the npn BJT emitter current. From Kirchoff's current law, the emitter current must equal the sum of the base and collector currents:

$$I = I_{ep} = I_{cp} + I_{bp}.$$

From the coupling relationships, it can be expressed as

$$I = I_{ep} = I_{cp} + I_{cn} = I_{bp} + I_{bn} = I_{cn} + I_{bn} = I_{en} = I.$$

Solving for the current as a function of the two collector relationships, the collector current can be expressed as a function of the emitter current,

$$I_{cp} = \alpha I_{ep} + I_{cp0},$$
$$I_{cn} = \alpha I_{en} + I_{cn0},$$

where the collector current is equal to the product of transport factor and the emitter current summed with the base–collector leakage. Solving for the current through the pnpn structure, the current can be expressed as

$$I = \frac{I_{cp0} + I_{cn0}}{1 - (\alpha_n + \alpha_p)}.$$

Given that the transport factor is less than 1, it can be shown that all states of the system must be contained within a rectangle in $\alpha_n$–$\alpha_p$ space. Figure 3.2 shows the 'alpha space' mapping with two regions. In the rectangular region, this can be divided into two triangles where the sum of the alpha terms is less than and greater than unity.

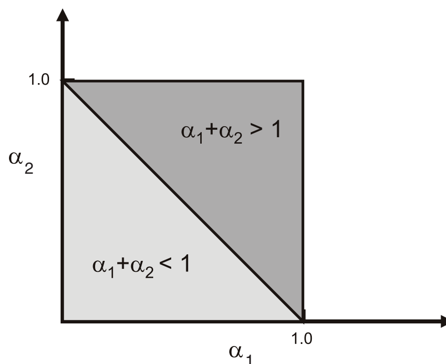



**Figure 3.2** Alpha space highlighting the rectangle of all states of the two-transistor system. The rectangular region is divided into two triangular regions.

### 3.1.2 Regenerative Feedback without Shunt Resistors and Beta Representation

The above current relationship can be expressed as a function of the bipolar current gain, substituting in for the collector-to-emitter transport [13]:

$$I = \frac{I_{cp0} + I_{cn0}}{1 - \left( \frac{\beta_n}{\beta_n + 1} + \frac{\beta_p}{\beta_p + 1} \right)}.$$

### 3.1.3 Regenerative Feedback with Shunt Resistors and Alpha Representation

The condition for latchup initiation is a function of the currents through the feedback elements. The intrinsic resistances in the collector, base and emitter structures can influence the regenerative feedback condition. The evaluation of the base resistance plays a critical role in the modulation of the regenerative feedback. The base resistances play a significant role in pnpn holding voltage $V_H$. In CMOS technology, the well and substrate resistances serve as the base resistances, where these resistances can be modified by both semiconductor process and design. It was shown that the total pnpn current equals the emitter current. The emitter current is equal to the sum of the collector currents, or the sum of the base currents.

$$I = I_{cp} + I_{cn},$$

where

$$I_{cp} = \alpha I_{ep} + I_{cp0},$$
$$I_{cn} = \alpha I_{en} + I_{cn0}.$$

The base resistance terms of the pnp and the npn transistor are outside of the collector–base current loop in the regenerative feedback, but they do play a role in the relationship between the emitter and the pnpn currents. Taking into account the current flow through the shunt resistors,

$$I = I_{ep} + I_w,$$
$$I = I_{en} + I_{sx}.$$

Then,

$$I = I_{cp} + I_{cn} = \alpha_p I_{ep} + I_{cp0} + \alpha_n I_{en} + I_{cn0}.$$

Rearranging the terms,

$$I = \alpha_p (I_{ep} + I_w) - \alpha_p (I_w) + I_{cp0} + \alpha_n (I_{en} + I_{sx}) - \alpha_n (I_{sx}) + I_{cn0}.$$

Again, grouping the terms,

$$I = (\alpha_p + \alpha_n) I - \alpha_p (I_w) - \alpha_n (I_{sx}) + (I_{cn0} + I_{cp0}).$$

From this form, the expression for current is

$$I = \frac{(I_{\mathrm{cn0}} + I_{\mathrm{cp0}}) - \alpha_{\mathrm{p}}(I_{\mathrm{w}}) - \alpha_{\mathrm{n}}(I_{\mathrm{s}x})}{1 - (\alpha_{\mathrm{p}} + \alpha_{\mathrm{n}})}.$$

In this form, two important points can be observed. First, the numerator has been modified to include the current through the well and substrate shunt resistances. Second, the denominator condition has not been modified (e.g. compared to the analysis without the well and substrate resistance terms).

From the equation on the total current relationship,

$$(\alpha_{\mathrm{p}} + \alpha_{\mathrm{n}})I - \alpha_{\mathrm{p}}(I_{\mathrm{w}}) - \alpha_{\mathrm{n}}(I_{\mathrm{s}x}) + (I_{\mathrm{cn0}} + I_{\mathrm{cp0}}) - I = 0.$$

Dividing by the current,

$$(\alpha_{\mathrm{p}} + \alpha_{\mathrm{n}}) - \alpha_{\mathrm{p}}\frac{(I_{\mathrm{w}})}{I} - \alpha_{\mathrm{n}}\frac{(I_{\mathrm{s}x})}{I} + \frac{(I_{\mathrm{cn0}} + I_{\mathrm{cp0}})}{I} - 1 = 0.$$

Assuming that the leakage current is significantly less than the current, the condition for latchup can be expressed as [13]

$$(\alpha_{\mathrm{p}} + \alpha_{\mathrm{n}}) = 1 + \alpha_{\mathrm{p}}\frac{(I_{\mathrm{w}})}{I} + \alpha_{\mathrm{n}}\frac{(I_{\mathrm{s}x})}{I}.$$

This equation shows that the criterion for instability is modified by the additional terms in the equation. When the well or substrate shunt resistor currents are zero, then the right-hand side of the equation is unity (e.g. the 'sum of the alphas' is unity). When the well or substrate current term is nonzero, then the latchup condition requires that the 'sum of the alphas' has to be greater than unity.

### 3.1.4 Regenerative Feedback with Shunt Resistors and Beta Representation

From the bipolar current gain expression, for the pnp BJT,

$$\alpha_{\mathrm{p}} = \frac{\beta_{\mathrm{p}}}{\beta_{\mathrm{p}} + 1}$$

and for the npn BJT,

$$\alpha_{\mathrm{n}} = \frac{\beta_{\mathrm{n}}}{\beta_{\mathrm{n}} + 1.}$$

Substituting the bipolar current gain term for the collector-to-emitter transport factors,

$$\left(\frac{\beta_{\mathrm{p}}}{\beta_{\mathrm{p}} + 1} + \frac{\beta_{\mathrm{n}}}{\beta_{\mathrm{n}} + 1}\right) = 1 + \frac{\beta_{\mathrm{p}}}{\beta_{\mathrm{p}} + 1}\frac{(I_{\mathrm{w}})}{I} + \frac{\beta_{\mathrm{n}}}{\beta_{\mathrm{n}} + 1}\frac{(I_{\mathrm{s}x})}{I}.$$

Multiplying both sides of the equation by the term $(\beta_n + 1)(\beta_p + 1)$, the equation can be expressed as

$$\beta_p(\beta_n + 1) + \beta_n(\beta_p + 1) = (\beta_n + 1)(\beta_p + 1) + \frac{(I_w)}{I}\beta_p(\beta_n + 1) + \frac{(I_{sx})}{I}(\beta_p + 1)\beta_n.$$

Factoring out the terms and eliminating of common terms on both sides of the equation, this can be expressed as a function of the forward bipolar current gain product,

$$\beta_p\beta_n = 1 + \frac{(I_w)}{I}\beta_p(\beta_n + 1) + \frac{(I_{sx})}{I}(\beta_p + 1)\beta_n.$$

In this form, it can be observed that in the case where the current flowing through the shunt elements approaches zero, this expression reduces to the product of the bipolar current gains equal to unity. Additionally, when these terms are nonzero, the requirement for latchup is that the product of the bipolar current gains must be greater than unity. Hence, it is clear that the presence of the shunt resistances leads to a higher requirement to initiate latchup. Also note that in these terms the RHS terms contain the bipolar current gain product and they are of the same order of magnitude as the term on the left-hand side (e.g. the terms are of order $O[\beta^2]$). As a result, when the ratio of the shunt currents is of the same order of magnitude as the current $I$ (e.g. $O[1]$), then the terms on the RHS are significant.

A key question is given a known bipolar current gain of the first transistor, what is the necessary condition of the bipolar current gain of the second transistor to initiate latchup? To determine the bipolar current gain required to satisfy this condition, knowing one of the bipolar current gains, we can solve for the current gain of the second transistor as a function of the first transistor. Factoring the above equation,

$$\beta_p\left[\beta_n - \frac{(I_w)}{I}(\beta_n + 1) - \frac{(I_{sx})}{I}\beta_n\right] = 1 + \frac{(I_{sx})}{I}\beta_n.$$

Factoring and solving for $\beta_{pnp}$ (denoted as $\beta_p$) as a function of $\beta_{npn}$ (denoted as $\beta_n$) and currents in the well resistor, the substrate resistor and the pnpn [13],

$$\beta_p \geq \frac{\frac{(I)}{\beta_n} + I_{sx}}{\left[I - I_w\frac{(\beta_n + 1)}{\beta_n} - I_{sx}\right]}.$$

To obtain the criterion as a function of the current gain product, we can multiply the expression by the npn BJT current gain,

$$\beta_n\beta_p \geq \frac{I + I_{sx}\beta_n}{\left[I - I_w\frac{(\beta_n + 1)}{\beta_n} - I_{sx}\right]},$$

or equivalently due to symmetry [13]

$$\beta_n\beta_p \geq \frac{I + I_w\beta_p}{\left[I - I_{sx}\frac{(\beta_p + 1)}{\beta_p} - I_w\right]}.$$

The well and substrate currents can be solved as a function of $V_{BE}$ divided by the well and substrate resistances, respectively:

$$I_w = \frac{(V_{BE})_{pnp}}{R_w} = \frac{V_0}{R_w} \ln\left[\frac{I - I_w}{(I_0)p}\right],$$
$$I_{sx} = \frac{(V_{BE})_{npn}}{R_{sx}} = \frac{V_0}{R_{sx}} \ln\left[\frac{I - I_{sx}}{(I_0)_n}\right].$$

## 3.1.5 Sensitivity Factors – Beta Product

The 'beta product' relationship $\beta_p\beta_n$ (or $\beta_{pnp}\beta_{npn}$) is a function of semiconductor process and design parameters. A latchup design practice is the evaluation of the beta product $\beta_p\beta_n$ to the fundamental parameters. In the development of a semiconductor technology and design rules, it is important to understand how this metric varies as a function of the independent variables. A latchup design practice is the evaluation of the partial derivatives of the beta product relationship. The beta product sensitivity relationship as a function of the bipolar npn and the bipolar pnp of the parasitic pnpn can be expressed as [13]

$$S_{\beta_n} = \frac{\beta_n}{\beta_n\beta_p}\frac{\partial}{\partial\beta_n}\{\beta_n\beta_p\}$$

and

$$S_{\beta_p} = \frac{\beta_p}{\beta_n\beta_p}\frac{\partial}{\partial\beta_n}\{\beta_n\beta_p\}.$$

From the relationship, other parameters can be defined. As an example, Estreich [13] noted that for an n-type substrate the beta product relationship can be defined as

$$\beta_n\beta_p \geq \frac{F + \left\{\frac{\beta_n^2}{\beta_n(K+1)+K}\right\}}{F-1},$$

where

$$F = \frac{I_{DD}}{I_{R_W} + I_{R_S}\left(\frac{\beta_n + 1}{\beta_n}\right)}$$

and

$$K = \frac{I_{R_S}}{I_{R_W}}.$$

From this form, it can be observed that the parameters $F$ and $K$ are on the RHS of the equation. As a result, the magnitude of the RHS of the equation is a function of the parameters $F$ and $K$. Estreich

defined sensitivity parameters as a function of these parameters. The parameter $K$ contains the current flowing through the well and substrate resistors and influences the triggering of latchup. A latchup design practice is the evaluation of the sensitivity parameters on the RHS of the beta product relationship [13]:

$$S_{\beta_{p}} = \frac{F}{\beta_{n}\beta_{p}} \frac{\partial}{\partial F} \{\beta_{n}\beta_{p}\}$$

and

$$S_{\beta_{p}} = \frac{K}{\beta_{n}\beta_{p}} \frac{\partial}{\partial K} \{\beta_{n}\beta_{p}\}.$$

Emitter resistance of the parasitic pnp and npn transistors can also influence the 'beta product' relationship. The sensitivity of $\beta_{pnp}\beta_{npn}$ to the emitter resistance can be evaluated from the following sensitivity factors for the pnp and npn emitter resistances, respectively [13]:

$$S_{r_{ep}} = \frac{r_{ep}}{\beta_{n}\beta_{p}} \frac{\partial}{\partial_{r_{ep}}} \{\beta_{n}\beta_{p}\},$$

$$S_{r_{en}} = \frac{r_{en}}{\beta_{n}\beta_{p}} \frac{\partial}{\partial_{r_{en}}} \{\beta_{n}\beta_{p}\}.$$

## 3.2 LATCHUP CRITERION WITH EMITTER RESISTANCE

Latchup is a function of the bipolar current gains as well as corresponding resistances. In addition, any other resistance term can influence the condition for latchup and the circuit stability. Latchup is a function of the following resistances:

- collector resistance;
- emitter resistance.

The intrinsic resistances in the emitter structure can influence the regenerative feedback latchup condition. In addition, emitter resistances influence both the electrical and the thermal stability of a network. The latchup triggering condition is a function of the currents through the feedback elements. Figure 3.3 shows the circuit schematic of the CMOS circuit with the inclusion of the shunt and the emitter resistances.

From the circuit schematic, it is clear that the emitter resistances are not present in the regenerative feedback loop (e.g. note that the regenerative feedback is associated with the cross-coupling of the base–collector junctions). But, the emitter resistances are present in the Kirchoff current loop that influences the forward bias condition of the emitter. In addition, the emitter resistance is also in the current path of the latchup current. Hence, the presence of the emitter resistances will play a role in 'debiasing' the emitter–base junction; in this fashion, it influences the feedback of the network.

In addition, the emitter resistors also alter the current flow through the shunt resistances. As the emitter resistances increase, more current will flow through the shunt resistors, and less through the bipolar junction transistor. In the case where the emitter resistances were infinite, the current would only flow through the well and substrate shunt elements. As a result, it is clear that the emitter resistance influences the shunt currents. Another means to observe this is that the emitter resistance and

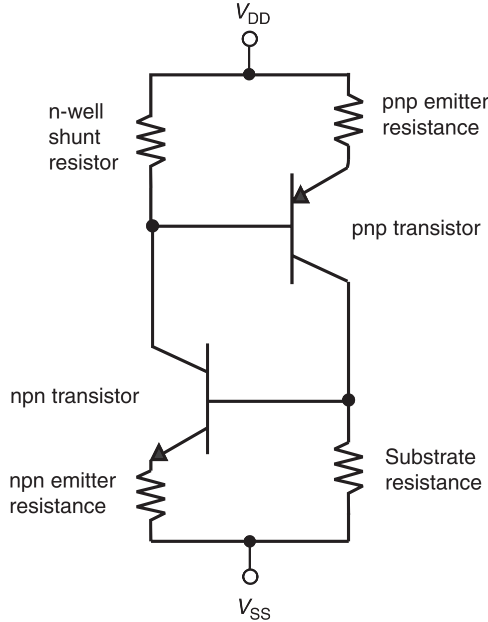



**Figure 3.3** CMOS latchup circuit schematic with the emitter resistances.

the shunt resistance are contained within the same Kirchoff current loop with the emitter–base junction of the respective transistor.

As shown above, to determine the bipolar current gain required to satisfy the latchup condition, knowing one of the bipolar current gains, the current gain of the second transistor as a function of the first transistor can be solved. $\beta_{\mathrm{pnp}}$ can be factored and solved as a function of $\beta_{\mathrm{npn}}$, well resistor current, substrate resistor current and the current through the pnpn structure. In the presence of the emitter resistances, the substrate and well resistances are modified. This can be shown as follows:

$$\beta_{\mathrm{p}} \geq \frac{\dfrac{(I)}{\beta_{\mathrm{n}}} + I'_{sx}}{\left[I - I'_{\mathrm{w}}\dfrac{(\beta_{\mathrm{n}}+1)}{\beta_{\mathrm{n}}} - I'_{sx}\right]},$$

and hence the product of the bipolar current gains can be expressed as the modified substrate and well currents,

$$\beta_{\mathrm{n}}\beta_{\mathrm{p}} \geq \frac{I + I'_{sx}\beta_{\mathrm{n}}}{\left[I - I'_{\mathrm{w}}\dfrac{(\beta_{\mathrm{n}}+1)}{\beta_{\mathrm{n}}} - I'_{sx}\right]},$$

or equivalently due to symmetry [13]

$$\beta_{\mathrm{n}}\beta_{\mathrm{p}} \geq \frac{I + I'_{\mathrm{w}}\beta_{\mathrm{p}}}{\left[I - I'_{sx}\dfrac{(\beta_{\mathrm{p}}+1)}{\beta_{\mathrm{p}}} - I'_{\mathrm{w}}\right]}.$$

The emitter resistance in the pnp transistor influences the voltage drops across the n-well resistor. Using Kirchoff's voltage law, the sum of the voltage drops in the current loop across the two resistor elements (well and emitter) and emitter–base voltage equals zero. The current flowing through the emitter resistor is the pnpn current minus the current through the n-well resistor; hence, the well resistor current is [13]

$$I'_{\mathrm{w}} = \frac{(V_{\mathrm{BE}})_{\mathrm{pnp}} + Ir_{\mathrm{ep}}}{R_{\mathrm{w}} + r_{\mathrm{ep}}}.$$

From this expression, as the emitter resistance approaches infinity, no current flows through the emitter resistor, and

$$I'_{\mathrm{w}} \cong I.$$

For the opposite case, where the emitter resistance approaches zero, this reduces to the well-known term,

$$I'_{\mathrm{w}} \approx \frac{(V_{\mathrm{BE}})_{\mathrm{pnp}}}{R_{\mathrm{w}}}.$$

Equivalently, the substrate resistor current is

$$I'_{sx} = \frac{(V_{\mathrm{BE}})_{\mathrm{npn}} + Ir_{\mathrm{en}}}{R_{sx} + r_{\mathrm{en}}}.$$

Focusing on one transistor only, for the case of the emitter resistance for the npn (ignoring the pnp emitter resistance), the latchup criterion is as follows [13]:

$$\beta_{\mathrm{n}}\beta_{\mathrm{p}} \geq \frac{I + \left(\dfrac{(V_{\mathrm{BE}})_{\mathrm{npn}} + Ir_{\mathrm{en}}}{R_{sx} + r_{\mathrm{en}}}\right)\beta_{\mathrm{n}}}{\left[I - I'_{\mathrm{w}}\dfrac{(\beta_{\mathrm{n}} + 1)}{\beta_{\mathrm{n}}} - \left(\dfrac{(V_{\mathrm{BE}})_{\mathrm{npn}} + Ir_{\mathrm{en}}}{R_{sx} + r_{\mathrm{en}}}\right)\right]}.$$

For the case of the emitter resistance for the pnp (ignoring the npn emitter resistance),

$$\beta_{\mathrm{n}}\beta_{\mathrm{p}} \geq \frac{I + \left(\dfrac{(V_{\mathrm{BE}})_{\mathrm{pnp}} + Ir_{\mathrm{ep}}}{R_{\mathrm{w}} + r_{\mathrm{ep}}}\right)\beta_{\mathrm{p}}}{\left[I - I'_{\mathrm{SX}}\dfrac{(\beta_{\mathrm{p}} + 1)}{\beta_{\mathrm{p}}} - \left(\dfrac{(V_{\mathrm{BE}})_{\mathrm{pnp}} + Ir_{\mathrm{ep}}}{R_{\mathrm{w}} + r_{\mathrm{ep}}}\right)\right]}.$$

For the effect of both transistors, the latchup criterion including the pnp and npn emitter resistances [13] is

$$\beta_{\mathrm{n}}\beta_{\mathrm{p}} \geq \frac{I + \left(\dfrac{(V_{\mathrm{BE}})_{\mathrm{npn}} + Ir_{\mathrm{en}}}{R_{sx} + r_{\mathrm{en}}}\right)\beta_{\mathrm{n}}}{\left[I - \left(\dfrac{(V_{\mathrm{BE}})_{\mathrm{pnp}} + Ir_{\mathrm{ep}}}{R_{\mathrm{w}} + r_{\mathrm{ep}}}\right)\dfrac{(\beta_{\mathrm{n}} + 1)}{\beta_{\mathrm{n}}} - \left(\dfrac{(V_{\mathrm{BE}})_{\mathrm{npn}} + Ir_{\mathrm{en}}}{R_{sx} + r_{\mathrm{en}}}\right)\right]}.$$

## 3.3 HOLDING POINT CONDITIONS

The holding point condition is a key latchup metric. It is a common latchup design practice to have the holding voltage ($V_H$) above the power supply voltage ($V_{DD}$) so that latchup cannot be maintained. In other applications, holding current $I_H$ is critical.

### 3.3.1 Holding Voltage

The holding voltage $V_H$ is the voltage across the pnpn structure after the structure switches from its blocking condition to its 'on' state. After the voltage exceeds the knee voltage, the pnpn undergoes regenerative feedback and transitions from a low-current/high-voltage state to a high-current/low-voltage state. As a rough approximation, $V_H$ is equal to the sum of the voltage drops across the structure when the transistors are forward active and the voltage across the n-well to substrate junction, $V_r$ [59],

$$V_H \approx (V_{be})_n + (V_{be})_p + V_r.$$

Experimentally, it can be observed that the S-type characteristic falls into a state close to $V_H = 1.2\,V$. Experimentally using external resistance elements, $V_H$ can be modulated to higher voltages. Figure 3.4 shows the holding voltage as a function of the bipolar product.

### 3.3.2 Holding Current

A key latchup design consideration is the holding current $I_H$. Given a constant $I_H$ value, for given $\beta_{pnp}$ and $\beta_{npn}$, the question is what are the conditions for the range of the well and substrate resistances for a constant $I_H$? The holding current can be represented by the following relationship [13]:

$$I_H = \frac{\beta_p(\beta_n + 1)I_w + (\beta_p + 1)\beta_n I_{sx}}{\beta_p\beta_n - 1},$$

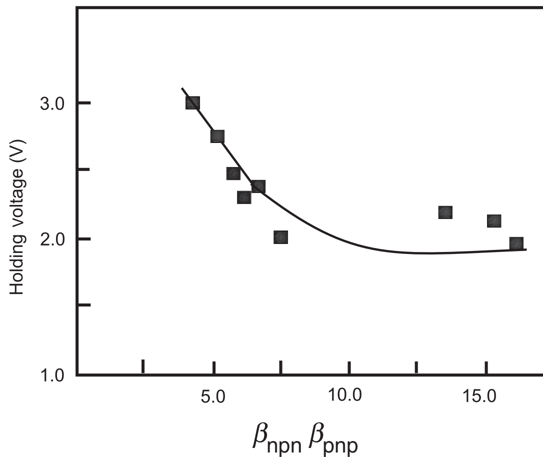



**Figure 3.4** CMOS latchup holding voltage versus bipolar product.

where the well and substrate currents can be solved as a function of the base–emitter voltage divided by the well and substrate resistances, respectively. In addition, for both formulations, the well and the substrate currents can be obtained as

$$I_{\mathrm{w}} = \frac{(V_{\mathrm{BE}})_{\mathrm{pnp}}}{R_{\mathrm{w}}} = \frac{V_0}{R_{\mathrm{w}}} \ln\left[\frac{I - I_{\mathrm{w}}}{(I_0)_{\mathrm{p}}}\right]$$

and

$$I_{sx} = \frac{(V_{\mathrm{BE}})_{\mathrm{npn}}}{R_{sx}} = \frac{V_0}{R_{sx}} \ln\left[\frac{I - I_{sx}}{(I_0)_{\mathrm{n}}}\right].$$

## 3.3.3 Holding Current Contours

The holding current condition can be structured as follows:

$$\beta_{\mathrm{p}}(\beta_{\mathrm{n}} + 1)I_{\mathrm{w}} + (\beta_{\mathrm{p}} + 1)\beta_{\mathrm{n}} I_{sx} = (\beta_{\mathrm{p}}\beta_{\mathrm{n}} - 1)I_{\mathrm{H}}.$$

Dividing by $(\beta_{\mathrm{p}} + 1)(\beta_{\mathrm{n}} + 1)$,

$$\frac{\beta_{\mathrm{p}}}{\beta_{\mathrm{p}} + 1} I_{\mathrm{w}} + \frac{\beta_{\mathrm{n}}}{\beta_{\mathrm{n}} + 1} I_{sx} = \frac{(\beta_{\mathrm{p}}\beta_{\mathrm{n}} - 1)}{(\beta_{\mathrm{p}} + 1)(\beta_{\mathrm{n}} + 1)} I_{\mathrm{H}}.$$

Substituting the resistance values,

$$\frac{\beta_{\mathrm{p}}}{\beta_{\mathrm{p}} + 1} \frac{(V_{\mathrm{be}})_{\mathrm{pnp}}}{R_{\mathrm{w}}} + \frac{\beta_{\mathrm{n}}}{\beta_{\mathrm{n}} + 1} \frac{(V_{\mathrm{be}})_{\mathrm{npn}}}{R_{sx}} = \frac{(\beta_{\mathrm{p}}\beta_{\mathrm{n}} - 1)}{(\beta_{\mathrm{p}} + 1)(\beta_{\mathrm{n}} + 1)} I_{\mathrm{H}}.$$

In this form, it can be observed that for constant npn and pnp bipolar current gains, and for a constant holding current, the equation RHS is a constant. As a result, the two terms on the equation LHS must balance each other as one increases or decreases. Hence, to satisfy the constant holding current, for a given structure size, as the shunt well resistance increases, the shunt substrate resistance must decrease, and vice versa. There is an interrelationship between the two shunt resistances in order to remain on a given $I_{\mathrm{H}}$ contour. For the case where the well resistance increases to infinity,

$$\frac{\beta_{\mathrm{n}}}{\beta_{\mathrm{n}} + 1} \frac{(V_{\mathrm{be}})_{\mathrm{npn}}}{R_{sx}} = \frac{(\beta_{\mathrm{p}}\beta_{\mathrm{n}} - 1)}{(\beta_{\mathrm{p}} + 1)(\beta_{\mathrm{n}} + 1)} I_{\mathrm{H}},$$

and solving for substrate resistance,

$$R_{sx} = \frac{(\beta_{\mathrm{p}} + 1)(\beta_{\mathrm{n}} + 1)}{(\beta_{\mathrm{p}}\beta_{\mathrm{n}} - 1)} \frac{\beta_{\mathrm{n}}}{\beta_{\mathrm{n}} + 1} \frac{(V_{\mathrm{be}})_{\mathrm{npn}}}{I_{\mathrm{H}}}.$$

For the opposite condition, where the substrate resistance is infinite, the condition for the well resistance for the same bipolar current gains and holding current is

$$R_w = \frac{(\beta_p + 1)(\beta_n + 1)}{(\beta_p\beta_n - 1)} \frac{\beta_p}{\beta_p + 1} \frac{(V_{be})_{pnp}}{I_H}.$$

Taking the natural logarithm of both sides,

$$\ln\left[\frac{\beta_p}{\beta_p + 1}\frac{(V_{be})_{pnp}}{R_w} + \frac{\beta_n}{\beta_n + 1}\frac{(V_{be})_{npn}}{R_{sx}}\right] = \ln\left[\frac{(\beta_p\beta_n - 1)}{(\beta_p + 1)(\beta_n + 1)} I_H\right].$$

From natural logarithm relationships, $\ln[ab] = \ln[a] + \ln[b]$,

$$\ln\left[\left(\frac{\beta_p}{\beta_p + 1}\right)\left(\frac{\beta_n}{\beta_n + 1}\right)\frac{(V_{be})_{pnp}(V_{be})_{npn}}{R_w R_{sx}}\right] = \ln\left[\frac{(\beta_p\beta_n - 1)}{(\beta_p + 1)(\beta_n + 1)} I_H\right].$$

Factoring out the well and substrate resistance terms,

$$\ln\left[\left(\frac{\beta_p}{\beta_p + 1}\right)\left(\frac{\beta_n}{\beta_n + 1}\right)(V_{be})_{npn}(V_{be})_{pnp}\right] - \ln[R_w R_{sx}] = \ln\left[\frac{(\beta_p\beta_n - 1)}{(\beta_p + 1)(\beta_n + 1)} I_H\right].$$

Then for a constant holding current, the well and substrate resistance conditions are related as the natural logarithm product of the well and shunt resistance terms,

$$\ln[R_w R_{sx}] = \ln\left[\left(\frac{\beta_p}{\beta_p + 1}\right)\left(\frac{\beta_n}{\beta_n + 1}\right)(V_{be})_{npn}(V_{be})_{pnp}\right] - \ln\left[\frac{(\beta_p\beta_n - 1)}{(\beta_p + 1)(\beta_n + 1)} I_H\right].$$

For constant holding current, bipolar current gains and forward active bias conditions, it is clear that this will provide a hyperbolic relationship on a logarithmic plot of the shunt well and substrate resistances.

Given a fixed semiconductor process and a fixed pnpn structure, given that different well and contact spacings are chosen, one can vary the well and substrate resistance conditions. Hence, $I_H$ and $V_H$ contours can be shown to be a function of $R_W$ and $R_{SX}$. This can be graphically observed by a representation of the locus of points in a log–log resistance space, where one axis is the logarithm of the well shunt resistance and the other axis is the logarithm of the shunt substrate resistance.

## 3.3.4 Sensitivity Factors – Holding Current

The holding current relationship is a function of semiconductor process and design parameters. A latchup design practice is the evaluation of the partial derivatives of the holding current relationship. The holding current sensitivity relationship can be defined as follows [13]:

$$S^{H}_{\beta_n} = \frac{\beta_n}{I_H}\frac{\partial I_H}{\partial \beta_n},$$

$$S_{\beta_p}^{H} = \frac{\beta_p}{I_H}\frac{\partial I_H}{\partial \beta_p},$$
$$S_{R_W}^{H} = \frac{R_W}{I_H}\frac{\partial I_H}{\partial R_W},$$
$$S_{R_S}^{H} = \frac{R_S}{I_H}\frac{\partial I_H}{\partial R_S}.$$

## 3.4 RESISTANCE SPACE

In latchup analysis, there are four fundamental variables that influence the latchup criterion. Given a fixed pnp base width transistor dimension (e.g. the p+ diffusion to n-well edge dimension) and a fixed npn base width transistor dimension (e.g. an n+ diffusion to n-well edge dimension), the p+/n+ space for the pnpn structure is fixed. In a given technology, these minimum dimensions are fixed by the technology generation and design 'ground rules'. Yet, there are the shunt resistances; there are two of the fundamental variables in the differential generalized tetrode relationship for latchup avoidance.

The latchup representation was shown as a function of the 'sum of the alphas' representation in 'alpha space' and 'generalized alpha space', which represented the regions of the latchup [59]. In this case, the region was triangular, and the alpha sum demarcation was represented as a line. In addition, the representation in 'beta space' is represented as the 'beta product', where graphically the transformation converted the contour into a hyperbola [59, 71, 72]. Another practical representation is the representation in 'shunt resistance space' or can also be referred to as 'log well–log substrate resistance space'. In this case, contours of constant $V_H$ or constant $I_H$ can be defined for a fixed structure size. This representation is very important in the latchup design discipline for the evaluation of the sensitivity of the pnpn structure as a function of the substrate and well contact spacing.

Figure 3.5 is an example of a design plot in resistance space, where a constant $V_H$ contour is shown [71,72]. The relationship shows that this follows a hyperbolic form in the resistance space. This contour is for a given holding voltage condition and is a function of the bipolar current gain of the pnp and npn transistors.

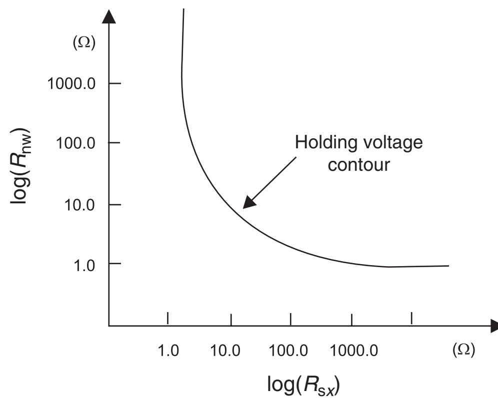



**Figure 3.5** Logarithm of shunt well resistance versus logarithm of shunt substrate resistance for a constant holding voltage contour.

A key latchup design practice is to choose a holding voltage $V_H$, that is above the application power supply voltage $V_{DD}$. In the latchup design practice, a trade-off exists between the substrate and the well resistance to remain on the $V_H$ contour. For example, Figure 3.5 can be used as a design curve for defining the semiconductor technology doping concentrations and physical dimensions.

### 3.4.1 Resistance Space – Constant Holding Condition Contours

A key latchup design consideration is the well and substrate conditions. Given we assume a constant value for the holding current, for given bipolar transistor current gains of the pnp and the npn, the question is what are the conditions for the range of the well and substrate resistances for a constant holding current? The holding current can be represented by the following relationship:

$$I_H = \frac{\beta_p(\beta_n + 1)I_w + (\beta_p + 1)\beta_n I_{sx}}{\beta_p\beta_n - 1},$$

where the well and substrate currents can be solved as a function of $V_{BE}$ divided by the well and substrate resistances, respectively:

$$I_w = \frac{(V_{BE})_{pnp}}{R_w} = \frac{V_0}{R_w}\ln\left[\frac{I - I_w}{(I_0)_p}\right]$$

and

$$I_{sx} = \frac{(V_{BE})_{npn}}{R_{sx}} = \frac{V_0}{R_{sx}}\ln\left[\frac{I - I_{sx}}{(I_0)_n}\right].$$

The $I_H$ condition can be structured as follows:

$$\beta_p(\beta_n + 1)I_w + (\beta_p + 1)\beta_n I_{sx} = (\beta_p\beta_n - 1)I_H.$$

Dividing by $(\beta_p + 1)(\beta_n + 1)$,

$$\frac{\beta_p}{\beta_p + 1}I_w + \frac{\beta_n}{\beta_n + 1}I_{sx} = \frac{(\beta_p\beta_n - 1)}{(\beta_p + 1)(\beta_n + 1)}I_H.$$

Substituting the resistance values,

$$\frac{\beta_p}{\beta_p + 1}\frac{(V_{be})_{pnp}}{R_w} + \frac{\beta_n}{\beta_n + 1}\frac{(V_{be})_{npn}}{R_{sx}} = \frac{(\beta_p\beta_n - 1)}{(\beta_p + 1)(\beta_n + 1)}I_H.$$

In this form, it can be observed that for constant $\beta_{npn}$ and $\beta_{pnp}$, and for a constant holding current, the equation RHS is a constant. As a result, the two terms on the equation LHS must balance each other as one increases or decreases. Hence, to satisfy the constant $I_H$, for a given structure size, as the shunt well

resistance increases, the shunt substrate resistance must decrease, and vice versa. Hence, there is an interrelationship between the two shunt resistances in order to remain on a given holding current contour. Taking the case where the well resistance increases to infinity,

$$\frac{\beta_n}{\beta_n + 1}\frac{(V_{be})_{npn}}{R_{sx}} = \frac{(\beta_p\beta_n - 1)}{(\beta_p + 1)(\beta_n + 1)} I_H$$

and

$$R_{sx} = \frac{(\beta_p + 1)(\beta_n + 1)}{(\beta_p\beta_n - 1)}\frac{\beta_n}{\beta_n + 1}\frac{(V_{be})_{npn}}{I_H}.$$

For the opposite condition, where the substrate resistance is infinite, the condition for the well resistance for the same bipolar current gains and holding current is

$$R_w = \frac{(\beta_p + 1)(\beta_n + 1)}{(\beta_p\beta_n - 1)}\frac{\beta_p}{\beta_p + 1}\frac{(V_{be})_{pnp}}{I_H}.$$

Taking the natural logarithm of both sides,

$$\ln\left[\frac{\beta_p}{\beta_p + 1}\frac{(V_{be})_{pnp}}{R_w} + \frac{\beta_n}{\beta_n + 1}\frac{(V_{be})_{npn}}{R_{sx}}\right] = \ln\left[\frac{(\beta_p\beta_n - 1)}{(\beta_p + 1)(\beta_n + 1)} I_H\right].$$

From natural logarithm relationships, $\ln[ab] = \ln[a] + \ln[b]$,

$$\ln\left[\left(\frac{\beta_p}{\beta_p + 1}\right)\left(\frac{\beta_n}{\beta_n + 1}\right)\frac{(V_{be})_{pnp}(V_{be})_{npn}}{R_w R_{sx}}\right] = \ln\left[\frac{(\beta_p\beta_n - 1)}{(\beta_p + 1)(\beta_n + 1)} I_H\right].$$

Factoring out the well and substrate resistance terms,

$$\ln\left[\left(\frac{\beta_p}{\beta_p + 1}\right)\left(\frac{\beta_n}{\beta_n + 1}\right)(V_{be})_{npn}(V_{be})_{pnp}\right] - \ln[R_w R_{sx}] = \ln\left[\frac{(\beta_p\beta_n - 1)}{(\beta_p + 1)(\beta_n + 1)} I_H\right].$$

Then for a constant $I_H$, the condition for the well and substrate resistances is related as the natural logarithm product of the well and shunt resistance terms,

$$\ln[R_w R_{sx}] = \ln\left[\left(\frac{\beta_p}{\beta_p + 1}\right)\left(\frac{\beta_n}{\beta_n + 1}\right)(V_{be})_{npn}(V_{be})_{pnp}\right] - \ln\left[\frac{(\beta_p\beta_n - 1)}{(\beta_p + 1)(\beta_n + 1)} I_H\right].$$

For constant holding current, bipolar current gains and forward active bias conditions, it is clear that this will provide a hyperbolic relationship on a logarithmic plot of the logarithm of the shunt well resistance and the logarithm of the substrate resistance.

Given a fixed semiconductor process and a fixed structure size of a pnpn structure, given that different well and contact spacings are chosen, one can vary the well and substrate resistance

conditions. Hence, it is anticipated for $I_H$ or $V_H$ conditions that constant contours can be shown as a function of well and substrate resistances. This can be graphically observed by a representation of the locus of points in log–log resistance space, where one axis is the logarithm of the well shunt resistance and the other axis is the logarithm of the shunt substrate resistance.

### 3.4.2 Resistance Space Plots

Each unique pnpn structure is a function of $\beta_{npn}$, $\beta_{pnp}$, $R_W$ and $R_{SX}$ [72]. For a fixed pnpn structure with a fixed p+ to n-well edge space and a fixed n-well to n+ space, there exists a unique state of the structure for a given substrate resistance and a given well resistance. The holding voltage $V_H$ is a function of all four variables. As the well and substrate resistances are varied, a constant $V_H$ contour can be mapped in a 'resistance space' of log ($R_{SX}$) versus log ($R_W$). Figure 3.6 shows a plot of a $V_H$ contour as a function of well and substrate resistances. For each unique p+/n+ space, there will be a different contour in the resistance space. The $V_H$ contour is a function of the p+/n+ spacing and temperature. Mapping of the $V_H$ contour in $\log(R_w)$ versus $\log(R_{sx})$ space provides all possible cases of the potential resistance values observed for a given p+/n+ space. Hence, these resistance plots can be utilized to quantify all possible design states observed in a semiconductor chip. Note that for each pnpn structure with a different bipolar current gain, the contours shift relative to the axis. A latchup semiconductor design methodology is as follows [72]:

- Choose a worst-case p+/n+ space for a given technology.
- Using external or internal resistances, map the holding voltage contour for a fixed holding voltage as a function of well and substrate resistances from zero to infinity (e.g. to high possible resistance values observed on chip).
- Repeat the process for all p+/n+ spaces in the technology and all temperatures of interest.
- Choose a $V_H$ to $V_{DD}$ margin design condition.
- Find the $V_H$ value for the worst-case p+/n+ space.

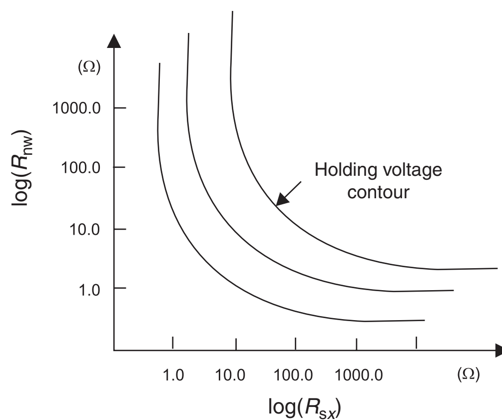



**Figure 3.6** Holding voltage contours in resistance space (well and substrate).

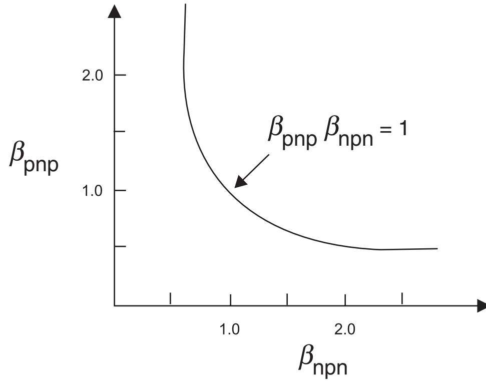



**Figure 3.7** Beta space contour of the latchup pnpn structure.

- Choose a worst-case well resistance (or substrate resistance) on the worst-case holding voltage contour.
- Choose a worst-case substrate resistance (or well resistance) on the worst-case $V_H$ contour for the given worst-case well resistance (or substrate resistance) value.

## 3.5 BETA SPACE

From the analysis, there exists a region in bipolar 'beta space' associated with the npn and the pnp bipolar current gain [13, 59]. Figure 3.7 shows a plot of the two domains of interest for the pnpn structure. From the bipolar criterion, it is known that the product of the npn and pnp bipolar current gains forms a hyperbola in 'beta space'. In the simplest case, when the beta product is less than unity, the pnpn circuit is in the blocking state and latchup does not occur. When the latchup criterion is satisfied, switching of the pnpn can occur. Hence, the latchup criterion forms a hyperbola in $\beta$-space, which separates the cases of latchup and no latchup. In the more general case, the hyperbola for switching shifts for each well and substrate resistance value. Note that for each pnpn structure with well and substrate resistances, the contours shift relative to the axis.

## 3.6 CMOS LATCHUP DIFFERENTIAL TETRODE CONDITION

In the latchup analysis, different formulations exist that address the dc conditions and the transient conditions. A more generalized form can be used for the evaluation of the stability criterion. In this form, latchup can be represented as a function of the effective large-scale injection factors and the effective transport factors [59].

Define the effective large-scale injection factors for the pnp and npn as the ratio of current flowing through the emitter compared to the total current flowing through the latchup structure, for the pnp and the npn, respectively,

$$\gamma_p^* = \frac{I_{ep}}{I}$$

and

$$\gamma_n^* = \frac{I_{en}}{I},$$

where $I_{ep}$ and $I_{en}$ are the emitter currents and $I$ is the total latchup current. The total latchup current $I$ is larger due to the current flowing through the bypass well and substrate resistors. The total current flowing through the latchup structure can be expressed as

$$I = I_n + I_p = \alpha_{fn} I_{en} + \alpha_{fp} I_{ep} + I_{nd} + I_{pd} + I_{sc},$$

where the first terms are the collector currents and the last terms are the total diffusion current at the well junction and the space charge generation current at the well junction. Expressing the current as a function of the effective injection factors,

$$I = I_n + I_p = \alpha_{fn}\gamma_n^* I + \alpha_{fp}\gamma_p^* I + (I_{nd} + I_{pd} + I_{sc}).$$

Solving for the total current, by factoring out the terms, the expression takes a familiar form in the alpha representation [59],

$$I = \frac{(I_{nd} + I_{pd} + I_{sc})}{1 - [\alpha_{fn}\gamma_n^* + \alpha_{fp}\gamma_p^*]}.$$

Defining an ‘effective bipolar transistor gain’ as

$$\alpha_{fn}^* = \alpha_{fn}\gamma_n^*,$$
$$\alpha_{fp}^* = \alpha_{fp}\gamma_p^*,$$

the latchup current relationship takes on the form of the earlier alpha relationship. Expressing the leakages as a sum of the total diffusion current at the well junction and the generation current within the well depletion region,

$$I_s = I_{nd} + I_{pd} + I_{sc},$$

the stability relationship takes on the familiar form as in the prior derivations, as the ratio of the total diffusion current and generation current within the well region, divided by the effective transistor gains:

$$I = \frac{I_s}{1 - [\alpha_{fn}^* + \alpha_{fp}^*]}$$

or

$$I = \frac{I_s}{1 - [\alpha_{fn}\gamma_n^* + \alpha_{fp}\gamma_p^*]}.$$

### 3.6.1 CMOS Latchup Differential Tetrode Condition – First Representation

Troutman [59] noted that this can be addressed looking at the differential response. From this relationship, it appears that the switching state will occur when the denominator is zero. Troutman pointed out that this cannot occur, that the usage of this expression for determining the switching state is inaccurate and that a generalized tetrode stability relationship must be represented as the derivative of this condition [59]. Rearranging the expression,

$$I\{1-[\alpha_{\mathrm{fn}}^{*}+\alpha_{\mathrm{fp}}^{*}]\}=I_{\mathrm{s}}.$$

Taking the derivative of the expression and factoring, as a function of the current $I_{\mathrm{s}}$,

$$\frac{\mathrm{d}}{\mathrm{d}I_{\mathrm{s}}}[I\{1-[\alpha_{\mathrm{fn}}^{*}+\alpha_{\mathrm{fp}}^{*}]\}]=\frac{\mathrm{d}}{\mathrm{d}I_{\mathrm{s}}}I_{\mathrm{s}},$$
$$\frac{\mathrm{d}}{\mathrm{d}I_{\mathrm{s}}}[\{I-I[\alpha_{\mathrm{fn}}^{*}+\alpha_{\mathrm{fp}}^{*}]\}]=\frac{\mathrm{d}}{\mathrm{d}I_{\mathrm{s}}}I_{\mathrm{s}},$$
$$\frac{\mathrm{d}I}{\mathrm{d}I_{\mathrm{s}}}-\frac{\mathrm{d}}{\mathrm{d}I_{\mathrm{s}}}\{I[\alpha_{\mathrm{fn}}^{*}+\alpha_{\mathrm{fp}}^{*}]\}=\frac{\mathrm{d}I_{\mathrm{s}}}{\mathrm{d}I_{\mathrm{s}}},$$

where finally this can be expressed as

$$\frac{\mathrm{d}I}{\mathrm{d}I_{\mathrm{s}}}-\frac{\mathrm{d}}{\mathrm{d}I_{\mathrm{s}}}\{I[\alpha_{\mathrm{fn}}^{*}+\alpha_{\mathrm{fp}}^{*}]\}=1.$$

From the second term in the above equation,

$$\frac{\mathrm{d}}{\mathrm{d}I_{\mathrm{s}}}\{I[\alpha_{\mathrm{fn}}^{*}+\alpha_{\mathrm{fp}}^{*}]\}=\frac{\mathrm{d}}{\mathrm{d}I}\{I[\alpha_{\mathrm{fn}}^{*}+\alpha_{\mathrm{fp}}^{*}]\}\frac{\mathrm{d}I}{\mathrm{d}I_{\mathrm{s}}}.$$

Substituting this term and factoring out the partial derivative,

$$\frac{\mathrm{d}I}{\mathrm{d}I_{\mathrm{s}}}\left\{1-\frac{\mathrm{d}}{\mathrm{d}I}\{I[\alpha_{\mathrm{fn}}^{*}+\alpha_{\mathrm{fp}}^{*}]\}\right\}=1.$$

In this form, the differential stability criterion can be established by taking the derivative of the expression with respect to the latchup current [59]:

$$\frac{\mathrm{d}I}{\mathrm{d}I_{\mathrm{s}}}=\frac{1}{1-\frac{\mathrm{d}}{\mathrm{d}I}[I(\alpha_{\mathrm{fn}}^{*}+\alpha_{\mathrm{fp}}^{*})]}.$$

Hence, the condition for instability for latchup occurs when the denominator is negative, or

$$\frac{\mathrm{d}}{\mathrm{d}I}[I(\alpha_{\mathrm{fn}}^{*}+\alpha_{\mathrm{fp}}^{*})]>1,$$

and stable when

$$\frac{d}{dI}[I(\alpha^*_{fn} + \alpha^*_{fp})] < 1.$$

A generalized differential latchup criterion can be expanded and expressed as [59]

$$(\alpha^*_{fn} + \alpha^*_{fp}) + I\frac{d}{dI}[(\alpha^*_{fn} + \alpha^*_{fp})] \geq 1.$$

From this form, it can be observed that if the last term is positive and greater than zero, the sum of the effective gains will always be less than unity.

### 3.6.2 CMOS Latchup Differential Tetrode Condition – Second Representation

Although accurate, the first generalized differential tetrode condition does not provide a convenient form with respect to the terminal conditions. The second representation lends itself to the terminal information. Troutman [59] noted that this can be expressed as

$$\frac{d}{dI}[I(\alpha^*_{fn} + \alpha^*_{fp})] > 1.$$

From the definition of the effective bipolar transistor gain,

$$\begin{aligned}
&\alpha^*_{fn} = \alpha_{fn}\gamma^*_n,\\
&\alpha^*_{fp} = \alpha_{fp}\gamma^*_p,\\
&\frac{d}{dI}[I(\alpha_{fn}\gamma^*_n + \alpha_{fp}\gamma^*_p)] \geq 1.
\end{aligned}$$

Substituting for the generalized $\gamma$ terms, from the terminal currents,

$$\frac{d}{dI}\left[I\left(\alpha_{fn}\left(\frac{I_{en}}{I}\right) + \alpha_{fp}\left(\frac{I_{ep}}{I}\right)\right)\right] \geq 1.$$

Simplifying,

$$\frac{d}{dI}(\alpha_{fn}I_{en} + \alpha_{fp}I_{ep}) \geq 1,$$

or in a two-term form, where the first term is associated with the npn transistor and the second term is associated with the pnp transistor in the latchup structure,

$$\frac{d}{dI}(\alpha_{fn}I_{en}) + \frac{d}{dI}(\alpha_{fp}I_{ep}) \geq 1.$$

For the npn transistor term,

$$\frac{d}{dI}(\alpha_{fn}I_{en}) = I_{en}\frac{d}{dI}(\alpha_{fn}) + \alpha_{fn}\frac{d}{dI}(I_{en}).$$

For the pnp transistor term,

$$\frac{d}{dI}(\alpha_{fp}I_{ep}) = I_{ep}\frac{d}{dI}(\alpha_{fp}) + \alpha_{fp}\frac{d}{dI}(I_{ep}).$$

Substituting into the general expression, the differential tetrode condition is expressed as four terms:

$$I_{en}\frac{d}{dI}(\alpha_{fn}) + \alpha_{fn}\frac{d}{dI}(I_{en}) + I_{ep}\frac{d}{dI}(\alpha_{fp}) + \alpha_{fp}\frac{d}{dI}(I_{ep}) \geq 1.$$

Define the collector-to-emitter current gain for the npn and pnp transistors, respectively,

$$\alpha_{fn} = \frac{I_{cn}}{I_{en}},$$
$$\alpha_{fp} = \frac{I_{cp}}{I_{ep}}.$$

From this, the small-signal forward collector-to-emitter transport factor can be defined as the derivative, for the npn and the pnp, respectively,

$$\alpha_{fns} = \frac{dI_{cn}}{dI_{en}},$$
$$\alpha_{fps} = \frac{dI_{cp}}{dI_{ep}}.$$

Then, let

$$I_{en}\frac{d}{dI_{en}}(\alpha_{fn}) = I_{en}\frac{d}{dI_{en}}\left(\frac{I_{cn}}{I_{en}}\right) = I_{en}\left(\frac{1}{I_{en}}\frac{dI_{cn}}{dI_{en}}\right) + I_{en}I_{cn}\left(-\frac{1}{I_{en}^2}\right),$$
$$I_{en}\frac{d}{dI_{en}}(\alpha_{fn}) = \left(\frac{dI_{cn}}{dI_{en}}\right) - \left(\frac{I_{cn}}{I_{en}}\right).$$

Substituting the definitions for the collector-to-emitter transport factors, for the npn transistor,

$$I_{en}\frac{d}{dI_{en}}(\alpha_{fn}) = \alpha_{fns} - \alpha_{fn}$$

or

$$-I_{en}\frac{d}{dI_{en}}(\alpha_{fn}) + \alpha_{fns} = \alpha_{fn}.$$

Hence, the same is true for the pnp transistors,

$$I_{ep}\frac{d}{dI_{ep}}(\alpha_{fp}) = \alpha_{fps} - \alpha_{fp},$$
$$-I_{ep}\frac{d}{dI_{ep}}(\alpha_{fp}) + \alpha_{fps} = \alpha_{fp}.$$

Then substituting these expressions into the differential tetrode condition,

$$I_{en}\frac{d}{dI}(\alpha_{fn}) + \left\{-I_{en}\frac{d}{dI_{en}}(\alpha_{fn}) + \alpha_{fns}\right\}\frac{d}{dI}(I_{en}) + I_{ep}\frac{d}{dI}(\alpha_{fp}) + \left\{-I_{ep}\frac{d}{dI_{ep}}(\alpha_{fp}) + \alpha_{fps}\right\}\frac{d}{dI}(I_{ep}) \geq 1.$$

Separating terms,

$$\left[I_{en}\frac{d}{dI}(\alpha_{fn}) + \left\{-I_{en}\frac{d}{dI_{en}}(\alpha_{fn})\right\}\frac{d}{dI}(I_{en})\right] + \left[I_{ep}\frac{d}{dI}(\alpha_{fp}) + \left\{-I_{ep}\frac{d}{dI_{ep}}(\alpha_{fp})\right\}\frac{d}{dI}(I_{ep})\right] + \{\alpha_{fns}\}\frac{d}{dI}(I_{en}) + \{\alpha_{fps}\}\frac{d}{dI}(I_{ep}) \geq 1.$$

Applying the chain rule and canceling common terms, the second representation of the differential tetrode relationship can be expressed as [59]

$$\{\alpha_{fns}\}\frac{d}{dI}(I_{en}) + \{\alpha_{fps}\}\frac{d}{dI}(I_{ep}) \geq 1.$$

In this form, the terms are a function of the emitter currents, the current through the structure and the small-signal collector-to-emitter transport factors for the npn and the pnp transistors.

### 3.6.3 CMOS Latchup Differential Tetrode Condition – Third Representation

The second representation of the differential tetrode relationship can be expressed as

$$\{\alpha_{fns}\}\frac{d}{dI}(I_{en}) + \{\alpha_{fps}\}\frac{d}{dI}(I_{ep}) \geq 1.$$

In this form, the terms are a function of the emitter currents, the current through the structure and the small-signal collector-to-emitter transport factors. In this form, the contributing term from each transistor is the product of the small-signal collector-to-emitter transport terms and the derivative of the current with respect to its emitter current. In this case, it is clear that the response of the circuit is a function of the transport properties as well as the percentage of the current through the emitter with

respect to changes in the current flowing through the complete circuit. The above equation can be written in the following form [59]:

$$\frac{\alpha_{\mathrm{fns}}}{\left\{\frac{1}{\mathrm{d}I_{\mathrm{en}}/\mathrm{d}I}\right\}}+\frac{\alpha_{\mathrm{fps}}}{\left\{\frac{1}{\mathrm{d}I_{\mathrm{ep}}/\mathrm{d}I}\right\}}\geq 1$$

or

$$\frac{\alpha_{\mathrm{fns}}}{\{\mathrm{d}I/\mathrm{d}I_{\mathrm{en}}\}}+\frac{\alpha_{\mathrm{fps}}}{\{\mathrm{d}I/\mathrm{d}I_{\mathrm{ep}}\}}\geq 1.$$

From our earlier development, it is understood that the injection factor is a function of the current flowing through the emitter divided by the total current flowing through the terminals. For the npn transistor, it is the ratio of the current that flows through its emitter and the current that flows through the shunt element (e.g. the substrate resistor),

$$\gamma_{\mathrm{n}}^{*}=\frac{I_{\mathrm{en}}}{I}=\frac{I_{\mathrm{en}}}{I_{\mathrm{en}}+I_{\mathrm{rs}}}.$$

For the pnp transistor, the injection factor is the ratio of the current flowing through the pnp transistor and the total current through the pnpn element. In this case, the current is associated with the current through the well resistor:

$$\gamma_{\mathrm{p}}^{*}=\frac{I_{\mathrm{ep}}}{I}=\frac{I_{\mathrm{ep}}}{I_{\mathrm{ep}}+I_{\mathrm{rw}}}.$$

Troutman defined a small-signal injection factor term for the npn and the pnp transistor, respectively,

$$\gamma_{\mathrm{ns}}^{*}=\frac{\mathrm{d}I_{\mathrm{en}}}{\mathrm{d}I},$$
$$\gamma_{\mathrm{ps}}^{*}=\frac{\mathrm{d}I_{\mathrm{ep}}}{\mathrm{d}I}.$$

To derive the differential injection factors,

$$\gamma_{\mathrm{ns}}^{*}=\frac{\mathrm{d}I_{\mathrm{en}}}{\mathrm{d}I}.$$

Let

$$I=I_{\mathrm{en}}+I_{\mathrm{rs}}.$$

Then

$$\frac{\mathrm{d}}{\mathrm{d}I}I=\frac{\mathrm{d}}{\mathrm{d}I}I_{\mathrm{en}}+\frac{\mathrm{d}}{\mathrm{d}I}I_{\mathrm{rs}}.$$

From the chain rule, this can be expanded as a function

$$\frac{\mathrm{d}}{\mathrm{d}I}I=\frac{\mathrm{d}}{\mathrm{d}I}I_{\mathrm{en}}+\frac{\mathrm{d}}{\mathrm{d}I_{\mathrm{en}}}I_{\mathrm{rs}}\frac{\mathrm{d}}{\mathrm{d}I}I_{\mathrm{en}}.$$

Factoring the terms, the differential injection factor for the npn transistor is

$$\gamma_{ns}^{*} = \frac{\mathrm{d}\,I_{en}}{\mathrm{d}\,I} = \frac{1}{1 + \dfrac{\mathrm{d}\,I_{rs}}{\mathrm{d}\,I_{en}}}.$$

Due to symmetry, the differential injection factor for the pnp transistor is

$$\gamma_{ps}^{*} = \frac{\mathrm{d}\,I_{ep}}{\mathrm{d}\,I} = \frac{1}{1 + \dfrac{\mathrm{d}\,I_{rw}}{\mathrm{d}\,I_{ep}}}.$$

From this form, the third differential tetrode relationship can be expressed as follows [59]:

$$\frac{\alpha_{fns}}{1 + \frac{\mathrm{d}\,I_{rs}}{\mathrm{d}\,I_{en}}} + \frac{\alpha_{fps}}{1 + \frac{\mathrm{d}\,I_{rw}}{\mathrm{d}\,I_{ep}}} \geq 1.$$

Hence, it is clear that we can derive a generalized forward collector-to-emitter differential transport factor for the npn transistor as [59]

$$\alpha_{fns}^{*} = \frac{\alpha_{fns}}{1 + \frac{\mathrm{d}\,I_{rs}}{\mathrm{d}\,I_{en}}}$$

and a generalized forward collector-to-emitter differential transport factor for the pnp transistor as

$$\alpha_{fps}^{*} = \frac{\alpha_{fps}}{1 + \frac{\mathrm{d}\,I_{rw}}{\mathrm{d}\,I_{ep}}},$$

with the generalized differential tetrode condition for latchup as follows:

$$\alpha_{fns}^{*} + \alpha_{fps}^{*} \geq 1.$$

For application of the generalized differential tetrode equation, the derivatives in the denominator must be evaluated. From the chain rule,

$$\frac{\mathrm{d}\,I_{en}}{\mathrm{d}\,I_{rs}} = \frac{\mathrm{d}\,I_{en}}{\mathrm{d}V_{ben}} \frac{\mathrm{d}V_{ben}}{\mathrm{d}\,I_{rs}},$$

and

$$\frac{\mathrm{d}\,I_{ep}}{\mathrm{d}\,I_{rw}} = \frac{\mathrm{d}\,I_{ep}}{\mathrm{d}V_{bep}} \frac{\mathrm{d}V_{bep}}{\mathrm{d}\,I_{rw}}.$$

Define a differential small-signal emitter resistance, for the npn and the pnp transistor, respectively,

$$r_{en} = \frac{\mathrm{d}V_{ben}}{\mathrm{d}\,I_{ben}}$$

and

$$r_{\mathrm{ep}} = \frac{\mathrm{d}V_{\mathrm{bep}}}{\mathrm{d}I_{\mathrm{bep}}}.$$

The current derivatives can then be expressed as a function of the small-signal emitter resistances,

$$\frac{\mathrm{d}I_{\mathrm{en}}}{\mathrm{d}I_{\mathrm{rs}}} = \frac{1}{r_{\mathrm{en}}}\frac{\mathrm{d}V_{\mathrm{ben}}}{\mathrm{d}I_{\mathrm{rs}}}$$

and

$$\frac{\mathrm{d}I_{\mathrm{ep}}}{\mathrm{d}I_{\mathrm{rw}}} = \frac{1}{r_{\mathrm{ep}}}\frac{\mathrm{d}V_{\mathrm{bep}}}{\mathrm{d}I_{\mathrm{rw}}}.$$

To evaluate the derivative, the differential voltage across the E–B junction of the npn BJT can be derived as a function of the emitter resistance, the substrate resistance and the base resistance,

$$\frac{\mathrm{d}V_{\mathrm{ben}}}{\mathrm{d}I_{\mathrm{rs}}} = R_{\mathrm{s}} - [R_{\mathrm{n}} + R_{\mathrm{bn}}(1-\alpha_{\mathrm{fns}})]\frac{\mathrm{d}I_{\mathrm{en}}}{\mathrm{d}I_{\mathrm{rs}}},$$

and for the pnp transistor,

$$\frac{\mathrm{d}V_{\mathrm{bep}}}{\mathrm{d}I_{\mathrm{rw}}} = R_{\mathrm{w}} - [R_{\mathrm{p}} + R_{\mathrm{bp}}(1-\alpha_{\mathrm{fps}})]\frac{\mathrm{d}I_{\mathrm{ep}}}{\mathrm{d}I_{\mathrm{rw}}}.$$

Troutman [59] showed that this can be substituted into the alpha relationship, where

$$\alpha^*_{\mathrm{fns}} = \frac{R_{\mathrm{s}}\alpha_{\mathrm{fns}}}{R_{\mathrm{s}} + r_{\mathrm{en}} + \mathrm{R}_{\mathrm{en}} + [R_{\mathrm{bn}}(1-\alpha_{\mathrm{fns}})]}$$

and equivalently, by analogy,

$$\alpha^*_{\mathrm{fps}} = \frac{R_{\mathrm{w}}\alpha_{\mathrm{fps}}}{R_{\mathrm{w}} + r_{\mathrm{ep}} + \mathrm{R}_{\mathrm{ep}} + [R_{\mathrm{bp}}(1-\alpha_{\mathrm{fps}})]}.$$

From this form, the differential tetrode relationship can be expressed as

$$\frac{R_{\mathrm{s}}\alpha_{\mathrm{fns}}}{R_{\mathrm{s}} + r_{\mathrm{en}} + \mathrm{R}_{\mathrm{en}} + [R_{\mathrm{bn}}(1-\alpha_{\mathrm{fns}})]} + \frac{R_{\mathrm{w}}\alpha_{\mathrm{fps}}}{R_{\mathrm{w}} + r_{\mathrm{ep}} + \mathrm{R}_{\mathrm{ep}} + [R_{\mathrm{bp}}(1-\alpha_{\mathrm{fps}})]} \geq 1,$$

where normalizing the terms to the respective resistances,

$$\frac{\alpha_{\mathrm{fns}}}{1 + \dfrac{1}{R_{\mathrm{s}}}\{r_{\mathrm{en}} + \mathrm{R}_{\mathrm{en}} + [R_{\mathrm{bn}}(1-\alpha_{\mathrm{fns}})]\}} + \frac{\alpha_{\mathrm{fps}}}{1 + \dfrac{1}{R_{\mathrm{w}}}\{r_{\mathrm{ep}} + \mathrm{R}_{\mathrm{ep}} + [R_{\mathrm{bp}}(1-\alpha_{\mathrm{fps}})]\}} \geq 1.$$

In this development, when the differential resistance is much greater than the emitter resistance and the base resistances, the expression simplifies to the form [59]

$$\frac{\alpha_{\mathrm{fns}}}{1 + \dfrac{r_{\mathrm{en}}}{R_{\mathrm{s}}}} + \frac{\alpha_{\mathrm{fps}}}{1 + \dfrac{r_{\mathrm{ep}}}{R_{\mathrm{w}}}} \geq 1.$$

In this form, it is clear that the collector-to-emitter transport factors, the differential emitter resistances, the substrate resistance and the well resistance all play a role in the stability of the circuit and participate in the latchup general criterion. The second key point is that the influence of the well and substrate resistances on the latchup stability is a function of its magnitude relative to the differential emitter resistance of the respective transistor.

## 3.7 CMOS LATCHUP DIFFERENTIAL HOLDING CURRENT RELATIONSHIP

Another form of the holding current relationship, which is consistent with the differential analysis developed by Troutman [59], is

$$I_{\mathrm{h}} = I_{\mathrm{to}} + \{I_{\mathrm{to}}^2 - I_{\mathrm{m}}^2\}^{1/2},$$

where the turn-off current has the familiar form and the expression is a function of the reverse substrate and well currents, $I_{\mathrm{rs}}$ and $I_{\mathrm{rw}}$, respectively,

$$I_{\mathrm{to}} = \frac{\alpha_{\mathrm{fn}} I_{\mathrm{rs,to}} + \alpha_{\mathrm{fp}} I_{\mathrm{rw,to}}}{\{\alpha_{\mathrm{fn}} + \alpha_{\mathrm{fp}} - 1\}}$$

and the second term

$$I_{\mathrm{m}}^2 = \frac{M_1 I_{\mathrm{rs,to}}^2 + M_2 I_{\mathrm{rw,to}}^2 + M_3 I_{\mathrm{rs,to}} I_{\mathrm{rw,to}}}{\Delta\{\alpha_{\mathrm{fp}} + \alpha_{\mathrm{fn}} - 1\}},$$

where within the expression the variables are factors associated with the forward and reverse characteristics of the pnpn structure,

$$M_1 = \alpha_{\mathrm{fn}}\{(1 - \alpha_{\mathrm{fp}}\alpha_{rp})[(1 - \alpha_{\mathrm{fn}}\alpha_{\mathrm{rn}}) + \alpha_{\mathrm{rp}}(\alpha_{\mathrm{fn}} - \alpha_{\mathrm{fp}}) - \alpha_{\mathrm{fn}}\alpha_{\mathrm{rn}}\alpha_{\mathrm{rp}}(1 - \alpha_{\mathrm{fp}})]\},$$
$$M_2 = \alpha_{\mathrm{fp}}\{(1 - \alpha_{\mathrm{fn}}\alpha_{\mathrm{rn}})[(1 - \alpha_{\mathrm{fp}}\alpha_{\mathrm{rp}}) + \alpha_{\mathrm{rn}}(\alpha_{\mathrm{fp}} - \alpha_{\mathrm{fn}}) - \alpha_{\mathrm{fp}}\alpha_{\mathrm{rp}}\alpha_{\mathrm{rn}}(1 - \alpha_{\mathrm{fn}})]\}$$

and

$$M_3 = 1 - (\alpha_{\mathrm{fn}}\alpha_{\mathrm{rn}} + \alpha_{\mathrm{fp}}\alpha_{\mathrm{rp}}) + (1 - \alpha_{\mathrm{fp}}\alpha_{\mathrm{rp}})(2\alpha_{\mathrm{fn}} - 1)\alpha_{\mathrm{fp}}\alpha_{\mathrm{rp}} + (1 - \alpha_{\mathrm{fn}}\alpha_{\mathrm{rn}})(2\alpha_{\mathrm{fn}} - 1)\alpha_{\mathrm{fn}}\alpha_{\mathrm{rn}},$$

as well as

$$\Delta = [(1 - \alpha_{\mathrm{fn}}\alpha_{\mathrm{rn}}) - \alpha_{\mathrm{rp}}(1 - \alpha_{\mathrm{fn}})][(1 - \alpha_{\mathrm{fp}}\alpha_{\mathrm{rp}}) - \alpha_{\mathrm{rn}}(1 - \alpha_{\mathrm{fp}})].$$

In this form, it can be observed that the turn-off holding current term is a function of the difference between the 'alpha sum' and unity. As the sum of the alphas approaches unity, the holding current increases significantly. Note that the turn-off holding current term $I_{\mathrm{to}}$ is only a function of the forward collector-to-emitter transport factors. Hence, the turn-off current is only related to the forward device characteristics of the pnp and the npn.

The second term $I_m$ in its denominator is also a function of the difference between the forward 'alpha sum' and unity, as well as the second term that contains product terms of the reverse and forward characteristics. In the numerator, all the terms contain product terms of the forward and the reverse characteristics of the device. The complexity of this relationship is related to the interaction of the forward and reverse characteristics.

Given the reverse characteristic transport factors approach zero, the terms reduce to

$$\begin{aligned} M_1 &= \alpha_{\mathrm{fn}}, \\ M_2 &= \alpha_{\mathrm{fp}}, \\ M_3 &= 1, \\ \Delta &= 1, \end{aligned}$$

with a holding current relationship [59]

$$\begin{aligned} I_{\mathrm{h}} &= I_{\mathrm{to}} + \{I_{\mathrm{to}}^2 - \mathrm{I}_{\mathrm{m}}^2\}^{1/2}, \\ I_{\mathrm{to}} &= \frac{\alpha_{\mathrm{fn}} I_{\mathrm{rs,to}} + \alpha_{\mathrm{fp}} I_{\mathrm{rw,to}}}{\{\alpha_{\mathrm{fn}} + \alpha_{\mathrm{fp}} - 1\}}, \\ I_{\mathrm{m}}^2 &= \frac{\alpha_{\mathrm{fn}} I_{\mathrm{rs,to}}^2 + \alpha_{\mathrm{fp}} I_{\mathrm{rw,to}}^2 + I_{\mathrm{rs,to}} I_{\mathrm{rw,to}}}{\{\alpha_{\mathrm{fp}} + \alpha_{\mathrm{fn}} - 1\}}. \end{aligned}$$

From this form, it can be observed that the holding current is a function of the forward transport factors and the well and substrate currents in the turn-off state.

## 3.8 CMOS LATCHUP DIFFERENTIAL HOLDING VOLTAGE RELATIONSHIP

In this section, the differential holding voltage $V_h$ addresses the dynamic holding voltage. When a structure undergoes latchup, the holding voltage is the low-voltage state in the S-type characteristic. In an S-type $I$–$V$ characteristics, there are at least two stable states for a given voltage state. At the holding point, the structure can undergo a high-current state (e.g. the holding current). The holding voltage $V_H$ is one of the stable state conditions. In latchup, the voltage on the pnpn must increase to the knee voltage (e.g. avalanche voltage condition). In this state, it is at low current. After the knee condition, the latchup structure passes the negative resistance state and falls toward the holding point. Given that the power supply voltage is above the holding state, a low-voltage condition at the holding can occur. But, if the holding voltage exceeds the power supply voltage, the circuit will fall back to the low-current stable voltage state below the knee voltage. The differential holding voltage $V_h$ can be expressed as [59]

$$V_{\mathrm{h}} = (V_{\mathrm{be}})_{\mathrm{n}} + (V_{\mathrm{be}})_{\mathrm{p}} - V_{\mathrm{r}} + \left\{(1 - \alpha_{\mathrm{fn}}) I_{\mathrm{h}} + \alpha_{\mathrm{fn}} \frac{(V_{\mathrm{be}})_{\mathrm{n}}}{R_{\mathrm{s}x}}\right\} R_{\mathrm{s}2},$$

where the voltages are associated with the forward bias voltages across the two transistors, $I_h$ is the holding current, $R_{sx}$ is the substrate shunt resistance, and $R_{s2}$ is the resistance of the substrate resistance between the pnp collector and the npn base region within the regenerative feedback loop. It

is assumed that the second resistance $R_{w2}$, the well resistance within the regenerative feedback loop, is negligible:

$$V_h = (V_{be})_n + (V_{be})_p - V_r + \left\{ (1 - \alpha_{fn}) I_h + \alpha_{fn} \frac{(V_{be})_n}{R_{sx}} \right\} R_{s2}.$$

In the case where both $R_{w2}$ and $R_{s2}$ are small, the equation reduces to

$$V_h = (V_{be})_n + (V_{be})_p - V_r,$$

where $V_r$ is the forward bias at the pnpn center junction. In the case where the forward bias on the pnpn center junction is negligible,

$$V_h = (V_{be})_n + (V_{be})_p.$$

Hence, the latchup design practice is as follows:

- The holding voltage magnitude is set such as to maximize the holding voltage.
- The holding voltage is chosen as a semiconductor design point and physical dimensions whose worst-case pnpn network has a holding voltage state that exceeds the power supply voltage.

## 3.9 CMOS LATCHUP DIFFERENTIAL RESISTANCE RELATIONSHIP

The differential resistance can be derived from the derivative of the current–voltage relationship. Using the derivative of the *I–V* characteristic, key metric conditions can be solved at the knee and the holding points. Under these conditions, the derivative of voltage with respect to current is $\mathrm{d}V_r/\mathrm{d}I = 0$. Troutman [59] noted that in the blocking state, the differential resistance between cathode and anode is expressed as

$$\frac{\mathrm{d}V}{\mathrm{d}I} \approx \frac{r_r[1 - (\alpha^*_{fns} + \alpha^*_{fps})]}{1 - \alpha_{rns}\alpha^*_{fns} - \alpha_{rps}\alpha^*_{fps}}.$$

It was also noted that at the switching point, since the voltage is equal to the sum of the forward biases $V_{be}$ and $V_r$, the differential resistance is related to the derivative of these terms. By definition, it was stated that $\mathrm{d}V_r/\mathrm{d}I = 0$; hence, the differential resistance at the switching condition is only a function of the differential voltage of the emitter base as a function of current. In this condition, note that the differential resistance is related to the parallel configuration of the emitter resistance and the shunt resistance. Hence, it can be shown that

$$\frac{\mathrm{d}V}{\mathrm{d}I} = \frac{\mathrm{d}V_{ben}}{\mathrm{d}I} + \frac{\mathrm{d}V_{bep}}{\mathrm{d}I} = R_{nq} + R_{pq},$$

where the parallel resistance term of the npn is the parallel resistance of the substrate shunt and npn emitter resistances,

$$R_{\mathrm{nq}} = \frac{r_{\mathrm{en}} R_{\mathrm{s}}}{r_{\mathrm{en}} + R_{\mathrm{s}}},$$

and by symmetry, the parallel resistance term of the pnp is the parallel resistance of the well shunt and pnp emitter resistances,

$$R_{\mathrm{pq}} = \frac{r_{\mathrm{ep}} R_{\mathrm{w}}}{r_{\mathrm{ep}} + R_{\mathrm{w}}}.$$

For the holding voltage condition, $\mathrm{d}V/\mathrm{d}I = 0$, and the differential resistance where the transistors are in a saturation mode can be expressed as

$$r_{\mathrm{r,h}} = C_{\mathrm{n}} R_{\mathrm{nq,h}} + C_{\mathrm{p}} R_{\mathrm{pq,h}},$$

and normalized amplification factor for the npn transistor is

$$C_{\mathrm{n}} = \frac{(1 - \alpha^{*}_{\mathrm{fps}} \alpha_{\mathrm{rps}}) - \alpha_{\mathrm{rns}} (1 - \alpha^{*}_{\mathrm{fps}})}{(\alpha^{*}_{\mathrm{fns}} + \alpha^{*}_{\mathrm{fps}} - 1)}$$

and normalized amplification factor for the pnp transistor is

$$C_{\mathrm{p}} = \frac{(1 - \alpha^{*}_{\mathrm{fns}} \alpha_{\mathrm{rns}}) - \alpha_{\mathrm{rps}} (1 - \alpha^{*}_{\mathrm{fns}})}{(\alpha^{*}_{\mathrm{fns}} + \alpha^{*}_{\mathrm{fps}} - 1)}.$$

In the holding voltage condition, the response is a function of the forward amplification terms and the parallel resistances.

## 3.10 DIFFERENTIAL GENERALIZED ALPHA SPACE RELATIONSHIP

From the third form of the generalized differential tetrode relationship, the following inequality exists [59,72]:

$$\frac{\alpha_{\mathrm{fns}}}{1 + \dfrac{r_{\mathrm{en}}}{R_{\mathrm{s}}}} + \frac{\alpha_{\mathrm{fps}}}{1 + \dfrac{r_{\mathrm{ep}}}{R_{\mathrm{w}}}} \geq 1.$$

This is true under the assumption that the emitter and base resistances are small compared to the emitter dynamic resistance (for each respective transistor). In this form, it is clear that the collector-to-emitter transport factors, the differential emitter resistances, the substrate resistance and the well resistance all play a role in the circuit stability. Define the generalized collector-to-emitter transport factors for the npn transistor as follows:

$$\alpha^{*}_{\mathrm{fns}} = \frac{\alpha_{\mathrm{fns}}}{1 + \dfrac{r_{\mathrm{en}}}{R_{\mathrm{s}}}},$$

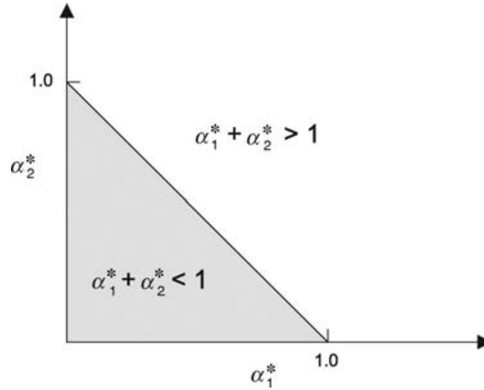



**Figure 3.8** Generalized alpha space.

and for the pnp transistor,

$$\alpha_{\text{fps}}^* = \frac{\alpha_{\text{fps}}}{1 + \dfrac{r_{\text{ep}}}{R_{\text{w}}}}.$$

In this form, it is clear that the generalized forward collector-to-emitter transport factors are a function of the forward collector-to-emitter transport factor, the differential emitter resistances, and the respective shunt resistance. Replacing these expressions, the generalized differential tetrode condition for CMOS latchup is as follows:

$$\alpha_{\text{fns}}^* + \alpha_{\text{fps}}^* \geq 1.$$

This is best observed as a graphical representation (Figure 3.8). A graphical representation is made, where the $x$ and $y$ axes are the two generalized differential forward collector-to-emitter transport factors. From the equality, the locus of points that satisfy

$$\alpha_{\text{fns}}^* + \alpha_{\text{fps}}^* = 1$$

is a line in the generalized $\alpha$-space. The locus of points that satisfy

$$\alpha_{\text{fns}}^* = 1 \quad \text{when} \quad \alpha_{\text{fps}}^* = 0$$

and

$$\alpha_{\text{fps}}^* = 1 \quad \text{when} \quad \alpha_{\text{fns}}^* = 0$$

hits the axis.

The locus of points that satisfy this is the demarcation state between the stable and unstable states of the system. For the cases where the inequality is not true,

$$\alpha_{\text{fns}}^* + \alpha_{\text{fps}}^* < 1,$$

this region will be referred to as the SAFE region [59,72]. The states contained within this region form a triangular space between the axis and the demarcation line. For the cases where

$$\alpha^*_{\mathrm{fns}} + \alpha^*_{\mathrm{fps}} \geq 1,$$

the space is in the plane outside of the triangular area; these cases are where CMOS latchup can occur.

From this equation, it is possible to pictorially observe the motion of the state of the circuit within the generalized alpha space. Note that there are at least six variables that influence the state of the circuit (e.g. ignoring the dc emitter resistance and the base resistance).

## 3.11 HIGH-LEVEL INJECTION

The high-level injection phenomenon influences the CMOS stability condition due to modulation of the metallurgical junctions and differential resistance in the pnpn network.

### 3.11.1 High-Level Injection – Base Width Modulation

As the voltage across the parasitic transistor increases, the base–collector (B–C) metallurgical junction depletion width increases. With the widening of the depletion width, the depletion region extends into the base region. As the width of the base decreases, the amount of stored charge in the base region decreases. In addition, the emitter current increases since the slope of the minority population decreases. In low-level current injection, the emitter–base current in a transistor is equal to

$$I_{\mathrm{e}} = I_{\mathrm{s}}\left\{\exp\left(\frac{qV_{\mathrm{be}}}{kT}\right) - 1\right\}.$$

At high-level injection of current, this can be expressed as [59]

$$I_{\mathrm{e}} = \frac{I_{\mathrm{s}}}{\chi}\left\{\exp\left(\frac{qV_{\mathrm{be}}}{kT}\right) - 1\right\},$$

where $\chi$ is a correction factor for the high-level injection correction to the physical model. Note that in a transistor structure there are both forward and reverse characteristics. As a result, there is a correction for both the forward and reverse characteristics, associated with an Early effect for the B–C junction, as well as the E–B junction. For high-level injection correction factor for the Early effect (base width modulation), where $V_{\mathrm{A}}$ is the forward Early voltage and $V_{\mathrm{B}}$ is the reverse Early voltage,

$$\chi_{\varepsilon} = 1 + \frac{V_{\mathrm{be}}}{|V_{\mathrm{A}}|} + \frac{V_{\mathrm{bc}}}{|V_{\mathrm{B}}|},$$

where the Early effect is dominant, the emitter current can be expressed as

$$I_{\mathrm{e}} = \frac{I_{\mathrm{s}}}{\chi_{\varepsilon}}\left\{\exp\left(\frac{qV_{\mathrm{be}}}{kT}\right) - 1\right\}.$$

## 3.11.2 High-Level Injection – Knee and Base Width Modulation Correction

In the case of high-level injection, the first effect is associated with base width narrowing, known as the Early effect. The second effect is also evident associated with the collector current. Defining the term $\chi_\kappa$ as the ratio of the forward active low-level current and the knee current [59]

$$\chi_\kappa = \frac{I_{eo}}{I_{kf}},$$

and with the Early effect correction,

$$\chi_\varepsilon = 1 + \frac{V_{be}}{|V_A|} + \frac{V_{bc}}{|V_B|},$$

this can be expressed as

$$\chi = \frac{1}{2}\{\chi_\varepsilon + (\chi_\varepsilon^2 + 4\chi_\kappa)^{1/2}\}$$

and

$$I_e = \frac{I_s}{\chi}\left\{\exp\left(\frac{qV_{be}}{kT}\right) - 1\right\}.$$

For the case where $\chi_\varepsilon \gg \chi_\kappa$, then $\chi = \chi_\varepsilon$:

$$I_e = \frac{I_s}{\chi_\varepsilon}\left\{\exp\left(\frac{qV_{be}}{kT}\right) - 1\right\}.$$

## 3.11.3 High-Level Injection – CMOS Latchup Differential Tetrode Condition

High-level injection influences the differential latchup criterion through the influence on the small-signal emitter resistance. High-level injection leads to an increase in the small-signal emitter resistance terms. The third differential tetrode relationship can be expressed as follows:

$$\frac{\alpha_{fns}}{1 + \dfrac{d\,I_{rs}}{d\,I_{en}}} + \frac{\alpha_{fps}}{1 + \dfrac{d\,I_{rw}}{d\,I_{ep}}} \geq 1.$$

Hence, it is clear that we can derive a generalized forward collector-to-emitter differential transport factor for the npn transistor as

$$\alpha_{fns}^* = \frac{\alpha_{fns}}{1 + \dfrac{d\,I_{rs}}{d\,I_{en}}}$$

and a generalized forward collector-to-emitter differential transport factor for the pnp transistor as

$$\alpha^*_{\mathrm{fps}} = \frac{\alpha_{\mathrm{fps}}}{1 + \dfrac{\mathrm{d}\,I_{\mathrm{rw}}}{\mathrm{d}\,I_{\mathrm{ep}}}},$$

with the generalized differential tetrode conditions for latchup as follows:

$$\frac{\mathrm{d}\,I_{\mathrm{en}}}{\mathrm{d}\,I_{\mathrm{rs}}} = \frac{\mathrm{d}\,I_{\mathrm{en}}}{\mathrm{d}V_{\mathrm{ben}}} \frac{\mathrm{d}V_{\mathrm{ben}}}{\mathrm{d}\,I_{\mathrm{rs}}}$$

and

$$\frac{\mathrm{d}\,I_{\mathrm{ep}}}{\mathrm{d}\,I_{\mathrm{rw}}} = \frac{\mathrm{d}\,I_{\mathrm{ep}}}{\mathrm{d}V_{\mathrm{bep}}} \frac{\mathrm{d}V_{\mathrm{bep}}}{\mathrm{d}\,I_{\mathrm{rw}}},$$

or the current derivatives expressed as [59]

$$\frac{\mathrm{d}\,I_{\mathrm{en}}}{\mathrm{d}\,I_{\mathrm{rs}}} = \frac{1}{r_{\mathrm{en}}} \frac{\mathrm{d}V_{\mathrm{ben}}}{\mathrm{d}\,I_{\mathrm{rs}}},$$
$$\frac{\mathrm{d}\,I_{\mathrm{ep}}}{\mathrm{d}\,I_{\mathrm{rw}}} = \frac{1}{r_{\mathrm{ep}}} \frac{\mathrm{d}V_{\mathrm{bep}}}{\mathrm{d}\,I_{\mathrm{rw}}},$$

with differential small-signal emitter resistances, for the npn and the pnp transistor, respectively, defined as follows:

$$r_{\mathrm{en}} = \frac{\mathrm{d}V_{\mathrm{ben}}}{\mathrm{d}\,I_{\mathrm{en}}},$$
$$r_{\mathrm{ep}} = \frac{\mathrm{d}V_{\mathrm{bep}}}{\mathrm{d}\,I_{\mathrm{ep}}}.$$

With high-level injection effects, the base width and knee current modulation effects influence the interrelationship of the E–B voltage and the emitter current.

From the relationship

$$I_{\mathrm{e}} = \frac{I_{\mathrm{s}}}{\chi}\left\{\exp\left(\frac{qV_{\mathrm{be}}}{kT}\right) - 1\right\},$$
$$\chi = \frac{1}{2}\{\chi_\varepsilon + (\chi_\varepsilon^2 + 4\chi_\kappa)^{1/2}\},$$
$$\chi_\varepsilon = 1 + \frac{V_{\mathrm{be}}}{|V_{\mathrm{A}}|} + \frac{V_{\mathrm{bc}}}{|V_{\mathrm{B}}|},$$
$$\chi_\kappa = \frac{I_{\mathrm{eo}}}{I_{\mathrm{kf}}},$$

it can be observed that terms in the high-level injection parameter are a function of the E–B voltage. Taking the derivative with respect to the emitter–base voltage,

$$\frac{\mathrm{d}}{\mathrm{d}V_{\mathrm{be}}} I_{\mathrm{e}} = \frac{I_{\mathrm{s}}}{\chi}\frac{\mathrm{d}}{\mathrm{d}V_{\mathrm{be}}}\left\{\exp\left(\frac{qV_{\mathrm{be}}}{kT}\right) - 1\right\} + I_{\mathrm{s}}\left\{\exp\left(\frac{qV_{\mathrm{be}}}{kT}\right) - 1\right\}\frac{\mathrm{d}}{\mathrm{d}V_{\mathrm{be}}}\left(\frac{1}{\chi}\right),$$

then

$$\frac{\mathrm{d}}{\mathrm{d}V_{\mathrm{be}}} I_{\mathrm{e}} = \frac{q}{kT}\frac{I_{\mathrm{s}}}{\chi}\left\{\exp\left(\frac{qV_{\mathrm{be}}}{kT}\right)\right\} - I_{\mathrm{s}}\left\{\exp\left(\frac{qV_{\mathrm{be}}}{kT}\right) - 1\right\}\frac{1}{\chi 2}\frac{\mathrm{d}}{\mathrm{d}V_{\mathrm{be}}}(\chi).$$

Approximating the expression, by factoring out the diode relationship,

$$\frac{\mathrm{d}}{\mathrm{d}V_{\mathrm{be}}} I_{\mathrm{e}} \approx \frac{I_{\mathrm{s}}}{\chi}\left\{\exp\left(\frac{qV_{\mathrm{be}}}{kT}\right) - 1\right\}\left[\frac{q}{kT} - \frac{1}{\chi}\frac{\mathrm{d}}{\mathrm{d}V_{\mathrm{be}}}(\chi)\right].$$

Factoring out the thermal voltage,

$$\frac{\mathrm{d}}{\mathrm{d}V_{\mathrm{be}}} I_{\mathrm{e}} \approx I_{\mathrm{e}}\left(\frac{q}{kT}\right)\left[1 - \frac{kT}{q}\frac{1}{\chi}\frac{\mathrm{d}}{\mathrm{d}V_{\mathrm{be}}}(\chi)\right].$$

Then the resistance term can be expressed as

$$r_{\mathrm{e}} \approx \frac{\mathrm{d}V_{\mathrm{be}}}{\mathrm{d}I_{\mathrm{e}}} \approx \frac{\dfrac{kT}{q}}{I_{\mathrm{e}}\left[1 - \dfrac{kT}{q}\dfrac{1}{\chi}\dfrac{\mathrm{d}}{\mathrm{d}V_{\mathrm{be}}}(\chi)\right]},$$

where [59]

$$r_{\mathrm{e}} \approx \frac{\mathrm{d}V_{\mathrm{be}}}{\mathrm{d}I_{\mathrm{e}}} \approx \frac{\left\{\dfrac{kT}{q}\right\}}{\{1 - H\}I_{\mathrm{e}}},$$

where the high-current injection factor is expressed as

$$H = \frac{kT}{q}\frac{1}{\chi}\frac{\mathrm{d}}{\mathrm{d}V_{\mathrm{be}}}(\chi).$$

Given the high-level injection term is associated with the base width modulation and knee current term,

$$\frac{\mathrm{d}}{\mathrm{d}V_{\mathrm{be}}}\chi = \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}V_{\mathrm{be}}}\{\chi_{\varepsilon} + (\chi_{\varepsilon}^2 + 4\chi_{\kappa})^{1/2}\},$$

where

$$\chi_{\varepsilon} = 1 + \frac{V_{\mathrm{be}}}{|V_{\mathrm{A}}|} + \frac{V_{\mathrm{bc}}}{|V_{\mathrm{B}}|},$$
$$\chi_{\kappa} = \frac{I_{\mathrm{eo}}}{I_{\mathrm{kf}}}.$$

Troutman [59] derived the value of $H$ as

$$H = \frac{\zeta}{2[1 + \zeta + \sqrt{(1 + \zeta)}]},$$

where

$$\zeta = \frac{4\chi_\kappa}{\chi_\varepsilon^2}.$$

From this, the generalized tetrode relationship is modified by substituting of the high-current resistance expressions into the generalized differential tetrode relationship,

$$\alpha_{\text{fns}}^* = \frac{R_\text{s}\alpha_{\text{fns}}}{R_\text{s} + r_{\text{en}} + \text{R}_{\text{en}} + [R_{\text{bn}}(1 - \alpha_{\text{fns}})]}$$

and equivalently, by analogy,

$$\alpha_{\text{fps}}^* = \frac{R_\text{w}\alpha_{\text{fps}}}{R_\text{w} + r_{\text{ep}} + \text{R}_{\text{ep}} + [R_{\text{bp}}(1 - \alpha_{\text{fps}})]},$$

where the high-level injection terms are placed in the expression

$$\frac{\alpha_{\text{fns}}}{1 + \dfrac{r_{\text{en}}}{R_\text{s}}} + \frac{\alpha_{\text{fps}}}{1 + \dfrac{r_{\text{ep}}}{R_\text{w}}} \geq 1.$$

## 3.12 TRANSIENT LATCHUP

Transient latchup had a growing interest when it was found that CMOS products were failing in systems during transient power-up. Troutman and Zappe [26, 29, 30], Nelsen [27], Rung and Momose [36], Hamdy and Mohsen [41] and other researchers were interested in the problem of the relationship of power-up and excitation of transient latchup.

### 3.12.1 Transient Latchup – The Four States

Transient latchup is an important topic that addresses the response of a circuit to a transient state. Transient states can occur from ionizing radiation, injection sources, and power-up and power-down states. Hence, there is interest in the general excitation of a pnpn (or pnp and npn transistors). The states of the network are as follows [30]:

- State 1: Both npn and pnp transistors are ‘off’.
- State 2: One transistor is ‘on’ and the other transistor is ‘off’.
- State 3: Both npn and pnp transistors are ‘forward active’.
- State 4: Both npn and pnp transistors are in saturation.

In each one of these cases, to understand the circuit response, a different representation is needed for the transistor elements in different modes of operation.

*First state*: The first state, in order to initiate latchup, does not occur until the circuit transitions to the forward biasing of the transistors (hence must be made to transition to the other states). Hence, at the initial time, the network is ‘off’ and there is not enough current to flow through the network, and it does

not satisfy the latchup forward bias transistor criterion (e.g. the voltage is such that both transistors are forward biased).

*Second state*: In the second state, the condition is that the magnitude of the current is significant enough to establish one of the two transistors to forward bias. The transistor with the larger shunt resistance will undergo forward bias first. For example, given the substrate shunt resistance is larger than the n-well shunt resistance, the npn transistor will achieve the forward bias state first. When one of the transistors is forward active, then the bipolar current gain is $\beta$ and the collector-to-emitter gain is $\alpha$.

*Third state*: In the third case, given the current magnitude is large enough, the second transistor will also achieve forward bias. When the second transistor is forward active, then this transistor will also achieve a bipolar current gain $\beta$, and the collector-to-emitter gain is $\alpha$. In this third case, the sum of the alpha products can be below or above unity. When above unity, regenerative feedback exists.

*Fourth state*: When the fourth state occurs, both transistors saturate, and the current through the structure is a constant. The increase will only occur due to the Early effect (base width narrowing).

### 3.12.2 Transient Latchup – Generalized Excitation Differential Equation and Analysis of the Four States

Zappe and Troutman simplified the analysis where the response of the latchup circuit is represented as a simple differential equation [29, 30]. First assume the current through the structure follows the relationship

$$C\frac{\mathrm{d}}{\mathrm{d}t}V(t) = I_{\mathrm{C}},$$

where $C$ is the capacitor across the pnpn latch and $I_{\mathrm{C}}$ is the current through the capacitor element. The voltage across the latch is the excitation voltage $E(t)$, $I(t)$ is the current through the terminals, and $R$ is the total series resistance of the structure:

$$V(t) = E(t) - RI(t).$$

Zappe and Troutman established a general circuit equation that addresses all four states of the latchup pnpn structure, where $I^*$ is the latchup holding current, $G$ is the gain factor of the circuit (e.g. in its given state), and $R$ is the equivalent resistance in series with the pnpn latch capacitance $C$ [30]:

$$C\frac{\mathrm{d}}{\mathrm{d}t}[E(t) - RI(t)] = I^* + (1 - G)I(t).$$

Figure 3.9 shows a general simple circuit diagram to characterize the state of the system.

In the first region, when the transistors are 'off', as in the first region, the gain factor $G = 0$ and the holding current $I^*$ is also zero, simplifying the differential equation.

In the second region, one of the transistors is 'on' and hence $G = \alpha$, where it is the current gain of the 'on' transistor. Initially, this does not occur until adequate current is established in the network. Hence, the circuit must transition from the first region to the second region.

Assume a certain minimum voltage condition, which will be referred to as $V_{\mathrm{B}}^*$, the voltage that is needed to establish the onset of forward active; then the current that is needed to achieve this at some time $t = t_1$ is

$$I(t = t_1) = \frac{V_{\mathrm{B}}^*}{R_1},$$

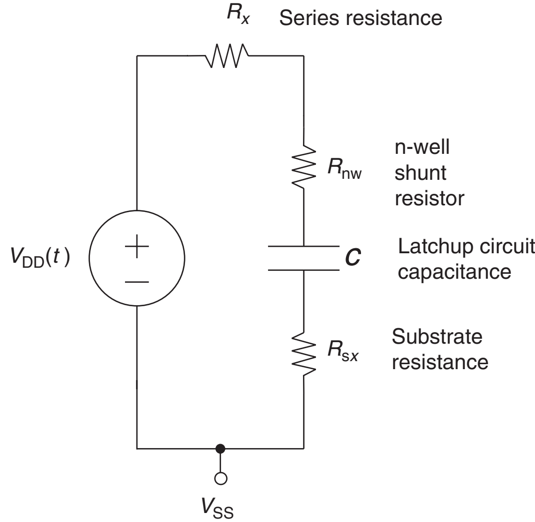



**Figure 3.9** Circuit schematic of the power supply and latchup circuit.

where $R_1$ is the shunt resistance of the first transistor that turns 'on'. The current required to maintain this state (e.g. region 2) is

$$I^* = I_1^* = \alpha_1 \frac{V_B^*}{R_1},$$

which is the current flowing through the 'on' transistor, whose gain is $\alpha_1$, in the third region, at a time $t = t_2$, the second transistor also turns on, and hence

$$I(t = t_2) = \frac{V_B^*}{R_2},$$
$$I^* = \left\{ \frac{\alpha_1}{R_1} - \frac{\alpha_2}{R_2} \right\} V_B^*,$$
$$G = \alpha_1 + \alpha_2.$$

Hence, the solution to the differential equation can be solved using this general method, where it is valid when either transistor turns on first, independent of whether the latchup is 'pnp initiated' or 'npn initiated'.

To solve the differential equation, we can divide by $RC$ [30],

$$\frac{C}{RC} \frac{d}{dt} [E(t) - RI(t)] = \left\{ \frac{1}{RC} \right\} I^* + \left\{ \frac{(1-G)}{RC} \right\} I(t),$$

and place it in the following form:

$$\frac{1}{R} \frac{d}{dt} [E(t)] - \frac{d}{dt} [I(t)] = \left\{ \frac{1}{RC} \right\} I^* + \left\{ \frac{(1-G)}{RC} \right\} I(t),$$
$$\frac{1}{R} \frac{d}{dt} [E(t)] - \frac{d}{dt} [I(t)] = \left\{ \frac{1}{RC} \right\} I^* + \frac{1}{\tau} I(t),$$

where

$$\tau = \frac{RC}{\{1 - G\}},$$

$$\frac{1}{R}\frac{\mathrm{d}}{\mathrm{d}t}[E(t)] - \frac{\mathrm{d}}{\mathrm{d}t}[I(t)] = \left\{\frac{1}{RC}\right\}I^* + \frac{1}{\tau}I(t).$$

This can be put in the form

$$\frac{\mathrm{d}}{\mathrm{d}t}I(t) + \frac{1}{\tau}I(t) + \left\{\frac{1}{RC}\right\}I^* = \frac{1}{R}\frac{\mathrm{d}}{\mathrm{d}t}[E(t)].$$

In this form, the current is a first-order differential equation, with a ‘drive term’ associated with the excitation voltage. The Laplace transform of the equation is

$$\{sI(s) - I(t_0)\} + \frac{1}{\tau}I(s) + \frac{1}{s}\left\{\frac{1}{RC}\right\}I^* = \frac{1}{R}\{sE(s) - E(t_0)\}.$$

Defining time constants for the different cases where $R_x$ is a series resistance within the network, then from the above general term, the following time constants exist in the different domains:

$$\tau_1 = \{R_1 + R_2 + R_x\}C,$$

$$\tau_2 = \frac{\{R_2 + R_x\}C}{1 - \alpha_1}$$

and

$$\tau_3 = \frac{\{R_x\}C}{1 - \alpha_1 - \alpha_2}.$$

The solutions to the differential equation in the different regions [30] are as follows:

- In the first region, $0 < t < t_1$,

$$I(t) = \frac{1}{\{R_x + R_2 + R_1\}}\int_0^t \mathrm{d}t' E'(t') \exp\left\{-\frac{t - t'}{(R_x + R_1 + R_2)C}\right\},$$

$$V(t) = \{R_1 + R_2\}I(t) + V_\mathrm{C}(t),$$

$$V_\mathrm{C}(t) = \frac{1}{C}\int I(t)\mathrm{d}t.$$

- In the second region, $t_1 < t < t_2$,

$$I(t) = \frac{V_\mathrm{B}^*}{R_1}\left(\frac{\alpha_1}{\alpha_1 - 1}\right) + \frac{V_\mathrm{B}^*}{R_1(1 - \alpha_1)}\exp\left\{-\frac{(1 - \alpha_1)(t - t_1)}{(R_x + R_2)C}\right\}$$
$$+ \frac{1}{\{R_x + R_2\}}\int_{t_1}^t \mathrm{d}t' E'(t') \exp\left\{-\frac{(1 - \alpha_1)(t - t')}{(R_x + R_2)C}\right\},$$

$$V(t) = V_B^* + \{R_2\}I(t) + V_C(t),$$
$$V_C(t) = V_C(t_1) + \frac{\alpha_1 V_B^*}{R_1 C}(t - t_1) + \frac{V_B^*}{R_1(1-\alpha_1)} + \frac{1-\alpha_1}{C}\int_{t_1}^{t} dt' I(t').$$

- For the third region, $t_2 < t < t_3$,

$$I(t) = -\frac{V_B^*}{(1-\alpha_1-\alpha_2)}\left(\frac{\alpha_1}{R_1}+\frac{\alpha_2}{R_2}\right) + \frac{V_B^*}{(1-\alpha_1-\alpha_2)}\left\{\frac{1-\alpha_1}{R_2}+\frac{\alpha_1}{R_1}\right\}\exp\left\{-\frac{(1-\alpha_1-\alpha_2)(t-t_2)}{(R_x)C}\right\}$$
$$+\frac{1}{\{R_x\}}\int_{t2}^{t} dt' E'(t')\exp\left\{-\frac{(1-\alpha_1-\alpha_2)(t-t')}{(R_x)C}\right\},$$
$$V(t) = 2V_B^* + V_C(t),$$

$$V_C(t) = V_C(t_2) + \left\{\frac{\alpha_1}{R_1}+\frac{\alpha_2}{R_2}\right\}\frac{V_B^*}{C}(t-t_2) + \frac{1-\alpha_1-\alpha_2}{C}\int_{t2}^{t} dt' I(t').$$

## 3.12.3 Transient Latchup – Power-Up Excitation

In the previous section, the voltage excitation was of the general form $E(t)$ to represent any type of excitation process. Different excitation processes are of interest:

- linear power-up positive voltage ramp with single polarity;
- linear power-down positive voltage ramp with single polarity;
- damped oscillation voltage ramp with a positive first transition;
- damped oscillation voltage ramp with a negative first transition.

The first two cases of a linear voltage ramp provide a general condition of interest in semiconductor chips of the power-up and power-down of a power supply. The second two cases are of interest in transient latchup. The last case of a damped oscillation is of interest in the ESD Association transient latchup specification [63–66, 68, 69]. In this section, the derivation of the first case will be shown based on the transient latchup analysis.

For the case of a linear ramp rate, it will be assumed that a constant ramp rate $K$ is applied to the power supply voltage $V_{DD}$ rail, where $V_{DD}(t) = K\ t$ between the initial time $t = 0$ and the power-up time $t = t_{pu}$, or

$$\frac{d}{dt}V_{DD}(t) = K, \qquad 0 \le t \le t_{pu},$$

and for times after the ‘power-up time’, $t > t_{pu}$, the power supply is stable and constant,

$$V_{DD}(t) = V_P, \qquad t \ge t_{pu}.$$

The ramp rate $K$ is then defined as

$$K = \frac{V_{DD}}{t_{pu}}.$$

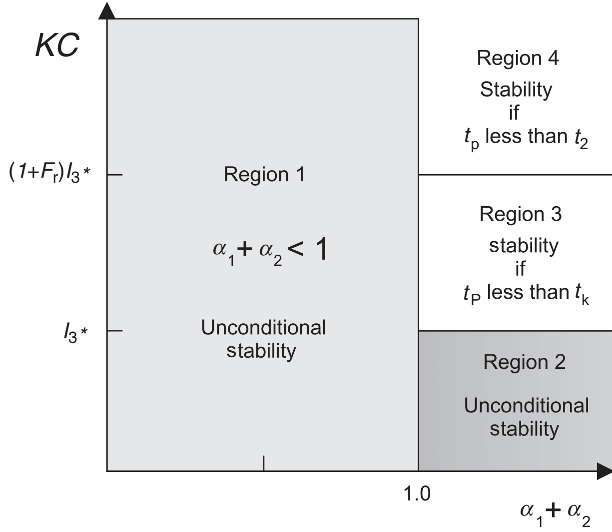



**Figure 3.10** Circuit transition state diagram.

Zappe noted that the response of the latchup structure is a function of the state of the circuit, the ramp rate and the ramp voltage termination [30]. This is can be divided into different cases. Figure 3.10 highlights four cases of transition associated with the ramp rate and sum of the alphas. The cases are as follows:

- *Case I*: Slow ramp rate, with voltage ramp termination prior to the first transistor achieving 'turn on'.
- *Case II*: Faster long ramp rate, with the first transistor achieving 'turn-on' but voltage ramp termination prior to the second transistor achieving 'turn-on'.
- *Case III*: Faster long ramp rate, with both transistors becoming forward active. In this case, there are two subclasses where
  1. Circuit satisfies forward active condition, but does not satisfy the regenerative feedback latchup requirement.
  2. Circuit satisfies forward active condition and the regenerative feedback requirement, but voltage ramp termination point occurs prior to achieving the knee voltage, as well as an additional condition. In this case, 'dynamic recovery' occurs.
- *Case IV*: Fast ramp rate, with long ramp, the latchup conditions are satisfied, the voltage termination value exceeds the knee condition and latchup occurs.

For simplicity, let us define the power-up time $t_{pu}$ as a rise time $t_r$ (note that $t_{pu} = t_r$, where it is understood not the industry-accepted rise time variable of 10–90 % but 0–100 %).

*Case I*: In the first case, neither transistors turn on during the voltage ramp. As a result, during the voltage ramp, the capacitor of the pnpn latch begins to charge, without any of the transistors becoming

forward active. If the magnitude of the ramp rate $K$ is too slow, or if the time of the ramp, $t_r$, is too short, the voltage is not reached that leads to forward biasing of either of the transistors [30]:

$$I(t) = KC\left[1 - \exp\left\{-\frac{t}{\tau_1}\right\}\right], \qquad 0 \le t \le t_r,$$
$$I(t) = KC\left[1 - \exp\left\{-\frac{t_r}{\tau_1}\right\}\right] \exp\left\{-\frac{t - t_r}{\tau_1}\right\}, \qquad t > t_r.$$

*Case II*: In the second case, if the ramp rate $K$ is fast and the time of the ramp is long enough, at least one of the transistors becomes forward active. Prior to achieving forward bias, the first transistor is charging; after forward bias occurs, the current gain $\alpha$ of the first transistor to turn on is on, leading to a change in the gain $G$ of the network. This is addressed in the second equation (e.g. note the presence of $\alpha_1$ in the equation) [30]:

$$I(t) = KC\left[1 - \exp\left\{-\frac{t}{\tau_1}\right\}\right], \qquad 0 \le t \le t_1,$$
$$I(t) = I_1^* + \left[\frac{KC - I_1^*}{1 - \alpha_1}\right]\left[1 - \exp\left\{-\frac{t - t_1}{\tau 2}\right\}\right], \qquad t_1 \le t \le t_r.$$

Given the voltage ramp is completed, then there is no further charging of the pnpn network and there is a time when the first transistor that was forward active begins to again shut off. The time when the transistor ceases to be forward active is $t = t_a$:

$$I(t) = \frac{\alpha_1 I_1^*}{1 - \alpha_1} + \frac{1}{1 - \alpha_1}\left[I_1^* + KC\left(\exp\left\{\frac{t_r - t_1}{\tau_2}\right\} - 1\right)\right] \exp\left\{-\frac{t - t_1}{\tau_2}\right\}, t_r \le t \le t_a.$$

For times after $t = t_a$, the current gain of the network is not a function of the bipolar collector-to-emitter current gain ($\alpha_1$) [30]:

$$I(t) = I_1^* \exp\left\{-\frac{t - t_a}{\tau_1}\right\}, \qquad t \ge t_a,$$

where

$$t_a = t_1 + \tau_2 \ln\left[1 + \frac{KC}{I_1^*}\left(\exp\left\{\frac{t_r - t_1}{\tau_2}\right\} - 1\right)\right].$$

*Case III*: In the third case, the first equation is the supply passing through the forward bias state of the first transistor,

$$I(t) = KC\left[1 - \exp\left\{-\frac{t}{\tau_1}\right\}\right], \qquad 0 \le t \le t_1.$$

The voltage ramp continues, where one transistor is forward biased, and the gain of the circuit is dependent on that given transistor,

$$I(t) = I_1^* + \left[\frac{KC - I_1^*}{1 - \alpha_1}\right]\left[1 - \exp\left\{-\frac{t - t_1}{\tau_2}\right\}\right], \qquad t_1 \le t \le t_2.$$

At a later time, both transistors become forward active, and the gain of the circuit is dependent on the sum of the two collector-to-emitter gains, where $\alpha_1 + \alpha_2 = \alpha_{\rm pnp} + \alpha_{\rm npn}$:

$$I(t) = I_2^* + \left[\frac{KC - I_3^*}{1 - \alpha_1 - \alpha_2}\right]\left[1 - \exp\left\{-\frac{t - t_2}{\tau 3}\right\}\right], \qquad t_2 \le t \le t_{\rm r}.$$

Zappe noted that in this case, although both transistors are forward active, this does not guarantee that latchup occurs [30]. Zappe noted the first case where if the ‘beta product’ or ‘alpha sum’ is less than unity, then the network does not lead to latchup. In addition, there is the second case, where the ‘alpha sum’ criterion is satisfied; but if the ramp rate is terminated prior to the time to reach the knee condition (e.g. knee time),

$$t_{\rm r} < t_{\rm K},$$

and

$$KC < I_3^*\left[1 - \exp\left\{\frac{t_{\rm r} - t_2}{\tau 3}\right\}\right] - 1,$$

then latchup does not occur. Zappe and Troutman referred to this state as the ‘dynamic recovery’ [26, 29, 30]. Hence, there is a time when both transistors are forward active and the equation is a function of the alpha sum,

$$I(t) = \left[\frac{1}{1 - \alpha_1 - \alpha_2}\right]\left[\left[KC\left(\exp\left\{\frac{t_{\rm r} - t_2}{\tau 3}\right\} - 1\right) + I_3^*\right]\exp\left\{-\frac{t - t_2}{\tau 3}\right\} - I_k^*\right], \qquad t_{\rm r} \le t \le t_{\rm b}.$$

After time $t = t_{\rm b}$, the latched network undergoes recovery where the second transistor shuts off, leading to a gain associated with only the first transistor initiated,

$$I(t) = \left[I_2^* + \frac{\alpha_1 I_1^*}{1 - \alpha_1}\right]\exp\left\{-\frac{t - t_b}{\tau_2}\right\} - \frac{\alpha_1 I_1^*}{1 - \alpha_1}, \qquad t_{\rm b} \le t \le t_{\rm c},$$

and at a some later time, both transistors are off, leading to a discharging of the network as

$$I(t) = I_1^* \exp\left\{-\frac{t - t_{\rm c}}{\tau_1}\right\}, \qquad t \ge t_{\rm c},$$

where the time of the shut off of the second transistor is

$$t_{\rm b} = t_2 + \tau_3 \ln\left[1 + \frac{KC}{I_3^*}\left(\exp\left\{\frac{t_r - t_2}{\tau_3}\right\} - 1\right)\right]$$

and the time of shut off of both transistors is [30]

$$t_{\rm c} = t_{\rm b} + \tau_2 \ln\left[1 + (1 - \alpha_1)\left(\frac{R_{\rm s}}{R_{\rm w}} - 1\right)\right].$$

*Case IV*: The last case highlighted by Zappe [30] was the one where the voltage ramp time extends beyond the time needed to reach the knee voltage, the knee time,

$$t_r > t_K,$$

or a ramp rate condition of

$$KC > I_3^*\left[1 - \exp\left\{\frac{t_r - t_2}{\tau_3}\right\}\right] - 1.$$

In this case, all the prior times undergo the same time sequence of the first transistor turn 'on', followed by the second transistor turn 'on'. Given both transistors are forward active, where $t_3$ is larger than the ramp rate,

$$I(t) = \left[\frac{1}{1 - \alpha_1 - \alpha_2}\right]\left[\left[KC\left(\exp\left\{\frac{t_r - t_2}{\tau_3}\right\} - 1\right) + I_3^*\right]\exp\left\{-\frac{t - t_2}{\tau_3}\right\} - I_k^*\right], \qquad t_r \leq t \leq t_3.$$

In the case where the transistors are saturated, but the voltage ramp condition is ongoing, we can defined a time $t_4$, associated with the time that the transistors undergo saturation [30],

$$t_4 = t_2 - \tau_3 \ln\left[1 + \frac{(\alpha_1 + \alpha_2 - 1)}{KC - I_3^*}\left(\frac{Kt_4 - V_B^*}{R_x} - I_2^*\right)\right].$$

## 3.13 EXTERNAL LATCHUP

In a semiconductor chip environment, an external injection outside of the CMOS inverter circuit (e.g. parasitic pnpn) can initiate latchup. An injection of minority carriers from a separate circuit can influence both the well and substrate regions and serve as a 'injection source' for latchup. A typical injection source can be a n-MOSFET source, or n-diffusion or n-well diode structure in a substrate region [73, 74].

### 3.13.1 External Latchup Diode Injection Source Analysis

In the case of a diode-type injection phenomenon, the analysis of the injection source can be quantified by the solution of the diffusion equation in space and time. First, the number of carriers can be quantified from the high-level injection physics. Under high-level injection, the excess minority carrier populations can be expressed as

$$\Delta p_n = \frac{p_n(e^{V_j}/kT - 1)\left(1 + \frac{n_n}{p_p}e^{q(V_j - \phi_i)/kT}\right)}{1 - e^{2q(V_j - \phi_i)/kT}}$$

and

$$\Delta n_p = \frac{n_p(e^{V_j}/kT - 1)\left(1 + \frac{p_p}{n_n}e^{q(V_j - \phi_i)/kT}\right)}{(1 - e^{2q(V_j - \phi_i)/kT})}.$$

From this expression, the diode current equation for an n+ diffusion in high-level injection at an initial position can be expressed as

$$I = I_s \frac{(e^{V_j}/kT - 1)\left(1 + \eta e^{q(V_j - \phi_i)/kT}\right)}{1 - e^{2q(V_j - \phi_i)/kT}},$$
$$\eta = \frac{N_d}{N_d},$$
$$I_s = qA\left\{\frac{D_p}{L_p} p_{n0} + \frac{D_n}{L_n} n_{p0}\right\},$$

where $I$ is the diode current and $I_s$ is the reverse saturation current [73, 74]. The injection source can be represented as a step function initial condition in time, where the injection is the current or the carrier density. Applying the step function of carrier density to the carrier diffusion equation, the diffusion equation can be solved in frequency domain using the Laplace transform for a step impulse. The solution for the diffusion equation in space and time for the carrier propagation is the complementary error function. To obtain the current of the primary source, using the diffusion expression, the current can be obtained. In this model, the spatial and temporal responses are coupled. For simplicity, we will show the 1D solution. Define a transform integral pair of temperature using the Laplace transform pair:

$$F(x,s) = \int_{t=0}^{t=\infty} e - stN(x,t)dt \; N(x,t) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} e^{st} F(x,s) ds,$$

where the function $N(x,t)$ is the minority carrier injection and $F(x,s)$ is its transform pair that only exists in the right half-plane (RHP) of the $s$-space where we have the real part of $s$ satisfying $\mathrm{Re}\{s\} > \alpha$. From the time-dependent diffusion equation with constant coefficients

$$\frac{\partial^2 N(x,t)}{\partial x^2} = \frac{1}{D} \frac{\partial N(x,t)}{\partial t},$$

the function can be expressed as a function of the transform pair $F(x,s)$. In the $s$-domain, this can be expressed as

$$\frac{\partial^2 N(x,s)}{\partial x^2} - s\frac{1}{D} N(x,s) = -N(x,0).$$

The homogeneous equation has the general solution of

$$N(x,s) = N^+ \exp\{-\alpha x\} + N^- \exp\{\alpha x\},$$

where

$$\alpha = \sqrt{\frac{s}{D}}.$$

With an initial condition of minority injected carriers, at time $t = 0$,

$$N(x = 0, t) = N_0 u(t),$$

and for the assumption of an infinite domain, the physical solution is

$$N(x,s) = \frac{N_0}{s}\exp\{-\alpha x\}.$$

From the inverse transform, we can solve for the number of carriers in space and time; the solution can be expressed as the complementary error function,

$$N(x,t) = N\left[1 - \operatorname{erf}\left\{\frac{x}{2}\sqrt{\frac{1}{Dt}}\right\}\right]u(t) = Nu(t)\operatorname{erfc}\left\{\frac{x}{2}\sqrt{\frac{1}{Dt}}\right\}.$$

For a more physical interpretation, the solution can be substituted for the diffusion length $L_D$ and the recombination time $\tau$ [73, 74],

$$N(x,t) = N_0\left[1 - \operatorname{erf}\left\{\frac{x}{2}\sqrt{\frac{\tau}{L_D^2\, t}}\right\}\right]u(t)$$

or

$$N(x,t) = N_0 u(t)\operatorname{erfc}\left\{\frac{x}{2L_D}\sqrt{\frac{\tau}{t}}\right\}.$$

In this form, the physical competition between the spatial diffusion of the source charge and the associated diffusion length, as well as the temporal competition between the recombination time and the propagation time, is evident. From this form, the primary injection source can be solved, using the diffusion relationship,

$$I_{\text{inj}}(x,t) = -qD\frac{\partial}{\partial x}N(x,t).$$

In this fashion, the solution of an injection source can be evaluated in space and time. Taking the first derivative, the current can be expressed as [73, 74]

$$I(x,t) = qD\frac{N_0}{L_D}\sqrt{\frac{\tau}{\pi t}}\exp\left\{-\left(\frac{x}{2L_D}\right)^2\left(\frac{\tau}{t}\right)\right\}u(t).$$

Hence, from these developments, the spatial relationship and temporal response are solved, where the number of carriers injected is known. The current propagation in space and time can be understood.

## 3.13.2 External Latchup Criterion

To evaluate latchup and the regenerative feedback from an external current source, the fundamental relationships must incorporate the external current source in the analysis. Voldman applied the external source analysis into the latchup criterion and derived the external current magnitude necessary to initiate latchup [73, 74]. The two nodal equations can be expressed as

$$I_{cp} = I_{bn},$$
$$I_{cn} = I_{bp}$$

and

$$I = I_{ep} = I_{cp} + I_{bp}.$$

From the coupling relationships, we can express this as

$$I = I_{ep} = I_{cp} + I_{cn} = I_{bp} + I_{bn} = I_{cn} + I_{bn} = I_{en} = I.$$

Solving for the current as a function of the two collector relationships, we can relate the collector current as a function of the emitter current,

$$I_{cp} = \alpha I_{ep} + I_{cp0},$$
$$I_{cn} = \alpha I_{en} + I_{cn0}.$$

The condition for triggering of the parasitic pnpn is a function of the currents through the feedback elements; these include the well and substrate resistances.

$$I = I_{ep} + I_{w},$$
$$I = I_{en} + I_{sx}.$$

Then

$$I = I_{cp} + I_{cn} = \alpha_p I_{ep} + I_{cp0} + \alpha_n I_{en} + I_{cn0}.$$

From an external source mechanism, electrons enter the base of the pnp transistor. Electrons that traverse the substrate-to-well junction are collected at the base node of the pnp transistor in the current loop; this will be expressed as $I_{in}$. Hence, at the first nodal point in the regenerative feedback loop, the current continuity is expressed as

$$I = \{I_{cp} + I_{cn}\} + I_{in}.$$

In the derivation, it is assumed that both the electron and hole populations can influence latchup [73, 74]. At the same time, holes influence the local substrate potential serving as base drive for the parasitic npn transistor. This can be incorporated as a current source $I_{out}$. Current continuity at the second nodal point of the regenerative feedback loop can be expressed as

$$I = \{I_{cp} + I_{cn}\} + I_{out}.$$

Substituting collector current terms, we can express the current in the pnpn structure as follows:

$$I = \{\alpha_p I_{ep} + I_{cp0}\} + \{\alpha_n I_{en} + I_{cn0}\} + I_{in}.$$

Substituting the current through the well shunt resistance,

$$I = \alpha_p\{I_{ep} + I_w\} - \alpha_p I_w + \alpha_n\{I_{en} + I_s\} - \alpha_n I_s + \cdots$$
$$+ \{I_{cp0} + I_{cn0}\} + I_{in}.$$

The last three terms in the expression are associated with the injection and leakage currents within the regenerative feedback loop:

$$I = (\alpha_p + \alpha_n)I - \alpha_p(I_w) - \alpha_n(I_{sx}) + (I_{cn0} + I_{cp0}),$$

$$I = \frac{(I_{cn0} + I_{cp0} + I_{in}) - \alpha_p(I_w) - \alpha_n(I_{sx})}{1 - (\alpha_p + \alpha_n)}.$$

Normalizing the expression to the current, the form of the equation can be expressed as

$$I\left\{(\alpha_p + \alpha_n) - 1 - \alpha_p\frac{(I_w)}{I} - \alpha_n\frac{(I_{sx})}{I} + \frac{I^*}{I}\right\} = 0,$$

and define the new term [73,74]

$$I^* = I_{cn0} + I_{cp0} + I_{in}.$$

Hence, the new condition for latchup is [73, 74]

$$(\alpha_p + \alpha_n) = 1 + \alpha_p\frac{(I_w)}{I} + \alpha_n\frac{(I_{sx})}{I} - \frac{I^*}{I}.$$

This latchup criterion is the relationship of the sum of the alphas in the presence of both leakage currents and an external source. From this relationship, the necessary current needed to initiate latchup can be obtained. Hence, the necessary current for latchup initiation from an external current source occurs when [73, 74]

$$I^* = I\left\{1 + \alpha_p\left[\frac{(I_w)}{I} - 1\right] + \alpha_n\left[\frac{(I_{sx})}{I} - 1\right]\right\},$$

where the well and substrate current expressions are

$$I_w = \frac{(V_{BE})_{pnp}}{R_w} = \frac{V_0}{R_w}\ln\left[\frac{I - I_w}{(I_0)_p}\right],$$

$$I_{sx} = \frac{(V_{BE})_{npn}}{R_{sx}} = \frac{V_0}{R_{sx}}\ln\left[\frac{I - I_{sx}}{(I_0)_n}\right].$$

Hence, the condition for latchup from an external source can also be expressed as

$$I^* = I\left\{1 + \frac{\beta_p}{\beta_p + 1}\left[\frac{I_w}{I} - 1\right] + \frac{\beta_n}{\beta_n + 1}\left[\frac{I_{sx}}{I} - 1\right]\right\}$$

or

$$I_{inj} = I\left\{1 + \frac{\beta_p}{\beta_p + 1}\left[\frac{I_w}{I} - 1\right] + \frac{\beta_n}{\beta_n + 1}\left[\frac{I_{sx}}{I} - 1\right]\right\} - I_{cn0} - I_{cp0}.$$

### 3.13.3 External Latchup Propagation

In all the analyses, it is assumed that the CMOS circuit is in an isolated environment. In a semiconductor chip, there are many circuits in a given area that can potentially undergo latchup. The primary source injection of minority carriers will extend under these networks in space and time. In advanced CMOS technology, the high circuit density is such that the minority carrier injection will traverse many circuits. As the carriers diffuse, through the substrate, they influence the substrate potential as well as the n-well potential of the circuitry. In this scenario, the primary source injection will be migrating under circuits that may potentially latchup the circuits, depending on the satisfying of the latchup criterion.

The latchup 'domino effect' is effective where the primary source in combination with the latched circuits leads to latchup of additional networks. Analogous to the game of dominos, a primary source initiates the first response of the first circuit when the switching threshold of the first circuit is achieved; this initiates the response of the second circuit when its threshold is achieved, and this continues to propagate until the threshold for switching of the last circuit in the chain is not achieved. The distinction in this environment has the following differences:

- The primary source can be involved in space and time over a plurality of circuits (e.g. not just a single first circuit).
- The secondary sources have two conditions – a lower level pnp injection and a higher level pnpn latchup injection.
- The phenomenon is coupled through the substrate.
- The secondary sources can be coupled through common wells but may be in independent well structures and different power supply rails.
- The propagation phenomenon is two dimensional in space along a chip surface.
- The latchup initiation is a function of the satisfying of the four variables needed to initiate latchup (well resistance, substrate resistance, npn current gain and pnp current gain).
- The time constant for initiation is associated with the diffusion time across the substrate from one source to the other source.
- The time constant for secondary sources switching is associated with the latchup switching time of the pnpn circuit and the diffusion time from source to source.
- The time constant for the power supply voltage response is a function of the power bus distribution (e.g. *IR* voltage drops) and the number of latched networks on a given power bus.
- The truncation of latchup propagation is associated with the lack of satisfying the forward active voltage condition and the latchup 'alpha' or 'beta' criterion.
- The domino effect is cumulative in that the circuits remain in the latched state, until the power supply is reduced below the holding voltage condition of that pnpn element.
- The truncation along a power bus can be associated with the final voltage along a position in the power grid, which is less than the sum of the forward voltages.

If the external current magnitude is high enough to satisfy the 'alpha' relationship or the 'beta' relationship, external latchup will occur in the first device. The primary injection pulse has a time

constant associated with the physical distance the carriers need to diffuse to reach the latchup-sensitive circuit of interest. From the argument of the external latchup pulse, the propagation in space and time can be understood by fixing the argument of the external source as a constant and switching reference frames. From the injected carrier density, we can let the argument be a constant and move along the top of the pulse as it propagates in time and space:

$$\xi \equiv \frac{x}{2L_{\mathrm{D}}}\sqrt{\frac{\tau}{t}},$$

where

$$N(x,t) \equiv N_0 \mathrm{erfc}(\xi)u(t) = \mathrm{const.}$$

In the frame of the pulse source, we can move along the pulse source at a constant injection level and observe the carriers (e.g. current) propagate. The constant value with the fixed argument moves in space $x$ at time $t$ according to

$$x \equiv \frac{2\xi L_{\mathrm{D}}}{\sqrt{\tau}}\sqrt{t}.$$

The condition to initiate latchup in a given circuit requires that the voltage across the latched state establishes a forward active voltage across the emitter–base junction of both the pnp transistor and the npn transistor. This precondition for the establishment of latchup is critical in the propagation problem. Hence, the sum of the two forward bias voltages is the first precondition for the initiation of latchup in a circuit. The forward active condition for latchup initiation requires

$$(V_{\mathrm{BE}})_{\mathrm{pnp}} \geq 0.7\,\mathrm{V},$$
$$(V_{\mathrm{BE}})_{\mathrm{npn}} \geq 0.7\,\mathrm{V},$$

where

$$(V_{\mathrm{BE}})_{\mathrm{pnp}} = I_{\mathrm{w}}R_{\mathrm{w}} = V_0 \ln\left[\frac{I - I_{\mathrm{w}}}{(I_0)_{\mathrm{p}}}\right],$$
$$(V_{\mathrm{BE}})_{\mathrm{npn}} = I_{\mathrm{s}x}R_{\mathrm{s}x} = V_0 \ln\left[\frac{I - I_{\mathrm{s}x}}{(I_0)_{\mathrm{n}}}\right].$$

For the specific circuit to undergo latchup, it was shown that the sum of the alphas has to be equal to or greater than the following form with the extrinsic current source evident on the RHS:

$$(\alpha_{\mathrm{p}} + \alpha_{\mathrm{n}}) = 1 + \alpha_{\mathrm{p}}\frac{(I_{\mathrm{w}})}{I} + \alpha_{\mathrm{n}}\frac{(I_{\mathrm{s}x})}{I} - \frac{I^*}{I}.$$

This can be written in an alternate form,

$$\left(\alpha_{\mathrm{p}}\left\{1 - \frac{I_{\mathrm{w}}}{I}\right\} + \alpha_{\mathrm{n}}\left\{1 - \frac{I_{\mathrm{s}x}}{I}\right\}\right) = 1 - \frac{I^*}{I},$$

or equivalently in the generalized alpha form, it can be expressed as [73]

$$(\alpha_p * + \alpha_n *) = 1 - \frac{I^*}{I},$$

where the transport factors are generalized to include the terms in the parentheses. In this generalized form, when there is no injection phenomenon, the latchup occurs when the sum of the generalized alphas is greater than unity. When the sum of the generalized alphas is less than unity, no latchup occurs. In a generalized alpha space, this forms a triangular region of the SAFE zone where latchup occurs outside the SAFE zone (triangular region) and does not occur when within. Hence, when the injection from the primary source does occur, the demarcation line between latchup and no latchup shifts, decreasing the demarcation line toward the *x*–*y* zero intercept. As the injection increases, the demarcation line is driven toward zero-axis intersection. For a given circuit, we can also state that the excitation current needed to latchup a given circuit from a single or plurality of sources can be expressed as [73, 74]

$$I^* = I\{1 - \alpha_p * - \alpha_n *\},$$
$$\left(\alpha_p\left\{1 - \frac{I_w}{I}\right\} + \alpha_n\left\{1 - \frac{I_{sx}}{I}\right\}\right) = 1 - \frac{I^*}{I},$$
$$\alpha_p^* = \alpha_p\left\{1 - \frac{I_w}{I}\right\},$$
$$\alpha_n^* = \alpha_n\left\{1 - \frac{I_{sx}}{I}\right\}.$$

For the same circuit, this can be expressed as a function of the bipolar current gains. Substituting the relationship between the collector-to-emitter transport and the bipolar current gain, the beta product relation for undergoing latchup in the environment of an external injection can be written in the following form [73, 74]:

$$\beta_n\beta_p = \frac{1 + \beta_p\left[\frac{I_w - I^*}{I}\right] + \beta_n\left[\frac{I_s - I^*}{I}\right] - \left[\frac{I^*}{I}\right]}{\left[1 - \frac{I_w + I_s - I^*}{I}\right]},$$

where

$$I^* = I_{cp0} + I_{cn0} + I_{inj}(x, t),$$
$$I_w = \frac{(V_{BE})_{pnp}}{R_w} = \frac{V_0}{R_w}\ln\left[\frac{I - I_w}{(I_0)_p}\right],$$
$$I_{sx} = \frac{(V_{BE})_{npn}}{R_{sx}} = \frac{V_0}{R_{sx}}\ln\left[\frac{I - I_{sx}}{(I_0)_n}\right].$$

From this criterion, in the beta product relationship, latchup occurs when the RHS > LHS of the equation. When the well, substrate and external injection approach zero, the RHS of the equation equals unity. Without the external latchup injection, the RHS is larger; hence, a larger bipolar current gain product is needed to initiate latchup. As the external injection current increases, the RHS terms decrease in magnitude, making it easier for the circuit to undergo latchup. When the current magnitude

of the LHS exceeds the RHS, latchup occurs. Graphically, in beta product space, the area between the hyperbola and the axis is the condition where latchup does not occur, and outside of it, latchup occurs. As the external injection increases, the locus of points demarking the latchup condition decreases, making the SAFE region in beta product space decrease.

In the environment of multiple circuits, when the first circuit undergoes latchup, current can be injected into the substrate and be additive to the primary source. For the first circuit element that undergoes latchup, we know the current injection initiated by this secondary source of current, from the holding current relationship. Hence, for a given circuit $j$th element that has undergone a latched state (where $j = 1$ for the first circuit) for a specific well and contact spacing, the holding current relationship is

$$(I_{\mathrm{H}}(x_j, t_j))_j = \frac{\beta_{\mathrm{p}}(\beta_{\mathrm{n}} + 1)I_{\mathrm{w}} + (\beta_{\mathrm{p}} + 1)\beta_{\mathrm{n}} I_{\mathrm{s}x}}{\beta_{\mathrm{p}}\beta_{\mathrm{n}} - 1},$$

$$(I_{\mathrm{w}})_j = \frac{(V_{\mathrm{BE}})_{\mathrm{pnp}}}{(R_{\mathrm{w}})_j} = \frac{V_0}{R_{\mathrm{w}}} \ln\left[\frac{I - I_{\mathrm{w}}}{(I_0)_{\mathrm{p}}}\right],$$

$$(I_{\mathrm{s}x})_j = \frac{(V_{\mathrm{BE}})_{\mathrm{npn}}}{(R_{\mathrm{s}x})_j} = \frac{V_0}{R_{\mathrm{s}x}} \ln\left[\frac{I - I_{\mathrm{s}x}}{(I_0)_{\mathrm{n}}}\right].$$

This first circuit that undergoes latchup switches from a low- to a high-current state with a current magnitude of the holding current. Its injection into the substrate will be lower than the holding current of that specific circuit, and this holding state is established as long as the power supply voltage is maintained. This holding current becomes the initial current condition at the location of the specific circuit and is associated with the secondary substrate injection at a position $x = x_j$. The primary source is at $x = x_0$ and the physical distance is $x = x_j - x_0$.

The current injection from the first latched circuit has a time response associated with two physical time constants: the time from the initial injection source and the time associated with initiated latchup in that specific circuit $(j = 1)$. The first time constant is associated with the time for the current to flow from the primary injection source to the $j$th circuit $(j = 1)$. Additionally, the injection from the $j = 1$ latched circuit is associated with the pnpn switching time of that specific circuit. The pnpn parasitic circuit has a finite response time associated with the pnp and the npn transistor. The time response of the pnpn structure is the sum of the pnp base transit time $\tau_{\mathrm{pnp}}$ and the npn base transit time $\tau_{\mathrm{npn}}$. The pnp base transit time is equal to

$$(t_{\mathrm{pnp}})_j = \left(\frac{(W_{\mathrm{bp}})^2}{2D_{\mathrm{p}}}\right)_j$$

and the npn base transit time is

$$(t_{\mathrm{npn}})_j = (\frac{(W_{\mathrm{bn}})^2}{2D_{\mathrm{n}}})_j.$$

Hence, the response time of the pnpn is the sum of the transit times,

$$(t_{\mathrm{pnpn}})_j = (t_{\mathrm{pnp}} + t_{\mathrm{npn}})_j = \left\{\frac{(W_{\mathrm{bp}})^2}{2D_{\mathrm{p}}} + \frac{(W_{\mathrm{bn}})^2}{2D_{\mathrm{n}}}\right\}_j.$$

As a result, after the time of the injection into the physical structure, there is a finite time in order to initiate the pnpn latching secondary circuit. Hence, in the propagation analysis, the temporal dynamic of the latchup domino effect, and the time for the transport to the first circuit and the initiation of the first latchup event occur. From, the argument, from the initial source injection position at $(x_0, t_0)$, the time to propagate can be shown as

$$x_j = x_0 \equiv \frac{2\xi L_{\mathrm{D}}}{\sqrt{\tau}} (\sqrt{t_j - t_0})$$

or for the first time constant

$$t_j - t_0 = \frac{\tau}{4\xi^2 L_{\mathrm{D}}^2} (x_j - x_0)$$

and the time of injection of the secondary is then

$$t_j = (t_j - t_0) + (t_{\mathrm{pnpn}})_j.$$

To further address the latchup domino effect, once latchup occurs in the first circuit, the holding current flowing through the 'latched circuit' also injects current into the substrate and well regions. Hence, the linear superposition of currents applies, and the 'injected current relationship' at any point in space and time is a function of the primary injection source and additional secondary sources. The time of initiation of the second circuit (e.g. $j = 2$) is the sum of the following times:

- time to transport from the primary injection location to the $j = 2$ location:

$$t_j - t_0|_{j=2} = \frac{\tau}{4\xi^2 L_{\mathrm{D}}^2} (x_j - x_0)|_{j=2};$$

- time to transport from the primary injection location to the $j = 1$ location:

$$t_j - to|_{j=1} = \frac{\tau}{4\xi^2 L_{\mathrm{D}}^2} (x_j - x_0)|_{j=1};$$

- latchup switching time of the $j = 1$ circuit:

$$(t_{\mathrm{pnpn}})_1 = \left\{ \frac{(W_{\mathrm{bp}})^2}{2D_{\mathrm{p}}} + \frac{(W_{\mathrm{bn}})^2}{2D_{\mathrm{n}}} \right\}_{j=1};$$

- time to transport from the $j = 1$ to the $j = 2$ circuit:

$$t_j - t_1|_{j=2} = \frac{\tau}{4\xi^2 L_{\mathrm{D}}^2} (x_j - x_1)|_{j=2}.$$

The secondary current source is also then added to the primary source in the stability of the sequence to the additional second $j = 2$ circuit, associated with the current observed at the $j = 2$ location,

associated with the injection of the primary source, and the $j = 1$ substrate current injection at the location of the $j = 2$ circuit,

$$(I_{\mathrm{inj}}(x_2, t : x_0, t_0))_2 = [I_{\mathrm{inj}}(x_2 - x_0, t)] + \sum_{j=1}^{j=1} [I_{\mathrm{s}x}(x_2 - x_j, t_2 - t_j)].$$

But note that the same diffusion process and solution occur for the secondary sources but at a new place in time and space. This can be generalized to the $j = n$th circuit that undergoes latchup in the form of

$$(I_{\mathrm{inj}}(x_n, t : x_0, t_0))_n = [I_{\mathrm{inj}}(x_n - x_0, t)] + \sum_{j=1}^{j=n-1} [I_{\mathrm{s}x}(x_n - x_j, t_n - t_j)],$$

where the total is the primary source at $x = x_0$ and the summation is over the secondary sources of $n - 1$ circuits that have undergone latchup at the locations $x = x_j$. The secondary injected current source is from the holding current of the latched circuit, which is then used as the initial condition. (*Note*: Current is injected into substrate only.) The solution is used as the initial condition, and the space and time propagation of the secondary source is proportional to the derivative of the error function for a step impulse $u(t - t_0)$ at a displaced time $t = t_j$ for the time of the latched state to occur. Hence, for a given circuit $j$th element that has undergone a latched state, for a specific well and contact spacing, the holding current relationship is

$$(I_{\mathrm{H}})_j = \frac{\beta_{\mathrm{p}}(\beta_{\mathrm{n}} + 1)I_{\mathrm{w}} + (\beta_{\mathrm{p}} + 1)\beta_{\mathrm{n}}I_{\mathrm{s}x}}{\beta_{\mathrm{p}}\beta_{\mathrm{n}} - 1},$$

$$(I_{\mathrm{w}})_j = \frac{(V_{\mathrm{BE}})_{\mathrm{pnp}}}{(R_{\mathrm{w}})_j} = \frac{V_0}{R_{\mathrm{w}}} \ln\left[\frac{I - I_{\mathrm{w}}}{(I_0)_{\mathrm{p}}}\right],$$

$$(I_{\mathrm{s}x})_j = \frac{(V_{\mathrm{BE}})_{\mathrm{npn}}}{(R_{\mathrm{s}x})_j} = \frac{V_0}{R_{\mathrm{s}x}} \ln\left[\frac{I - I_{\mathrm{s}x}}{(I_0)_{\mathrm{n}}}\right].$$

## 3.14 ALPHA PARTICLE INDUCED LATCHUP

When the alpha particle traverses silicon, energy is absorbed by silicon, leading to the generation of electron–hole pairs (EHPs) along the alpha particle track. In semiconductor materials, the amount of energy required to generate an EHP is typically three to four times the material band gap. In silicon, one EHP is formed for every 3.5 eV. As the alpha particle loses energy, the EHP generation increases. At the beginning of its track, it generates approximately on the order of 5000 EHPs, which increases to 15 000 EHPs at the end of its track. The range of the alpha particle in silicon can be expressed as

$$R(E) = 1.53\, E^{1.67},$$

where $R(E)$ is the range in microns and $E$ is the energy in MeV. The energy range of alpha particle in the uranium and thorium spectra is less than 8.8 MeV. An 8.8-MeV alpha particle can traverse approximately 80 μm in a silicon wafer. The amount of charge can be calculated from the energy of the alpha particle. Since it is known that a single electron equals $1.6 \times 10^{-19}$ fC, an alpha particle can generate on the order of 400 fC [76–83].

### 3.14.1 Charge Collection Process and Analysis

As the EHPs are generated, they randomly diffuse from the particle track in the silicon. EHPs can recombine or become collected at metallurgical junctions. Figure 3.11 shows an example of an alpha particle track penetrating the silicon devices, highlighting the EHP generation.

Alpha particles can initiate latchup if the charge collected within the CMOS pnpn structure is enough to switch the parasitic pnpn state. Electron–hole pairs within the regenerative feedback loop can lead to switching; EHPs can be generated in the emitter, base and collector regions of the pnpn; this includes the p− substrate, the n-well, the p+ source/drain diffusion and the n+ source/drain diffusion. For the case of an EHP generated in the substrate, electron–hole pairs that diffuse to the n-well-to-substrate metallurgical junction lead to the electron being collected in the n-well electrical node, leaving the hole behind in the substrate (Figure 3.12). From a circuit perspective, the electron is pulling the n-well 'low' (e.g. toward logical '0') and the hole is raising the p− substrate 'high' (toward logical '1') (Figure 3.13). In the second case, electron–hole pairs that diffuse to the n+ diffusion-to-substrate metallurgical junction lead to electron being collected by the pnpn cathode and the hole remains in the local substrate region. From a circuit perspective, the electron is pulling the cathode 'low' (e.g. toward logical '0') and the hole is raising the p− substrate 'high' (toward logical '1').

For the case of an EHP generated in the n-well, electron–hole pairs that diffuse to the n-well-to-substrate metallurgical junction lead to the hole being collected in the n-well electrical node, leaving the electron behind in the n-well. From a circuit or electrical potential perspective, the electron is pulling the n-well 'low' (e.g. toward logical '0') and the hole is raising the p− substrate 'high' (toward logical '1'). For the case of an electron–hole pair generated in the n-well, electron–hole pairs can also diffuse to the p+ diffusion–n-well metallurgical junction; in this case, the hole is collected in the pnpn anode, leaving the electron behind in the n-well. This case pulls the anode potential 'high' and the n-well potential 'low'. As a result, the electrical potential is perturbed in potentially all physical regions of the pnpn from the alpha particle event. For latchup to occur, the charge collected has to be enough to change the state of the pnpn structure.

The time response and the total charge collected are a function of α-particle characteristics as well as the circuit layout that it traverses. Alpha particles are emitted in random angles of incidence relative to the silicon surface. As a result, the collection process of the carriers is a function of the emission energy, initial point of the emission and the angle of incidence. For vertical angles of incidence, the

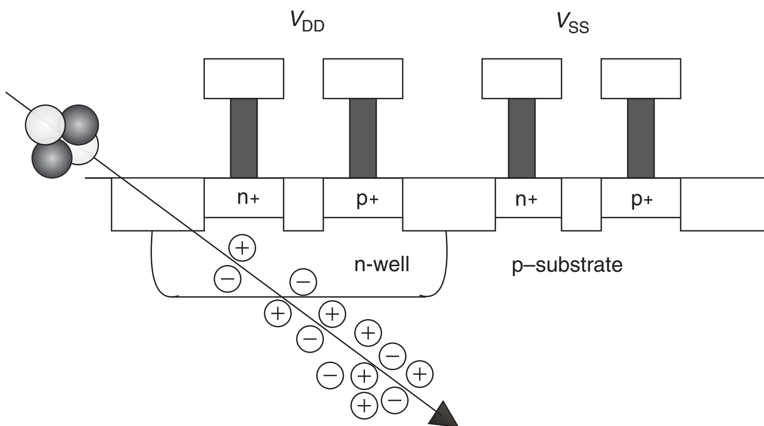



**Figure 3.11** Alpha particle track in a silicon device.

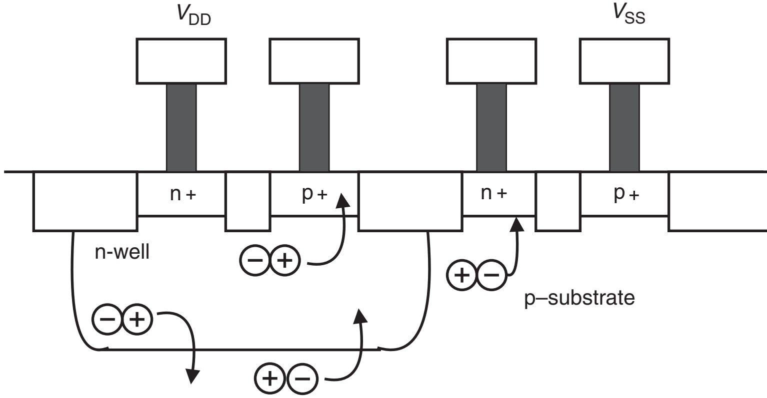



**Figure 3.12** Electron–hole pair charge transfer that influences CMOS latchup.

carrier generation occurs deep in the substrate region. EHPs generated deep in the substrate have a higher probability of recombining or diffusing radially outward. In heavily doped substrates, the electron–hole pairs recombine prior to reaching the silicon surface devices. In lightly doped substrates, the charge generation influences the electrostatic potentials, forming an effect known as the 'funneling effect'. Figure 3.14 shows a sequence of the potential fields as an $\alpha$-particle penetrates vertically through a modern CMOS process. The structure consists of a p+ diffusion, a retrograde n-well and a p+ substrate [75]. These simulation results were achieved using a 3D mesh in a 3D device simulator (FIELDAY) [75].

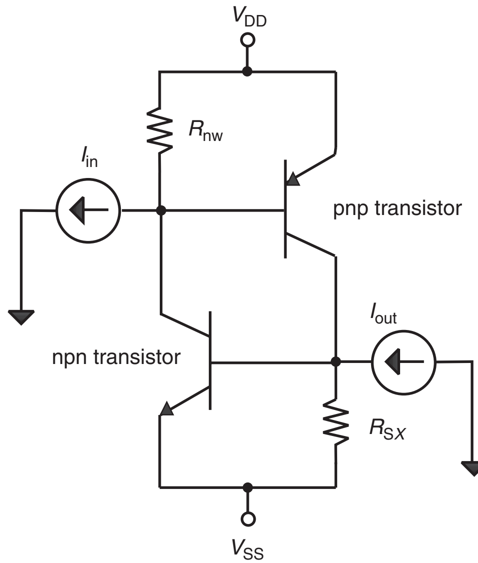



**Figure 3.13** Circuit schematic highlighting the charge transfer processes with electron–hole pair generation.

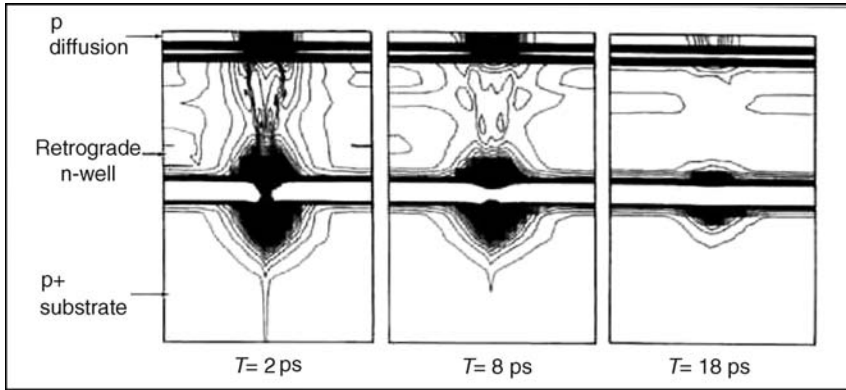



**Figure 3.14** Semiconductor device simulation time sequence of an alpha particle through a CMOS process (e.g. profile is a retrograde n-well and a p++ substrate).

## 3.14.2 Circuit Analysis and Monte Carlo Simulation

To understand the response of the circuit to particle events, circuit simulation can be performed. To simplify the response of a circuit to an alpha particle, the waveform can be modeled as a two-state response. The first form is a rapid response associated with the charge collected in the metallurgical

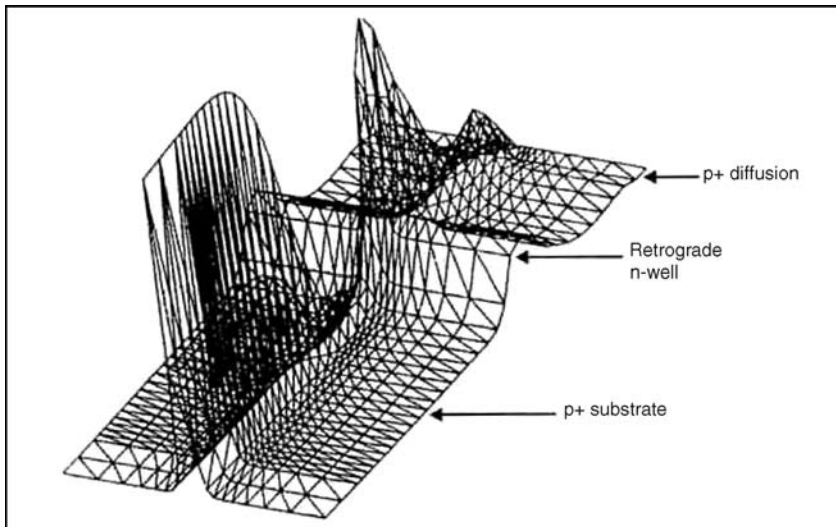



**Figure 3.15** Semiconductor device simulation of the electric potentials as an alpha particle penetrates through a retrograde n-well and p++ substrate region.

junction and the ‘funnel’. Estimates of this time can be obtained analytically or by semiconductor device simulation. The waveform can consist of a rapid pulse, followed by a decay characteristic associated with the diffusion time of the electron–hole pairs. Troutman simplified the problem by assuming that the worst-case alpha particle is a Dirac delta function; in this fashion, analytical circuit analysis using Laplace transforms can be used for the evaluation of the circuit response. To evaluate the statistical probability that latchup can occur from a particle source, Monte Carlo particle simulators can be used for the analysis. Monte Carlo particle simulators can provide the computation of the probability that a given structure collects a given amount of charge for randomly emitted emissions. Given the probability of charge collected is less than the charge to initiate latchup, no latchup will occur. Given the probability of the charge collected exceeds the charge needed to initiate latchup, latchup will occur. Hence, from the cumulative distribution, knowing the needed charge to flip the specific circuit, one can determine the probability of a latchup event from the particle source. Note that this methodology can be applied to other single event effects (SEEs) to evaluate single event latchup (SEL), where the method would be modified to the specific source event interaction and the linear energy transfer (LET). A latchup design practice can be summarized as follows:

- Particle-induced latchup can be evaluated using manual calculations of the amount of charge collected along an particle track, knowing the physical layout of interest.
- Worst-case analysis can be completed by assuming (1) 100 % charge collection along the track (e.g. collected in a critical junction), (2) the complete track fits within a given circuit layout, and (3) collection time is the collection time across a depletion region (e.g. n-well junction) or a Dirac delta function.
- Particle-induced latchup can be evaluated using semiconductor simulators to evaluate the collected charge using semiconductor device simulators.
- Accurate particle analysis that evaluates the statistics of a latchup event can be obtained using Monte Carlo particle random walk simulators.

## 3.14.3 Maximum Collection Evaluation in a Parallelpiped Region

To evaluate the maximum charge collection of a particle through a structure that can initiate latchup, a simple analytical expression can be defined. Assume an analytical relationship of the range of a particle can be expressed as follows:

$$R(E) = a\,E^b,$$

where $R(E)$ is the range of the particle, $E$ is the energy, and $a$ and $b$ are two constant parameters. A region can be represented as a parallelopiped of width $W$, length $L$ and depth $t$. If we assume that the particle traverses the parallelopiped, the largest path through the parallelopiped is the main diagonal. The maximum particle track in a parallelopiped region is then

$$R = \left(W^2 + L^2 + t^2\right)^{1/2}.$$

For a particle whose energy is sufficient to traverse the region, and which ends at the corner of the parallelopiped, we can equate with the range energy equation

$$R = (W^2 + L^2 + t^2)^{1/2} = R(E) = a\,E^b.$$

Solving for the particle energy of this event,

$$E = \left\{ \frac{1}{a} (W^2 + L^2 + t^2)^{1/2} \right\}^{1/b}.$$

Hence, the total energy absorbed in the medium to generate electron–hole pairs is the complete energy of the alpha particle. The number of EHPs, $N_{\text{EHP}}$, can be evaluated assuming it takes ε energy to create one EHP. Then, the number of EHPs in the parallelopiped is

$$N_{\text{EHP}} = n' = p' = \frac{E}{\varepsilon} = \frac{1}{\varepsilon} \left\{ \frac{1}{a} (W^2 + L^2 + t^2)^{1/2} \right\}^{1/b}.$$

The total charge of a given carrier (e.g. either the electron or the hole) can be calculated from the product of the charge per carrier, $q$, and the number of carriers, $N_{\text{EHP}}$:

$$Q = qN_{\text{EHP}} = qn' = qp' = \frac{qE}{\varepsilon} = \frac{q}{\varepsilon} \left\{ \frac{1}{a} (W^2 + L^2 + t^2)^{1/2} \right\}^{1/b}.$$

This chord length analysis can also be applied to other events using the linear energy transfer and the chord length through a collection volume. In this process, the charge in the volume would be associated with the ‘collected charge’ and compared to the ‘critical charge’ to initiate single event latchup.

## 3.15 SUMMARY AND CLOSING COMMENTS

This chapter discussed latchup theory. Different venues of the latchup criterion from a design space perspective are also viewed – resistance space, beta product and alpha sum conditions are visualized by plotting design curves and the state of the network for a given semiconductor process design point and design spacings. This allowed visualization from the theory to graphical methods to apply to characterization techniques.

In Chapter 4, latchup characterization and test structures are addressed. The latchup theory is applied to simple test structures and the characterization techniques incorporate the concepts of the graphical methods.

## PROBLEMS

1. Given a standard CMOS structure, as the well resistance is increased from zero to infinity, show the trajectory of the technology state in $\log(R_{\text{well}})$– $\log(R_{sx})$ plot. Show the trajectory of holding the well resistance constant, as the substrate resistance is increased from zero to infinity. Also show the trajectory in the alpha space plot.
2. Given a standard CMOS process, assume a shallow trench isolation structure is formed. Develop a model of the transport of the lateral and vertical bipolar transistors as a function of the shallow trench isolation depth. Assume the shallow trench isolation is deeper than the junctions, but shallower than the n-well junction. How does the isolation change the state of the technology in alpha and beta spaces as the depth increases? Show the trajectory of states as a function of the shallow trench isolation depth.

3. Given a standard CMOS process, with a dual depth isolation, where the first isolation has the first depth and the second isolation has the second depth, develop the model for the lateral and vertical bipolar transistors as a function of both the first depth and the second depth. Assume the shallow trench isolation is deeper than the junctions, but shallower than the n-well junction. Assume the second isolation depth can extend beyond the n-well-to-substrate junction. How does the isolation change the state of the technology in alpha and beta spaces as the depth increases? Show the trajectory of states as a function of the shallow trench isolation depth and the second isolation.

4. Given a deep trench (DT) isolation that extends beyond the n-well region. Explain what is the influence on the npn and the pnp parasitic transistors.

5. Explain how the silicide film changes CMOS latchup. What does it influence? Show in the equations how it influences the latchup state. How does this change with the silicide sheet resistance? The effective junction depth? Junction scaling?

6. Electron–hole pair recombination influences the bipolar current gain of the parasitic npn and pnp transistors. Show a plot of the Shockley–Read–Hall (SRH) recombination and Auger recombination as a function of doping concentration. At what doping concentration is SRH recombination dominant? At what doping concentration is Auger recombination dominant? Calculate the diffusion length as a function of doping concentration assuming typical silicon diffusion coefficients, and SRH and Auger recombinations.

7. Recombination centers have been formed in CMOS for the purpose of increasing the electron–hole pair recombination to minimize latchup. Gold and other elements are placed in the lattice to reduce the recombination time. The recombination time is a function of the density.

8. Heavily doped buried layers (HBDLs) introduce damage at the end of range of the implant. These damage sites form recombination centers in the lattice. Explain the impacts of the use of an HDBL in a semiconductor process. Show in the equations how the damage sites influence the bipolar gain characteristics, and how the HDBL influences the latchup substrate shunt resistance.

9. Using an HDBL region can influence the substrate resistance and the recombination physics. Derive a model for the substrate and HDBL.

10. HDBL regions can also influence MOSFET snapback. Derive a relationship of MOSFET snapback as a function of the HDBL dose. Assume the substrate and HDBL implant can be modeled as parallel resistor elements. How does this influence ESD protection assuming MOSFET-based ESD devices (e.g. grounded gate NMOS)? How does it affect ESD protection with a diode-based ESD implementation?

## REFERENCES

1. W. Shockley. Theories of high values of alpha for collector contacts on germanium. *Physical Review*, **78**, 1950, 294–295.
2. W. Shockley, M. Sparks and G.K. Teal. P–N junction transistors. *Physical Review*, **83**, 1951, 151–162.
3. J.J. Ebers. Four terminal pnpn transistors. *Proceedings of the IRE*, **40**, 1952, 1361–1364.
4. J.L. Moll, M. Tannenbaum, J.M. Goldey and N. Holonyak. P-N-P-N transistor switches. *Proceedings of the IRE*, **44**, 1956, 1174–1182.
5. I.M. Mackintosh. The electrical characteristics of silicon pnpn switches. *Proceedings of the IRE*, **46**, 1958, 1229–1235.
6. F.E. Gentry, F.W. Gutzwiller, N. Holonyak and E.E. von Zastrow. *Semiconductor Controlled Rectifiers: Principles and Applications of p-n-p-n Devices*. Englewood Cliffs, NJ: Prentice-Hall; 1964.

7. J.F. Gibbons. A critique of the theory of pnpn devices. *IEEE Transactions on Electron Devices*, **11**, 1964, 406–413.
8. J. Lindmayer and C.Y. Wrigley. *Fundamentals of Semiconductor Devices*, Princeton, NJ: Van Nostrand; 1965.
9. I.E. Getreau. *Modeling the Bipolar Transistor*. New York: Elsevier Scientific; 1978.
10. M. Koyomasu, T. Araki, T. Ohtsuki and M. Nakayama. Analysis of latchup phenomena in CMOS IC. *Transaction IECE Japan*, **E61**(2), 1978, 109–110.
11. D.B. Estreich, A. Ochoa Jr. and R.W. Dutton. An analysis of latchup prevention in CMOS IC's using an epitaxial-buried layer process. *International Electron Device Meeting (IEDM) Technical Digest*, December 1978. p. 230–234.
12. A. Ochoa, W. Dawes and D. Estreich. Latchup control in integrated circuits. *IEEE Transaction on Nuclear Science*, **26**, 1979, 5065–5068.
13. D.B. Estreich. *The physics and modeling of latch-up and CMOS integrated circuits*. Technical Report No. G-201-9, Integrated Circuits Laboratory, Stanford Electronic Laboratories, Stanford University, Stanford, CA, November 1980.
14. P.V. Dressendorfer and M.G. Armendariz. A SEM technique for experimentally locating latchup paths in integrated circuits. *IEEE Transactions on Nuclear Science*, **27**, 1980, 1688–1693.
15. J.E. Schroder. A. Ochoa and P.V. Dressendorfer. Latchup elimination in bulk CMOS LSI circuits. *IEEE Transactions on Nuclear Science*, **27**, 1980, p. 1735–1738.
16. R.S. Payne, W.R. Grant and W.J. Bertram. Elimination of latchup in bulk CMOS. *International Electron Device Meeting (IEDM) Technical Digest*, December 1980. p. 248–251.
17. W.D. Raburn. A model for parasitic SCR in bulk CMOS. *International Electron Device Meeting (IEDM) Technical Digest*, 1980. p. 252–255.
18. R.D. Rung, C.J. Delloca and L.G. Walker. A retrograde p-well for higher density CMOS. *IEEE Transactions on Electron Devices*, **28**(10), 1980, 1115–1119.
19. T. Iizuka and J. Moll. A figure of merit for CMOS latchup tolerance. *The 1981 CMOS Workshop*, San Francisco Airport Hilton, San Francisco, May 18, 1981.
20. P.V. Dressendorfer and A. Ochoa. An analysis of modes of operation of parasitic SCR's. *IEEE Transactions on Nuclear Science*, **28**(6), 1981, 4288–4291.
21. A. Ochoa and P.V. Dressendorfer. A discussion of the role of distributed effects in latchup. *IEEE Transactions on Nuclear Science*, **28**(6), 1981, 4292–4294.
22. S.R. Combs. Scaleable retrograde p-well CMOS technology. *International Electron Device Meeting (IEDM) Technical Digest*, December 1981. p. 346–349.
23. A.W. Wieder, C. Werner and J. Harter. Design model for bulk CMOS scaling enabling accurate latchup prediction. *International Electron Device Meeting (IEDM) Technical Digest*, 1981. p. 354–357.
24. G.J. Hu, M.R. Pinto and S. Kordic. Two-dimensional simulation of latchup in CMOS structure. *Device Research Conference Proceedings*, VA-5, June 21–23, 1982.
25. R.D. Rung and H. Momose. Improved modeling of CMOS latchup and VLSI implications. *Symposium on VLSI Technology*, September 1982. p. 52–53.
26. R.R. Troutman and H.P. Zappe. Power-up triggering condition for latchup in bulk CMOS. *1982 Symposium on VLSI Technology*, Oiso, Japan, September 1982.
27. D.E. Nelsen. Modelling latchup triggering in CMOS circuits. *MIT VLSI Seminar*, Massachusetts Institute of Technology, Cambridge, MA, October 19, 1982.
28. C.C. Huang, M.D. Hartanft, N.F. Pu, C. Yue, C. Rahn, J. Schrankler, G.D. Kirchner, F.L. Hampton and T.E. Hendrickson. Characterization of CMOS latchup. *International Electron Device Meeting (IEDM) Technical Digest*, 1982. p. 454–457.
29. R.R. Troutman and H.P. Zappe. A transient analysis of latchup in bulk CMOS. *IEEE Transactions on Electron Devices*, **30**, 1983, 170–179.
30. H.P. Zappe. *A transient analysis and characterization of latchup in bulk CMOS*. BS and MS thesis, Massachusetts Institute of Technology, Cambridge, MA, February 1983.
31. J. Manoliu, F.H. Tseng, B.J. Woo and T.J. Meier. High density and reduced latchup susceptibility CMOS technology for VLSI. *IEEE Electron Device Letters*, **4**, 1983, 233–235.
32. L. Wakeman. Silicon gate CMOS chips gain immunity to SCR latchup. *Electronics*, **56**(16), 1983, 136–140.
33. R.R. Troutman and M.J. Hargrove. Transmission line model for latchup in CMOS circuits. *Symposium on VLSI Technology*, September 13–15, 1983. p. 56–59.

34. U. Schwabe, H. Herbst, E.J. Jacobs and D. Takacs. N- and p-well optimization for high-speed n-epitaxy CMOS circuits. *IEEE Transactions on Electron Devices*, **30**, 1983, 1339–1344.
35. R.R. Troutman. Epitaxial layer enhancement of n-well guard rings for CMOS circuits. *IEEE Electron Device Letters*, **4**, 1983, 438–440.
36. R.D. Rung and H. Momose. DC holding and dynamic triggering characteristics of bulk CMOS latchup, *IEEE Transactions on Electron Devices*, **30**, 1983. 1647–1655.
37. D.L. Wollesen, J. Haskell and J. Yu. N-well and p-well performance comparison. *International Electron Device Meeting (IEDM) Technical Digest*, 1983. p. 31–34.
38. D. Takacs, J. Harter, E.P. Jacobs, C. Werner, U. Schwabe, J. Winnerl and E. Lange. Comparison of latchup in n-well and p-well CMOS circuits. *International Electron Device Meeting (IEDM) Technical Digest*, 1983. p. 159–164.
39. Y. Niitsu, H. Nihara, K. Kanzaki and S. Kohyama. Resistance modulation effect due to current injection and CMOS latchup. *International Electron Device Meeting (IEDM) Technical Digest*, 1983. p. 164–167.
40. G. Goto, H. Takahashi and T. Nakamura. Latchup immunity against noise pulses in a CMOS double well structure. *International Electron Device Meeting (IEDM) Technical Digest*, 1983. p. 168–171.
41. E. Hamdy and A. Mohsen. Characterization and modeling of transient latchup in CHMOS technology. *International Electron Device Meeting (IEDM) Technical Digest*, 1983. p. 172–175.
42. G.J. Hu, Y. Taur, R.H. Dennard, L.M. Terman and C.Y. Ting. A self-aligned 1 μm CMOS technology for VLSI. *International Electron Device Meeting (IEDM) Technical Digest*, 1983. p. 731–734.
43. G.J. Hu. A better understanding of CMOS latchup. *IEEE Transactions on Electron Devices*, **31**, 1984, 62–67.
44. R.C. Fang and J.L. Moll. Latchup model for the parasitic path in bulk CMOS. *IEEE Transactions on Electron Devices*, **31**, 1984, 113–120.
45. J.G. Dooley and R.C. Jaeger. Temperature dependence of latchup in CMOS circuits. *IEEE Electron Device Letters*, **5**, 1984, 41–43.
46. W. Craig. Latchup test structures and their characterization. *IEEE VLSI Workshop on Test Structures*, San Diego, February 20, 1984.
47. D. Takacs, C. Werner, J. Harter and U. Schwabe. Surface induced latchup in VLSI CMOS circuits. *IEEE Transactions on Electron Devices*, **31**, 1984, 279–286.
48. R.R. Troutman and H.P. Zappe. Layout and bias considerations for preventing transiently triggered latchup in CMOS. *IEEE Transactions on Electron Devices*, **31**, 1984, 279–286.
49. I.H. Leventhal. *Comparison of dc latchup characterization techniques for CMOS technology.* BS thesis, Massachusetts Institute of Technology, June 1984.
50. G.J. Hu and R.H. Bruce. A CMOS structure with high holding voltage. *IEEE Electron Device Letters*, **5**, 1984, 211–214.
51. K.W. Terrill and C. Hu. Substrate resistance calculation for latchup modeling. *IEEE Transactions on Electron Devices*, **31**, 1984, 1152–1155.
52. A.G. Lewis. Latchup suppression in fine dimension shallow p-well CMOS circuits. *IEEE Transactions on Electron Devices*, **31**, 1984, 1472–1481.
53. K.W. Terrill, P.F. Byrne, H.P. Zappe, N.W. Cheung and C. Hu. A new method for preventing CMOS latchup. *International Electron Device Meeting (IEDM) Technical Digest*, 1984. p. 406–409.
54. U. Schwabe, E.P. Jacobs, D. Takacs, J. Winnerl and E. Lange. Reduced n+/p+ spacing with high latchup hardness in self-aligned double well CMOS technology. *International Electron Device Meeting (IEDM) Technical Digest*, 1984. p. 410–413.
55. M.R. Pinto and R.W. Dutton. Accurate trigger condition analysis for CMOS latchup. *IEEE Electron Device Letters*, EDL-6 volume 2, 1985, 100–102.
56. K.Y. Fu. Transient latchup in bulk CMOS with a voltage-dependent well–substrate junction capacitance. *IEEE Transactions on Electron Device Letters*, **32**, 1985, 717–720.
57. J.E. Hall, J.A. Seitchik, L.A. Arledge and P. Yang. An improved circuit model for CMOS latchup. *IEEE Electron Device Letters*, **6**, 1985, 320–322.
58. S. Odanaka, M. Wakabayashi and T. Ohzone. The dynamics of latchup turn-on behavior in scaled CMOS. *IEEE Transactions on Electron Devices*, **32**, 1985, 1334–1340.
59. R. Troutman. *Latchup in CMOS Technology: The Problem and the Cure*. New York: Kluwer Academic Press; 1986.
60. JEDEC Standard No. 17. *Latchup in CMOS Integrated Circuits*, August 1988.
61. JEDEC EIA/JESD78. *IC Latchup Test*, March 1997.

62. IEEE Latchup Standards Committee. *Latchup Test Method for Process Characterization*. 1988.
63. E.J. Chwastek. A new method for assessing the susceptibility of CMOS integrated circuits to latchup: the system-transient technique. *Proceedings of the Electrical Overstress/Electrostatic Discharge (EOS/ESD) Symposium*, 1989. p. 149–155.
64. G. Weiss and D. Young. Transient induced latchup testing of CMOS integrated circuits. *Proceedings of the Electrical Overstress/Electrostatic Discharge (EOS/ESD) Symposium*, 1995. p. 194–198.
65. M. Mahanpour and I. Morgan. The correlation between latchup phenomenon and other failure mechanisms. *Proceedings of the Electrical Overstress/Electrostatic Discharge (EOS/ESD) Symposium*, 1995. p. 289–294.
66. R.J. Consiglio. AC and transient latchup characteristics of a twin-well CMOS inverter with load capacitance. *Proceedings of the Electrical Overstress/Electrostatic Discharge (EOS/ESD) Symposium*, 1993. p. 93–101.
67. M.D. Ker and S.F. Hsu. Evaluation on efficient measurement setup for transient-induced latchup with bipolar trigger. *Proceeding of the International Reliability Physics Symposium (IRPS)*, 2005. p. 121–128.
68. M. Kelly, L.G. Henry, J. Barth, G. Weiss, M. Chaine, H. Gieser, D. Bonfert, T. Meuse, V. Gross, C. Hatchard and I. Morgan. Developing a transient induced latchup standard for testing integrated circuits. *Proceedings of the Electrical Overstress/Electrostatic Discharge (EOS/ESD) Symposium*, 1999. p. 178–189.
69. I. Morgan, M. Mahanpour and C. Hatchard. Transient latchup using an improved bipolar trigger. *Proceedings of the Electrical Overstress/Electrostatic Discharge (EOS/ESD) Symposium*, 1999. p. 190–202.
70. K. Domanski, S. Bargstadt-Franke, W. Stadler, M. Streibl, G. Steckert and W. Bala. Transient-latchup failure analysis of the integrated circuits, methods of investigation and computer aided simulation. *Proceeding of the International Reliability Physics Symposium (IRPS)*, 2004. p. 370–374.
71. S. Voldman. *ESD and latchup in advanced technologies*. Tutorial, Tutorial Notes of the International Reliability Physics Symposium (IRPS), April 25, 2004.
72. S. Voldman. *Latchup physics and design*. Tutorial Notes, ESD Tutorials of the EOS/ESD Symposium, September 20, 2004.
73. S. Voldman. Cable discharge event and CMOS latchup. *Proceedings of the Taiwan ESD Conference (T-ESDC)*, 2005. p. 63–68.
74. S. Voldman. Latchup and the domino effect. *Proceedings of the International Reliability Physics Symposium (IRPS)*, 2005. p. 145–156.
75. G.A. Sai-Halasz and M.R. Wordeman. Monte Carlo modeling of the transport of ionizing radiation created carriers in integrated circuits. *IEEE Electron Device Letters*, **1**(10), 1980, 211–213.
76. C.H. Hsieh, P.C. Murley and R.R. O'Brien, A field-funneling effect on the collection of alpha particle generated carriers in silicon devices. *IEEE Electron Device Letters*, vol. **2**, 1981, 103–105.
77. C. Hu. Alpha particle-induced field and enhanced collection of carriers. *IEEE Transactions on Electron Letters*, **3**(2) 1982, 31–34.
78. G. Sai-Halasz, M.R. Wordeman and R.H. Dennard. Alpha-particle-induced soft error rate in VLSI circuits. *IEEE Transactions on Electron Devices*, **29**, 1982, 725–731.
79. G. Sai-Halasz and D.D. Tang. Soft errors rates in static bipolar RAMs. *International Electron Device Meeting (IEDM) Technical Digest*, Session 20.7, 1983. p. 344–347.
80. S. Voldman and L. Patrick. Alpha particle induced single event upsets in bipolar static emitter coupled logic (ECL) cells. *IEEE Nuclear and Space Radiation Conference*, Boulder, CO, 1983.
81. S. Voldman and L. Patrick. Alpha particle induced single event upsets in bipolar static emitter coupled logic (ECL) cells. Special issue on 1984 Annual Conference on Nuclear and Space Radiation Effects. *IEEE Transactions on Nuclear Science*, **31**(6), 1984, 1196–1200.
82. S. Voldman, L. Patrick and D. Wong. Alpha particle effects on bipolar emitter-coupled logic static arrays. *Proceedings of the International Solid State Circuit Conference (ISSCC)*, Vol. XXVII, 1985. p. 262–263.
83. S. Voldman, P. Corson, L. Patrick, K. Nguyen, L. Gilbert, R. Goodwin, T. Maffit and S. Murphy. CMOS SRAM alpha particle modeling and experimental results. *International Electron Device Meeting (IEDM) Technical Digest*, Session 20.7, 1987. p. 518–521.

# 4 Latchup Structures, Characterization and Test

The analysis of the characterization of latchup involves test structures, circuits, I/O designs, semiconductor packaged products and systems. Latchup is unique in semiconductor development since academic test structures are used to characterize the semiconductor technology [1–31]. For test structures, there are two classes of structures. The latchup testing design practice constructs and evaluates 'pnpn' test structures as basic benchmark test structures [8–30]. Evaluation of the 'worst-case' pnpn structures allows quantification of the worst-case latchup parasitic and is used to develop the semiconductor technology design rules. Latchup is quantified for dc, ac and transient phenomena [31–47]. A latchup design practice utilizes many different failure analysis techniques to visualize the interactions and propagation of latchup in a semiconductor chip environment [54–81].

Additionally, the latchup testing practice also constructs and evaluates 'guard ring' structures [1–7]. By evaluating the guard rings, both the semiconductor technology and the ability to prevent interaction between devices, circuits and chip-level subfunctions are evaluated. At the product level, fully integrated designs 'I/O books' that contain the peripheral off-chip drivers, receivers, ESD networks and guard rings are also evaluated. In addition, the full product is tested on commercial testers to evaluate satisfaction of and compliance with the latchup standard specification. To have a full appreciation of what the test structures represent, the role of guard rings and how they integrate into the test structure will be discussed. From this, the decisions of the type of pnpn structures formed, as well as the decisions of guard ring will be clear. We will first discuss the guard rings then return to the discussion on the pnpn latchup structures and characterization, and then focus on the independent guard ring test structures and characterization.

## 4.1 GUARD RINGS

Guard rings are essential for design integration of digital, analog and radio frequency (RF) application circuitry where semiconductor devices and circuits need to be electrically isolated from adjacent circuits, noise and CMOS latchup concerns. Guard rings serve the purpose of providing both electrical


© 2007 John Wiley & Sons, Ltd

and spatial isolations between adjacent circuit elements and decreasing the risk of intradevice, intracircuit and inter-circuit interactions.

Two types of guard rings are used in the semiconductor industry for improving in latchup. Guard rings can be segmented into two different classes [1–30, 48–53]:

- The first kind of guard ring, an integrated guard ring, is a shape or structure on the perimeter or edge of another region serving as a border to change the characteristics into or out of the region.
- The second kind of guard ring is a separate ring structure that is separate from another region (e.g. not on the perimeter or border, but a different physical shape and region outside the perimeter of the region).

An example of the first kind of guard ring is an n+ diffusion that can be placed to border a region on another n-type region, changing the transport in or out of the n-type region [2,28–30]. As a second example, a trench region can be placed on the border of a well or subcollector edge as will be shown [21–26]. In this first kind of guard ring, the guard ring can also be interwoven in the interior of the region as well. The guard ring can be placed around a single shape or plurality of shapes. For example, each p+ diffusion within a well is surrounded by an n+ ring; this serves different purposes of:

- providing a bounded region for each physical shape minimizing lateral current flow;
- providing a low-resistance region around each physical shape;
- preventing local interaction between each physical shape in the region.

In this type of integrated guard ring structure, the space between the physical shape and the edge is also increased by placement of a guard ring on the edge of each perimeter of the region. This increases the physical spacing between the structure and the edge of the region (e.g. placement of an n+ guard ring on the edge of the n-well region, increases the space between the p+ diffusion region and the n-well edge). In this case, the guard ring is contained within the pnpn structure, as is key to the understanding of its role in 'internal latchup'. These implant structures were important in earlier generations of CMOS technology where these p+ and n+ guard rings were deeper than the isolation structure (e.g. recessed oxide and LOCOS isolation generations). Today, LOCOS isolation is still being practiced in smart power and high-voltage CMOS technologies, where they still provide value for latchup robustness at the semiconductor wafer surface. Trench structures used in BiCMOS technology, trench DRAM and power technologies can be placed on the perimeter of an n-well or subcollector providing improved latchup robustness. We can distinguish a first guard ring type as a guard ring contained within the pnpn structure (e.g. for the n+ and trench contained within or bordering the n-well region).

The second kind of guard ring is a separate ring structure that is separate from another region (e.g. not on the perimeter or border but a different physical shape and region outside the perimeter of the region). In this case, the guard rings are separate physical domains. For example, an n-well guard ring can be placed around each n-well that contains p-type devices. In this fashion, the separate n-well isolates the p+ devices in the n-well and the n-well itself from interaction with other devices. In this case, the guard ring will be outside a pnpn structure, and this is the key to the understanding and quantification of 'external latchup' interactions.

Guard ring structures can consist of 'passive' or 'active' guard ring structures. With high-level integration of system-on-chip (SOC) and nonnative voltages, efforts have significantly increased on the usage of new guard ring concepts. Examples of passive guard rings are as follows:

- n-well guard rings;
- n-type triple-well implant;

- n-well and triple-well implant;
- n-well and subcollector;
- trench isolation (TI) guard rings;
- deep trench (DT) guard rings;
- plurality of DT guard rings.

These passive guard ring structures can also be integrated with process solutions to influence the guard ring characteristics. The following concepts are used:

- p− epitaxy/p++ substrate step–junction integrated with n-well/n+ buried layer guard ring.
- p− epitaxy/p++ substrate step–junction integrated with deep trench array.

With the use of $p-/p++$ substrate wafer, the lateral current is reflected upward, enhancing the guard ring collection of an n-well region. A second effect is that the lateral $\beta_{npn}$ roll-off can occur at high currents (e.g. the lateral bipolar current gain is a function of the collector current).

Typically, passive guard ring are biased on the power supply voltage, $V_{DD}$. Another class of guard rings are 'active guard rings'. Active guard rings are electrically connected to a 'soft ground' or an electrical circuitry to influence the minority carrier transport or substrate potential. Examples of 'active' guard ring concepts are as follows:

- n-Well structure electrically connected to a 'soft ground' p+ substrate contact to lower the substrate potential locally to prevent local forward biasing or influence the local potential.
- n-Well structure electrically connected to a 'soft ground' p+ substrate contact to induce a lateral electric field, providing 'electric field assist' opposite to the lateral current flow in a parasitic lateral bipolar transistor (e.g. see Chapter 3).
- First and second substrate contacts electrically connected to an input and output of an inverter circuit to 'invert' the polarity of the injection phenomenon.

## 4.1.1 Intradevice Integrated Guard Ring

Given a p-type semiconductor device in an n-type well, an n+ ring can be placed inside the perimeter of the well region. A second means of placement is to have the outer edge of the n+ ring at the edge of the n-well region. A third means of placement is to have the outside edge of the n+ region extend past the n-well edge. In a multiple-domain p-type device, the n+ shape can be placed between different p− shapes, either allowing for electrical isolation between the p-regions or serving as a low-resistance contact shunt.

## 4.1.2 Intracircuit Guard Ring

Guard rings are placed between individual semiconductor device elements in a circuit where interaction is not desirable. For example, it is common to have guard rings between p- and n-channel transistors in an inverter circuit to avoid latchup between the two transistors. Guard rings are placed

between the two transistors to avoid formation of a parasitic pnpn (e.g. a cross-coupled pnp and npn parasitics formed in the bulk substrate and well regions).

### 4.1.3 Intercircuit Guard Ring

Guard rings are placed between circuits within a common network. For example, in a CMOS I/O circuit, guard rings exist between the off-chip driver output stage, its pre-drive circuitry, the receiver, its ballast resistors and the electrostatic discharge (ESD) networks [6, 7]. Guard rings are used within the ESD device itself and in all structures connected to the signal pad.

Guard rings can also be placed between different circuit functions. Guard rings are placed between mixed voltage application domains, such as array or peripheral I/O circuit blocks, core logic and memory on a common chip. In mixed voltage chips, it is desirable for higher voltage regions to be electrically and spatially isolated from the low-voltage and core-voltage regions. These higher voltage chip domains may require different design 'groundrules'.

### 4.1.4 Interchip Sector Guard Ring

Guard rings are placed between digital, analog and RF chip sectors in a mixed signal chip and SOC applications to minimize noise injection of the digital domain on the analog or RF circuitry. In a mixed signal environment, it is not desirable to have the analog and RF circuits near the digital logic networks that are injecting minority carriers into the substrate. The injection of minority carriers can influence the substrate potential of the bulk locally or globally.

High-voltage applications also exist that integrate power devices with standard digital and analog circuitries. In smart power technology, low-voltage CMOS circuits (e.g. 0–5 V power supplies) and high-voltage power devices (20–120 V power supplies) are contained on a common substrate wafer. The high-voltage power devices can be power transistors, such as lateral diffused MOSFETs (LDMOS), drain-extended NMOS (DEMOS) and insulated gate bipolar trench (IGBT) devices. During inductive load switching transitions, large power devices can inject electrons into the chip substrate. These smart power devices must be adequately isolated from low-voltage CMOS circuits to prevent external latchup. Guard rings are placed between the high-voltage power device sector and the CMOS low-voltage circuitry. These guard ring structures can be both 'passive' and 'active'.

### 4.1.5 ESD Guard Rings – Usage of Guard Rings in ESD Protection

In ESD design, guard rings are used to isolate the ESD network or can serve as part of the ESD strategy [4]. In some cases, the guard rings are placed between the ESD circuit elements to prevent their interaction. In other cases, the guard rings can serve as current-collecting structures and as a means of sinking electrical overstress (EOS) events, transient phenomena, ESD events and other voltage or current disturbances. At the same time, the guard rings are being used as a means to discharge current from the n-well diode to the guard ring structure itself. These are best evaluated using ESD and latchup testing of the specific structure.

## 4.2 LATCHUP CHARACTERIZATION STRUCTURES – SINGLE- AND DUAL-WELL CMOS PNPN TEST STRUCTURES

For CMOS latchup characterization, it is a common latchup design practice to utilize a four-stripe test structure for evaluation of the latchup sensitivity of a semiconductor technology [2]. Use of

standardised simple latchup structures allows the ability to compare the latchup robustness of a technology; these standardized test structures can be used as a methodology for a latchup technology benchmarking practice and quantify scaling implications. As a result, it is very important to establish a latchup test structure strategy that addresses both issues.

### 4.2.1 Basic pnpn Structure with Single Substrate Contact Spacing

A CMOS latchup design practice for test structures is as follows [2,28–30]:

- A single-stripe p+ diffusion is placed into an n-well region.
- A single-stripe n+ diffusion is placed into a substrate region.
- A single-stripe n+ diffusion is placed into an n-well region to serve as the n-well contact.
- A single-stripe p+ diffusion is placed into a substrate region to serve as the substrate contact.

Figure 4.1 shows the cross section of the four-stripe structure. Figure 4.2 shows the layout design of the four-stripe structure. The four-stripe structure represents a CMOS inverter circuit at the edge of n-well and p-well regions that is most vulnerable to latchup events. The four-stripe structure also represents the 'worst-case' structure allowed in a given technology. In Chapter 5, these structures will be used to characterize the CMOS process technology. In the layout of the design, the n-well contact and the p− substrate contact are placed on the outer edges of the physical design. In this physical design, a single-well contact and a single-substrate contact are placed adjacent to the structure. The lateral pnp transistor is represented by the p+ diffusion, the n-well and the p− substrate region. The lateral npn transistor is represented by the n+ diffusion, the p− substrate and the n-well. The base width of the pnp transistor is defined as the spacing between the p+ diffusion and the n-well edge. The base width of the npn transistor is defined as the spacing between the n+ diffusion and the n-well edge. The minimum 'p+/n+ spacing' is the sum of the two base widths. In order for this structure to represent the 'worst-case pnpn structure' allowed in the technology design rules, it is a CMOS latchup test site design practice to provide the following:

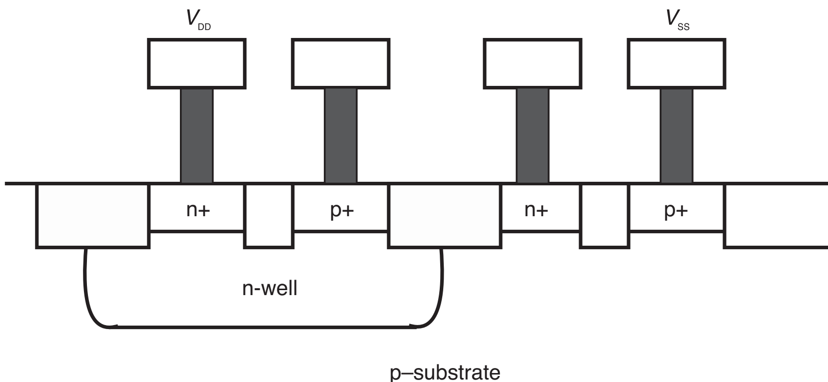



**Figure 4.1** Four-stripe CMOS latchup test structure cross section.

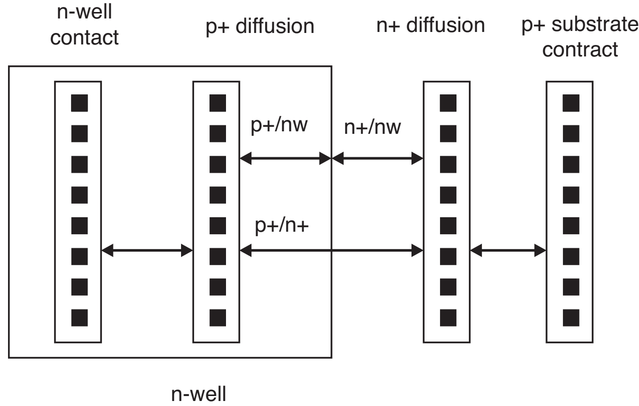



**Figure 4.2** Four-stripe CMOS latchup test structure.

- At least one pnpn test structure is constructed with the minimum ‘p+/n+ spacing’.
- At least one pnpn test structure is constructed with the maximum-allowed well and substrate spacings to verify the design rules.
- At least one pnpn test structure is constructed below the minimum ‘p+/n spacing’ to evaluate the technology variations, margin and sensitivity below the acceptable ground rules.

In the physical design of the test structure, it is important to place full contact density along the length of the test structure with a minimum contact-to-contact spacing. In addition, the wiring must not induce lateral voltage drops across the test structure. As a result, it is important to have wide metal wires entering the test structure, so the metal design does not introduce nonuniform current flow, voltage drops or metal failure prior to latchup characterization.

## 4.2.2 Basic pnpn Structure with Internal Well Guard Ring Structure

An alternative latchup test structure places the n+ substrate contact between the p+ diffusion and the n-well edge (within the n-well) [2, 28–30]. Figures 4.3 and 4.4 show the cross section and the design layout, respectively. Troutman constructed this additional structure are part of the set of benchmark structures for evaluation of the effectiveness of placing an n+ diffusion between the p+ diffusion and the n-well to evaluate the effectiveness of ‘n+ guard ring’ structures for the n-well region. A CMOS latchup design practice for test structures is the following test structure:

- A single-stripe p+ diffusion is placed into an n-well region.
- A single-stripe n+ diffusion is placed into a substrate region.
- A single-stripe n+ diffusion is placed into an n-well region to serve as the n-well contact, where the n+ diffusion is placed between the p+ diffusion (in the n-well) and the n-well edge.
- A single-stripe p+ diffusion is placed into a substrate region to serve as the substrate contact.

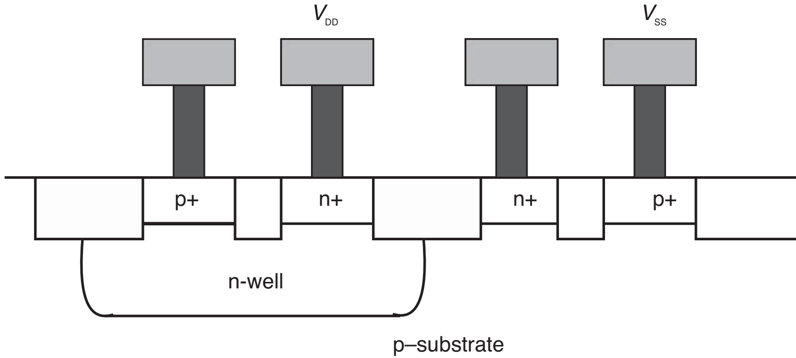



**Figure 4.3** Four-stripe CMOS latchup test structure cross section with internal n-well contact.

This four-stripe structure represents a CMOS inverter circuit at the edge of n-well and p-well regions where the n+ well contact is placed between the p-channel MOSFET and the n-channel MOSFET. In the layout of the design, the p- substrate (or p-well contact) is placed on the outer edges of the physical design. The lateral pnp transistor is represented by the p+ diffusion, the n-well and the p-substrate region. The lateral npn transistor is represented by the n+ diffusion, the p− substrate and the n-well. The base width of the pnp transistor is defined as the spacing between the p+ diffusion and the n-well edge, where the n+ diffusion is placed between the p+ diffusion and the n-well edge. The placement of the n+ diffusion well contact between the p+ diffusion and the n-well creates a larger space (this leads to a non-minimum 'p+/n+ spacing'. It is a CMOS latchup test site design practice to provide the following:

- At least one pnpn test structure is constructed with the minimum 'p+/n+ spacing' under the condition that the n+ diffusion n-well contact dimensions and all other dimensions are minimum.

In a LOCOS technology, the placement of the n+ diffusion between the p+ diffusion and the n-well lowers the pnp lateral bipolar transistor current gain. For LOCOS isolation, it introduces a region of higher minority bulk recombination and higher minority carrier surface recombination, eliminating

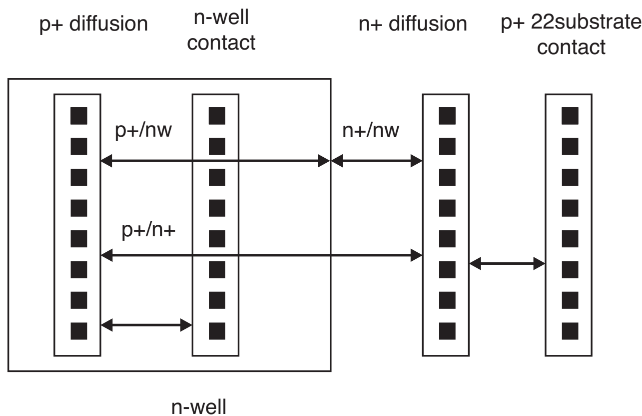



**Figure 4.4** Four-stripe CMOS latchup test structure layout with internal n-well contact.

latchup surface effects. In addition, due to the ground rules and physical design dimensions, it also increases the base width of the lateral pnp BJT. In addition, this also provides the closest contact possible spacing of the n+ well contact. Hence, this structure evaluates the worst-case pnpn structures with the minimum or 'best-case' well contact placement. Today, in shallow trench isolation (STI), the n+ diffusion does not interfere with the diffusion of carriers from the p+ diffusion and n-well since it is shallower than the isolation structure itself; hence, it serves in lowering the well series resistance and widening the base width.

In practice, as discussed with the internal substrate contact, one of the primary reasons this layout design practice is not followed in CMOS technology for inverter gate design, is that the polysilicon MOSFET gate structure crosses perpendicular to the n-well to p− substrate metallurgical boundary. Hence, the introduction of the n-diffusion contact on the inner edge. In addition, the extra space is not desirable for dense designs. But, for regions of a semiconductor chip that are sensitive to latchup, trading-off area versus CMOS latchup robustness is diserable.

## 4.2.3 Basic pnpn Structure with Internal Substrate Guard Rings

An alternative structure places the p+ substrate contact between the n-well and the n+ diffusion (in the substrate). Figures 4.5 and 4.6 show the cross section and the design layout, respectively. Troutman constructed this benchmark structure for evaluation of the effectiveness of placing a p+ substrate contact between the n+ diffusion and the n-well as a one within a set of test structures to evaluate the minimum sensitivity and effectiveness of 'guard ring' structures. These latchup test structures (LUTS) became part of the CMOS latchup design practice. A CMOS latchup design practice for test structures is the following test structure:

- A p+ diffusion stripe is placed into an n-well region.
- An n+ diffusion stripe is placed into a substrate region.
- An n+ diffusion stripe is placed into an n-well region to serve as the n-well contact.
- A p+ diffusion stripe is placed into a substrate region to serve as the substrate contact, where the p+ diffusion is placed between the n+ diffusion (in the substrate) and the n-well edge.

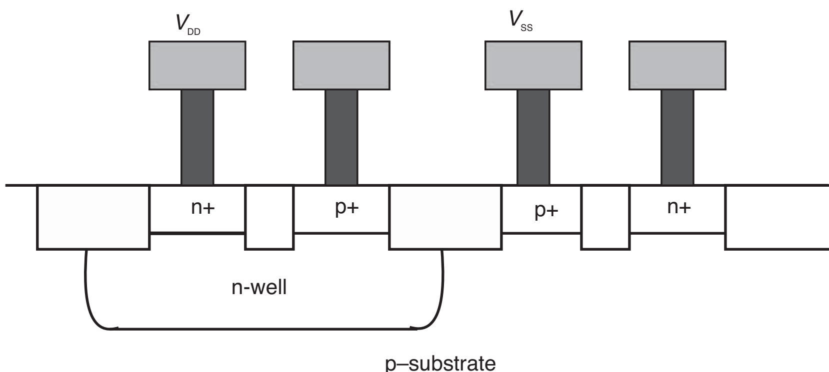



**Figure 4.5** Four-stripe CMOS latchup test structure cross section with internal substrate guard ring.

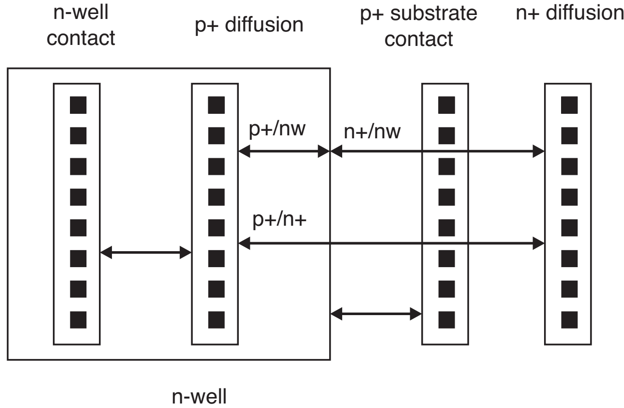



**Figure 4.6** Four-stripe CMOS latchup test structure layout with internal substrate guard ring.

This four-stripe structure represents a CMOS inverter circuit at the edge of n-well and p-well regions where the p+ substrate contact is placed between the p- and n-channel MOSFETs. It is a latchup test site design practice to provide the following:

- At least one pnpn test structure is constructed with the minimum 'p+/n+ spacing' under the condition that the p-diffusion substrate contact dimensions and all other dimensions are minimum.
- At least one pnpn test structure is constructed below the minimum 'p+/n spacing' to evaluate the technology variations, margin and sensitivity below the acceptable ground rules.

In a LOCOS technology, the placement of the p+ diffusion between the n+ diffusion and the n-well lowers the npn lateral bipolar transistor current gain. For LOCOS isolation, it introduces a region of higher bulk minority carrier recombination and higher minority carrier surface recombination; this eliminates latchup surface effects. In addition, due to the ground rules and physical design dimensions, it also increases the base width of the lateral npn. In addition, this also provides the closest contact possible spacing of the p+ substrate contact. Hence, this structure evaluates the worst-case pnpn structures with the minimum or 'best-case' contact placement.

## 4.2.4 Basic pnpn Structure with Internal Well Guard Ring and Substrate Guard Ring

In the set of CMOS latchup benchmark structures, the combination of the p+ substrate contact and the n-well contact can be placed adjacent to the n-well/p-well junction edge. Figures 4.7 and 4.8 show the cross section and the design layout, respectively. In 1982, Troutman constructed this additional structure as a part of the set of benchmark structures for evaluation of the effectiveness of placing an n+ diffusion between the p+ diffusion and the n-well to evaluate the effectiveness of 'n+ guard ring' structures for the n-well region and the effectiveness of placing a 'p+ substrate guard ring' for the p− substrate (e.g. or p− well). A CMOS latchup design practice for test structures is the following test structure:

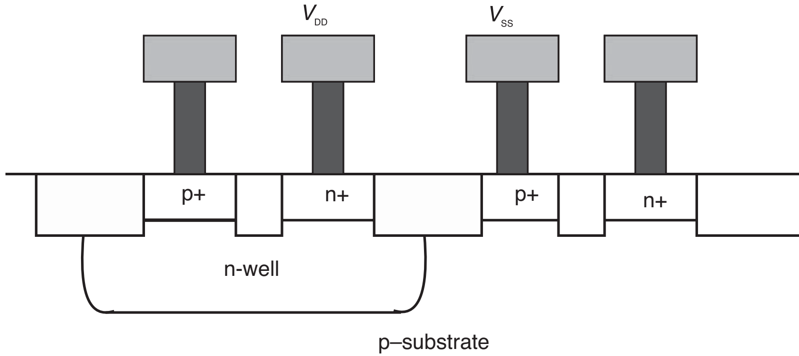



**Figure 4.7** Four-stripe CMOS latchup test structure cross section with internal p-well contact and n-well contact.

- A p+ diffusion stripe is placed into an n-well region.
- A n+ diffusion stripe is placed into a substrate region.
- A n+ diffusion stripe is placed into an n-well region to serve as the n-well contact, where the n+ diffusion is placed between the p+ diffusion (in the n-well) and the n-well edge.
- A p+ diffusion stripe is placed into a substrate region to serve as the substrate contact where the p+ diffusion is placed between the n+ diffusion (in the p-well) and the n-well edge.

It is a CMOS latchup test site design practice to provide the following:

- At least one pnpn test structure is constructed with the minimum 'p+/n+ spacing' under the condition that the n+ diffusion n-well contact and p-well contact dimensions are minimum and contained within the internal area, and all other dimensions are minimum.

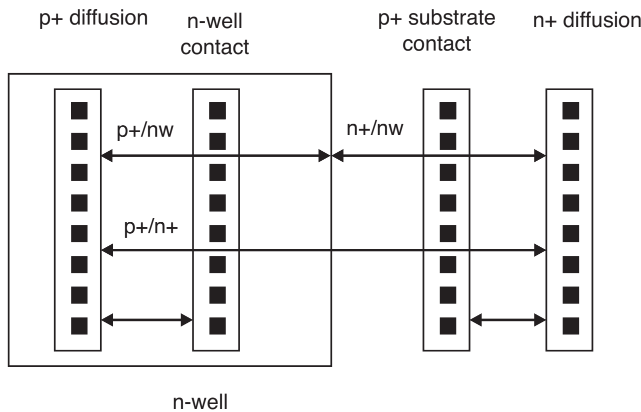



**Figure 4.8** Four-stripe CMOS latchup test structure layout with internal p-well contact and n-well contact.

### 4.2.5 Basic pnpn Structure with Multiple p+/n+ Spacings

For development of the CMOS latchup design rules, it is valuable to have test structures where the p+/n+ spacing is varied. In addition, for the development of the latchup design rules for the substrate and well contacts, it is valuable to have test site structures with a plurality of well and substrate contact spacings. A CMOS latchup design practice for test structures is as follows:

- A plurality of p+ diffusions are placed into a common n-well, where there is an electrical pad connection to each independent p+ diffusion. Each p+ diffusion has a varied spacing relative to the n-well junction.
- A plurality of n+ diffusions are placed into a substrate region. Each n+ diffusion has a varied spacing relative to the n-well to p-well junction.
- The p+/n+ spacings were paired in the test structure in the same dimension.
- A single-stripe n+ diffusion is placed into an n-well region to serve as the n-well contact.
- A single-stripe p+ diffusion is placed into a substrate region to serve as the substrate contact.

Figure 4.9 shows the cross section of the four-stripe structure with the plurality of p+ and n+ stripes. In 1983, Craig developed these CMOS latchup test structures for ground rule development of the p+/n+ spacing requirements. Using an automated test system, the well and substrate resistances can be modified by using external resistances. Using an automated test system developed by Troutman and this multiple p+/n+ stripe test structure, both the design spaces were evaluated. The advantage of this CMOS latchup test structure is the ability to evaluate a large number of p+/n+ spacing combinations in a common test structure (e.g. a $1 \times 25$ pad set test vehicle).

### 4.2.6 Basic pnpn Structure with Multiple Well and Substrate Contact Spacings

It is a common latchup design practice to utilize a four-stripe test structure for the development of the CMOS latchup design rules for the substrate and well contacts, where a plurality of well and substrate contact spacings is established. A CMOS latchup design practice for test structures is as follows:

- A p+ diffusion is placed into a common n-well region, where there is an electrical pad connection to each independent n+ diffusion. Each n+ diffusion has a varied spacing relative to the p+ diffusion.
- An n+ diffusion is placed into a substrate region. Each p+ diffusion substrate contact has a varied spacing relative to the n-well to p-well junction.
- The p+/n+ spaces chosen are minimum spacings.

Figure 4.10 shows the cross section of the four-stripe structure with the plurality of well and substrate contacts, with a fixed p+/n+ space. The advantage of this structure is the ability to provide an actual spacing of the well and contacts without depending on a resistance value. The disadvantage of this method is that only one or two single p+/n+ spacings are evaluated per large test structure in a common test structure (e.g. a $1 \times 25$ pad set test vehicle).

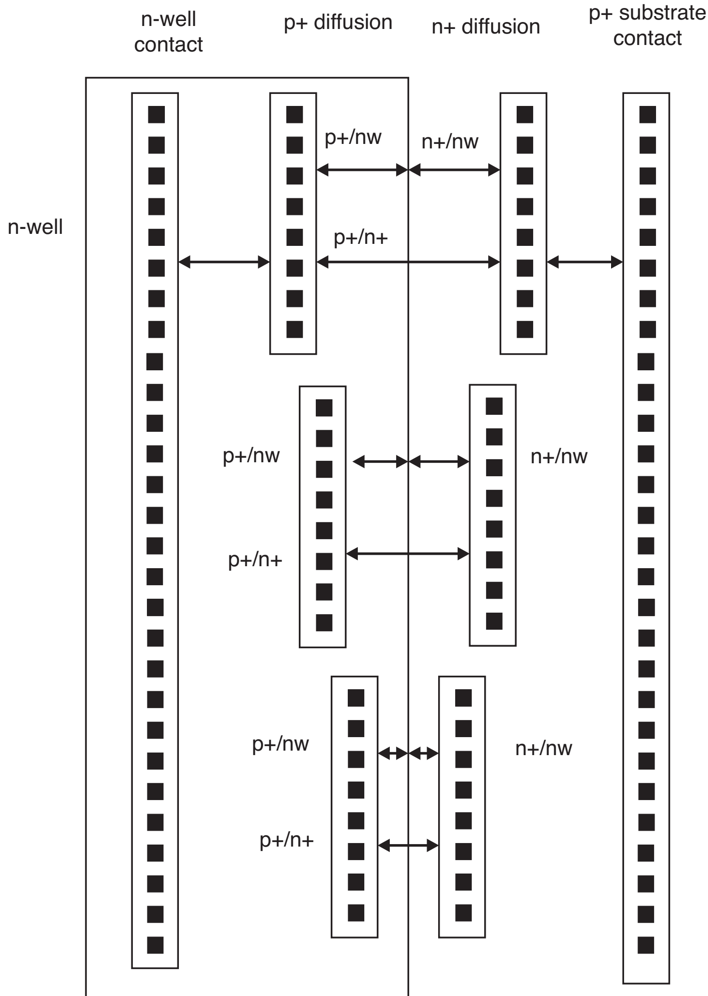



**Figure 4.9** Four-stripe CMOS latchup test structure design layout with a plurality of p+ and n+ spacings.

## 4.2.7 Basic pnpn with Large Substrate Resistance Measurements

In the evaluation of the technology latchup design point, it is important to evaluate the holding voltage for a wide range of substrate and well resistances to evaluate the holding voltage contour in $\log(R_{\text{well}})$ versus $\log(R_{sx})$ design plots. One method of providing that for a four-stripe structure is to add external resistors in the latchup network representing the shunt and well resistance values. In that fashion, a wide range of resistances can be evaluated. The advantage of that method is the ability to sweep a large resistance range and only four test pads are used to achieve it. A second advantage is that the actual resistance values are recorded (e.g. not model-based resistance values). One of the disadvantages is that

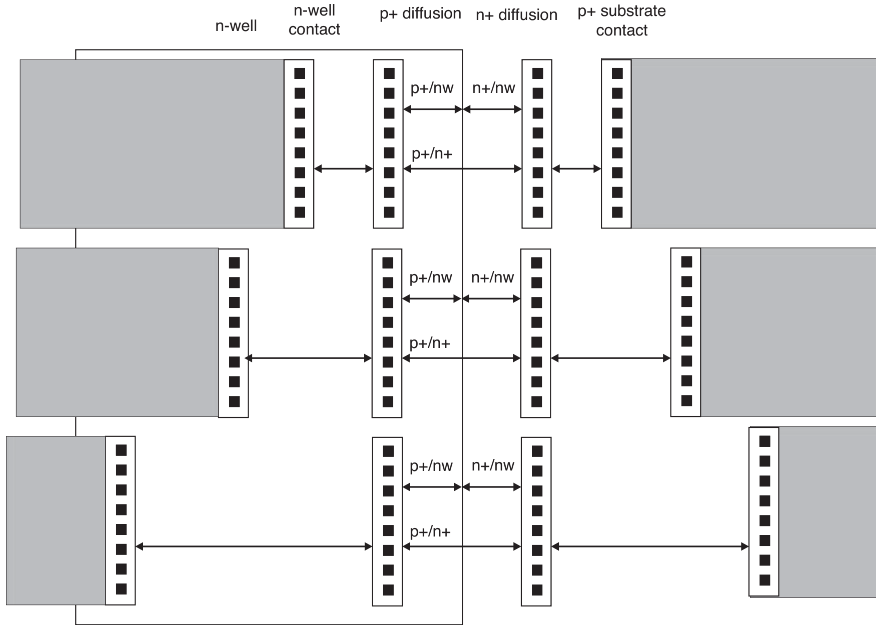



**Figure 4.10** Four-stripe CMOS latchup test structure design layout with a plurality of well and substrate contact spacings.

the measurement is not the actual wafer environment but an electrical element. A semiconductor test method to provide large substrate resistance values is as follows:

- First measurement: Using a four-stripe pnpn test structure, the p+ substrate stripe is used for the first data point for the 'near' substrate contact.
- Second measurement: In the same test macro, there exist multiple pnpn elements, where each element has a p+ substrate contact space. Using the three stripes of the pnpn element and the substrate contact of the second structure set, the p+ substrate spacing is varied.
- Third measurement: Continue this process with the next set of structures, where each time the new p+ substrate contact is used with the three stripes of the pnpn.
- Plot the data as a function of the substrate spacings (or equivalent resistance values).

## 4.2.8 Basic pnpn with External Injection Source

External latchup can occur from an injection source outside the basic pnpn structure. Test structures are needed to evaluate the external latchup injection effect. External latchup from a transient source is a function of the circuit dimensions, the injection source and the relative positioning. External latchup is a function of the following variables:

- the 'victim' circuit;
- injection source structure type;

- injection source size;
- injection source to 'victim' circuit distance;
- relative orientation of the injection source to the 'victim' circuit.

The physical dimensions of the sensitive circuit (e.g. representative pnpn) influence the latchup sensitivity as well as the injector variables. The physical dimensions of the victim circuit influence the stored charge of the pnpn latch circuit (pnpn latch capacitance to switch the circuit). In addition, the larger the circuit, the larger the physical area to 'collect' charge on its n-well junction. As the circuit increases in area, the solid angle formed between the injection source and the victim circuit increases. The simplest case of a latchup victim circuit can be the four-stripe pnpn test structure.

An example of an injection source type that injects electrons into a p- substrate is an n+ diffusion and an n-well diffusion (note that a p+ diffusion can also serve as an injection source to evaluate the effect of vertical pnp substrate injection). In CMOS technology, the injection sources can be as follows:

- n+ diffusions;
- n-type resistors;
- n-wells;
- triple-well n-regions.

With single-, dual- or triple-well processes, the forward biasing on the well structure can lead to electron injection into the p− substrate. These structures can be electrically connected to input nodes or to a power supply rail (e.g. $V_{DD}$). In all cases, these can initiate electron injection into the substrate. In standard CMOS, an n+ diffusion of a MOSFET, an n-type resistor or an ESD network can lead to electron injection. In these cases, the injector is relatively small and has high current density. In a masked triple-well technology, some chip sectors are within triple well and some are within the semiconductor p− substrate (not all MOSFETs are contained within an isolated epitaxial region). For masked triple well, the entire triple-well region can be forward bias injecting into the substrate; in this case, the area is large with a low current density.

In a BiCMOS technology, the injector structures can be as follows:

- n+ diffusion;
- n-wells;
- n++ subcollectors;
- trench-bound n++ subcollectors (e.g. deep trench (DT) and trench isolation (TI)).

In a BiCMOS process, the injection structures can be MOSFETs or bipolar elements. The injection sources can be passive elements, active elements and ESD networks. Bipolar transistor elements and ESD networks can contain either trench-bound and nontrench-bound subcollectors. In the case of power amplifiers, these structures can be large. Negative transients on the $V_{CC}$ power supply can lead to injection.

In smart power applications, in addition, the injector structures can be as follows:

- DeMOS power devices;

- LDMOS power devices;
- IGBT power devices.

In smart power applications, the DeMOS, LDMOS and IGBT devices can lead to injection into the p−substrate. In many power applications, the 'high-voltage' module can be integrated with low-voltage CMOS. In smart power applications, the LDMOS transistors can be a large percentage of an integrated high-voltage system and chip area. With large inductive loads, the switching characteristics can lead to injection into the substrate and the low-voltage CMOS chip section. In all these cases, the injection properties are a function of the physical size of the victim and the injection source and the relative distance and orientation between the two elements. Hence, test structures are need to evaluate these variables.

A first simple means to evaluate the external latchup issue is to utilize a family of basic four-stripe structures. In test sites with $1 \times 25$ macros containing multiple four-stripe pnpn structures, the adjacent structures are used to evaluate both the injection type and spacing. Given six pnpn structures (in a $1 \times 25$ pad set), one of the elements can serve as the injection source. In addition, there exists five different n+ diffusions and n-well shapes; utilizing these adjacent structures, the 'adjacent structures' can serve as injection sources. This provides five injection spacings and two injection types (e.g. n+ diffusion and n-wells). For external latchup characterization, external latchup test structures can also be constructed by designing a plurality of injection shapes and a single pnpn structure (e.g. or any of the 'victim circuits' discussed in this chapter). In a test site, a plurality of injection spacings relative to the victim circuit can be chosen to evaluate the design variable of injection-to-victim (pnpn) spacing.

In these structures, the relative orientation and the spacing is an issue. Figure 4.11 shows an example of a four-stripe pnpn latchup structure and injection source. Hence, test structures can be designed where the orientation of the injection source can be rotated relative to the victim circuit. Note that in this test structure, the total interaction between the two is a function of both the size of the injection source and the victim circuit. First, the distance between the injection source and the 'victim' circuit is defined. The relative orientation angle, $\theta$, between the source and the pnpn structure influences the minority carrier collection at the source location. For every incremental region of the injection source (e.g. $dx$), there is a solid angle, $d\Omega$, of integration to 'victim circuit'. The orientation of the injection source relative to the victim circuit (e.g. the pnpn element) changes the solid angle, $d\Omega$, of the electron collection. To fully evaluate the transfer of carrier from the injector to the 'victim' circuit, a double integration can be performed between the injection area and the collection area. At large distances, these can be reduced to simple solutions (e.g., areas or point source).

Three simple cases can be evaluated (Figure 4.12). Special cases of interest are as follows:

- Parallel injection source (e.g. $\phi = 0$), where injector is on the n-channel MOSFET side.
- Parallel injection source (e.g. $\phi = 180$), where the injector is on the p-channel MOSFET side.
- Perpendicular injection source (e.g. $\phi = 90$), where the injector is broadside to the pnpn structure.

These are the special cases of the general case. Hence, note that these injection test structure concepts can be constructed with all of the structures discussed in Sections 4.1–4.4. One of the problems with the analysis at close distance is that these four-stripe academic circuits do not represent the real or true orientation that will be utilized in an actual ASIC design methodology and only represent the worst-case evaluation at the edge. In an actual ASIC gate array environments, the orientation of the p-channel MOSFET and n-channel MOSFET gate (and hence the MOSFET source/drain diffusions) are perpendicular to the n-well to p-well border. As a result, for better understanding

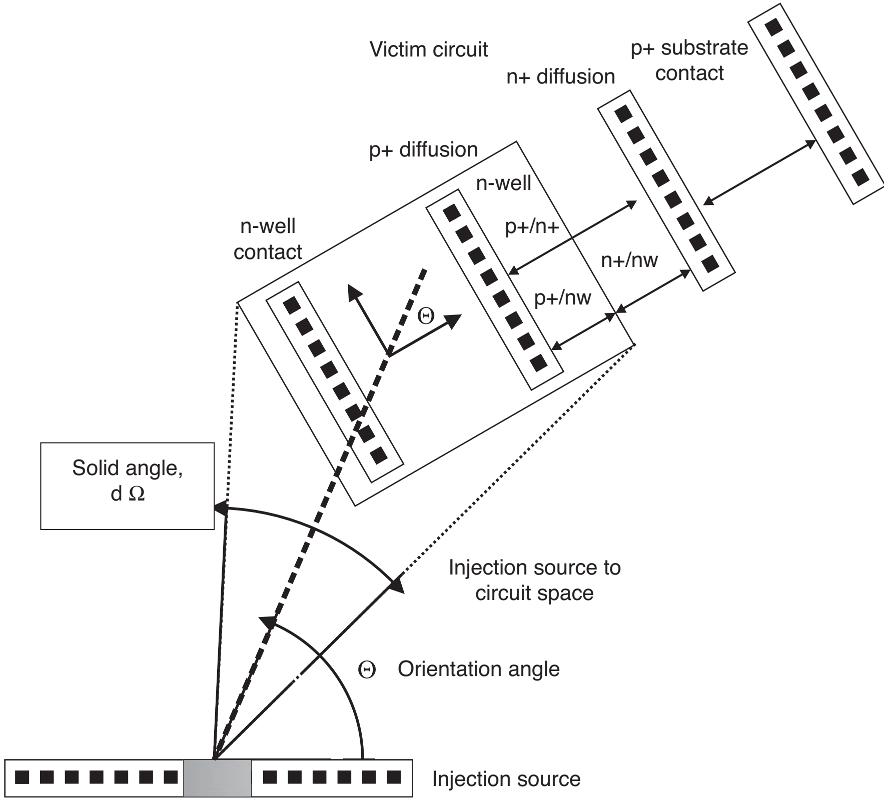



**Figure 4.11** pnpn test structure with injection structure at a relative orientation angle.

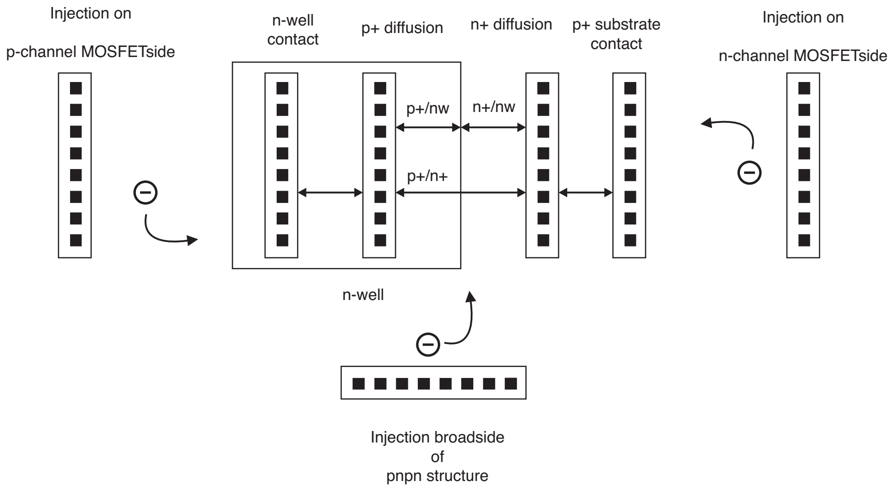



**Figure 4.12** pnpn test structure with injection structure for three special cases of orientation.

of the orientation influence, a test structure with the correct orientation and well sizes may be preferred to evaluate the relative orientation issue.

## 4.3 LATCHUP CHARACTERIZATION – BASIC TRIPLE-WELL pnpn LATCHUP TEST STRUCTURES

For triple-well technology, the structure must be modified to address the extra region, when the triple well forms an isolated p-well epitaxial region (referred to as 'merged triple well'). Voldman constructed triple-well CMOS test structures that would preserve the commonality to the test structures with the following CMOS design practice objectives [25–30]:

- Construct simple characteristic structures that represent the minimum design dimensions.
- Provide structures that represents the allowable triple-well technology semiconductor design rules.
- Provide structure that represent actual design practice of CMOS circuit designs.
- Maintain the CMOS latchup technology benchmarking strategy.
- Allow for comparison of the dual-well and merged triple-well technology.

For merged triple-well test structures, the concept of the four-stripe test structure was preserved within the allowable ground rule design dimensions and modified accordingly. Figures 4.13 and 4.14 show the cross section and the design layout of the five-stripe and the four-stripe triple-well test structures respectively. A CMOS latchup design practice for triple-well test structures is as follows [26]:

- A single-stripe p+ diffusion is placed into an n-well region.
- A single-stripe n+ diffusion is placed into a p-well region.
- A single-stripe n+ diffusion is placed into an n-well region to serve as the n-well contact.
- A single-stripe p+ diffusion is placed into the isolated region to serve as the substrate contact, where the p+ diffusion is placed between the n+ diffusion (in the substrate) and the n-well edge.

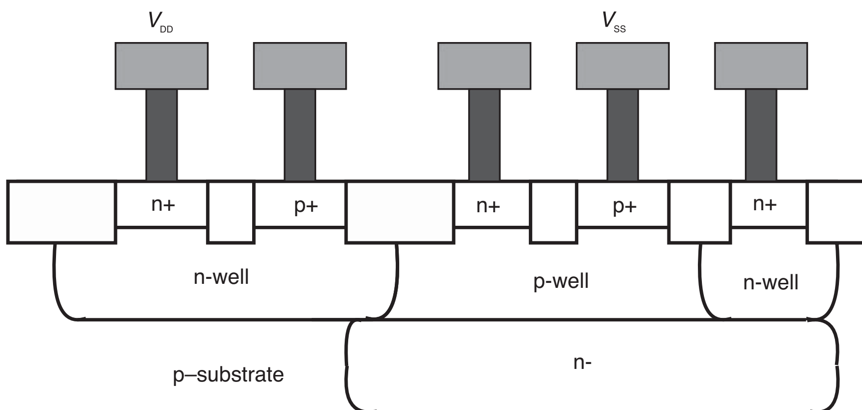



**Figure 4.13** Five-stripe triple-well latchup test structure cross section (substrate stripe not shown).

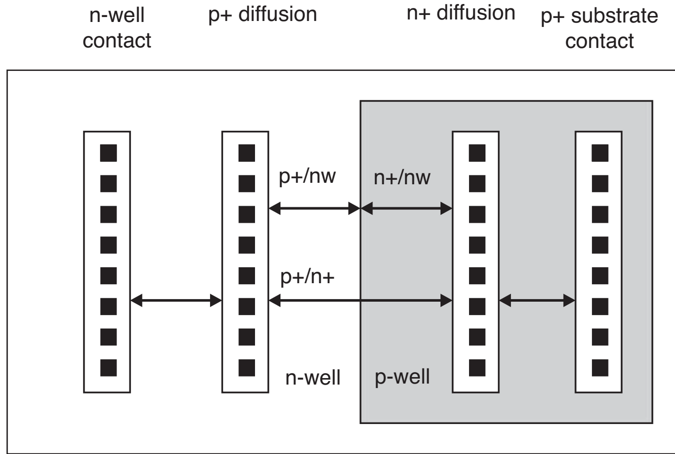



**Figure 4.14** Four-stripe triple-well latchup test structure layout and design (fifth substrate stripe not shown).

- An n-well region extends around the p-well contact stripe and n+ diffusion within the p-well forming a 'ring' around the isolated epitaxial p-well region.
- An implant level is placed under the p-well region to isolate the epitaxial region whose overlap spacings are in conformance to the design layout rules of the technology.
- A fifth stripe p+ diffusion is placed outside the n-well ring to provide electrical contact with the semiconductor chip substrate.

In this structure, the vertical npn transistor is formed between the n+ diffusion, the isolated p-well and the n-well region. In a merged transistor structure, the vertical and lateral npn transistors are not electrically isolated and distinguishable. The pnp transistor base width is defined as the spacing between the p+ diffusion and the n-well edge. The npn transistor base width is defined as the spacing between the n+ diffusion and the n-well edge. In the design of the triple-well n-well ring, it is important that the spacing of the edge of the n+ stripe within the p-well does not introduce a lateral npn (e.g. the spacing between the n-well ring and the n+ stripe should be significantly greater than the spacing between the n-well edge and the n+ stripe). A second CMOS latchup design practice issue is the placement of the triple-well implant. Two different CMOS triple-well bench structures can be constructed [26, 28–30]:

- Buried layer is placed only under the p-well region to electrically connect with the n-well ring (Figure 4.14).
- Buried layer is placed under both the p-well and n-well regions (Figure 4.15).

## 4.4 LATCHUP CHARACTERIZATION TECHNIQUES – pnpn STRUCTURES WITH DEEP TRENCH

For latchup characterization of different semiconductor structures, as shown above, the four-stripe test structure must be modified. For example, DT isolation is a technology feature in CMOS trench DRAM, advanced bipolar and bipolar/CMOS (BiCMOS) technologies, as will be discussed in Chapter 6.

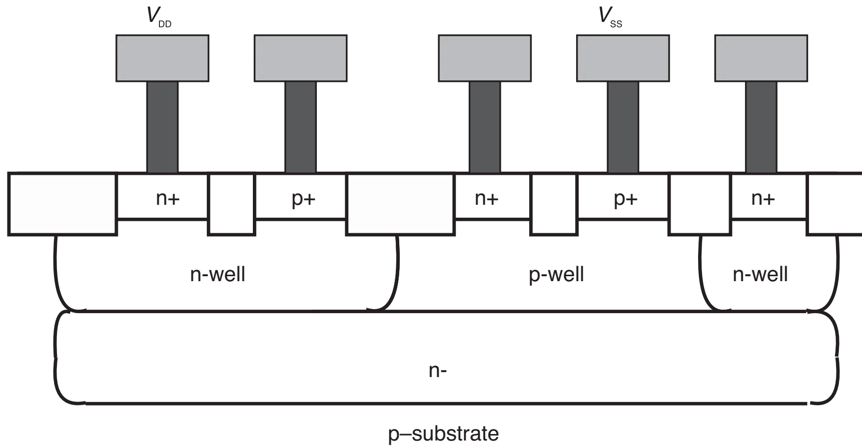



**Figure 4.15** Five-stripe triple-well latchup test structure layout design with buried layer under both p-well and n-well regions.

Voldman constructed DT latchup structures for the evaluation of the latchup CMOS latchup characterization [6, 7, 21–24, 26, 28–30]. The DT latchup test structures preserve the commonality to the standard pnpn test structures, with the following CMOS design practice objectives:

- Construct simple characteristic structures that represent the minimum design dimensions.
- Provide structures that represents the allowable DT technology semiconductor design rules.
- Maintain the CMOS latchup technology benchmarking strategy.
- Allow for comparison of deep trench CMOS technology and the standard CMOS technology.

A latchup design practice for test structure to quantify DT isolation is as follows:

- A single-stripe p+ diffusion is placed into an n-well region.
- A single-stripe n+ diffusion is placed into a substrate region.
- A single-stripe n+ diffusion is placed into an n-well region to serve as the n-well contact.
- A single-stripe p+ diffusion is placed into a substrate region to serve as the substrate contact.
- A DT was placed on the perimeter of the n-well region, where the trench center is coincident with the n-well edge definition (the DT borders the n-well edge).

Figures 4.16 and 4.17 show the cross section and layout of the four-stripe DT latchup structure, respectively. It is a CMOS latchup test site design practice to provide the following:

- At least one pnpn test structure is constructed with the minimum 'p+/n+ spacing' that is allowable with the minimum trench width.

With integration of DT and the triple-well structures, the n- and p-well regions can be separated, preventing the merging of the CMOS pnpn, as will be shown in Chapter 6 [26]. In this case, an

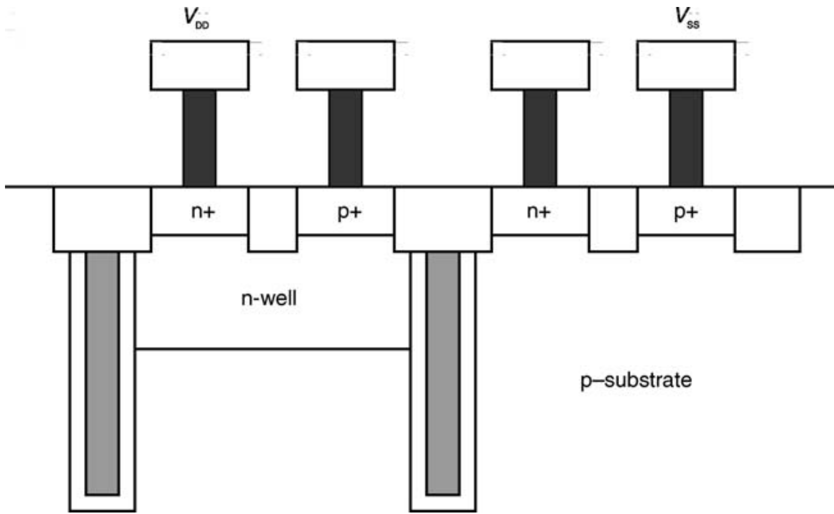



**Figure 4.16** Four-stripe BiCMOS/CMOS latchup deep trench (DT) test structure cross section.

additional electrical connection is needed for the isolation n-well ring around the p-well region (Figure 4.18). In some technologies, DT structures are polysilicon filled. As will be discussed in Chapter 6, the DT structure can be biased. In this case, a five-stripe latchup test structure was developed for biasing the DT polysilicon-filled structure (Figure 4.19) [24]. In all cases, the basic CMOS latchup design structure strategy is preserved in order to maintain a consistency with the four-stripe dual-well CMOS test structure strategy.

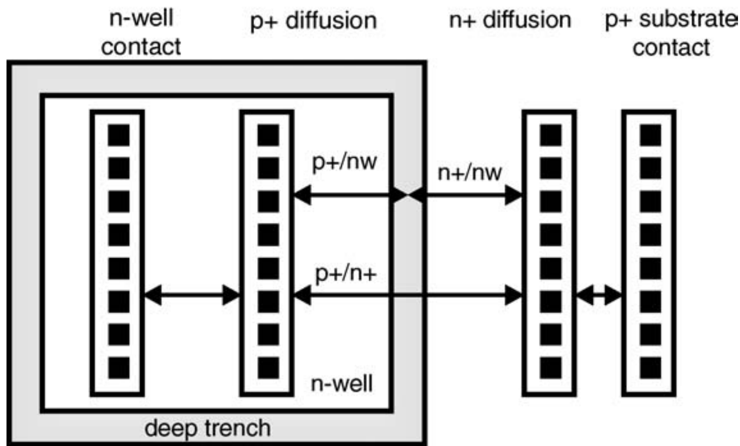



**Figure 4.17** Four-stripe BiCMOS/CMOS latchup deep trench (DT) test structure.

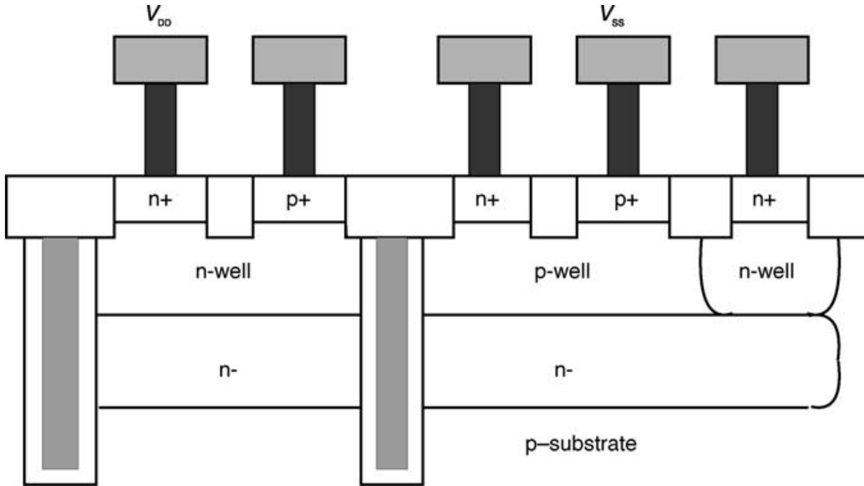



**Figure 4.18** BiCMOS/CMOS latchup triple-well deep trench (DT) test structure.

## 4.5 LATCHUP CHARACTERIZATION AND TESTING – NONAUTOMATED TEST SYSTEMS AND METHODOLOGY

For latchup characterization, it is advantageous to evaluate the *I–V* characteristic of the latchup structures over the complete design space. The latchup *I–V* parameters of interest include the following:

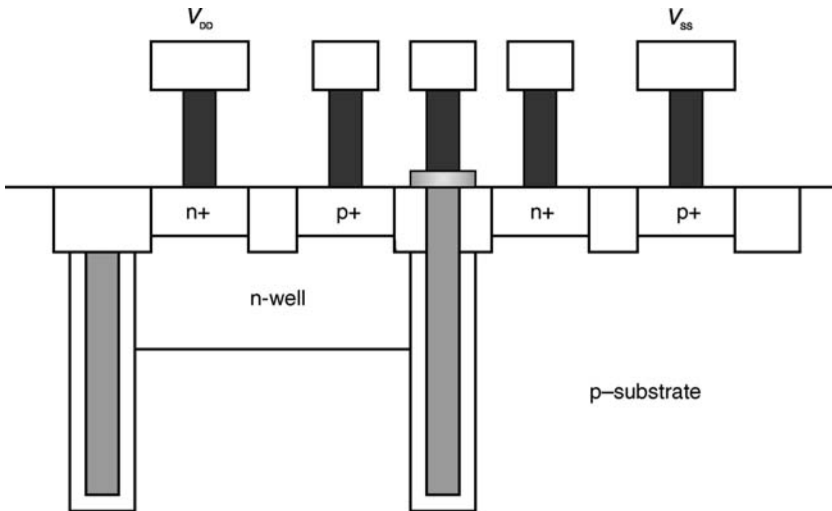



**Figure 4.19** BiCMOS/CMOS latchup DT test structure with electrical contact.

- $I$–$V$ characteristics;
- turn-on voltage, $V_{ON}$;
- trigger voltage, $V_{TR}$, and current, $I_{TR}$ (also referred to as knee voltage and current);
- holding voltage, $V_H$, and holding current, $I_H$;
- latchup on-resistance in the different regions (e.g. blocking state and holding state);
- latchup dynamic on resistance (e.g. blocking state and holding state).

For latchup development, it is desirable to obtain these parameters as a function of the following design parameters:

- a wide range of p+/n+ spacings from sub-minimum, minimum to large values;
- a full range of the substrate resistance from a short to an open circuit ($R_{SX} = 0$ to infinite).
- A full range of the well resistance from a short to an open circuit ($R_W = 0$ to infinite).

## 4.5.1 Latchup Testing – dc Test Methods

For dc latchup analysis, it is a latchup design practice to utilize the following test system and configuration:

- four-stripe CMOS latchup test structures;
- wafer-level bench-station with four probes (e.g. for the four different electrical nodes);
- a decade resistor box for the external substrate resistance (e.g. 0.1 Ω to 1 MΩ);
- a decade resistor box for the external well resistance (e.g. 0.1 Ω to 1 MΩ);
- a Tektronix 576 bipolar curve tracer (e.g. with oscilloscope and voltage generator).

Figure 4.20 shows the electrical test system configuration. The electrical connections are established between the Tektronix 576 bipolar curve tracer, the resistance boxes and the latchup four-stripe test structures. As according to the latchup circuit configuration with the $R_W$ and $R_{SX}$, the resistors are placed into the electrical circuit to provide additional series shunt resistances. In one test mode, the anode of the pnpn structure (e.g. p+ stripe) is electrically connected to the collector electrode of the bipolar curve tracer and the cathode of the pnpn (e.g. n+ stripe) is electrically connected to the emitter electrode of the bipolar curve tracer. To establish a high voltage across the test structure, the voltage level can be overcome using the override function (e.g. it is well known to use a pen and duct tape to override the over-voltage safety protection mechanisms).

In this test methodology, there are some significant advantages in latchup evaluation:

- The variations of the $V_H$ and $I_H$ as functions of the well and substrate resistances can be visualized.
- In the evaluation of the $V_H$ contours in resistance space ($R_W - R_{SX}$), significant data can be obtained to sweep out the characteristics at different voltage magnitudes. In the production of $V_H$ contours in resistance space, the quality of the measurements is significantly better than can be achieved in automated systems.

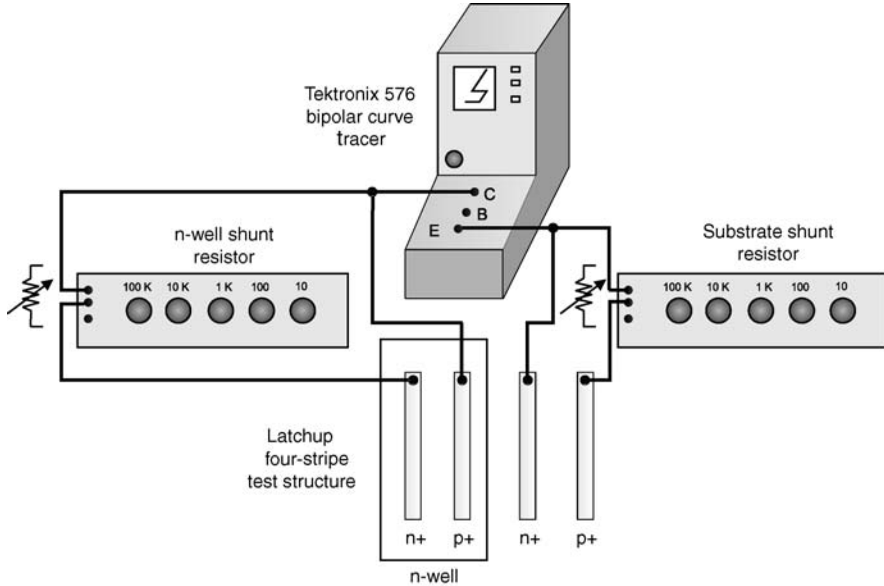



**Figure 4.20** Electrical test system using bipolar curve tracer and resistance decade boxes (for shunt resistance modification).

For the evaluation of the holding voltage, $V_H$, the use of decade resistor boxes allows the sweeping over 4 orders of magnitude of variation from conditions; the decade boxes allow for providing accurate experimental mapping of the holding voltage contours by visually observing the $V_H$ point on the bipolar curve tracer. To manually sweep the $V_H$ contours, as one shunt resistance is increased, the second shunt resistance must be decreased in order to maintain a common holding voltage contour; this is then completed by changing through all the resistance values on the resistance space contour. Experimentally, as one resistor box is set, one must manually 'tune' the second resistor box in order to sweep out the same holding voltage contour. This latchup design methodology can be completed for each four-stripe or five-stripe latchup structure. This latchup method can be applied to all the latchup test structures [28–30].

## 4.5.2 Latchup Testing – Pulsed ac Latchup Test Systems and Methodology

For latchup wafer-level characterization of transient latchup, the test system and methodology are applied to evaluate the latchup characteristics under pulsed condition and dc voltage bias. For transient analysis, latchup parameters are studied as a function of the pulse width and pulse rise time. For ac latchup analysis, it is a latchup design practice to utilize the following test system and configuration:

- four-stripe latchup test structures;
- Wafer-level bench-station with four probes (e.g. for the four different electrical nodes);
- a dc power supply suitable to source semiconductor chip application voltage levels;

- a single shot rectangular pulse source with variable pulse parameters (e.g. rise and fall time, and pulse width);
- a decade resistor box for the external substrate resistance (e.g. from 0.1 Ω to 1 MΩ);
- a decade resistor box for the external well resistance (e.g. from 0.1 Ω to 1 MΩ);
- at least one current probe to measure current transients.

Figure 4.21 shows the electrical test system configuration. The electrical connections are established between the dc power supply, the rectangular pulse source, the resistance boxes and the latchup four-stripe test structures. The voltage pulse source and current probes can be a Hewlett-Packard 8160 voltage pulse source and a Tektronix CT-1 current probe, respectively.

In this analysis, it is desirable to quantify the latchup sensitivity to positive overshoot and negative undershoot phenomena. For positive overshoot analysis, the voltage source can be placed on the anode of the pnpn structure and the current probe is placed on the cathode. In one method for positive overshoot analysis, the external well resistor and the ground of the voltage source are placed at a ground potential. A positive pulse is applied to the anode. The cathode node is biased to a negative $V_{DD}$ dc voltage source. The current source is placed on the connection to the negative $V_{DD}$ dc power supply. The external resistance for the substrate is also electrically connected to the $V_{DD}$ dc voltage source. In this same methodology, for undershoot analysis, the dc power supply voltage is placed on the pnpn anode and the pulse source is placed on the pnpn cathode. The current probe is placed on the

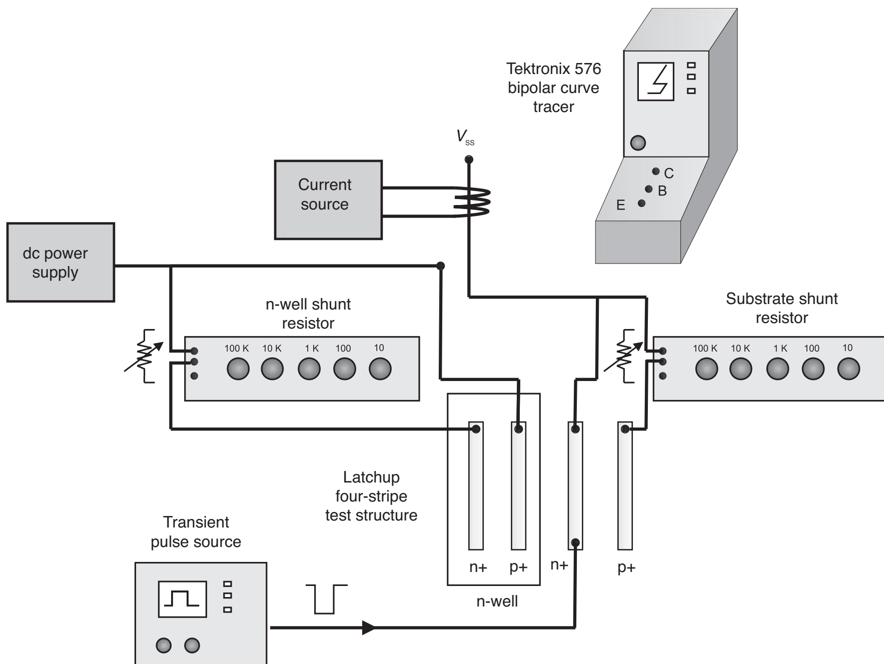



**Figure 4.21** Electrical test system for transient latchup analysis (ac analysis) highlighting the dc power supply, the rectangular pulse source and the resistance decade boxes for shunt resistance modification.

**Table 4.1** Example of transient latchup overvoltage response ($V_{DD} = -3.6$ V).

| Pulse width (ns) | Pulse height (V) |
|---|---|
| 10 | 3.70 |
| 15 | 2.96 |
| 20 | 2.65 |
| 30 | 2.32 |
| 50 | 1.97 |
| 100 | 1.74 |

External resistors well: 220 kΩ. Substrate: 10 kΩ.

connection to the pnpn anode. External resistors are connected in parallel with the anode and cathode through the $V_{DD}$ and $V_{SS}$ connections, respectively. Table 4.1 is an example of experimental data from such a test system using a four-stripe latchup structure. External resistors of 220 and 10 kΩ were placed as the shunt resistors, and a $V_{DD}$ voltage of $-3.6$ V was placed on the substrate. Experimental results show that as the pulse width increases, the magnitude of the pulse height needed to initiate latchup decreases.

The value of transient latchup testing as a function of the pulse width and the rise time provides a complete understanding of the CMOS circuitry under transient overshoot and undershoot environments. With a transient analysis system, the holding current versus pulse width study provides an understanding of the relationship of the circuit response. In addition, the analysis of the rise time evaluates the sensitivity of the circuits to ramp-up and ramp-down conditions during power-up and power-down sequencing conditions.

## 4.6 LATCHUP CHARACTERIZATION AND TESTING – AUTOMATIC TEST SYSTEMS

For CMOS latchup characterization of semiconductor technologies, automated test systems were constructed utilizing the CMOS latchup test structures. The four-stripe CMOS latchup test structures were placed into test macros for electrical characterization and design rule development on a wafer level. There are advantages and disadvantages of automated systems. These will be discussed in the next sections.

### 4.6.1 Noncommercial Wafer-level Automatic Test Systems

To evaluate the CMOS latchup design point, the evaluation of the CMOS latchup as a function of the p+/n+ spacing, n-well contact spacing and p-well (e.g. substrate) contact spacing. In test structures, with a fixed well and substrate contact spacing, automated test systems were first developed to evaluate the latchup sensitivity as a function of external resistors. In 1983, Troutman constructed a latchup automated test system that introduced different resistances to evaluate the design characteristics in resistance space. The automated CMOS latchup design system consisted of the following:

- semiconductor wafer prober;
- voltage and current sources;

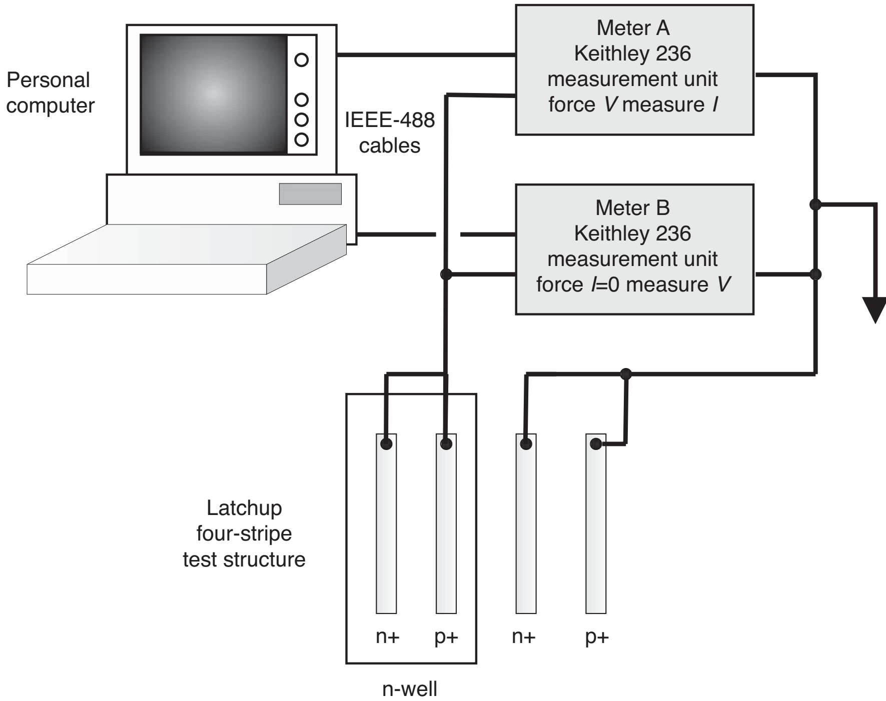



**Figure 4.22** Automated electrical test system for latchup analysis (noncommercial).

- switch matrix;
- two sets of single-component resistors of values in decades (e.g. 1 Ω to 1 kΩ);
- parameter analyzer and computer;
- software.

The latchup test system integrated the decade values of the resistor elements into the electrical circuit using the resistors as external resistor elements to vary the shunt impedances in the four-stripe circuit.

Figure 4.22 shows an example of an automated noncommercial test system used today for electrical characterization of latchup. The wafer-level latchup system of Liao [74] has current and voltage biasing capabilities to the four-stripe pnpn latchup structure for evaluation of undershoot and overshoot phenomena.

## 4.6.2 Commercial Automated Product Test System

Latchup evaluation can be addressed at the wafer level or product level. Commercial latchup test systems are typically designed only to address latchup at the product level. Due to the commonality of latchup and ESD testing, some commercial-level latchup testers perform both latchup testing and ESD

**Figure 4.23** State-of-the-art commercial latchup test system (KeyTek Zapmaster MK2, presently Thermo Fisher Scientific).

testing. Commercial latchup test systems were first introduced in the industry in the 1990s to evaluate latchup in packaged semiconductor products. Figure 4.23 shows an example of a present day commercial latchup test system. These commercial test systems address JEDEC's EIA/JESD 78 and AEC Q100-004 standards [43]. As the size of semiconductor chips increase, the number of I/O pins are increasing. As a result, the commercial latchup testers are increasing the ability to test larger semiconductor chips. Hence, new latchup test systems are being developed as the I/O count increases, as well as the package complexity. Figure 4.23 shows the KeyTek Zapmaster MK2 latchup tester (presently Thermo Fisher Scientific). Figure 4.24 shows a state-of-the-art commercial latchup tester, known as the MK4 latchup tester (presently Thermo Fisher Scientific). A commercial system contains a latchup pulse generator according to the latchup specifications. In this design, input pin drivers are used for latchup testing and parametric measurements. The designs have up to 10 MHz vector rate that are programmed with internal clock systems.

The latchup test systems include a bipolar source and sink trigger supply. These supply triggers can supply typically 50 V and 2 A capabilities. In these test systems, they allow for semiconductor chip

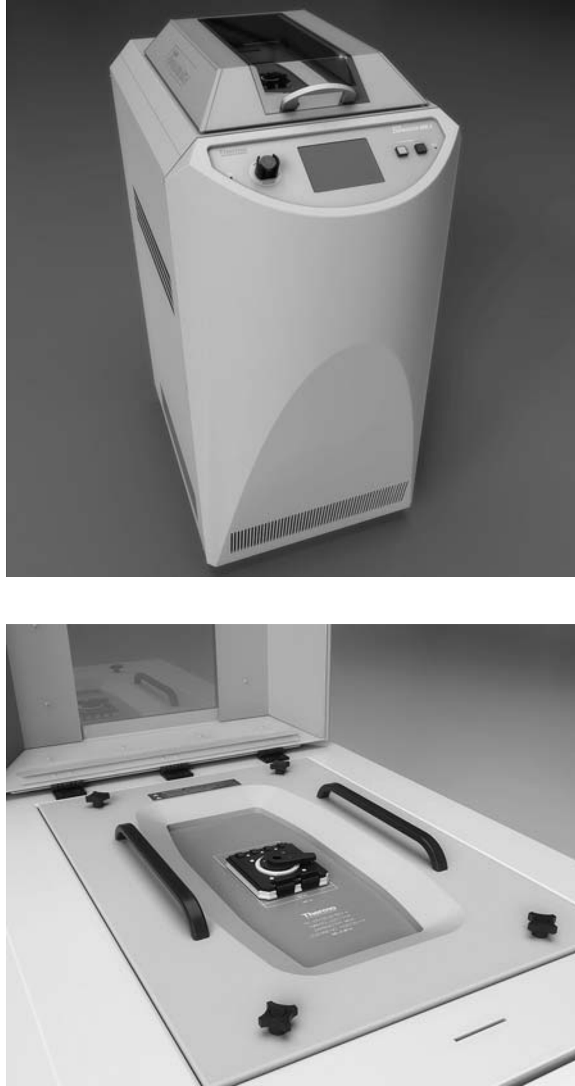



**Figure 4.24** State-of-the-art commercial latchup test system (Thermo Fisher Scientific, MK4).

pre-conditioning power supplies up to 15 V and current levels of 4 A. These commercial test systems have vector depths of 2048 for all the physical pins on the device under test (DUT), with vector rates on to a frequency of 4 kHz.

Figure 4.25 shows a third commercial test system that provides product-level evaluation of latchup according to the JEDEC latchup specification. This system introduces mechanical armatures for testing. One advantage of this system is that the pulse source is modular. For example, alternate pulse sources can be utilized for other HBM test sources in the same system as latchup testing. For example, Gross utilized this commercial latchup/ESD test system to evaluate cable discharge events (CDE) by

**Figure 4.25** Commercial latchup test system (Oryx Instruments 11000 EX, presently Thermo Fisher Scientific).

exchanging the HBM standard source with a 2000 pF capacitor source equivalent to a charged cable and using JEDEC latchup test methodologies. This was used to evaluate the effect of CDE on ESD structures, as well as the effect of introducing latchup.

## 4.7 LATCHUP CHARACTERIZATION – WAFER-LEVEL TEST PROCEDURES

For latchup characterization, different methods exist for test procedures. In the four-stripe latchup test structure, there are four independent electrodes allowing for different test procedures representing different test modes. Hence, there are different modes of evaluation [2].

### 4.7.1 Latchup Characterization Test – Ramped Power Supply Voltage Study

In the latchup characterization practice, the following test procedure is performed:

- the p+ anode and the n-well are electrically connected;
- the n+ cathode and the p− substrate (e.g. or p-well) are electrically connected;

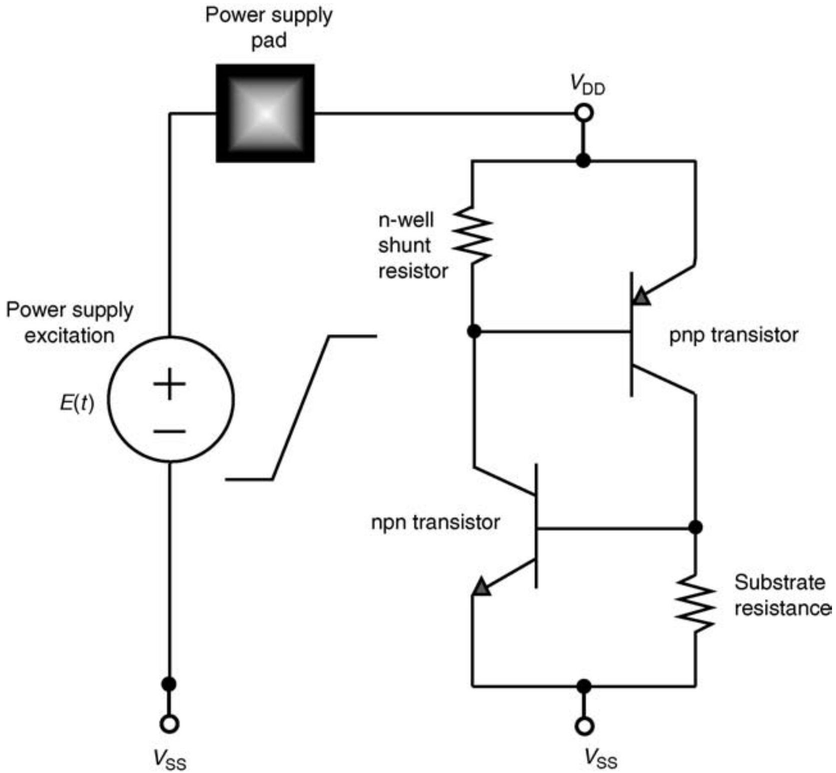



**Figure 4.26** CMOS latchup characterization test with a positive applied voltage ramp to n-well and the p+ anode.

- the n+ cathode and the p− substrate are electrically grounded;
- the p+ anode and the n-well are swept with an applied positive voltage ramp.

Figure 4.26 shows the example of the test configuration and the latchup structure circuit schematic. In this latchup characterization procedure, well shunt resistors can be placed in series with the n-well electrical connection (e.g. n+ diffusion stripe in the n-well). Additionally, a substrate shunt resistor can be placed in series with the substrate contact (e.g. p+ diffusion stripe in the p− substrate). In this case, the test represents a semiconductor chip whose n-well and p-channel MOSFET device are electrically connected. This is the case when the p-channel MOSFET device is contained within the n-well region. The power supply (e.g. $V_{DD}$) and the p-channel MOSFET are ramped with a positive polarity event.

## 4.7.2 Latchup Characterization Test – p+ Anode Power Supply Overshoot Study

In the latchup design practice, it is important to evaluate electrical overshoot. Electrical overshoot of the p+ channel MOSFET can occur when the MOSFET source voltage responds to a signal pad event

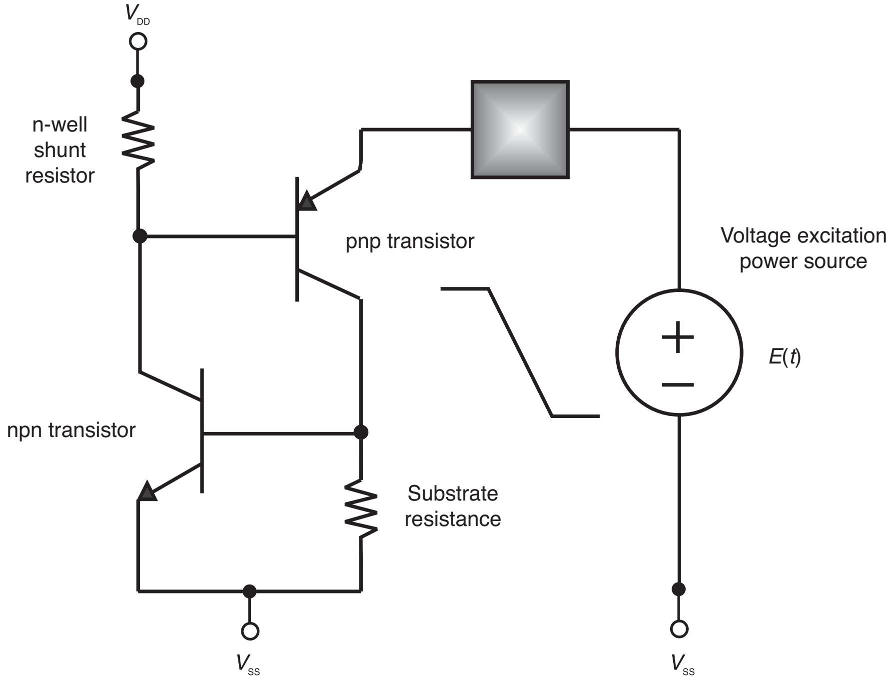



**Figure 4.27** CMOS latchup characterization test with a positive applied voltage event to the p+ anode.

[2]. In the event that the power supply $V_{DD}$ is slow to respond and the individual circuit overshoots the power supply voltage, forward biasing of the PMOSFET can occur. The following test procedure is performed as a latchup test practice [2]:

- the p+ anode is electrically connected to a voltage or a pulse source;
- the n-well is electrically connected to a dc power supply voltage;
- the n+ cathode and the p− substrate (e.g., or p-well) are electrically connected;
- the n+ cathode and the p− substrate are electrically grounded;
- the p+ anode is swept with an applied positive voltage ramp or current source.

Figure 4.27 shows the test configuration and the latchup structure circuit schematic. This is the case when the p-channel MOSFET is contained within the n-well region. The voltage or current source is increased until latchup occurs. When latchup triggering is initiated, the 'overshoot' condition on the p-stripe is recorded; this is an 'overshoot voltage' or an 'overshoot current' condition.

## 4.7.3 Latchup Characterization Test – Ramped Negative Power Supply Voltage Study

For the evaluation of substrate negative transients, a ramped negative transient method can be used. In the four-stripe latchup test structure, there are four independent electrodes allowing for different test

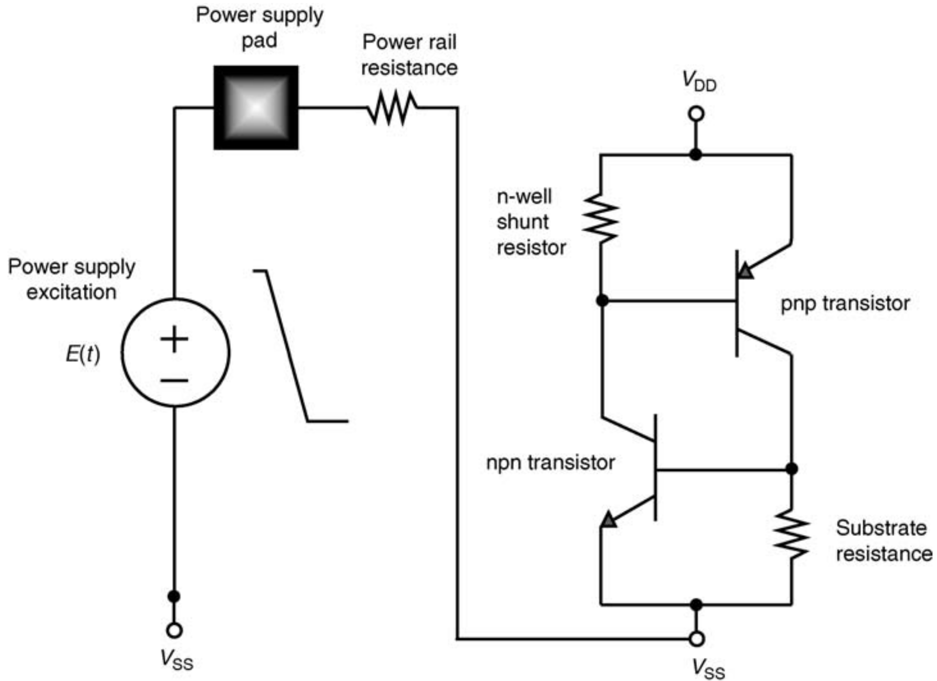



**Figure 4.28** CMOS latchup characterization test with a negative applied voltage ramp to substrate and the n+ cathode.

procedures representing different test modes. In the CMOS latchup characterization practice, the following test procedure is performed [2]:

- the p+ anode and the n-well are electrically connected;
- the n+ cathode and the p− substrate (e.g. or p-well) are electrically connected;
- the n+ cathode and the p− substrate are electrically grounded;
- the n+ anode and the p-well are swept with an applied negative voltage ramp.

Figure 4.28 shows the test configuration and the circuit schematic. In this case, the test represents a semiconductor chip whose p− substrate (e.g. or p-well and n-channel MOSFET device are electrically connected). This is the case when the n-channel MOSFET is contained within the p− substrate (e.g. or p-well region). The power supply $V_{SS}$ and the n-channel MOSFET are ramped with a negative polarity event.

### 4.7.4 Latchup Characterization Test – n+ Cathode Ground Power Supply Undershoot Study

In the latchup design practice, it is important to evaluate electrical undershoot. Electrical undershoot of the n-channel MOSFET can occur when the NMOSFET source voltage responds to a

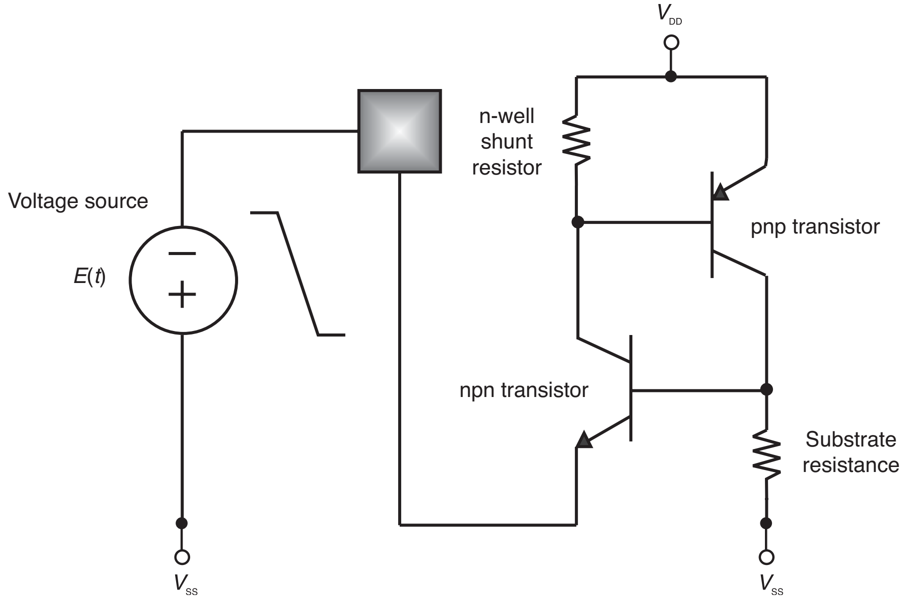



**Figure 4.29** CMOS latchup characterization test with a negative pulse to the n+ anode.

negative pulse signal pad event. In the event that the power supply $V_{SS}$ is slow to respond and the individual circuit undershoots the power supply voltage, forward biasing can occur. Note that this can also occur on an n-channel MOSFET drain node from a signal-pad overshoot or voltage transient. It is also possible that the n-channel MOSFET or n-type device is not within the power domain of the p-well tub or $V_{SS}$ supply. The following test procedure is performed as a latchup test practice [2]:

- the p+ anode is electrically connected to the n-well stripe;
- the n+ cathode is electrically connected to a power supply or pulse source;
- the p− substrate (e.g. or p-well) is electrically grounded;
- the n+ cathode is swept with an applied negative voltage ramp or current source.

This is the case when the n-channel MOSFET is contained within the p-well region (Figure 4.29). The voltage or current source magnitude is increased until latchup occurs. When latchup triggering is initiated, the 'undershoot' condition on the n-stripe is recorded; this is an 'undershoot voltage' condition or an 'undershoot current' condition.

## 4.8 LATCHUP CHARACTERIZATION TECHNIQUES – WAFER-LEVEL TRANSMISSION LINE PULSE METHODOLOGY

Pulsed testing of semiconductor devices was a common practice in early research and development of power electronics, pulsed power electronics and military applications (e.g. electromagnetic pulse

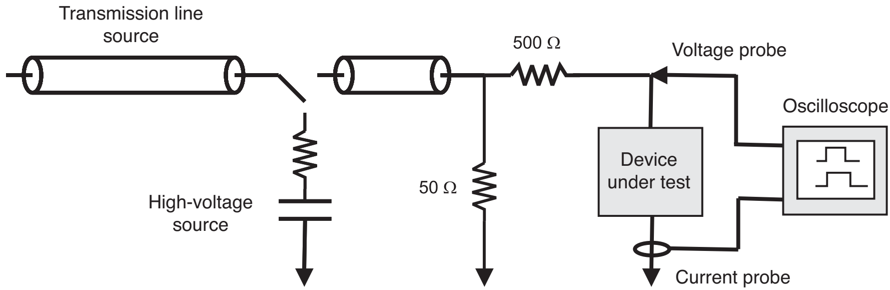



**Figure 4.30** Schematic of current source TLP system.

(EMP) tolerance). Early researchers used pulsed testing for evaluation of the power-to-failure of electronics. Transmission line pulse (TLP) testing provides the ability to understand the terminal current and voltage conditions on a semiconductor device or system, and can be used to generate pulsed I–V characteristics [48]. Standardization of TLP testing methodology has become a standard practice in the semiconductor industry for ESD phenomenon [49, 50]. This same practice, with modifications associated with latchup phenomenon, can be applied to latchup [4].

Latchup pnpn test structures can utilize TLP testing methodology for the latchup characterization of a semiconductor technology. TLP testing will provide an S-type pulsed I–V characteristics of the pnpn under a sequential pulsed state. Using the TLP test method, the latchup characteristic terms can be evaluated (e.g. trigger voltage and current, holding voltage and current, and the dynamic on-resistance). Additionally, applying this technique to latchup, this TLP technique allows the evaluation of the pnpn breakdown, saturation effects, thermal breakdown and pnpn device failures as a function of the pulse parameters (e.g. pulse width, pulse rise time, pulse fall time, etc.). For the evaluation of latchup phenomena, it is suitable for obtaining a significant understanding of single devices or multiple devices connected to a single pad. TLP systems can be designed in at least four different configurations, where the device response is extracted from the measurement of the transmitted and reflected signals.

Figure 4.30 shows an example of current source TLP system. This TLP system is a 500 Ω based TLP system [49,50]. The system consists of a transmission line (TL) cable, a TLP charging power supply source, a high impedance 10–100 MΩ resistor and a first switch to isolate the charging source from the TL cable. The system also contains a delay line, a 50 Ω shunt resistor, a large series resistor element, a current probe and a voltage probe. In this system, the latchup structure under test would be placed between the incoming signal and the ground potential.

A preferred system is the 50 Ω based system is the time domain reflection (TDR) TLP system, referred to as the TDR-TLP system [49, 50]. Figure 4.31 shows an example of a TDR–TLP system. The system consists of a TL cable, a TLP charging power supply source, a high impedance resistor and a first switch to isolate the charging source from the TLP cable. The system also contains an attenuator, a delay line, a current probe and a voltage probe. In this system, the latchup structure under test would be placed between the incoming signal and ground potential. The purpose of the attenuator is to reduce the reflections before they can reach the pnpn device under test. In this case, a 20-dB attenuator will provide a 10× reduction in the reflected signal.

Using the TLP methodology for latchup evaluation, the key objective is to provide a pulsed I–V characteristics that can best represent the structure response with pulse width associated with latchup events. Using a rectangular waveform, one can evaluate the key responses of the structure and quantify the key I–V characteristic variables. The key I–V metrics of interest are the avalanche voltage ($V_{AV}$),

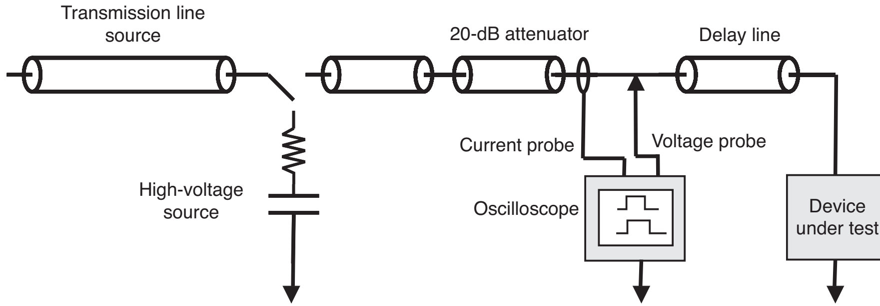



**Figure 4.31** Schematic of time domain reflectometer (TDR) TLP system (with 20 dB attenuator).

avalanche current ($I_{AV}$), holding voltage ($V_H$), holding current ($I_H$) and the on-resistance, $R_{ON}$. The pulse parameters that may influence the latchup I–V characteristic are as follows:

- pulse width;
- pulse rise time;
- pulse fall time;
- maximum peak current overshoot;
- maximum ringing duration;
- measurement window.

The pulse width chosen for latchup characterization should be associated with the pulse widths of interest that may occur during latchup events. In the analysis of the TLP I–V response, it is possible to evaluate a wide range of pulse widths from short pulse to long pulse response. The rise time of the TLP pulse may be important due to latchup sensitivity to voltage perturbations, $dV/dt$, or current perturbations, $dI/dt$.

In the latchup testing, a step stress of test sequence is initiated to sweep the I–V characteristic of the pnpn structure [49, 50]. Figure 4.32 shows the TLP test sequence. In the test process, the resistances of the system can be nullified prior to measurement of the step stress. This is also referred to as a calibration and verification procedure. To sweep the latchup characteristic, a pulse is generated successively where the leakage is tested between step stress to determine if the structure has undergone failure. During each pulse, voltage and current measurements are taken within the measurement window of the pulse. These data are averaged to reduce statistical noise. Given that the pnpn device is not destroyed, the stress is continued. In this fashion, the latchup I–V transitions can be characterized for a given pulse width. Note that in this methodology, there is no applied dc voltage to the pnpn structure (as a result, this test procedure is distinct from other transient latchup test processes). Figure 4.33 shows an example of a latchup pulsed I–V characteristic using a TLP test system.

## 4.9 LATCHUP CHARACTERIZATION – TRANSIENT LATCHUP

Today, transient latchup is evaluated according to wafer-level methods developed by Troutman and Zappe [33–35]. But today, new methods have been developed using transient latchup (TLU) amplifiers as a source for evaluation of new methods.

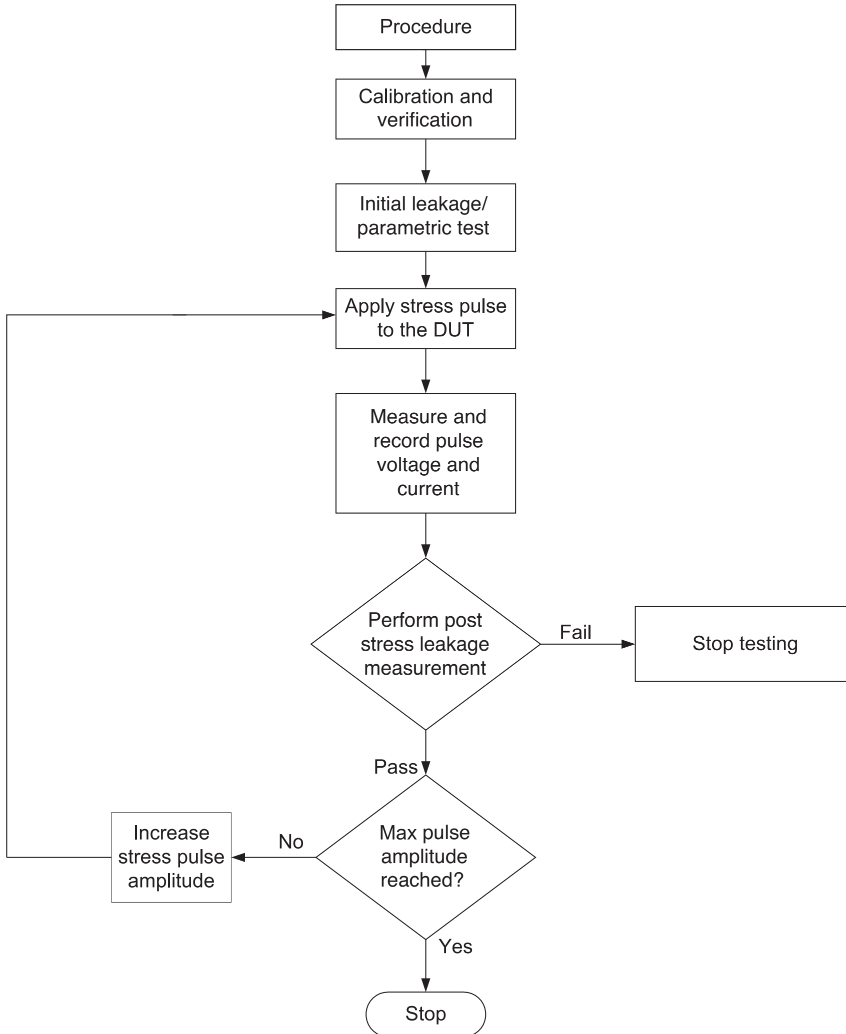



**Figure 4.32** TLP testing procedure.

## 4.9.1 Latchup Characterization – Wafer-level Transient Latchup

For the evaluation of transient latchup, a transient latchup test system is required. Figure 4.34 shows the transient latchup characterization system. The test system requires a dc power supply, an oscilloscope, probe station and latchup test structure. In addition, a current or voltage pulse source as well as current and voltage probes are used to evaluate the device response. In the transient latchup characterization practice, the following test procedure is performed:

- The test structure is biased with a dc power supply voltage at some voltage level (e.g. $V_{DD}$).
- The n+ cathode and the p− substrate (e.g. or p-well) are electrically grounded (e.g. $V_{SS}$).

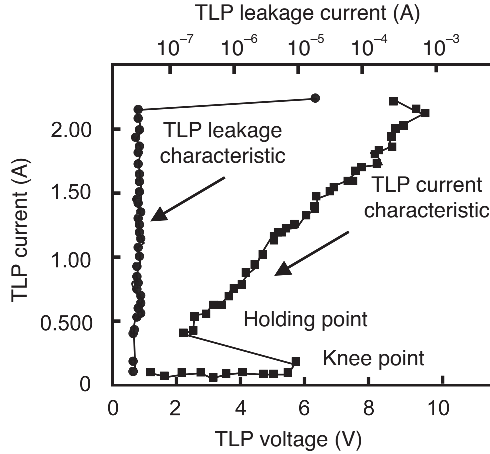



**Figure 4.33** Example CMOS latchup TLP pulsed *I–V* characteristic.

- An ac pulse of a given width and current (voltage) level is applied to the device under test.
- An oscilloscope is used to evaluate the device response. For evaluation of current and voltage, a two-channel oscilloscope is required.
- A step stress is achieved by increasing the magnitude of the pulse until latchup is evident from the pulse event.

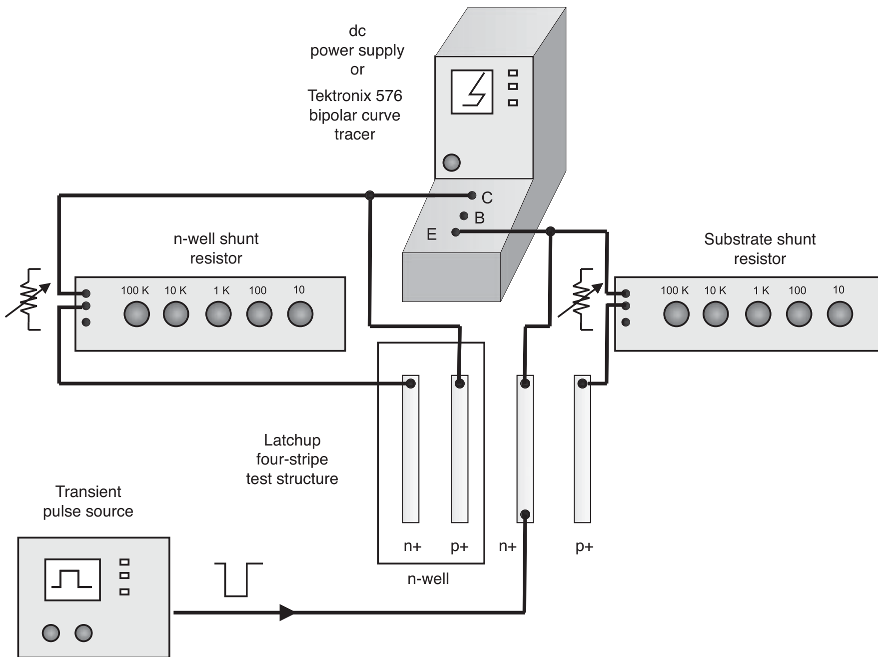



**Figure 4.34** Transient latchup characterization test system.

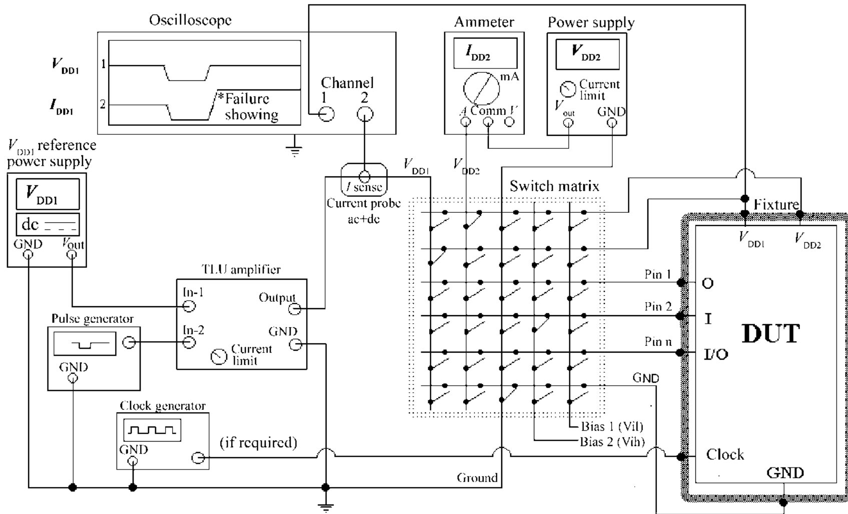



**Figure 4.35** Transient latchup test system with transient latchup (TLU) amplifier and signal applied to the input pins. Reproduced by permission of ESDA.

The pulse source can apply positive or negative polarity events on any of the electrical nodes of the latchup structure or device under test.

## 4.9.2 Latchup Characterization – Transient Latchup (TLU) Amplifier Product-level Test

Latchup testing methods must capture all types of potential latchup events that may occur on a chip or system level. A transient latchup upset (TLU) test was established to address all forms of transients observed on a semiconductor component. A commercial transient latchup amplifier source was constructed to address positive and negative voltage transitions that may occur. A primary goal was to design a latchup test to address short-duration transient electrical disturbances. Different transients phenomena were addressed (e.g. fast slew rate pulses, long-duration ringing events, short-duration supply bounce, EMI and RFI-induced noise). These waveforms must be controlled and reliable waveforms that could be simulated on latchup test equipment. Using a broadband power amplifier (TLU amplifier), it was possible to create specific transient waveforms to better understand the latchup response. Figure 4.35 shows the test configuration using a TLU amplifier. The test system consists of the TLU amplifier, power supply and diagnostics such as current probes and oscilloscope. The source is connected to the device under test through a switching matrix. In this figure, the TLU amplifier applies a pulse to the signal pins. Figure 4.36 shows the case where the signal is applied to the power supply.

A single cycle sine wave was also used to initiate latchup. Different frequencies, phases and half sine waveforms can be used (Figure 4.37).

The TLU amplifier characteristics have been defined in the TLU methodology. It can be described as a low-output impedance signal amplifier that is capable of superimposing a reference voltage source and a rectangular pulse from the pulse generator. The TLU amplifier has the following criteria:

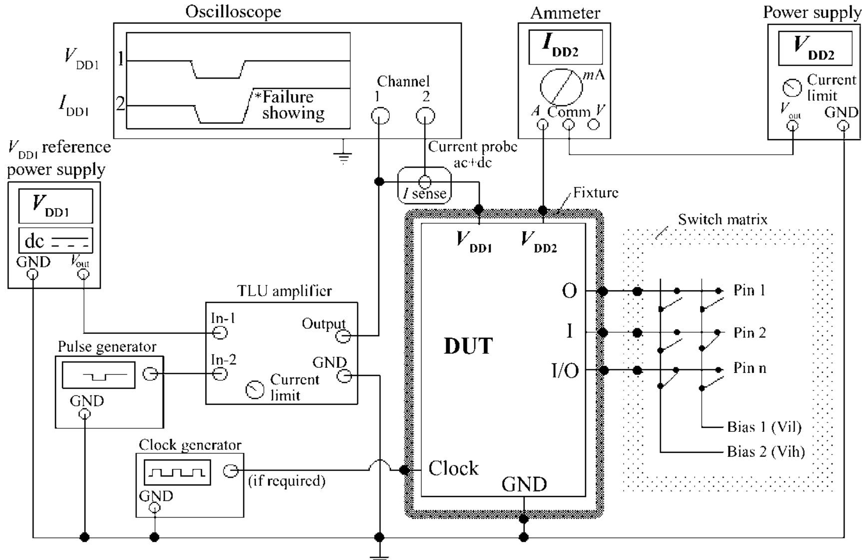



**Figure 4.36** Transient latchup test system with transient latchup (TLU) amplifier and signal applied to the power supply. Reproduced by permission of ESDA.

*Output voltage range*: greater of $\pm[3 \times$ maximum DUT operating voltage ($V_{DDMAX}$)] or $\pm 30$ V.
*Output current range*: greater of $\pm[3 \times$ failure current ($I_{FAIL}$)] or $\pm 3$ A.
*Output voltage regulation*: The dc output voltage is within 5 % of the input voltage.
*Output impedance*: Less than or equal to 1 $\Omega$ (at dc voltage).
*Output slew rate*: Greater than 10 V/μs with an open load (this value is sufficient to produce a 15 $V_{P-P}$, 300 kHz sine wave).

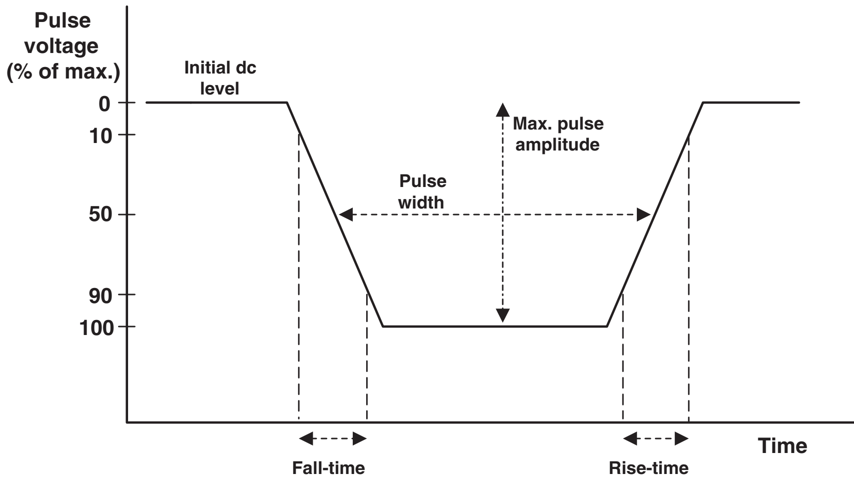



**Figure 4.37** Transient latchup test waveform. Reproduced by permission of ESDA.

**Table 4.2** Transient latchup amplifier waveform parameters.

| TLU amplifier load | Measured parameter | Acceptance criteria |
|---|---|---|
| Test load 1 (5 Ω) | Current clamp activation time | $< 1$ ms |
| Test load 2 (10 Ω) | Initial dc voltage | $10 \pm 0.2$ V |
| Test load 2 (10 Ω) | Maximum pulse amplitude | $10 \pm 0.2$ V |
| Test load 2 (10 Ω) | Pulse width | $20 \pm 1$ μs |
| Test load 2 (10 Ω) | Fall time ($10\,V_{\mathrm{P-P}}$) | $<2$ μs |
| Test load 2 (10 Ω) | Rise time ($10\,V_{\mathrm{P-P}}$) | $<2$ μs |

*Output current limiting* ± [output current range]: The value may be continuously variable or have preset increments.
*Current limit response*: The output current is limited to the final value within 1 ms.
*Amplifier voltage gain*: The output waveform is within 5 % of the expected amplitude with respect to the input.
*Amplifier bandwidth*: The amplifier transmits a signal without distortion from DC (0 Hz) to 300 kHz or the desired frequency of the transient signals, whichever is greater.
*Amplifier output function*: The output is the sum of the two inputs.
*Amplifier inputs*: The amplifier has two inputs, one for $V_{\mathrm{DD}}$ and one for the transient signal source. One or both of the inputs may be scaled (amplified) to compensate for small input signals. Alternatively, the amplifier may have only one input and the summing of the two signals is provided by a separate circuit.
*Amplifier input impedance*: The transient signal input shall have an impedance of 50 Ω and matching the output impedance of the pulse generator. The $V_{\mathrm{DD}}$ input shall have an impedance of 50 Ω or more to prevent excessive current load on the $V_{\mathrm{DD}}$ reference supply.
Table 4.2 shows the TLU amplifier waveform parameters and load conditions.
Figure 4.38 shows an example of the TLU test with a negative polarity pulse. The voltage magnitude was −1.1 V. The lower trace is the response of the semiconductor DUT to the half sinusoid TLU pulse. Figure 4.39 shows the same test with a larger pulse magnitude. At a magnitude of −1.32 V, the DUT on the second trace undergoes a −80 mA surge. At higher pulse magnitudes of −1.75 V, this response increased to −200 mA surge, indicating the onset of latchup.

The latchup mechanism for this latchup event (e.g. triggering latchup from a negative polarity pulse event on $V_{\mathrm{DD}}$) is referred to as an 'electron flood' event. Electron flood events are initiated by the following latchup procedure:

- A semiconductor component is powered to normal steady-state voltage conditions (e.g. $V_{\mathrm{DD}}$).
- A transient is placed on the power supply $V_{\mathrm{DD}}$ with a negative polarity below the ground potential (e.g. by at least one diode $V_{\mathrm{BE}}$).
- The transient is followed by return to the normal steady-state voltage conditions (e.g. $V_{\mathrm{DD}}$).

During the reduction of the power supply to negative polarity, the supply current is reversed and flows into ground due to the forward biasing of the well–substrate junctions (e.g. p− substrate); this leads to significant minority carrier injection into the substrate. As the power supply voltage rail returns to its steady state $V_{\mathrm{DD}}$, the injected charge carriers leave the p− substrate creating current flows to the ground $V_{\mathrm{SS}}$ and $V_{\mathrm{DD}}$ pads. The magnitude of this current is proportional to the trigger voltage (undershoot on $V_{\mathrm{DD}}$) and the rate at which $V_{\mathrm{DD}}$ returns to normal functional state (power supply rise time). This current flows through the substrate and other structures and creates across-chip voltage gradients and transient fluctuations. If this voltage gradient is such as to invert a metallurgical junction, the resulting current flow can initiate and sustain latchup. Since this injected current is not

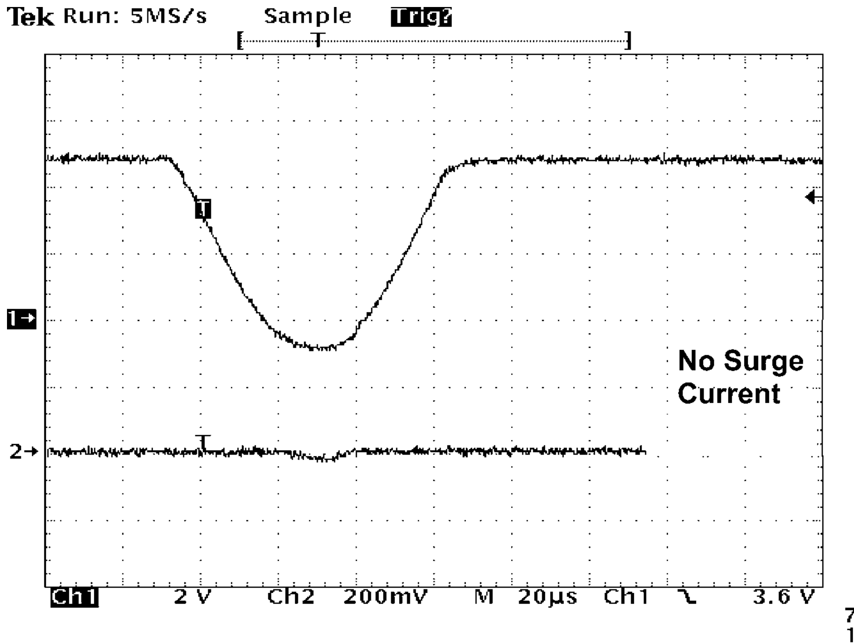



**Figure 4.38** TLU test highlighting the TLU waveform and the device under test (DUT) chip response (−1.1 V waveform pulse). Reproduced by permission of ESDA.

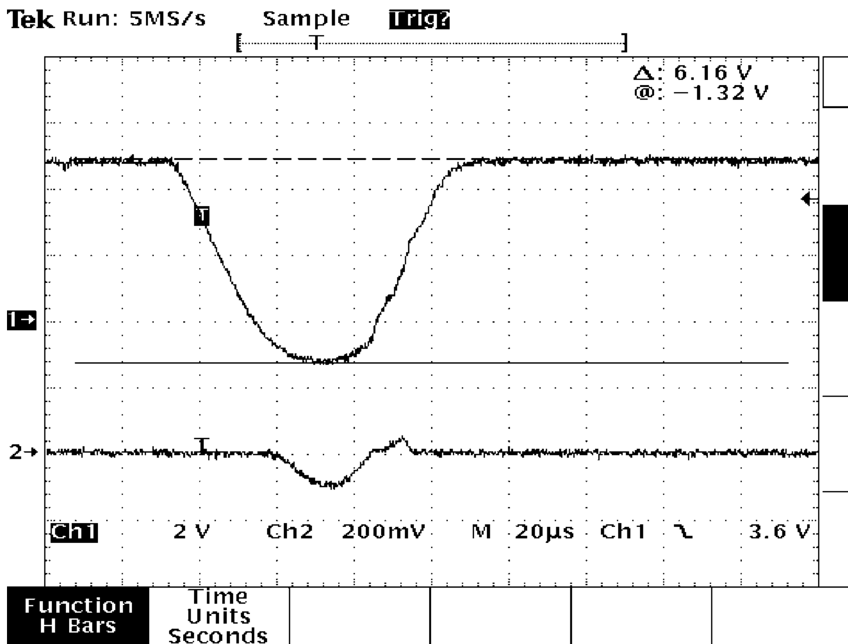



**Figure 4.39** Transient latchup (−1.32 V pulse) highlighting an −80 mA DUT response.

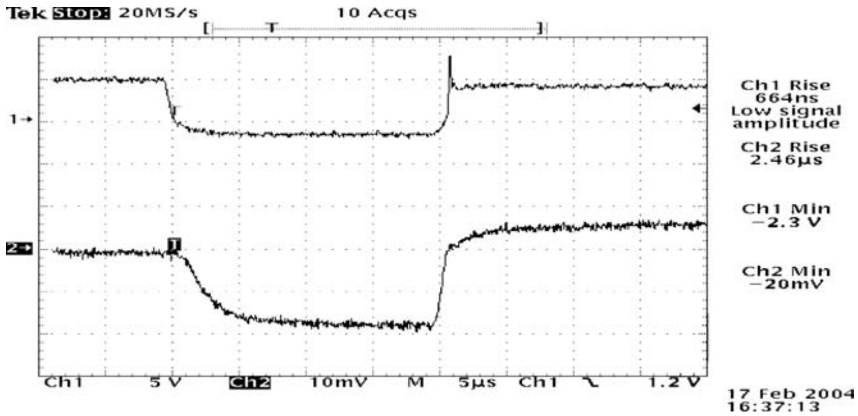



**Figure 4.40** Oscilloscope waveform showing TLU failure. The top trace is the power supply voltage $V_{DD}$ and the lower trace is the power supply current $I_{DD}$ (500 mA/div).

localized in the substrate, the latchup could potentially be triggered locally or globally in the semiconductor chip.

Figure 4.40 shows an example of a latchup failure. The top oscilloscope trace is the power supply voltage $V_{DD}$ and the lower trace is the power supply current $I_{DD}$ (500 mA/div). In this event, latchup occurred.

These types of events are sensitive to ramp rate, capacitance and slew rates. As the bandwidth of the latchup broadband amplifier increases, results show that latchup failure occurs at smaller trigger level. Investigations showed a strong correlation of latchup sensitivity to power supply rise time and the voltage undershoot amplitude.

## 4.10 GUARD RING CHARACTERIZATION

Guard rings serve the purpose of providing electrical and spatial isolations between adjacent circuit elements, preventing interaction between devices and circuits that may undergo latchup. This is achieved by the prevention of minority carriers from within a given circuit or the prevention of minority carriers from entering a sensitive circuit. In the first case, the role of the guard ring structure is to prevent the minority carriers from leaving the region of the circuit and influencing the surrounding circuitry. In the second case, the injection is external to the circuit, and the objective is to prevent the minority carriers from influencing the circuit of interest. Hence, the point of view is relevant. Hence, in the discussion of internal latchup, the role of the guard ring is to provide electrical isolation between the pnp and the npn structure. In this case, the role of the guard ring is to minimize electrical coupling and to prevent regenerative feedback from occurring between the pnp and the npn; stated otherwise, the role of the guard ring is to lower the gain of the feedback loop by reducing the parasitic current gain. Guard rings within the pnpn structures lower the parasitic bipolar gain by the following means:

- increases the base width of the parasitic pnp or npn structure;
- provides a region of collection of the minority carriers to the substrate or power supply electrodes, ‘collecting’ the minority carrier via a metallurgical junction or electrical connection;

- provides a region of heavy doping concentration to increase the recombination within the parasitic, 'capturing' the minority carrier via electron–hole pair recombination.

In external latchup, the role of the guard ring is to provide electrical isolation between the first and second regions; the role of the guard ring is not to minimize the electrical coupling between the local pnp and the npn transistor but to prevent injection mechanisms traversing from the first to the second region, where the concers are:

- lowering the minority carrier injection from the first region of interest to any exterior region;
- lowering the minority carrier injection from an exterior region into the circuit region of interest.

In external latchup, there is a greater focus on the transport of minority carrier injection into the region with the initiation of a local pnpn. As a result, the issue is not the regenerative feedback within a local pnpn but the interaction of an injection source and its electrical stimulus of a pnpn parasitic device in both intra- and intercircuit cases. Typically, the focus is an the issue of external initiation of latchup of an intracircuit pnpn parasitic structure, where no internal guard rings are present.

In both cases, the guard ring structure serves as a means of lowering the effectiveness of the minority carriers transport via recombination, collection and spatial separation. As such, the relationship between the injection source and the collection region can be viewed in a generalist view as a bipolar transistor itself. In a bipolar transistor, there is an emitter of carriers that are then transported to the collector of the carrier through a physical region. A bipolar transistor is a minority carrier device. In the transport of minority carriers, there is an efficiency at which the carriers transport between the injection region (e.g. the emitter) and the collection region (e.g. the collector). Hence, in the quantification of the guard ring structure, the 'guard rings' are perceived to serve as a means to mitigate the transport of minority carriers. Hence, the effectiveness of the guard ring to lower the minority carrier transport via collection, recombination or other means is of interest. Additionally, the 'inside–outside' perspective also indicates that quantification in a 'forward' or a 'reverse' operation is also of interest.

## 4.10.1 Guard Ring Efficiency

A metric for evaluation and characterization of the guard ring is the 'guard ring efficiency'. The ratio of the captured current in the guard ring structure and the injected current is a measure of its guard ring efficiency. In this development, two metrics will be defined. Let us define first as an injection metric $F$, where $F = I$ (injector)/$I$ (collector) [1–3, 6, 7, 21].

$$F = \frac{I_{\text{inj}}}{I_{\text{coll}}}.$$

Injection metric $F$ is the ratio of the current injected into the structure inside the guard ring and the current collected outside the guard ring [1, 3, 6, 7]. If every minority carrier that is injected is collected, the injection ratio would equal unity. As the number of collected minority carriers decreases, the factor $F$ increases above unity. This metric increase is a measure of the guard ring's ability to minimize the transport out of the guard ring to a region of interest where the carriers are collected. This can be trivially quantified on a test structure or simulated to quantify the effective transport ratio. Since this term is not normalized to the injection current, the inverse relationship is better viewed for a probability

perspective. The inverse of $F$, $1/F$, will be defined as the 'transport factor'. The transport factor $1/F$ is the ratio of the collected current normalized and the injection current.

$$\frac{1}{F} = \frac{I_{\text{coll}}}{I_{\text{inj}}}.$$

From the probability view, by normalizing to the injection carrier current, the collected current is the probability, of given an injection current level, that the current reaches the structure of interest. The inverse term, $1/F$, is the escape probability (e.g. $P(E)$) in the approximation that the number of collected electrons outside the ring is significantly greater than the number of electrons that recombine outside the guard ring. This interpretation can change based on the structures between the injecting and collecting structures. Hence, the 'collected current' normalized to the 'injected current' can be quantified using an 'injecting structure' and the 'collecting structure'.

The escape probability, or $P(E)$, is then related to the following design parameters:

- physical dimensions of the injection source (e.g. width, length and depth);
- semiconductor process of the injection source (e.g. n+ diffusion, n-well and n-channel transistor);
- physical separation between the injection source and the collecting region;
- physical dimensions of the collecting region (e.g. width length and depth);
- bias conditions on the collecting region (e.g. n-well to substrate bias);
- current magnitude;
- all structures and guard ring structures between the injection source and the collection source.

Figure 4.41 shows an example of a guard ring test structure. In the structure is an 'injecting' structure and a 'collecting' structure. In between the injecting and collecting structures are the structures that influence the transport of minority carriers. Quinke [3] defined a guard ring efficiency,

$$\Psi_{\text{GRE}} = 1 - \frac{I_{\text{coll}}}{I_{\text{inj}}}$$

or

$$\Psi_{\text{GRE}} = 1 - \frac{1}{F}.$$

In this case, if no current reaches the collecting structure, then the guard ring has captured all the current and the guard ring efficiency is unity (e.g. 100 %). If the guard ring is ineffective and all of the injected current is collected by the outer structure, the guard ring efficiency would be zero.

### 4.10.2 Guard Ring Theory – A Generalized Bipolar Transistor Perspective

The effectiveness of the guard ring structures to prevent latchup can be evaluated from different perspectives. Taking a first perspective, the problem can be viewed as a quantification of a generalized bipolar transistor. First, there is an injection structure, a collecting structure and a region of transport in between; this forms a lateral bipolar transistor. From this perspective, the lateral bipolar

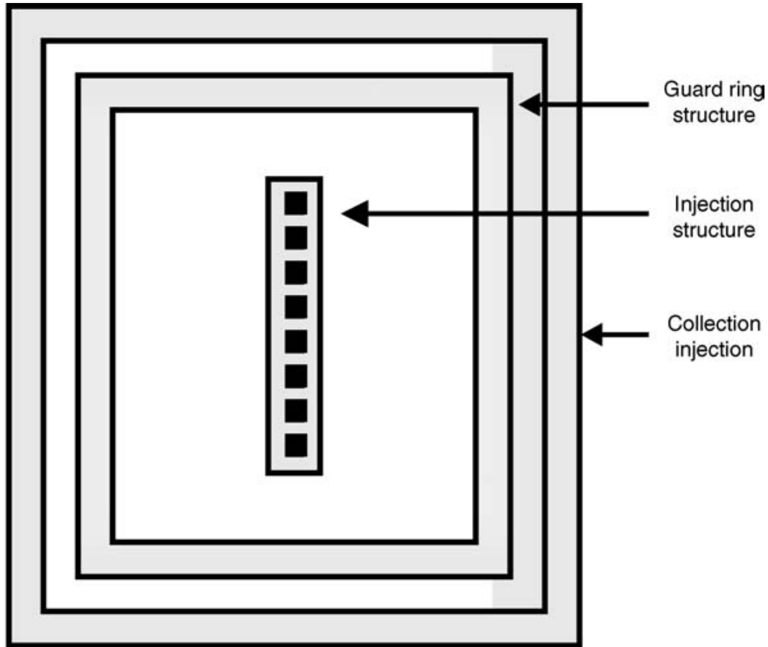



**Figure 4.41** Guard ring efficiency test structure.

characteristics can be quantified; the forward and reverse bipolar current gains can be evaluated as if this generalized region was actually a transistor structure [6, 7, 21]. Note that the n+ injector is enclosed inside the guard rings and n-well ring leading to asymmetrical results, just as is true that bipolar transistors have different forward and reverse characteristics. In this process, additional structures are placed in the 'base' region between the emitting and collecting structures. From this perspective, the cumulative structures between the emitter and collector are evaluated as '$\beta$ spoilers'. In this fashion, a bipolar-like model can be used to quantify the relationship between the injection structure and the collecting structure. From a bipolar perspective, the guard ring structure influences the 'base defect' term.

### 4.10.3 Guard Ring Theory – A Probability of Escape Perspective

A second perspective is a more phenomenological formulation from a probability view. The sum of the probability that an electron is collected plus the probability that an electron recombines equals unity

$$P(\text{recombine}) + P(\text{collected}) = 1.$$

Another perspective is that a minority carrier is either trapped or escapes from the guard ring. The sum of the probability that an electron is trapped and the probability that an electron escapes equals unity [1, 6, 7, 21].

$$P(\text{trapping}) + P(\text{escape}) + 1.$$

The probability that an electron is trapped is the probability of recombining or collecting within the guard ring structure (spatial region within the guard ring). As noted by Troutman, the probability to escape is the probability that an electron collects, or recombines outside the guard ring structure [1]. The probability a guard ring collecting an electron by a double guard ring structure is the current measured at the local p+ substrate ring and an n-type guard ring normalized to the injection current. The probability of an electron to escape from a guard ring, or series of guard rings, is the current measured at an additional ring outside the guard rings and the p+ substrate contact outside the guard ring normalized by the injection current.

## 4.10.4 Guard Ring Characterization – Electrical Measurements

In our methodology of evaluation of the guard ring characterization, the guard ring efficiency was of primary interest [1, 3, 6, 7]. The parameters that influence the measurement results are the following:

- the injection source type of event (e.g. diode injection, ionizing radiation, etc.);
- the injection source semiconductor region (e.g. n+ diffusion, n-well and sub-collector);
- the physical design of the injection source (e.g. dimensions of width and length);
- the injection source current and voltage magnitude;
- the collecting structure semiconductor region (e.g. n+ diffusion, n-well and subcollector);
- the collecting structure semiconductor design (e.g. width, length and width of the ring structure);
- the bias condition of the collecting structure (e.g. magnitude of the reverse bias junction).

The value of the transport factor (and the guard ring efficiency) is a function of the injection source type and its physical dimensions. In the experimental testing, the guard ring efficiency will also will be a function of whether one forces a current or a voltage during the sweeping of injection source. In the collection structure, the type of structure, bias conditions and physical design also influence the experimental results.

In this work, a test procedure needed to be defined to evaluate the measurement process. In our experimental work, the procedure was as follows [6, 7, 21]:

- Choose an injection source dimension. This remains ‘fixed’.
- Choose a collection region design/dimension. This remains ‘fixed’.
- Choose a guard ring structure.
- Sweep a forced current in the injection structure (with the collecting structure having a fixed dc voltage bias state).
- Repeat the forced current in the injection structure, with a new fixed dc voltage bias state (e.g. step different voltage states after each sweep of the forced current).
- Plot $F$ versus injection current for each fixed dc voltage bias step of the collection structure.

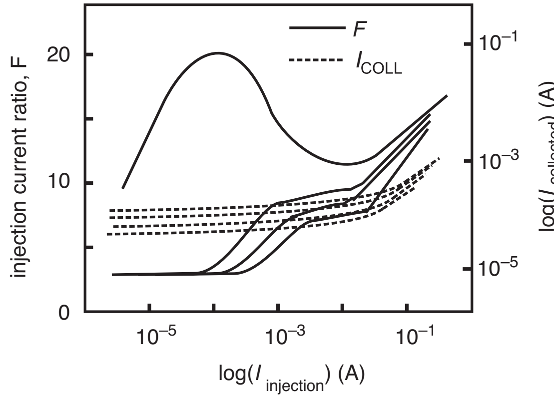



**Figure 4.42** Extraction of guard ring efficiency metric.

- Plot $1/F$ versus injection current for each fixed dc voltage bias step of the collection structure.
- Extract the value of $F$ and $1/F$ at 100 mA current level.

In the experimental work, the measurements were attempted by forward biasing the junction and injecting the current. The method of forcing a current was found to be a better methodology for initiation of the injection current. In each case, a fixed reverse-biased voltage was set on the reverse-biased collecting structure; this was stepped to a higher voltage with each successive sweep of the current. The injection ratio metric $F$ was evaluated on the left vertical axis and logarithm of the collector current on the right vertical axis, whereas the logarithm of the injected current was plotted on the $x$-axis. In guard ring evaluation, there are 'passive' and 'active' guard rings. Active guard ring structures and circuits will be discussed in Chapter 7. In some guard ring strategies, the guard ring is not biased. In case the guard ring itself can be biased, the extraction procedure of injection metric $F$ can include the biasing voltage of the guard ring structure. Examples of unbiased guard rings can include polysilicon-filled DT, 'floating' n-well regions and active guard rings not electrically connected to $V_{DD}$ or $V_{SS}$ (e.g. soft grounded circuits).

Figure 4.42 shows an example of the extraction of the $F$ for a guard ring structure. An injected current is initiated, discharging into the substrate region. The guard ring is placed between the injector and the collector. Different well bias values are chosen to evaluate the dependence on the guard ring well bias. The dashed line is the collected current, whereas the solid line is the injection ratio metric. As the value of the injected current increases, there is significant variation in the metric $F$. As the current increase, the metric $F$ follows a linear increase with increasing injection current. The slope of $F$ as a function of the injected current becomes a constant. As the reverse-bias voltage is modified, the metric $F$ also changes. At high injection currents, the difference in the metric $F$ as a function of the collector reverse bias is decreased. For our study, the 100 mA current level (e.g. JEDEC latchup current level) was chosen as an extraction value for the metric $F$.

Note, experimentally, we can establish the partial derivatives of the guard ring metric $F$. For example, the partial derivatives of interest are the change in the efficiency as a function of the injected current. A second partial derivative of interest is the change in the guard ring $F$ factor as a function of the bias on the collecting structure. Both these terms are evident experimentally in the testing of these guard ring test

**Table 4.3** Transient latchup results from a negative cathode pulse event with and without guard ring.

| Pulse width(ns) | Pulse height with n+ guard ring (V) | Pulse height without n+ guard ring (V) | Pulse height improvement with guard ring (V) |
|---|---|---|---|
| 10 | 2.85 | 1.2 | 1.65 |
| 12 | 2.2 | 0.9 | 1.3 |
| 15 | 1.8 | 0.8 | 1.0 |
| 20 | 1.45 | 0.7 | 0.75 |
| 30 | 1.1 | 0.67 | 0.43 |
| 50 | 0.75 | 0.65 | 0.1 |

structures. Evaluation of guard rings can also be done indirectly based on the influence of guard ring on the latchup measurements. The net improvement with the latchup guard rings can be evaluated based on 'with' and 'without' guard rings in the latchup structure. Tables 4.2 and 4.3 show the measurements from a transient latchup study where the pulse height to initiate latchup was evaluated as a function of pulse width, with and without a guard ring structure present. Table 4.2 contains transient latchup measurements for a negative polarity pulse on the n+ diffusion cathode within an p− epitaxial region. An external n-well resistor of 100 kΩ was placed on the n-well and no external resistor was placed on the substrate. Table 4.3 contains the transient latchup measurements for a positive polarity pulse on a p+ diffusion within a diffused n-well region. As can be observed, in the presence of a guard ring structure, a larger voltage pulse is required to initiate transient latchup. In both cases, a latchup can be defined for a positive and a negative polarity event. The metric can be a percent improvement or a voltage differential improvement. In Tables 4.3 and 4.4, the differential voltage improvements with the guard rings are shown. As the pulse width increases, the difference between the two cases decreases. Hence, if a 'guard ring improvement metric' is specified, it should be defined as a function of the pulse width.

## 4.11 LATCHUP FAILURE ANALYSIS TECHNIQUES

Failure analysis of latchup events is a key part of characterization and the understanding of semiconductor devices and electronics [54–82]. Latchup can occur in many locations within an integrated circuit. A key problem in the analysis of latchup is to determine the physical structures involved in the latchup event. For an example, in an peripheral I/O circuit, CMOS latchup can involve the I/O off-chip driver OCD, receiver or the ESD circuit.

Some latchup events are clear of the event without information in time. In some latchup events, it is also important to be able to watch the time evolution to determine how these structures interact and

**Table 4.4** Transient latchup results from a positive anode pulse event with and without guard ring.

| Pulse width(ns) | Pulse height with n+ guard ring (V) | Pulse height without n+ guard ring (V) | Pulse height improvement with guard ring (V) |
|---|---|---|---|
| 10 | $> 3$ | 2.5 | |
| 15 | $> 3$ | 1.6 | |
| 20 | $> 3$ | 1.45 | |
| 30 | $> 3$ | 1.1 | |
| 50 | 3.2 | 1.0 | 2.2 |
| 100 | 2.1 | 0.95 | 1.15 |

involve in the latchup process. In this case, a tool that provides both spatial and temporal visualizations may be critical to the understanding the latchup event. A characterization tool that can provide significant insight into the spatial and temporal dynamic responses of the latchup process in a complex network is critical in some integrated semiconductor chip design.

A second key problem in failure analysis and root cause evaluation is whether an event is an EOS or a latchup event. In latchup and EOS events, significant destruction of the module and package can prevent visualization of the location of the failure of the semiconductor. Latchup can lead to damage to the silicon chip as well as the package, preventing an evaluation of the process involved. The failure can be so severe that the package can melt or physically delaminate and lead to explosion. In smart power applications and power applications, voltage levels can exceed 10 kV voltage levels and 1 kA current levels. In multichip IGBT power applications, at high current, complete failure of the entire assembly occurs. In electrical characterization, the determination of whether the event is latchup or EOS can only be deciphered from the terminal conditions. Hence, having a characterization tool that provides full spatial and temporal understanding can allow us to discover interactions and the chain of events that lead to semiconductor chip failure. For transient latchup, the transient latchup event is initiated by a dynamic pulse event; as a result, the understanding of transient latchup may require both the spatial and the temporal understanding. The dependence of the ramp rate and the propagation of latchup may be critical to fully understand the latchup failure.

## 4.11.1 Failure Analysis – Optical Microscope

A standard device to evaluate latchup failures is optical microscope. Latchup events can be visualized using optical microscopes by delayering and removing various levels. Figure 4.43 is an example of an

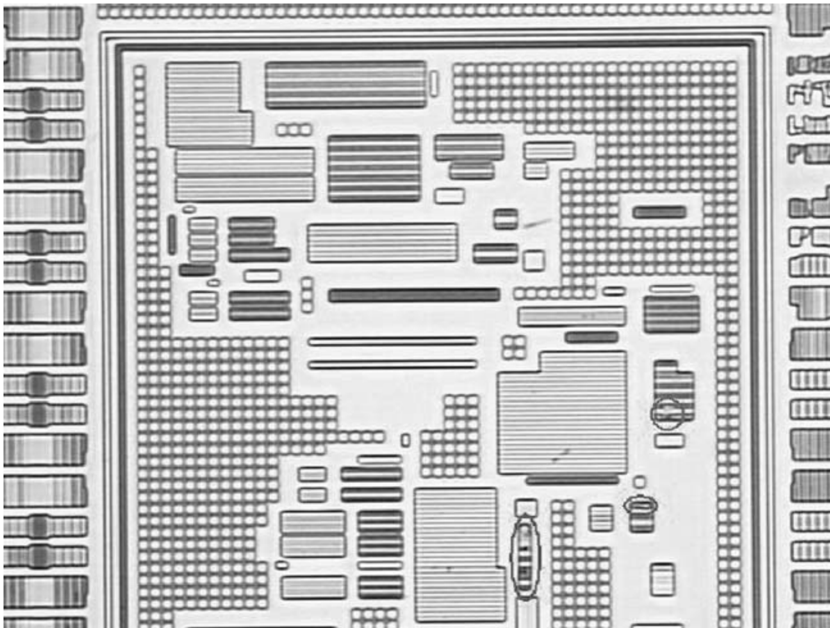


**Figure 4.43** Optical microscope image of CMOS latchup event damage. Reproduced by permission of IBM Corporation.

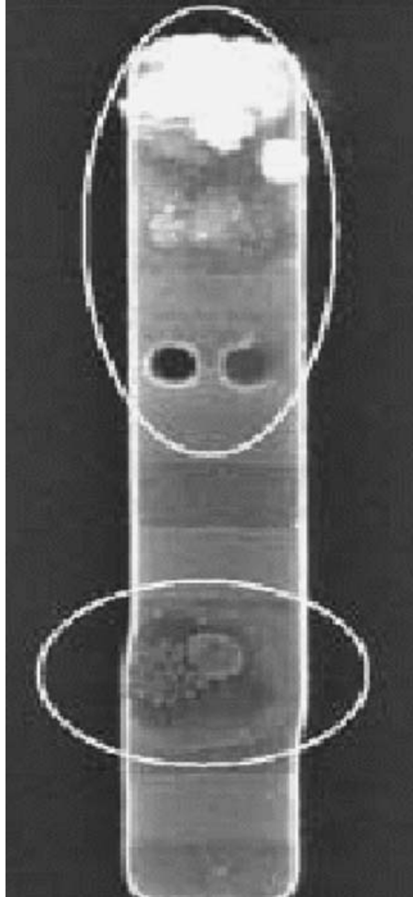


**Figure 4.44** Optical microscope image of CMOS latchup damage. Reproduced by permission of IBM Corporation.

optical image; all metal levels were removed to provide evidence of silicon failure. In the figure, it can be observed that three different circuits show evidence of silicon damage from a latchup event that was spatially distributed [62].

Figure 4.44 shows an example of one of the damaged transistor regions. In many cases, the latchup event damages the package, metallization levels and silicon such that determination of the latchup failure is not possible. Also in this example, three locations were determined that were involved in the event; from this methodology, the time evolution of the latchup event is not clear [62].

## 4.11.2 Failure Analysis – Emission Microscope (EMMI) Tool

A static technique that provides dc visualization is static emission microscopy. Static emission microscopy has been used since the early 1990s as a method of observing failure modes in integrated circuits of many types [54–74]. Many common failure modes result in an abundance of hot electrons,

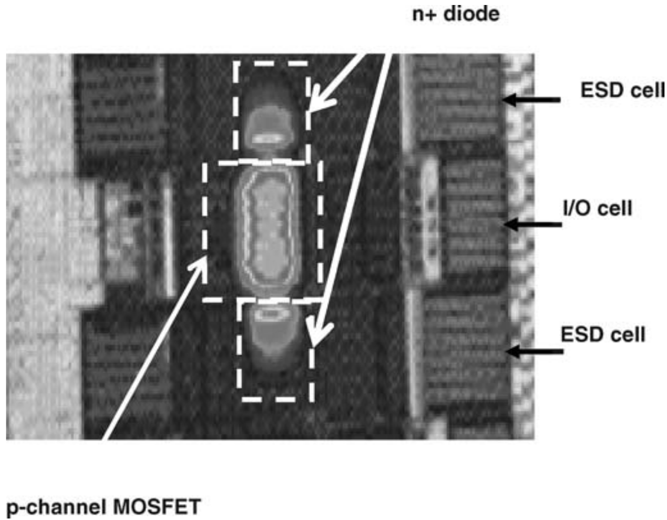



**Figure 4.45** EMMI tool optical emissions from a I/O-to-I/O CMOS latchup event. Reproduced by permission of ESD Association.

which in turn causes large numbers of near-infrared (NIR) photons to be emitted. When the circuit is observed under a microscope, this NIR emission is readily detectable by a CCD or similar device. Although little or no dynamic information is available, the confirmation and location of the defect is an enormous benefit. Figure 4.45 shows an example of optical emissions from an integrated circuit where latchup occurred between two adjacent circuits. Latchup occurred between a p-channel MOSFET and n+ diode elements; this occurred between neighboring peripheral I/O cells [75].

As a second EMMI tool image, a case of latchup between power supplies is shown in Figure 4.46. The photon emissions are evident in the photograph of the semiconductor chip (dashed rectangle is

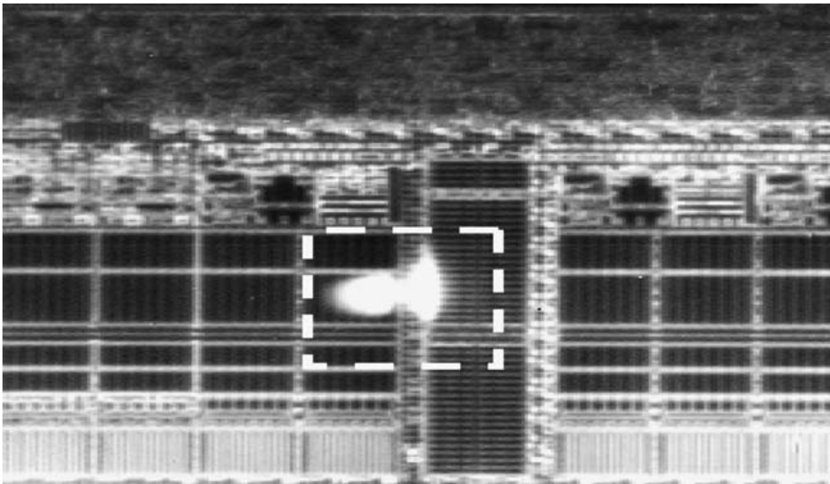


**Figure 4.46** Latchup event emissions between two devices at different power supply voltages. Reproduced by permission of ESD Association.

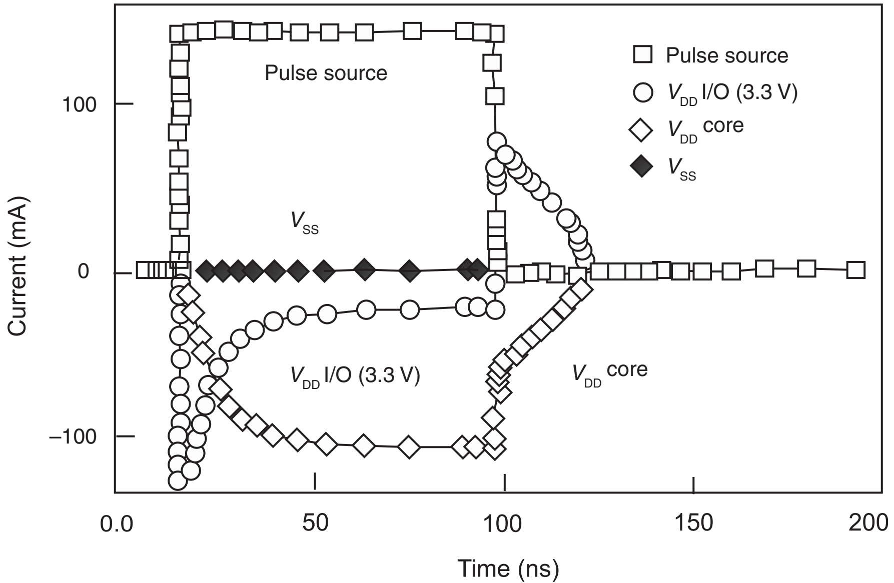



**Figure 4.47** Simulation of the CMOS latchup event between two devices after contacts were removed. Reproduced by permission of ESD Association.

around failure mechanism). With mixed signal design requirements, peripheral circuits powered at different power supply voltages are integrated into a common design. For instance, a chip can have a peripheral I/O power supply of 3.3 V and internal networks at 1.2 V. In core-limited design floor plans, I/O slots can be occupied by decoupling capacitors to reduce core power supply noise. In this example, the CMOS latchup event occurred between a p-channel MOSFET at 3.3 V and the n-well of a decoupling capacitor at 1.2 V (*note*: a p-type decoupling capacitor inside a 1.2 V n-well). In the figure, the lower left region is the p-channel MOSFET and the right side is the decoupling capacitor; the latchup event occurred between the two regions. To address the problem, contacts were removed from the decoupling capacitor n-well to increase the series resistance in the n-well; this introduces npn emitter series resistance [75]. Figure 4.47 shows the simulation of the interaction between the power supply rails after n-well contact removal.

Failure analysis and design layout review revealed that the OCD p-channel transistors pull-ups in the 3.3 V I/O cell and PMOS decoupling capacitor cell powered by 1.2 V form a parasitic pnpn structure. To verify this failure mechanism, latchup testing with varying core $V_{DD}$ voltage was performed at an ambient temperature of 25°C with an injection current of 100 mA. During the latchup verification, the core $V_{DD}$ power rail was increased. As the differential voltage between the two power rails decreased, the number of failed pins reduced (Table 4.5).

## 4.11.3 Failure Analysis Technique – EMMI Tool Stellari Animation Methodology

Stellari *et al.* [81] developed an EMMI methodology that allowed for better visual capability of the latchup. The JEDEC78 technique was modified to improve the time resolution of latchup evaluation,

**Table 4.5** Latchup test results with different core $V_{DD}$ bias conditions.

| I/O | Logic state | 1.2 V | 1.4 V | 1.5 V | 1.6 V |
|---|---|---|---|---|---|
| A | High | Fail | Pass | Pass | Pass |
| | Low | Fail | Fail | Pass | Pass |
| B | High | Fail | Pass | Pass | Pass |
| | Low | Fail | Fail | Fail | Pass |
| C | High | Fail | Pass | Pass | Pass |
| | Low | Fail | Fail | Fail | Fail |

similar to the method developed using the TLP–PICA methodology. Figure 4.48 shows the modified emission microscope test system. A CCD camera, power supply, pin current supply and a packaged configuration that allowed for flip-chip backside evaluation. The semiconductor chip was packaged on a temporary chip attach (TCA) carrier, allowing the flip-chip backside evaluation. The semiconductor chip was ground to 100 μm thickness to allow observation of the near-infrared photon emissions from the back of the wafer. The power supply current monitoring is achieved with a current meter, as the signal pins are stressed with a variable current source. Using a Hamamatsu C4880-21 back-illuminated CCD camera, time-integrated images of photon emissions from latchup can be evaluated.

In the testing of the semiconductor chip, the variable current source could evaluate the propagation of latchup spatially as a function of the applied current magnitude. Figure 4.49 shows an example of the photons mapped on the semiconductor design layout. In this fashion, one can visualize the location of the photon emissions and the semiconductor device and circuit locations. One of the difficulties with the methodology is the distinction of forward-biasing versus actual latchup. Note within the I/O region, all the regions that are emitting photons are forward biased. In the I/O physical region, an ESD element, a MOSFET and four n-doped ballast resistors are injecting current into the substrate due to forward biasing. Like in the work of Weger and coworkers [79, 81], this work shows that the location of the photon emissions in the logic area is in the region of the ESD n-well diode element. From this work, it is clear that the injection source initiates the onset of the forward biasing of the logic array, leading to the photon emissions; this is evidence of latchup propagation [80].

Stellari *et al.* also showed that the propagation occurs away from the local n-well contacts [82]. Figure 4.50 shows the photon emission intensity spatially mapped on the dense CMOS logic. The bars are the location of the n-well contacts. As the spacing increases from the n-well contacts, the photon emission intensity increases, indicating that the emissions are related to the spacing away from the n-well contacts in the CMOS inverter area.

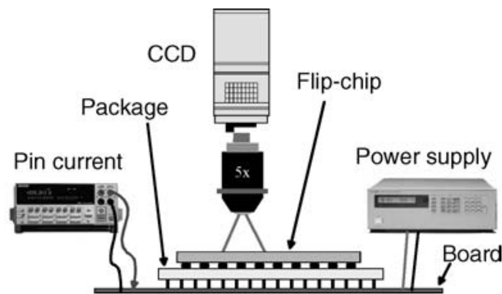



**Figure 4.48** Flip-chip EMMI tool test system. *Source*: [82]. Reproduced by permission of IBM.

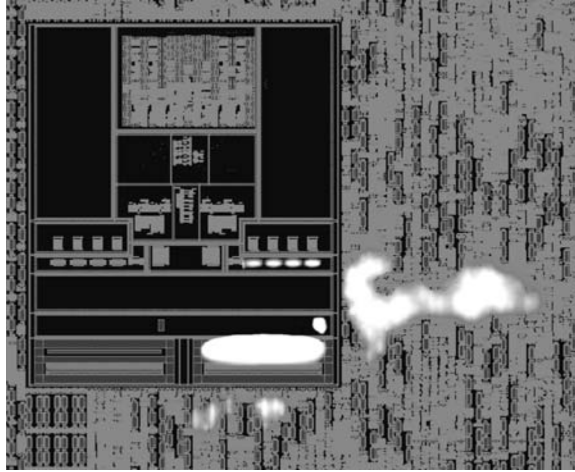


**Figure 4.49** Emission microscope (EMMI) tool example of latchup propagation. *Source*: [82]. Reproduced by permission of IBM.

### 4.11.4 Failure Analysis Technique – Scanning Superconducting Quantum Interference Device (SQUID)

A second technique that provides evidence of latchup events is the failure analysis tool called the scanning superconducting quantum interference device (SQUID). The SQUID technique provides a

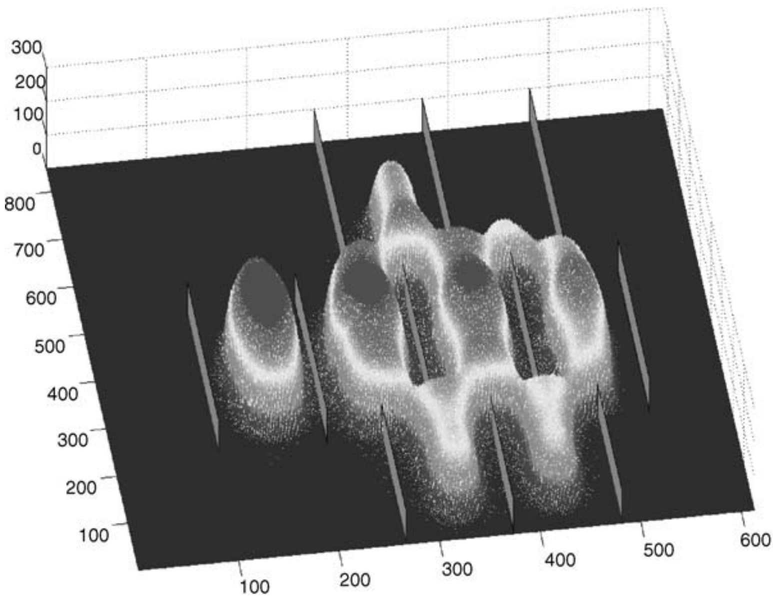



**Figure 4.50** Photon emissions mapped on the logic area. The n-well contacts are highlighted.

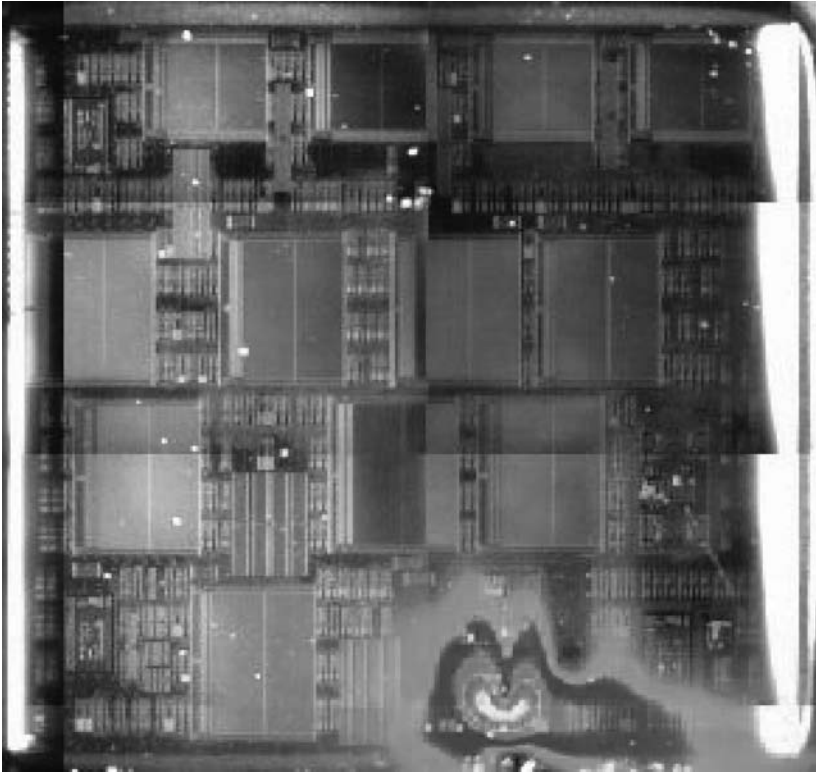


**Figure 4.51** Superconducting quantum imaging device image of a CMOS latchup event.

two-dimensional mapping of high currents in a semiconductor chip. Figure 4.51 shows the SQUID image. The light-colored region in the center is the region of highest field intensity. The extension of high field intensity beyond the peripheral circuit indicates that the latchup event extended beyond the initial I/O circuits. This technique allows for verification of the area of the high currents, but optical microscope and delayering are required to determine the exact network involved in latchup [62].

## 4.11.5 Failure Analysis Techniques – Modified Picosecond Imaging Circuit Analysis Tool – the TLP–PICA Methodology

An alternative tool that allows both spatial and temporal analyses is known as the picosecond imaging circuit analysis (PICA) tool (Figure 4.52). The PICA tool provides time-resolved emission microscopy [77–81]. PICA has been used extensively for characterization, timing verification and failure analysis of CMOS VLSI circuits. Mechanically, a PICA system may closely resemble a static emission microscope (EMMI tool), the difference lying in the use of a detector capable of resolving single photons on a picosecond timescale. A typical PICA system is used in one of the two modes. First, the DUT may be placed in a normal or standard mode of operation. Alternatively, the DUT may be

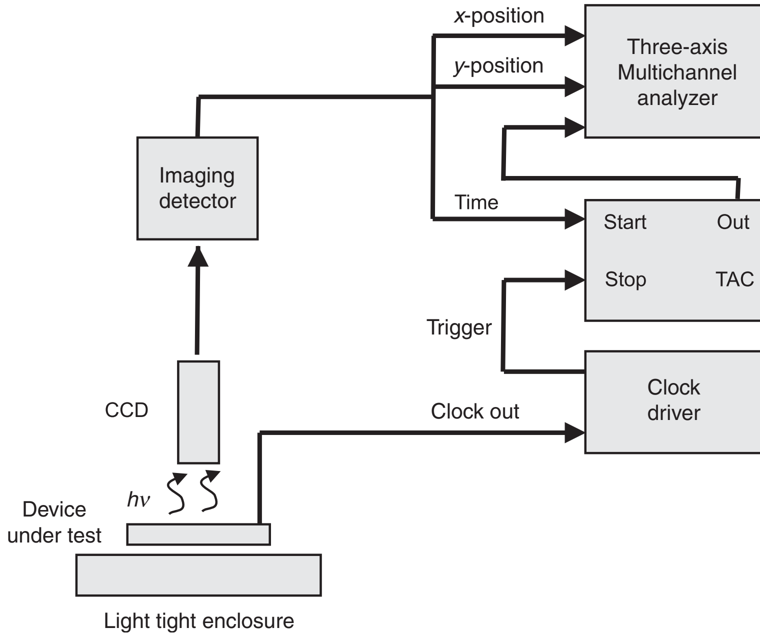



**Figure 4.52** Picosecond imaging analysis system.

attached to a tester and stimulated by the use of test vectors. In either case, the resulting periodic photon emissions are detected and correlated with a trigger such as a circuit output, divided clock or a tester output signal. In commercial practice, PICA systems tend to use one of the two types of detector: single element (single pixel) or imaging (area or multipixel). Each detector, and consequently the resulting PICA system, has advantages and disadvantages. Single-element detectors sometimes have higher quantum efficiency and superior time resolution, whereas imaging detectors permit identification of many distinct spatial emission sources in the same measurement.

In many cases, the exact spatial location within the DUT is known, and hence a single-point detector may be employed. In other cases, the source of the light emission is not known a priori or multiple points may emit at nearly the same time; in these cases, the visualization provided by an imaging detector may prove advantageous, especially when IR photon emission rates are high. In the imaging PICA system used in the test setup, the DUT is observed via a microscope objective. The spatial resolution of the imaging PICA system is a function of the spatial resolution of the Mepsicron detector and the optical properties of the microscope. A Mepsicron detector is very high quality detector with a sophisticated array of microchannel electron multipliers. First, the circuit is imaged on the photocathode of the Mepsicron. Timing is derived from the electrical feed to the internal microchannel plate electron multiplier. Finally, spatial position is derived from charge amplifiers at the four anode corners that are fed to a position analyzer that generates voltage signals proportional to the $x$- and $y$-coordinates of the detected photon on the photo cathode. Digital signals representing the output of the time-to-amplitude converter (TAC) and the $x$- and $y$-coordinates of the detected photon are stored in a three-parameter multichannel analyzer for subsequent display and analysis of optical waveforms and time-resolved images [77–80]. The 60-μm spatial resolution of the optical detector coupled with a diffraction-limited 100× microscope objective and light emission at a wavelength of approximately 1 μm results in spatial resolution on the order of 1 μm over a field of view of 0.25 mm.

The full PICA histogram typically uses 512 $x$- and $y$- coordinate values. The time resolution of 8192 time values can provide a time resolution of as little as 10 ps, corresponding to a bandwidth in the GHz range (10 GHz range).

One of the limitations of the CMOS latchup analysis is that the destructive nature of the CMOS latchup failure prevents an understanding of where the failure occurred and how it evolved. In severe latchup events, the silicon, metal and packaging materials are destroyed to the extent that failure analysis and the determination of the failure location is too difficult to observe. Additionally, this evolution in time of events that lead to latchup is also not observable from the terminal conditions.

Voldman proposed the application and integration of the PICA method with the TLP pulse method to address this problem of lack of spatial and temporal information [79]. Typical applications of the PICA tool is for semiconductor chip design at nominal voltage conditions where the chip is functionally exercised to evaluate timing issues. In the PICA tool, the functional signal patterns are processed in the semiconductor chip during a powered state. The repeated signal patterns allows for the collection of photons. During a TLP test, the semiconductor chip is typically in an un-powered state. The TLP method applies a high-current pulse to a device and measures the current- and voltage-state across the device under test. In this methodology, a step stress is practiced where each successive step increases the applied pulse magnitude; this continues to semiconductor device failure. This method, cannot be used for photon collection processes since the methodology does not apply a pulse train but only a single-pulse step-stress methodology for evaluation of the current and voltage across the semiconductor device.

To apply the PICA tool to latchup, a similar procedure is required. The pulse event must be a repeated event, to allow collection of the photon emissions. In the synthesis of these two methods, the PICA method and the TLP method, a new methodology and tool is developed for the evaluation of latchup. First, using the high-current pulse source from the TLP system, a high-current pulse train can be used to collect photon emissions at higher current magnitudes for states before latchup is initiated. Second, by increasing the current magnitude of the pulse train, the current can be increased at different steps to show the evolution of latchup. Third, with the high-current step-stress method, latchup can be evaluated in time. Fourth, since this method can be performed at a wafer level or product level, the combined tool will allow for visualization of the latchup event.

The TLP–PICA methodology allows for visualization of evolution of latchup events spatially and temporally by reconstruction of an animation movie in a picosecond time regime. Additionally, the synthesis of the TLP and the PICA method allows for the extension of the ‘ESD TLP methodology’ to spatial and time domain analysis for electrical characterization and reliability analysis. At the same time, the high-current pulsed source extends the utilization of the PICA methodology for failure analysis at wafer and chip levels.

The technology and science behind PICA rely on the dramatic improvement in measurement methods for temporally resolving extremely faint optical signals on picosecond timescales, as well as our rapidly expanding knowledge of ultrafast processes. For latchup, the PICA tool can capture both the electron–hole pair (EHP) recombination to parasitic bipolar switching. A hybrid TLP–PICA technique was developed to answer this need.

Since photon emission in PICA is a statistical process, it is necessary to collect time-correlated photons over many cycles of a periodic stimulus/response. If the response to a TLP-like condition is to be measured with PICA, a pulse train must be continuously applied to the DUT. Care must be taken to ensure that a persistent mode of failure such as latchup does not occur, as in that case the periodic nature of the DUT response would be destroyed and the resulting time-averaged data would not reveal the onset of the latchup event. In the case of current injection induced latchup, with judicious choice of terminal voltage $V_{DD}$, it is possible to enter a quasi-stable state that will persist only while the injection stimulus is applied. Once the electron injection ceases, a new equilibrium in charge, current and voltage distribution is established. If this state occurs and the injection pulse width and delay are

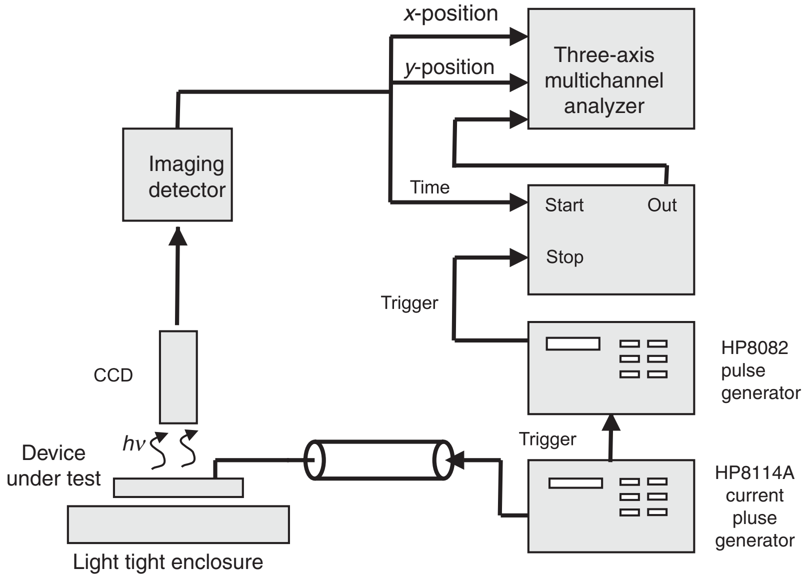



**Figure 4.53** Transmission line pulse–pico-second imaging circuit analysis tool.

chosen appropriately, both the onset and decay of the resulting latchup event can be captured by the TLP–PICA tool.

The first demonstration of this method was achieved by Weger *et al.* [79]. Figure 4.53 shows a block diagram of our TLP–PICA system [79,81]. An HP8114A high-current pulse source is used to establish a high-current pulse train. The HP8114A pulse source provides variable pulse widths at high currents with a fixed rise and fall time. To evaluate latchup sensitivity as a function of the rise time (e.g. slew rate), other equivalent pulse sources can be used. A pulse source with a variable rise time can be used to provide different ramp rates that influence transient latchup, as discussed in Chapter 3.

The resulting TLP–PICA system can be run according to a number of different test modes. The following test method can be utilized in different operational modes of the pulse train variables [79,81]:

- pulse number;
- pulse height (plateau);
- pulse rise time;
- time between successive pulses;
- pulse step height (e.g. increment levels).

To study latchup, a high-current pulse train was initiated on the DUT to evaluate the circuit response. The HP8114A pulse source was used as the high-current source in a high-output impedance mode driving a 50 Ω BNC coaxial cable into a 50 Ω resistor element in series with the input pin. A delayed trigger for the PICA was provided by a HP8082A network [79]. The test used a packaged

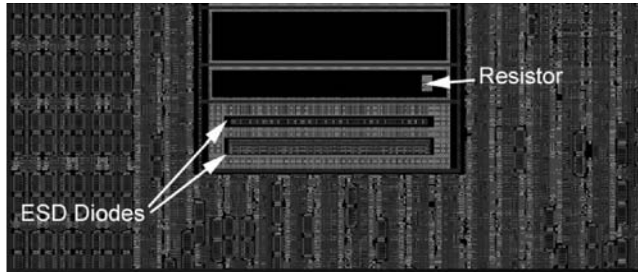



**Figure 4.54** CAD layout highlighting the ESD diodes, the I/O book and the CMOS logic circuitry.

product chip that was backside thinned to 70 μm of silicon to allow for evaluation of the photon emissions from the backside for evaluation of ESD-induced latchup.

For the evaluation of latchup, a 10 kHz pulse train was used with a range of current magnitude between 100 mA and 1 A. In our case, the pulse amplitude of interest was between 50 and 500 mA. In a given study shown below, the peak pulse amplitude was 370 mA and the latchup source pulse width was 500 ns. The photon detection rate through 70 μm of silicon using a Mepsicron with Super S25 photocathode and 100 × 0.5 numerical aperture (NA) objective was roughly 7000 photons per second or approximately 0.7 photons per current pulse. Notice that in this setup, as only one photon can be processed per trigger event, it is impossible to achieve more than 1.0 photons per current pulse. As the TAC range is set to 800 ns, a count rate below 12,500 photons/s is necessary to ensure that spectral distortion will not exceed 1 % [79].

In this methodology, the actual design of the semiconductor chip and the photon data can be overlaid to pinpoint the location of the CMOS latchup process. Figure 4.54 shows the computer-aided design of the semiconductor chip. In this example, the I/O book is an 'array I/O' design. In array I/O design practices, the I/O books are placed within the logic and not on the chip periphery. This physical design highlights dense CMOS logic circuits, with an I/O book containing dual-diode ESD elements and guard ring defining the book edge.

The pulsed input pin consisted of circuitry that included a dual-diode ESD circuit and the I/O OCD network. In our study, a negative pulse train was injected into a signal pin. The negative pulse leads to forward bias of all n-type diffusions and injection into the p− substrate. The n-diffusions in the circuit include the ESD n-well diode, the ballasting resistors and the off-chip driver n-channel MOSFETs. During the test, the evolution of the various elements can be observed in time.

A video animation was made to observe the temporal response of the network as well as the level of forward biasing of the structures in the chip substrate. For a negative polarity, the photon emissions can be observed surrounding the ESD n-well/substrate diode structure (Figure 4.55). The injection of the n-well diode was uniform across the entire structure. A Gaussian-like photon distribution was evident from the diode structure as a result of the current distribution and the recombination–generation statistics in the bulk substrate wafer.

As the electron injection pulse continues, an initiation of a latchup event becomes evident outside the I/O area (see Figures 4.56 and 4.57). Electrical testing indicated that latchup could be induced in the test structure if more than −400 mA was extracted from the I/O pin.

Figure 4.57 highlights the onset of latchup within the gate array area to the right of the I/O book ESD diode structure. At this time, the ESD diode structure injects current into the substrate and the injection process triggers latchup in adjacent circuitry. The latchup event occurred significantly after the onset of the photon emissions of the ESD element, as evidenced by the optical waveforms shown in Figure 4.58 [79, 81].

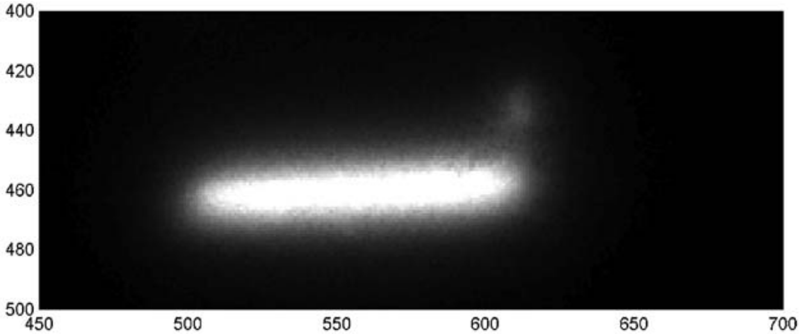



**Figure 4.55** TLP–PICA photon emissions of ESD n-well diode structure early in the injection process.

Figure 4.59 shows that after the pulse, the photon emissions in the ESD structure due to recombination–generation in the substrate is reduced, while the photon signal from the latchup process continues.

In this technique, additional information can be obtained by using optical frequency filters. Using optical frequency filters, the energy spectrum and the physical state (e.g. forward bias mid-band recombination–generation emissions versus avalanche breakdown emissions) can be distinguished. Hence using this method, one can distinguish the physical interaction that is occurring in latchup (whether it is a forward bias injection or breakdown phenomena).

In summary, a latchup characterization practice and methodology is as follows:

- backside grind semiconductor chip to visualize photon emissions;
- use a high-current pulse source with a given pulse characteristic (e.g. pulse width, frequency, plateau and rise and fall time) as an injection source to simulate CMOS latchup events;
- integrate high-current pulse source with an imaging detector, multichannel analyzer and scanning system;
- establish a trigger network to synchronize the pulse source and the imaging system;

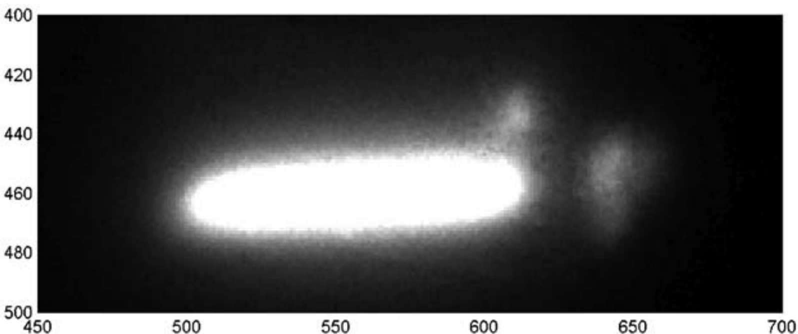



**Figure 4.56** TLP–PICA photon emissions of n-well diode and latchup, showing the onset of the latchup event (to the right-hand side of the ESD diode).

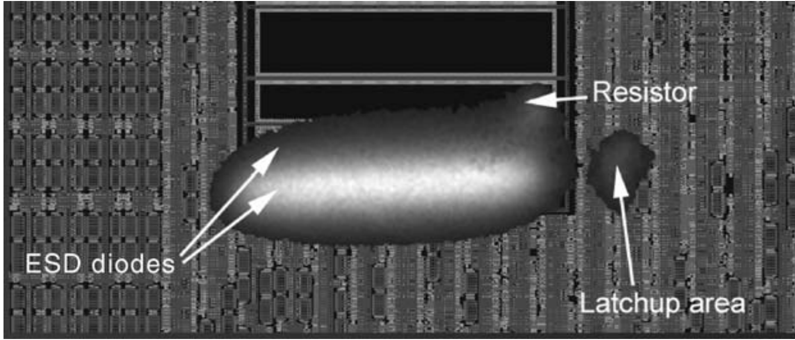



**Figure 4.57** Computer-aided design layout with photoemission of Figure 4.54 overlaid.

- provide a high-current 'pulse train' step-stress and store photon emission data;
- store information in animation video;
- store the computer-aided design layout plot;
- overlay the two images (CAD data layout design and photon emission mapping) to observe the location of the CMOS latchup event.

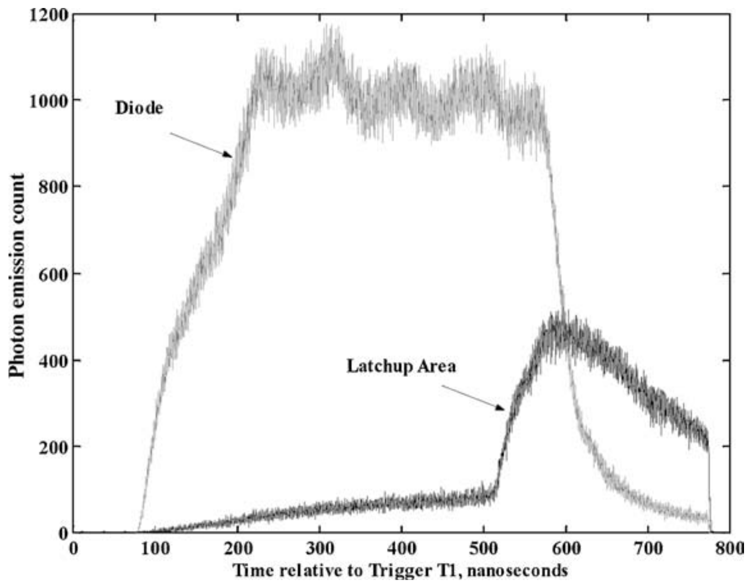



**Figure 4.58** Optical waveforms of light emission in ESD diode area versus that in the latchup area. Note a delay of about 400 ns prior to rapid rise in photoemission in latchup area.

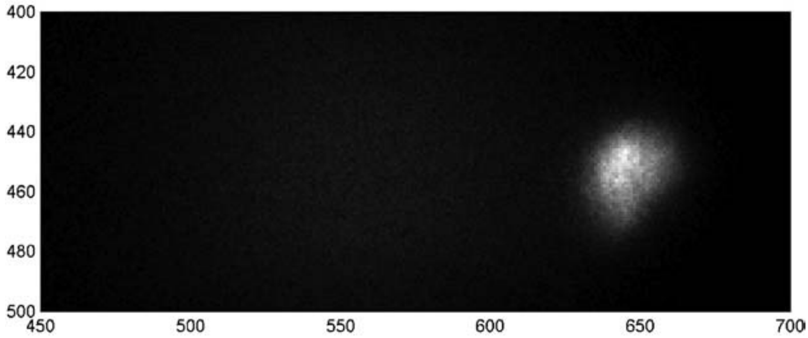



**Figure 4.59** TLP–PICA photon emissions showing the sustained latchup condition and low level bulk–substrate photon emissions after the TLP pulse response.

## 4.11.6 Failure Analysis – Thermal Imaging Techniques

For single-event latchup (SEL), it is difficult to determine the locations of latchup during testing. Thermal imaging techniques are utilized to determine the location of latchup during accelerated particle testing. This technique includes the following:

- infrared (IR) sensitive cooled detectors;
- microscope with an optical system that transmits IR signal;

One of the problems with this methodology is that the presence of the metallization impacts the emissivity. The lack of calibration prevents accurate determination of the temperature distribution, but latchup sensitive areas can be identified by using the variation in the thermal emissivity. By evaluating the thermal emissivity in powered and unpowered states, calibration can be achieved. Using two thermal maps, the hot areas associated with latchup can be identified. The difference between the two thermal images can be used to provide the actual temperature at the surface of the semiconductor device. Hence in SEL testing, the thermal characteristics and the mapping can determine the regions where latchup is occurring. Johnston and Miyahira used this technique to quantify SEL in an analog-to-digital converter circuit [83]. Irradiation was initiated with a Californium 252 ion source. After each latchup event, the sample was removed from a vacuum chamber with the chip. The chip was actually removed from the vacuum chamber in a powered condition to remain in latchup. The thermal mapping was then completed in an air environment that corresponded to the equilibrium condition in the nonvacuum environment.

# 4.12 SUMMARY AND CLOSING COMMENTS

This chapter discusses the latchup characterization structure used as a vehicle for latchup evaluation and benchmarking a semiconductor technology. The test structure and design methodology are discussed to quantify the latchup robustness of a technology.

In the next chapter, Chapter 5, semiconductor process solutions to improve the latchup robustness are discussed. Semiconductor process and structures will be shown that influence the latchup robustness of a technology; this includes a plethora of solutions that starts from the base substrate

wafer and works up to the n-wells, p-wells, buried implants, buried subcollectors, triple-well implants, isolation structures to silicide films. The chapter will address the introduction of the new physical regions and the evolution with technology scaling.

## PROBLEMS

1. Given a four-stripe latchup structure, latchup measurements were taken using external resistor elements and a bipolar curve tracer. The holding voltage, $V_H$, of 10 V was kept constant as the external resistor elements were swept. In the table, the external resistor values are shown for a given latchup structure (e.g. a fixed p+/n+ space) for the $V_H = 10$ V holding voltage. Plot the data on a log $R_{well}$ versus log $R_{substrate}$ plot.
2. Using the same structure as discussed in Problem 1, a second set of measurements was taken to evaluate the holding voltage contour at $V_H = 2.5$ V in resistance space. Draw the $V_H = 2.5$ V contour on the same plot as above. Explain the trend of the holding voltage contours in shunt resistance space. What happens to the contour as the p+/n+ spacing decreases or increases? Why?
3. Design the latchup pnpn structure for a test site for latchup characterization, assuming a technology with one level of metallurgy. How do you wire the latchup structure? What are the design issues? How does the current distribute in the structure and how does this effect the latchup results?
4. Design the latchup pnpn structure for a test site for latchup characterization, assuming a technology with more than one level of metallurgy. How do you wire the latchup structure? What are the design issues? How does the current distribute in the structure and how does this effect the latchup results? How do you wire the design so that the wire layout does not influence the latchup results?
5. Using 25 pads, design a set of pnpn structures of different spacings. How many different pnpn space sizes can be constructed within a $1 \times 25$ pad set? What do you do with the 25th pad?

**Table 4.6** External n-well resistance versus external substrate resistance for $V_H = 10$ V (Problem 1).

| External n-well resistance (k Ω) | External substrate resistance (k Ω) |
|---|---|
| 1000 | 1.7 |
| 500 | 1.7 |
| 200 | 1.7 |
| 50 | 1.8 |
| 20 | 2.0 |
| 10 | 2.5 |
| 5 | 4.5 |
| 4 | 7.5 |
| 3.75 | 9.6 |
| 3.50 | 13.5 |
| 3.25 | 25.0 |
| 3.0 | 500.0 |

**Table 4.7** External n-well resistance versus external substrate resistance for $V_H = 2.5\,V$ (Problem 2).

| External n-well resistance (k Ω) | External substrate resistance (k Ω) |
|---|---|
| 1000 | 5.58 |
| 500 | 5.63 |
| 200 | 5.80 |
| 100 | 6.07 |
| 50 | 6.70 |
| 20 | 9.9 |
| 15 | 13.5 |
| 10 | 46.0 |
| 9.5 | 70.0 |
| 9.3 | 90.0 |
| 9.0 | 160 |
| 8.8 | 360 |

6. As in Problem 3, how many different substrate spacings can be obtained when one utilizes the spacings of the next structure set (as discussed in Section 4.2.7, the p+ substrate of the next structure can be utilized for a different substrate contact spacing)?

7. Using 25 pads in a $1 \times 25$ macro, given one wants to vary the well and substrate spacings for a one fixed p+/n+ spacing, how many different well and substrate spacings can be constructed?

8. Explain the trade-offs (pros and cons) between the two test methods, the method discussed in Problem 5 versus the one discussed in Problems 3 and 4.

9. Design a triple-well latchup structure with a buried n− region that extends under the n-well. How many electrodes do you need? How do you wire the structure? How do you test it? What are the test conditions?

10. Design a latchup pnpn structure with a polysilicon deep trench structure. Assume that the polysilicon trench region is opened to electrically bias the polysilicon region, how do you design the structure? What masks do you modify? How do you wire the structure?

11. In the design of inverters, the orientation of the gate structure cross the n-well to p-well edge is perpendicular. Show an inverter circuit layout and the orientation relative to the perimeter. How does this real layout compare to the latchup test structures? Design a test structure that represents an inverter for latchup evaluation.

12. An n+ internal guard ring is constructed within an n-well for a p-channel MOSFET of an inverter. Explain the problem that occurs when the n+ internal guard ring is a ring around the entire n-well. Draw the entire inverter circuit and discuss the issue. Hint: draw all the design levels.

13. A p+ internal guard ring is constructed within a p-well for an n-channel MOSFET of an inverter. Explain the problem that occurs when the p+ internal guard ring is a ring around the entire p-well. Draw the entire inverter circuit and discuss the issue. Hint: draw all the design levels.

14. An n+ internal guard ring is constructed within an n-well for a p-channel MOSFET of an inverter. A p+ internal guard ring is constructed within a p-well for an n-channel MOSFET of an inverter.

Explain the problem in designing an inverter. How can one overcome the problem? What is the design impact?

## REFERENCES


1. R.R. Troutman. Epitaxial layer enhancement of n-well guard rings for CMOS circuits. *IEEE Electron Device Letters*, **4**(12), 1983, 438–440.
2. R.R. Troutman. *Latchup in CMOS Technology: The Problem and the Cure*. Boston: Kluwer Academic Publishers; 1986.
3. J. Quinke. Novel test structures for the investigation of the efficiency of guard rings used for I/O latchup prevention. *Proceedings of the International Conference on Microelectronic Test Structures* (*ICMTS*); 1990. p. 35–40.
4. D. Tremouilles, M. Bafluer, G. Bertrand and G. Nolhier. Latch-up ring design guidelines to improve electrostatic discharge (ESD) protection scheme efficiency. *IEEE Journal of Solid-State Circuits*, **39**(10), 2005, 1778–1782.
5. S. Voldman, C.N. Perez and A. Watson. Guard rings: theory, experimental quantification, and design. *Proceedings of the Electrical Overstress/Electrostatic Discharge (EOS/ESD) Symposium*, October 2005. p. 131–140.
6. S. Voldman, C.N. Perez and A. Watson. Guard rings: structures, design methodology, integration, experimental results, and analysis for RF CMOS and RF mixed signal silicon germanium technology. *Journal of Electrostatics*, **64**, 2006, 730–743.
7. D. Tremouilles, M. Scholz, G. Groseneken, M.I. Natarajan, N. Azilah, M. Bafluer, M. Sawada and T. Hasebe. A novel method for guard ring efficiency assessment and its applications for ESD protection design and optimization. *Proceedings of the International Reliability Physics Symposium (IRPS)*; 2007. p. 606–607.
8. T. Cavioni, M. Cecchetti, M. Muschitiello, G. Spiazzi, I. Vottre and E. Zanoni. Latch-up characterization in standard and twin-tub test structures by electrical measurements, 2-D simulations and IR microscopy. *Proceedings of the International Conference on Microelectronic Test Structures* (*ICMTS*); 1990. p. 41–46.
9. C. Mazure, W. Reczek, D. Takacs and J. Winnerl. Improvement of latching hardness by geometry and technology tuning. *IEEE Transactions on Electron Devices*, **35**(10), 1988, 1609–1615.
10. Y. Song, J.S. Cable, K.N. Vu and A.A. Witteles. The dependence of latchup sensitivity on layout features in CMOS integrated circuits. *IEEE Transactions of Nuclear Science*, **33**(6), 1986, 1493–1498.
11. R. Lohia and A. Ali. Parametric formulation of CMOS latchup as a function of chip layout parameters. *IEEE Journal of Solid State Circuits*, **23**(1), 1988, 245–250.
12. R. Menozzi, L. Selmi, E. Sangiorgi, G. Crisenza, T. Cavioni and B. Ricco. Layout dependence of CMOS latchup. *IEEE Transactions on Electron Devices*, **35**(11), 1988, 1892–1901.
13. W. Reczek, F. Bonner and B. Murphy. Reliability of latchup characterization procedures. *Proceedings of the International Conference on Microelectronic Test Structures* (*ICMTS*); 1990. p. 51–54.
14. W. Reczek, *et al.*, Latchup free VLSI CMOS circuits considering power-on transients. *Proceedings of the ESSDERC*; 1988. p. 49–52.
15. W. Reczek, *et al.*, Latchup free CMOS using buried polysilicon diodes. *Proceedings of the ESSDERC*; 1989. p. 67–682. (See also Siemens AG, Corporate Research and Development, Otto-Hahn-Ring 6, D-8000 Munich 83, FRG. p. 679–682.)
16. C. Cane. Optimitzacio de pous n per una tecnologia CMOS. Aplicacio a la prevencio del latch-up. Ph.D. Thesis. Barcelona: Polytechnics University of Catalonia; 1989.
17. C. Cane, M. Lozano, E. Cabruja, E. Lora-Tamayo and F. Serra-Mestres. A new test structure to characterize the latchup effect. *Proceedings of the International Conference on Microelectronic Test Structures* (*ICMTS*); 1990. p. 47–51.
18. T. Aoki. A new latchup test structure for practical design methodology for internal circuits in standard-cell-based CMOS/BiCMOS LSIs. *Proceedings of the International Conference on Microelectronic Test Structures* (*ICMTS*); 1992. p. 18–23.

19. S. Voldman. Test structures for analysis and parameter extraction of secondary photon-induced leakage currents in CMOS DRAM technology. *Proceedings of the International Conference on Microelectronic Test Structures* (*ICMTS*); 1992. p. 39–43.
20. H. Momose, T. Maeda, K. Inoue, Y. Urakawa and K. Maeguchi. Novel test structures for the characterization of latchup tolerance in a bipolar and MOSFET merged device. *Proceedings of the International Conference on Microelectronic Test Structures* (*ICMTS*); 1991. p. 225–230.
21. A. Watson, S. Voldman and T. Larsen. Deep trench guard ring structures and evaluation of the probability of minority carrier escape for ESD and latchup in advanced BiCMOS SiGe technology. *Proceedings of the Taiwan Electrostatic Discharge Conference* (*T-ESDC*); 2003. p. 97–103.
22. S. Voldman and A. Watson. The influence of deep trench and substrate resistance on the latchup robustness in a BiCMOS silicon germanium technology. *Proceedings of the International Reliability Physics Symposium* (*IRPS*); 2004. p. 135–142.
23. S. Voldman and A. Watson. The influence of polysilicon-filled deep trench and sub-collector implants on latchup robustness in RF CMOS and BiCMOS SiGe technology. *Proceedings of the Taiwan Electrostatic Discharge Conference* (*T-ESDC*); 2004. p. 15–19.
24. S. Voldman. The influence of a novel contacted polysilicon-filled deep trench (DT) biased structure and its voltage bias state on CMOS latchup. *Proceeding of the International Reliability Physics Symposium* (*IRPS*); 2006. p. 151–158.
25. S. Voldman, E.G. Gebreselasie, D. Hershberger, D.S. Collins, N.B. Feilchenfeld, S.A. St. Onge, A. Joseph and J. Dunn. Latchup in merged triple well technology. *Proceedings of the International Reliability Physics Symposium* (*IRPS*); 2005. p. 129–136.
26. S. Voldman and E.G. Gebreselasie. The influence of merged triple well, deep trench and subcollector on CMOS latchup. *Proceedings of the Taiwan Electrostatic Discharge Conference* (*T-ESDC*); 2006. p. 49–52.
27. D. Kontos, K. Chatty, K. Domanski, M. Muhammad, C. Sequin and R. Halbach. External latchup characteristics under static and transient conditions in advanced bulk CMOS technologies. *Proceedings of the International Reliability Physics Symposium* (*IRPS*); 2007.
28. S. Voldman. Latchup physics and design. *Tutorial Notes*, *ESD Tutorials of the Electrical Overstress/Electrostatic Discharge* (*EOS/ESD*) *Symposium*; September 20, 2004.
29. S. Voldman. ESD and latchup in advanced technologies. *Tutorial*, *Tutorial Notes of the International Reliability Physics Symposium* (*IRPS*); April 25, 2004.
30. S. Voldman. CMOS latchup. *Tutorial Notes of the International Reliability Physics Symposium* (*IRPS*); April 17, 2005.
31. E. Hamdy and A. Mohsen. Characterization and modeling of transient latchup in CHMOS technology. *International Electron Device Meeting* (*IEDM*) *Technical Digest*; 1983. p. 172–175.
32. R. Troutman. *Latchup in CMOS Technology: The Problem and the Cure.* New York: Kluwer Publications; 1985.
33. R.R. Troutman and H.P. Zappe. Layout and bias considerations for preventing transiently triggered latchup in CMOS. *IEEE Transactions on Electron Devices*, **31**, 1984, 279–286.
34. R.R. Troutman and H.P. Zappe. A transient analysis of latchup in bulk CMOS. *IEEE Transactions on Electron Devices*, **30**, 1983, 170–179.
35. H.P. Zappe. A transient analysis and characterization of latchup in bulk CMOS. B.S. and M.S. Thesis. Cambridge, MA: Massachusetts Institute of Technology; February 1983.
36. IEEE Latchup Standards Committee. *Latchup Test Method for Process Characterization*; 1988.
37. JEDEC Standard No. 17. *Latchup in CMOS Integrated Circuits*; August 1988.
38. E.J. Chwastek. A new method for assessing the susceptibility of CMOS integrated circuits to latchup: the system-transient technique. *Proceedings of the Electrical Overstress/Electrostatic Discharge* (*EOS/ESD*) *Symposium*; 1989. p. 149–155.
39. R.J. Consiglio. AC and transient latchup characteristics of a twin-well CMOS inverter with load capacitance. *Proceedings of the Electrical Overstress/Electrostatic Discharge* (*EOS/ESD*) *Symposium*; 1993. p. 93–101.
40. S. Voldman. MeV implants boost device design. *IEEE Circuits and Devices*, **11**(6), 1995, 8–16.
41. G. Weiss and D. Young. Transient-induced latchup testing of CMOS integrated circuits. *Proceedings of the Electrical Overstress/Electrostatic Discharge* (*EOS/ESD*) *Symposium*; 1995. p. 194–198.
42. M. Mahanpour and I. Morgan. The correlation between latchup phenomenon and other failure mechanisms. *Proceedings of the Electrical Overstress/Electrostatic Discharge* (*EOS/ESD*) *Symposium*; 1995. p. 289–294.

43. JEDEC EIA/JESD78. *IC Latchup Test*; March 1997.
44. M. Kelly, L.G. Henry, J. Barth, G. Weiss, M. Chaine, H. Gieser, D. Bonfert, T. Meuse, V. Gross, C. Hatchard and I. Morgan. Developing a transient induced latchup standard for testing integrated circuits. *Proceedings of the Electrical Overstress/Electrostatic Discharge* (*EOS/ESD*) *Symposium*; 1999. p. 178–189.
45. I. Morgan, M. Mahanpour and C. Hatchard. Transient latchup using an improved bipolar trigger. *Proceedings of the Electrical Overstress/Electrostatic Discharge* (*EOS/ESD*) *Symposium*; 1999. p. 190–202.
46. K. Domanski, S. Bargstadt-Franke, W. Stadler, M. Streibl, G. Steckert and W. Bala. Transient-latchup failure analysis of the integrated circuits, methods of investigation and computer aided simulation. *Proceeding of the International Reliability Physics Symposium* (*IRPS*); 2004. p. 370–374.
47. M.D. Ker and S.F. Hsu. Evaluation on efficient measurement setup for transient-induced latchup with bipolar trigger. *Proceeding of the International Reliability Physics Symposium* (*IRPS*); 2005. p. 121–128.
48. T. Maloney and N. Khurana and T. Maloney. Transmission line pulsing technique for circuit modeling of ESD phenomena. *Proceedings of the Electrical Overstress/Electrostatic Discharge* (*EOS/ESD*) *Symposium*; 1985. p. 49–54.
49. S. Voldman, R. Ashton, J. Barth, D. Bennett, J. Bernier, M. Chaine, J. Daughton, E. Grund, M. Farris, H. Gieser, L.G. Henry, M. Hopkins, H. Hyatt, M.I. Natarajan, P. Juliano, T. Maloney, B. McCaffrey, L. Ting and E. Worley. Standardization of the transmission line pulse (TLP) methodology for electrostatic discharge (ESD). *Proceedings of the Electrical Overstress/Electrostatic Discharge* (*EOS/ESD*) *Symposium*; 2003. p. 372–381.
50. ESD Association. DSP 5.5.1. Standard Practice Document – Electrostatic Discharge Sensitivity Testing: Transmission Line Pulse (TLP) Component Level; 2003.
51. S. Voldman, S. Strang and D. Jordan. A design system for auto-generation of ESD circuits. *Proceedings of the International Cadence Users Group* (*ICUG*); September 2002.
52. S. Voldman, S. Strang and D. Jordan. An automated electrostatic discharge computer-aided design (CAD) system with the incorporation of hierarchical parameterized cells in BiCMOS analog and RF technology for mixed signal applications. *Proceedings of the Electrical Overstress/Electrostatic Discharge* (*EOS/ESD*) *Symposium*; October 2002. p. 296–305.
53. C.N. Perez, S. Voldman, Method of forming a guard ring parameterized cell structure in a hierarchical parameterized cell design, checking and verification system. U.S. Patent Application 20040268284; December 30, 2004.
54. A. Chynoweth and K.G. McKay. Photon emissions from avalanche breakdown. *Physical Review*, **102**, 1956, 369.
55. S. Tam, F.C. Hsu, P.K. Ko, C. Hu and R.S. Muller. Spatially resolved observation of visible light emissions for Si MOSFETs. *IEEE Electron Device Letters*, **4**(10), 1983, 386–388.
56. S. Tam and C. Hu. Hot electron induced photon and photocarrier generation in silicon MOSFETs. *IEEE Transactions on Electron Devices*, **31**(9), 1984, 1264.
57. A. Toriumi and M. Yoshimo. Experimental determination of hot carrier energy distribution and minority carrier generation mechanism due to hot carrier effects. *International Electron Device Meeting* (*IEDM*) *Technical Digest*, 1985. p. 56–59.
58. A. Toriumi and M. Yoshimo, A study of photon emission from n-channel MOSFETs. *IEEE Transactions on Electron Devices*, **34**, 1987, 1051.
59. T. Aoki and A. Yoshii. Analysis of latchup-induced photoemissions. *International Electron Device Meeting* (*IEDM*) *Technical Digest*; 1989. p. 281–284.
60. T. Aoki and A. Yoshii. Analysis of latchup-induced photon emissions. *IEEE Transaction on Electron Devices*, **37**(9), 1990, 2080–2083.
61. N. Tsutsu, Y. Uraoka, S. Akiyama, Y. Nakata and H. Esaki, New detection method of hot carrier degradation using photon spectrum analysis of weak luminescence on CMOS VLSI. *Proceedings on the International Conference on Microelectronic Test Structures* (*ICMTS*); 1990. p. 143–148.
62. J. Mechler, C. Brennan, J. Massucco, R. Rossi and L. Wissel. Contention latchup. *Proceedings of the International Reliability Physics Symposium* (*IRPS*); 2004. p. 126–129.
63. T. Cavioni, M. Cecchetti, M. Muschitiello, G. Spiazzi, I. Vottre and E. Zanoni. Latch-up characterization in standard and twin-tub test structures by electrical measurements, 2-D simulations and IR microscopy. *Proceedings on the International Conference on Microelectronic Test Structures* (*ICMTS*); 1990. p. 41–46.
64. W. Reczek, F. Bonner and B. Murphy. Reliability of latchup characterization procedures. *Proceedings on the International Conference on Microelectronic Test Structures* (*ICMTS*); 1990. p. 51–54.

65. M.J. Chen, J.K. Jeng, P.N. Tseng, N.S. Tsai and C.Y. Wu. Photon emission identification of emitter resistance for CMOS latchup hysteresis. *Proceedings on the International Conference on Microelectronic Test Structures* (*ICMTS*); 1991. p. 231–236.
66. S. Voldman. Test structures for analysis and parameter extraction of secondary photon-induced leakage currents in CMOS DRAM technology. *Proceedings of IEEE International Conference on Microelectronic Test Structures* (*ICMTS*), Vol. 5; March 1992. p. 39–43.
67. Y. Uraoka, I. Miyanaga, T. Maeda and K. Tsuji. New failure analysis technique of ULSI circuits using photon emission methods. *Proceedings of IEEE International Conference on Microelectronic Test Structures* (*ICMTS*), Vol. 5; March 1992. p. 100–106.
68. R. Jenkins, R.W. Gould, Dale Gedcke. *Quantitative X-Ray Spectrometry*, 1st Ed. New York: Marcel Dekker; 1981, chapter 4.
69. T. Ohzone and H. Iwata. Three dimensional effects of latchup turn-on CMOS and forward biased n+ diode measured by photoemissions. *Proceedings of IEEE International Conference on Microelectronic Test Structures*, Vol. 5; March 1992. p. 115–120.
70. P. Salome, C. Leroux, J.P. Chante, P. Crevel and G. Reimbold. Study of a 3D phenomenon during ESD stress in deep submicron CMOS technologies using photon emission tool. *Proceedings of the International Reliability Physics Symposium* (*IRPS*); 1997. p. 325–332.
71. C. Canali, F. Corsi, M. Muschitiello and E. Zanoni. Infrared microscopy study of anonymous parasitic paths. *IEEE Transactions on Electron Devices*, **36**(5), 1989, 969–978.
72. T. Kessler and F.W. Wulfert. Diagnosing latch-up with backside emission microscopy. *Semiconductor International*, **7**(23), 2000, 313–316.
73. W.K. Chim, *Semiconductor Device and Failure Analysis: Using Photon Emission Microscopy.* John Wiley & Sons, Ltd.; 2000.
74. S. Liao C. Niou, W.T.K. Chien, A. Guo, W. Dong and C. Huang. New observance and analysis of various guard ring structures on latchup hardness by backside photoemission image. *Proceedings of International Reliability Physics Symposium* (*IRPS*); April 2003. p. 92–97.
75. Y. Huh, K. Min, P. Bendix, V. Axelrad, R. Narayan, J.W. Chen, L.D. Johnson and S. Voldman. Chip level layout and bias considerations for preventing neighboring I/O cell interaction-induced latchup and inter-power supply latchup in advanced CMOS technologies. *Proceedings of the Electrical Overstress/Electrostatic Discharge* (*EOS/ESD*) *Symposium*; 2005. p. 100–107.
76. C.T. Hsu, I.C. Lin, J.C. Tseng, M.D. Ker, Y.L. Chen, F.Y. Tsai, S.H. Yu, F.H. Chen and P. A. Chen. A pin latchup failure and the latchup trigger induced npn snapback effect in a high voltage IC product. *Proceeding of the Taiwan Electrostatic Discharge Conference* (*T-ESDC*); 2006. p. 53–56.
77. N. Goldblatt, M. Leibowitz and W. Lo. Unique and practical IC timing analysis tool utilizing intrinsic photon emission. *Microelectronic Reliability*, **41**(9–10), 2001, 1507.
78. J.A. Kash, J.C. Tsang. Full chip optical imaging of logic state evolution in CMOS circuits. *International Electron Device Meeting* (*IEDM*)*Technical Digest Late News Paper*, Vol. 1; 1996. p. 934–936.
79. A. Weger, S. Voldman, F. Stellari, P. Song, P. Sanda and M. McManus. A transmission line pulse (TLP) picosecond imaging circuit analysis (PICA) methodology for evaluation of ESD and latchup. *Proceedings of the International Reliability Physics Symposium* (*IRPS*); 2003. p. 99–104.
80. S. Voldman. Latchup and the domino effect. *Proceedings of the International Reliability Physics Symposium* (*IRPS*); 2005. p. 145–156.
81. N.P. Sanda, S.H. Voldman and A.J. Weger. Method and application of PICA (picosecond imaging circuit analysis) for high current pulsed phenomena. U.S. Patent No. 6,943,578; September 13th, 2005.
82. F. Stellari, P. Song, M. K. McManus, A. J. Weger, K. Chatty, M. Muhammad and P. Sanda. Study of critical factors determining latchup sensitivity in ICs using emission microscopy. *Proceeding of the International Symposium for Testing and Failure Analysis* (*ITSFA*); 2003. p. 19–24.
83. A.H. Johnston and T.F. Miyahira. Latchup test considerations for analog-to-digital converters. Jet Propulsion Laboratory, California Institute of Technology, National Aeronautics and Space Administration (NASA). *Proceedings of the Single Event Effects* (*SEE*) *Symposium*, Manhattan Beach, California; 2002.

# 5 CMOS Latchup Process Features and Solutions – Dual-Well and Triple-Well CMOS

## 5.1 CMOS SEMICONDUCTOR PROCESS SOLUTIONS AND CMOS LATCHUP

Semiconductor process solutions can provide the best answer for addressing latchup. In this chapter, the influence of the different CMOS semiconductor processes and structures on latchup will be highlighted. This chapter will discuss it in the progression of the historical transitions from single-well CMOS, then dual-well CMOS and then present-day triple-well CMOS technology. First, the substrate will be discussed. For single-well CMOS, the influence of diffused and retrograde n-well on latchup will be shown. This work will address the integration issues of retrograde well and p++ substrate effects. In the next section, for dual-well CMOS issues, the p-well integration with n-well, epitaxy and heavily doped substrate from a latchup perspective is then discussed. Isolation structures such as lateral oxide (LOCOS) and shallow trench isolation (STI) will be addressed. This chapter will conclude with discussions on triple-well integration, buried layers and buried grid structures; in these structures, there is some commonality of issues that are relevant to latchup.

## 5.2 SUBSTRATES

Substrates have a strong influence on latchup [1–4]. From a latchup perspective, the substrate resistance is one of the four fundamental variables that influences the latchup sensitivity of a technology, as discussed in Chapter 3. Substrate profile and doping concentration influence many technology choices in semiconductor development. The relationship between the latchup and the ESD response is dependent on the choice of ESD networks. Figure 5.1 shows the two-transistor representation of latchup, as well as

*Latchup* Steven H. Voldman
© 2007 John Wiley & Sons, Ltd

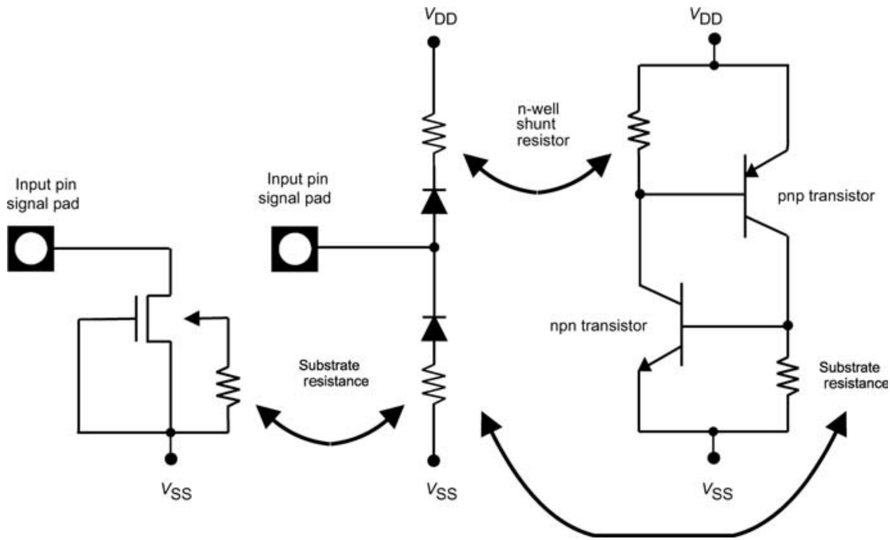



**Figure 5.1** Role of substrate resistance in latchup and ESD.

two ESD networks. The ESD networks consist of a grounded gate n-channel MOSFET ESD device and a dual-diode ESD network. In all three circuits, the substrate resistance influences their response to ESD events, as well as stimuli that can trigger latchup. As the substrate resistance decreases, latchup immunity improves. For the dual-diode ESD network, the substrate diode series resistance decreases, leading to an improvement in the ESD results for negative polarity ESD events. For the MOSFET ESD network, the snapback voltage increases, leading to a higher 'turn-on' voltage, which can lead to an ESD degradation.

The substrate doping concentration plays a large role in the minority carrier transport and the shunt resistance; both issues influence the latchup characteristics. The substrate doping concentration influences many variables in semiconductor process and device development. In early development, the substrate wafer was chosen to maximize yield and minimize cost. Wafers added epitaxial layers to CMOS processes for gate oxide integrity, yield and latchup [5–7]. Epitaxial layers on heavily doped wafers provided substrate wafer gettering. Heavily doped wafers provided a good ground plane for noise 'substrate bounce' in memory products [5–7]. The epitaxial layer and heavily doped wafer provided a high-capacitance n-well to p++ substrate metallurgical junction for stability from transient latchup as well as provided a low substrate shunt resistance. In addition, due to Auger recombination mechanism, minority carrier diffusion was limited to short distances (e.g. short diffusion length); this prevented latchup within a circuit, between different circuits and between circuit subfunctions. In this time frame, the guard rings consisted of n-well regions that extended to the heavily doped substrate region, leading to a very high guard ring efficiency [4]. From this period, significant understanding was achieved in latchup control. In the 1990s, CMOS memory and logic used the same base p++ substrate wafer [7, 8]. During this period, with the introduction of semiconductor foundries, DRAM and logic applications took different paths; CMOS development shifted toward what the needs would be in the sub-0.1-μm era [9]. The substrate trend is toward low-cost wafers, which provide high performance, and noise isolation accelerated anticipating high-level integration in the future.

With mixed signal applications, noise isolation is achieved using substrate wafers with low doping concentration. As the application frequency increases, noise coupling, crosstalk and noise rejection methods are required. For cost, the epitaxial growth process is being eliminated, with low-defect

wafers. With evolution from 0.25-μm to sub-45-nm technology, this trend to lower substrate resistance will continue, and with it, the reemergence of latchup concerns [10–16].

Although today utilization of p++ substrate wafers has decreased, substantial understanding was achieved that is relevant to future solutions for latchup. Although many will view this work as irrelevant, in fact it is quite relevant to guard ring analysis, high–low (H–L) and low–high (L–H) p−p+ transitions, heavily doped buried layers (HDBLs), triple well and BiCMOS subcollector understanding. Additionally, epitaxial wafers with p++ substrates are being used in image processing technologies, high-voltage CMOS and smart power technology. It is in the CMOS latchup discipline to utilize high-dose p++ substrate wafers for the following reasons:

- Auger recombination dominated providing a short recombination time; this provides a short diffusion length and low bipolar current gain.
- Auger recombination dominated short lifetime; this provides fast response time for transient latchup phenomenon.
- Auger-dominated short diffusion lengths and high guard ring efficiency; this minimizes intercircuit latchup.
- Introduces a high n-well to p++ substrate capacitance; this provides a high immunity to transient CMOS latchup.
- Heavily doped p++ substrate/$p^-$ epitaxy wafers introduce a vertical built-in electric field, enhancing the guard ring efficiency of both passive and active guard ring structures (see Chapter 7).

For a 0.8-μm technology, Troutman showed improvement of the guard ring efficiency for p− epitaxy/p++ substrate wafer [4]. With the n-well guard ring structure extending through the epitaxial region, abutting the p++ substrate made it difficult for carriers to leave the guard ring region and diffuse through the p++ substrate. Quinke also showed that the injection ratio is highly influenced with the usage of a $p^-$ epitaxial region on a p++ substrate wafer [17]. For a 0.5-μm CMOS technology, Voldman noted that the probability of escape, $P(E)$, is approximately 1 out of 1000 electrons using a guard ring structure.

## 5.2.1 Epitaxial Thickness

In early development, the epitaxy region thickness was wide to allow deep diffused well structures. In future generations, the epitaxial region was scaled to thinner dimensions and p++ substrates were introduced (Table 5.1).

With the epitaxial region thickness reduction (on a p++ substrate wafer), latchup was significantly influenced. With epitaxial region thickness scaling, the lateral and vertical bipolar transistors are influenced by the $p^-$/p++ doping step transition [18–23]. As noted by Estreich [1], this can be represented as a 'mirror reflection' boundary.

**Table 5.1** Technology features of epitaxial region thickness and n-well depth in single-well CMOS technology.

| Technology ground rule (μm) | Well depth (μm) | Epitaxy thickness (μm) |
|---|---|---|
| 1.2 | 8 | 12 |
| 0.75 | 1.5 | 4.0 |
| 0.5 | 1.1 | 2.2 |

**Table 5.2** n+ to n-well lateral npn bipolar current gain $\beta_{npn}$ versus epitaxial thickness.

| Epitaxial thickness (μm) | npn bipolar current gain, 0.4 μ m n+/nw (@ 25 °C), $\beta_{npn}$ | npn bipolar current gain, 0.275 μm n+/nw (@ 25 °C), $\beta_{npn}$ |
|---|---|---|
| 2.3 | 2.0 | 2.5 |
| 2.12 | 2.5 | 4.0 |
| 1.77 | 3.0 | 4.5 |

Experimental results on a p−/p++ substrate wafer in a 0.25-μm CMOS technology generation show that as epitaxial thickness decreases, the parasitic npn bipolar current gain $\beta_{npn}$ increases. Table 5.2 shows results of the parasitic npn transistor bipolar current gain as a function of epitaxial thickness. Measurements were taken for structures with an n+ to n-well space of 0.4 and 0.275 μm. For the 0.4-μm n+ to n-well space, the parasitic $\beta_{npn}$ increased from 2.0 to 3.0 as the epitaxy thickness decreased from 2.3 to 1.77 μm. For the 0.275-μm n+ to n-well space, the parasitic $\beta_{npn}$ increased from 2.5 to 4.5 as the epitaxy thickness decreased from 2.3 to 1.77 μm.

An explanation for the enhancement of the npn bipolar current gain is consistent with the model of the reflective (L–H) boundary condition at the $p^-$/p++ interface, which leads to a decrease in the vertical electron transport and enhancement of the lateral electron transport [18–23].

A second interesting result showed that as the n+ to n-well space increased from 0.4 to 1.4 μm, a much weaker dependence on the epitaxial thickness was observed (Figure 5.2). As the spacing

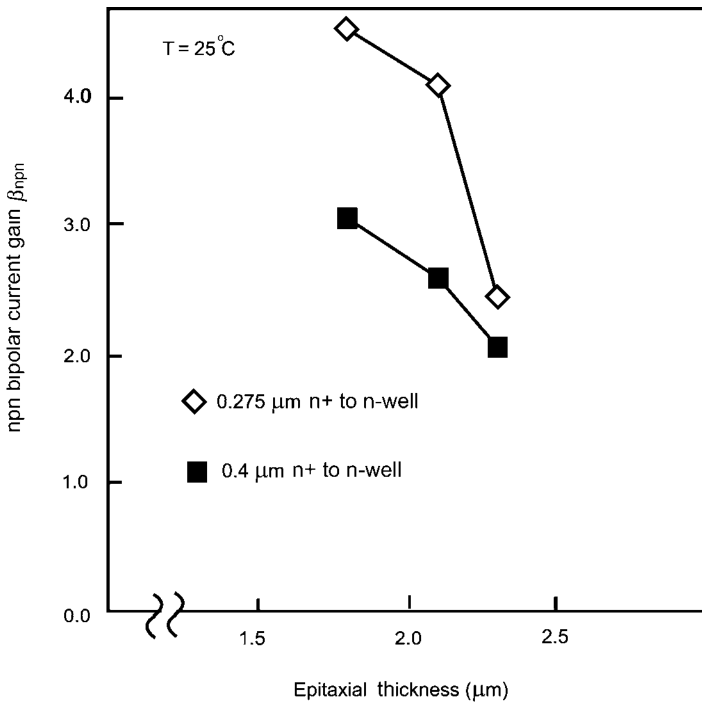



**Figure 5.2** npn bipolar current gain versus n+ to n-well space as a function of epitaxial thickness.

**Table 5.3** pnp bipolar current gain versus epitaxial thickness.

| Epitaxial thickness (μm) | pnp bipolar current gain, 0.45 μm p+/nw (@ 25 °C), $\beta_{pnp}$ | pnp bipolar current gain, 0.3 μm p+/nw (@ 25 °C), $\beta_{pnp}$ |
|---|---|---|
| 2.3 | 1.2 | 1.3 |
| 2.12 | 1.0 | 1.0 |
| 1.77 | 1.0 | 1.0 |

increased from 0.275 to 1.4 μm, the lateral $\beta_{npn}$ decreased from 4.5 to near unity, where there is little distinction between the epitaxial thickness splits.

Measurements were also taken to evaluate the change on the pnp transistor. Electrical measurements were taken for pnpn structures with a p+ to n-well space of 0.45 and 0.3 μm. For the 0.45-μm p+ to n-well space, the parasitic pnp bipolar current gain decreased from $\beta_{pnp} = 1.2$ to $\beta_{pnp} = 1.0$ as the epitaxy thickness decreased from 2.3 to 1.77 μm (Table 5.3). For the 0.3-μm n+ to n-well space, the parasitic $\beta_{pnp}$ decreased from 1.3 to 1.0. With the correct n-well depth, no strong n-well modulation effect was observed and only weak dependence on the lateral pnp transport was observed (e.g. the n-well peak was not compensated by epitaxy thinning).

In the latchup semiconductor development discipline, the following observations are made:

- Epitaxial scaling on a p−/p++ wafer can lead to enhancement of the vertical npn bipolar current gain $\beta_{npn}$, due to the reflective boundary at the p−/p++ step junction.
- At large n+ to n-well spacings in a p−/p++ heavily doped substrate wafer, the epitaxy effect on the bipolar npn current gain $\beta_{npn}$ has a lesser role as the spacing of the n+ to n-well approaches the epitaxial thickness.

## 5.3 n-WELLS

### 5.3.1 Diffused n-Well Design

Diffused wells were common in early low-voltage CMOS development prior to 1994. In the 1980s, only a few semiconductor corporations utilized the high-energy MeV implanters for formation of retrograde wells. Although significant development and research discussed retrograde wells, few corporations practiced it in commercial product development. In semiconductor technologies, isolation wells were formed to allow the placement of different semiconductor transistor types in a common substrate. But, from a latchup perspective, these isolation wells also lead to ‘parasitic’ devices. In the case of a p-channel MOSFET, a vertical and a lateral pnp are formed where the well region serves as a base region and the p− substrate serves as a collector region. As a result, the diffused well semiconductor process significantly influences latchup. Today, some technologies that require high breakdown voltages such as high-voltage CMOS still utilize diffused well regions.

To form a diffused well, a diffusion source was created that was doped or implanted to a certain dose level. Using thermal processing, the dopants diffuse into the substrate where the extension into the substrate is a function of the surface dopant concentration, the diffusion coefficient of the dopant in the material and the time of the process for the diffusion process. From the diffusion equation, the doping

profile can be expressed as a function of the complementary error function distribution, where $D$ is the diffusion constant of the element and $t$ is time:

$$C(x,t) = C_{\mathrm{S}}\mathrm{erfc}\left(\frac{x}{2\sqrt{Dt}}\right) = \frac{2C_{\mathrm{s}}}{\sqrt{\pi}}\int_{\frac{x}{2\sqrt{Dt}}}^{\infty} \mathrm{e} - v2_{\mathrm{d}v}.$$

To get the total dopant,

$$N' = \int_{0}^{\infty} C(x,t)\mathrm{d}x = 2\sqrt{\frac{Dt}{\pi}}C_{\mathrm{s}}.$$

Doping profiles can be represented as the doping concentration $N'$ followed by a hot process drive-in cycle, represented by a Gaussian distribution,

$$C(x,t) = \frac{N'}{\sqrt{\pi Dt}}\exp\left(-\frac{x^2}{4Dt}\right).$$

In a diffused well process, the doping gradient of the Gaussian distribution is such that it establishes a vertical built-in electric field. The built-in electric field in the dopant profile establishes a drift component in the current density constitutive equation. This built-in electric field drift component serves in transporting holes from the emitter to the collector. The built-in drift component leads to an increase in the base transport factor of the parasitic bipolar transistor element. From the constitutive relationships of hole and electron currents,

$$J_{\mathrm{p}} = \mu_{\mathrm{p}} p_0 E_x - qD_{\mathrm{p}}\frac{\mathrm{d}p_0}{\mathrm{d}x},$$

$$J_{\mathrm{n}} = \mu_{\mathrm{n}} n_0 E_x + qD_{\mathrm{n}}\frac{\mathrm{d}n_0}{\mathrm{d}x}.$$

From the principle of detailed balancing, these equations must balance independently. Letting the electron and hole currents equal zero, the built-in field can be

$$E_x = \frac{kT}{q}\frac{1}{p_0}\frac{\mathrm{d}p_0}{\mathrm{d}x}$$

or

$$E_x = -\frac{kT}{q}\frac{1}{n_0}\frac{\mathrm{d}n_0}{\mathrm{d}x}.$$

This expression relates the ‘built-in’ electrical field to the equilibrium electron and hole carrier concentrations in the doping concentration profile.

Since the diffused n-well is a function of position, its doping gradient will influence the minority carrier diffusion of holes through this region. In the case of a vertical bipolar transistor, the diffused n-well region serves as a base region. Minority carrier holes diffuse through the base region from the p+ diffusion at the device surface. In a diffused well, the doping profile is such that the doping concentration decreases from the device surface. From the above expression, given that the gradient of

the n-type dopant is negative, the built-in electric field is positive. The positive built-in electric field term leads to a positive hole current providing a drift component to the hole current density. The constitutive relationship, with built-in field $E_{x0}$, can be expressed as

$$J_{\mathrm{h}} = q\mu_{\mathrm{p}}(p_0 + p')E_{x0} - qD_{\mathrm{p}}\left(\frac{\mathrm{d}p_0}{\mathrm{d}x} + \frac{\mathrm{d}_{p'}}{\mathrm{d}x}\right),$$

$$J_{\mathrm{e}} = q\mu_{\mathrm{n}}(no + n')E_{x0} + qD_{\mathrm{n}}\left(\frac{\mathrm{d}n_0}{\mathrm{d}x} + \frac{dn'}{\mathrm{d}x}\right).$$

In CMOS latchup, these built-in fields modulate the vertical and lateral transport. It will be shown in this book that this concept can be utilized within latchup structures and with active guard ring structures.

The diffused well vertical bipolar pnp transistor has the property of an increased bipolar forward current gain as a result of the built-in field assisting the hole transport. Note that the diffused well built-in field also leads to a lower reverse bipolar current gain. For the diffused well bipolar transistor element, low diffused well doping concentrations lead to a low pnp base transistor. As the base doping concentration is decreased, the bipolar gain increases.

As a result of the doping profile, an enhanced parasitic bipolar pnp element current gain exists. In diffused well CMOS, the parasitic bipolar current gains typically were between $\beta_{\mathrm{pnp}} = 10$ and $\beta_{\mathrm{pnp}} = 30$ [24–31]. With these excessively high bipolar parasitic current gains, latchup was an issue in CMOS technology. This was also a significant issue in ESD networks and design. Maloney [24, 25] and Voldman and Gerosa [26–31] independently noted that the high parasitic pnp bipolar current gain $\beta_{\mathrm{pnp}}$ lowered the forward turn-on voltage in a diffused well p+/n-well ESD diode string; Voldman and Gerosa also noted that this effect could be alleviated through circuit elements [26–28, 31].

## 5.3.2 Retrograde n-Well Design

A solution for improving the latchup concern was the introduction of the retrograde n-well [2]. Retrograde well regions were accepted into the integration of CMOS technology because of its strategic value in CMOS performance and scaling. The retrograde well allowed for the ability to lower the junction capacitance, provide a low parasitic transistor pnp bipolar current gain and achieve a low n-well shunt resistance. In retrograde wells, the MOSFET p-diffusion to n-well capacitance can be lowered by lowering the implant dose near the wafer surface and providing the maximum doping concentration deep and away from the MOSFET transistor. Retrograde wells provide an extra degree of freedom in the semiconductor design since it can avoid influencing the transistor design point itself; in fact, it can avoid constraining the MOSFET design box [5, 6]. Second, the retrograde well introduces a retrograde electric field in the opposite direction of the hole current flow in the vertical pnp transistor lowering the forward collector-to-emitter transport. Finally, the implant dose and energy can be increased allowing a low resistance n-well shunt (e.g. in parallel with the pnp parasitic transistor). With the introduction of commercial high-energy MeV implanters in a semiconductor manufacturing line, it was possible to develop semiconductor manufacturing capability for retrograde wells [4–8].

Table 5.4 contains the latchup $\beta_{\mathrm{pnp}}$ as a function of the n-well dose (@ $T = 25\,^{\circ}\mathrm{C}$) in a 0.8-µm LOCOS-defined CMOS technology; this was the first commercial technology that integrated a retrograde n-well [5, 7]. The base wafer is a 2.5-µm thick p− epitaxial layer on a p++ wafer. With no retrograde well implant, but just the 'n-well fill' implant, the n-well sheet resistance is high. Without a retrograde well implant, the vertical $\beta_{\mathrm{pnp}}$ and n-well sheet resistance are high. With the retrograde n-well implant, the sheet resistance is significantly reduced, as well as the vertical $\beta_{\mathrm{pnp}}$ [32–35].

**Table 5.4** pnp bipolar current gain versus retrograde n-well dose.

| Retrograde n-well dose ($cm^{-2}$) | n-well sheet resistance (Ω/square) | Bipolar current gain $\beta_{pnp}$ |
|---|---|---|
| None | 2800 | 8 |
| $3 \times 10^{13}$ | 670 | 4 |
| $6 \times 10^{13}$ | 390 | 3 |

Figure 5.3 contains the results of a transient latchup holding voltage study for different retrograde n-well doses [36]. The study was performed as a function of the no retrograde well implant and increments of 1.0, 2.0 and $3.0 \times 10^{13}\ cm^{-2}$ retrograde n-well doses. Experimental studies of transient latchup were completed in the single-well, p++ substrate CMOS technology for different pulse widths. External resistors were chosen based on the n-well and substrate contact spacings. An external n-well shunt resistor of 200 kΩ and a substrate shunt resistor of 10 kΩ were placed in the circuit, while the rectangular pulse was applied to the anode of the four-stripe pnpn test structure. As the retrograde well dose increases, $V_H$ increases for short pulse widths [36].

Table 5.5 is an example of the parasitic $\beta_{pnp}$ versus n-well implant dose in a 0.5-μm ground rule (0.35-μm n-channel MOSFET) STI-defined CMOS technology on a 2.0-μm p− epitaxial region on a

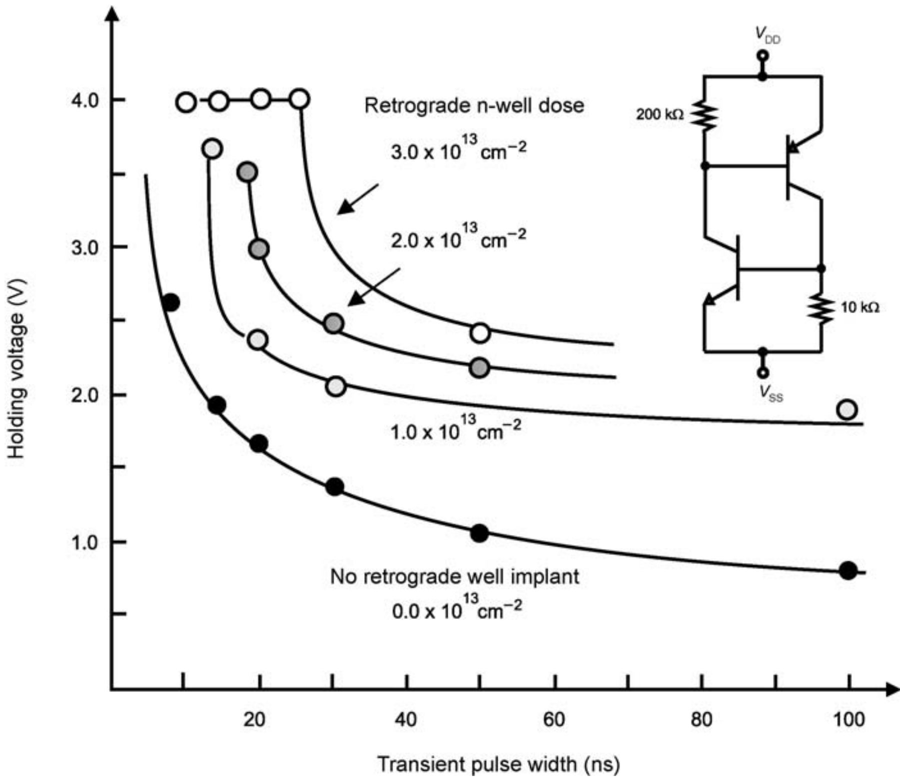



**Figure 5.3** Transient latchup holding voltage results versus pulse width as a function of retrograde n-well splits.

**Table 5.5** pnp bipolar forward current gain versus n-well dose (as a function of the p+ to n-well space).

| Retrograde n-well dose ($cm^{-2}$) | pnp forward bipolar current gain $\beta_{pnp}$ |
|---|---|
| $1 \times 10^{13}$ | 8 |
| $3 \times 10^{13}$ | 3.2 |
| $4 \times 10^{13}$ | 2.2 |
| $5 \times 10^{13}$ | 2.0 |

p++ substrate wafer [6–8]. The experiment evaluated n-well doses from $1 \times 10^{13}$ to $5 \times 10^{13}\ cm^{-2}$ [26–29]. The latchup structure is a four-stripe device with a minimum 0.6 μm p+ to n-well space.

Figure 5.4 shows $\beta_{pnp}$ versus n-well dose as a function of the p+ to n-well spacing. Experimental results show that in a STI technology, with a 0.55-μm STI depth, the lateral $\beta_{pnp}$ is a weak function of the p+ to n-well space [6–8].

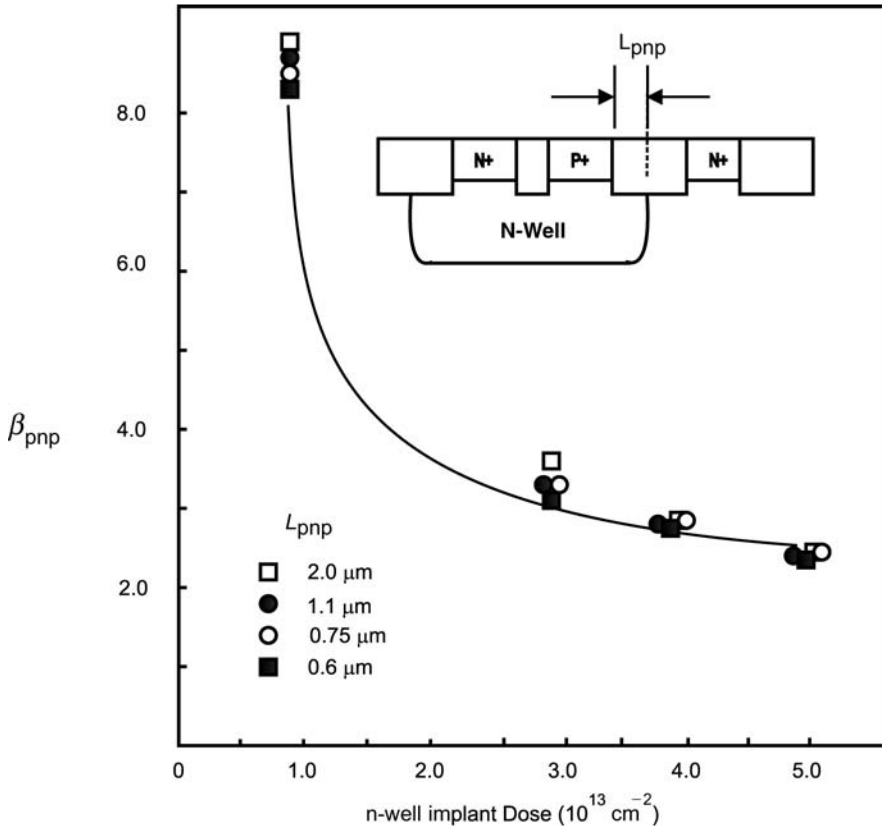



**Figure 5.4** Parasitic pnp bipolar current gain as a function of retrograde n-well implant dose.

### 5.3.3 Retrograde n-Well Design and Design Point Optimization

Choosing the n-well design point in a CMOS process requires the optimization of many variables required for the technology [5,6, 26–31, 34–36]. 'n-well engineering' requires the ability to optimize all the electrical parameters of interest associated with all devices within and adjacent to the n-well, and the n-well parameters themselves. For a CMOS logic technology, these parameters are as follows:

- p-channel MOSFET p-diffusion to n-well junction capacitance;
- p-channel MOSFET body effect;
- p-channel MOSFET gate-induced drain leakage mechanism;
- n-well sheet resistance;
- n-well to substrate junction capacitance;
- n-well to substrate junction leakage current;
- parasitic pnp bipolar current gain;
- parasitic pnp collector-to-emitter transport factor;
- n-well to substrate current injection.

Figure 5.5 shows a design box plot for the n-well design point [6]. In the plot, the parasitic $\beta_{pnp}$, n-well sheet resistance, p+/n-well junction capacitance and ESD protection results are shown. In the

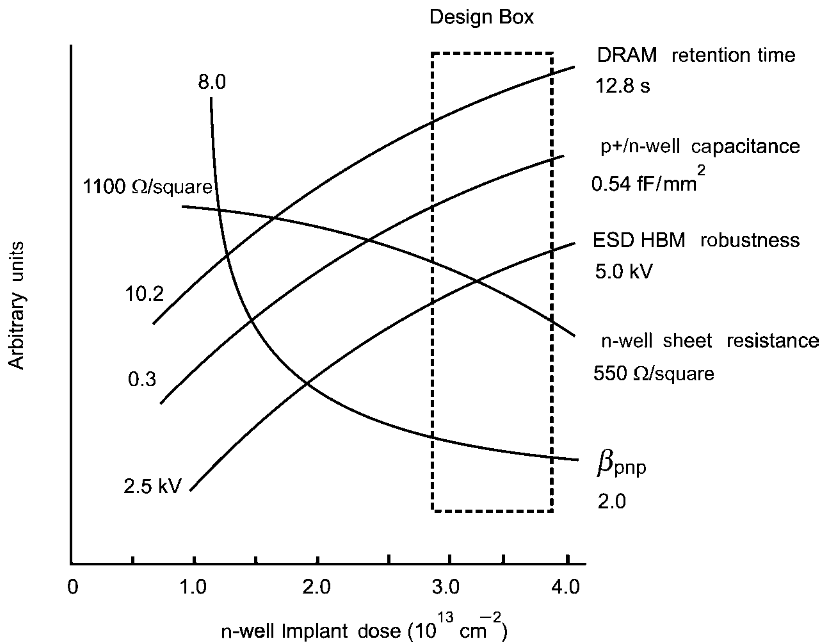



**Figure 5.5** n-well design plot for cooptimization of CMOS latchup and semiconductor device design point. The figure contains a design box with the acceptable n-well design point.

figure, a design box is drawn to indicate the acceptable semiconductor design center and the process variation. As the retrograde well dose increases, the p+/n-well junction capacitance increases and the vertical $\beta_{pnp}$ decreases from 8 to 2.

### 5.3.4 Retrograde n-Well Design for Latchup and ESD Design Synthesis

In semiconductor chip development, it is necessary to satisfy many technology objectives. A semiconductor design point is needed that allows for achieving both latchup and ESD objectives. Choosing an ESD strategy that is in conformance to the latchup objectives is key to achieve total product quality and reliability goals. In addition, it is also important that as the technology is migrated and scaled, this synergy between ESD and latchup objectives is maintained.

It was found that using an ESD strategy, of diode-based ESD input circuits, allowed for the capability to achieve both latchup and ESD objectives. It was also found that increasing the retrograde well dose improved the ESD protection results in products as well as increased the latchup robustness [6–8, 26–31, 34–36]. In the early 1990s, prior to implementation of retrograde well implants, semiconductor corporations that introduced ESD series diode (e.g. also known as ESD diode strings) discovered reverse amplification of the leakage currents and lower than ideal turn-on of the ESD networks [24–28]. With the migration to retrograde well implants and the higher retrograde well doses, this issue was not a concern [26–31].

### 5.3.5 n-Wells: n-Well to Substrate Modulation

Retrograde wells formed in a p++ heavily doped substrate with an epitaxial region, or abutting a HDBL implant can undergo modulation of the retrograde well profile. This effect is known as the retrograde well–substrate modulation effect [5, 6, 30, 31]. This effect has considerable influence on both CMOS latchup and ESD protection. The retrograde well profile can be represented as a Gaussian implant at $x_{w0}$ and n-well fill doping profile [30],

$$N_D(x) = N_{fill} + N_D \exp\left\{ -\left[\frac{x - x_{w0}}{\sqrt{2}\sigma}\right]^2 \right\},$$

where the effective epitaxial thickness [30] is

$$T_{eff} = T_{epi} + \lambda,$$

where the effective thickness is determined by the epitaxial flat zone at the end of the surface implant (or p-well) and the beginning of the p+ substrate (or p++ buried layer). In the p++ substrate model, as discussed in the previous chapter, the effective depth factor $\lambda$ addresses the extra effective thickness between the epitaxial flat zone and the heavily doped flat zone of the p++ wafer to add an increase in the epitaxial effective width. For the p++ wafer, a concentration profile can be assumed to be equal to

$$N_{sub}(x) = \frac{N_{sub}}{2} \operatorname{erfc}\left[\frac{x - x_e}{x_0}\right],$$

where we determine the epitaxial depth position from the above expression, letting $x_2 = x_e$. This can be expressed as the effective depth parameter

$$\lambda = \frac{x_0}{2} \operatorname{erfc}^{-1}\left\{\frac{2N_{epi}}{N_{sub}}\right\}.$$

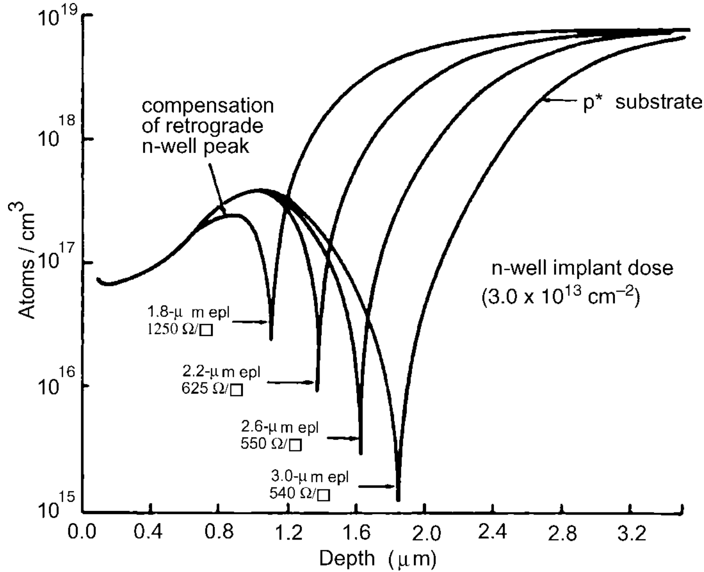



**Figure 5.6** Vertical profiles of a retrograde n-well as a function of epitaxy thickness on a p++ substrate wafer.

As the epitaxial thickness $T_{epi}$ decreases, the flat zone of the p− region decreases. Variation in the epitaxial thickness is common due to the control of incoming wafers. With the variation, the width term $\lambda$ does not change but is shifted toward the device surface. Additionally, the MeV retrograde well implant is also implanted deeper toward the p++ substrate. The MeV implant dose is significantly lower than that of the p++ substrate wafer, leading to strong modulation of the profile, the capacitance and the n-well sheet resistance. As the p++ substrate dopants impinge on the n-well region, the physical well depth decreases and the n-well sheet resistance increases. Figure 5.6 shows an example of the n-well doping profile modulation by the p++ heavily doped substrate.

For a p-channel MOSFET structure, this leads to a higher n-well shunt resistance series. Additionally, the vertical bipolar pnp current gain increases as a result of the smaller base width and the reduction of the Gummel number (e.g. integral of the dopants in the base region). Experimental results have shown that the n-well sheet resistance can vary from 700 to 1500 Ω/square and the vertical $\beta_{pnp}$ can increase from 2 to 8 (Table 5.6) [6].

As the epitaxial thickness is varied, for a fixed well dose and energy implant, experimental results show the n-well sheet resistance from 300 to 1000 Ω/square. With the decrease in the epitaxial

**Table 5.6** Vertical pnp bipolar current gain as a function of n-well sheet resistance.

| n-well sheet resistance (Ω/square) | pnp bipolar current gain ($\beta_{pnp}$) |
|---|---|
| 330 | 2 |
| 550 | 4 |
| 1000 | 7 |
| 1500 | 8 |

thickness, the effective film thickness of the series resistance decreases, as well as the total doped elements. Evaluating the n-well doping profiles, as the epitaxial region is too thin, the peak n-well concentration is consumed by the p++ substrate region, significantly changing the doping profile in the n-well. Although the base width decreases the vertical $\beta_{pnp}$ increases to approximately 8.

The effect of implant compensation of oppositely doped regions is important for single-well, dual-well and triple-well CMOS for manufacturing control of CMOS latchup. As one implanted region influences another, this modulation effect influences the physical thickness of a region, wherein this influences both the parasitic bipolar transistor gains and the parasitic bipolar series shunt resistance. This is true in wells, epitaxy regions and isolated epitaxial regions (e.g. triple-well regions).

### 5.3.6 n-Well Depth Scaling

In CMOS technology scaling, the n-well depth will be scaled in each successive generation. Figure 5.1 shows the n-well scaling trend. In dimensional scaling, the vertical and lateral dimensions are scaled to maintain dimensional similitude. In addition, as the technology dimensions are scaled, the p+/n+ space and isolation depth will also be reduced to achieve circuit density. In this scaling process, both the ESD protection and the latchup design point become a function of the retrograde n-well design point and the isolation design point [37–41].

## 5.4 p-WELL

In this section, the latchup impact of p-well regions will be discussed. In this section, it is assumed that the addition of the p-well assumes that the n-well is also present, leading to a dual-well CMOS semiconductor process.

### 5.4.1 p-Well Design and Design Point Synthesis

In dual-well CMOS, the p-well design point is established based on a number of semiconductor device constraints [44, 45]. The p-well design point will be influenced by the n-channel MOSFET junction capacitance, MOSFET device ‘punch-through’, MOSFET junction leakage, as well as integration with the isolation and n-well regions.

From a latchup perspective, the p-well influences the lateral npn structure and the ‘substrate’ shunt resistance. The shunt resistance is the parallel resistance of the substrate and the p-well regions. The p-well implant and the design point of the dose and energy must be established to also allow the p-well to n-well junction to be placed under the isolation region; this constraint is critical for latchup since it influences both the pnp and npn base widths. As a result, n-well, p-well and isolation must be integrated together; this places a constraint on the p-well design point.

### 5.4.2 p-Well and p++ Substrates

With dual-well CMOS scaling, the p-well depth is reduced with the n-well depth for integration under STI and p-well to n-well regions; the integration of both wells is required for cosynthesis. For achieving lateral cosynthesis of p-well and n-well regions, on a p++ substrate, the epitaxial thickness

must also be chosen to avoid breakdown of the n-well to p++ substrate metallurgical junction and p++ substrate modulation of the n-well sheet resistance. In addition to the poor control of epitaxial thickness, additional space is required to avoid p-well and n-well variations and compensations over the semiconductor process window. In addition, in foundry environments or semiconductor fabricators with many semiconductor technology types and generations, the starting wafer specification remains the same for the wafer and epitaxial thickness (as well as doping concentrations). In this fashion, the identical wafer specification can be used for many processes independent of the technology generation; this also has a practical opportunity to change the routing for one type of technology or another during the processing flow. For example, the same p− epitaxial thickness is used on a p++ substrate wafer and used for multiple semiconductor technologies (e.g. 0.5–0.22 μm). With the usage of a common base wafer, this can lead to latchup issues [46].

## 5.4.3 p-Well Connecting Implants (Epitaxial Buried Implants)

The issue of a low-doped gap between the p-well and the p++ substrate can be addressed with additional implants [46]. Without an additional implant, a resistive region exists between the p-well and the p++ substrate. Figure 5.7 shows an example of a profile of a scaled p-well on a p++ substrate wafer. In the definition of this technology, the p-well profile was scaled to accommodate scaling implications, but the base wafer remained the same.

As a result, a low-doped p− region exists between the p++ substrate and the p-well. This low-doped p− region introduces a vertical series component between the p-well and the p++ substrate, which influences the technology latchup robustness. Brown first noted the impact of the resistive region on the latchup results [47]. Figure 5.8 shows the cross section highlighting the vertical resistance element. A p-doped implant can be used to electrically ‘connect’ the p-well region and the p++ substrate, similar to the concept of a ‘sinker’ or ‘reachthrough’ implant. In this case, using a ‘connecting implant’, specifically added for latchup improvements, can provide a low-resistance region [47].

To address the resistive issue, instead of scaling the epitaxial region, a p-type ‘connecting implant’ was integrated into the semiconductor process to lower the effective substrate resistance. The p-type connecting implant was found to improve the latchup results in a 0.22-μm technology [47]. Figure 5.9 shows the cross section with the latchup connecting implant layer.

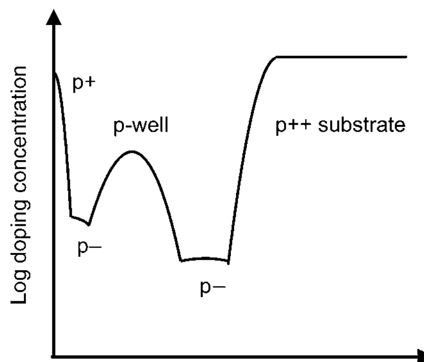



**Figure 5.7** Vertical doping profile of a scaled p-well on a p++ substrate (e.g. without p− epitaxial thickness scaling).

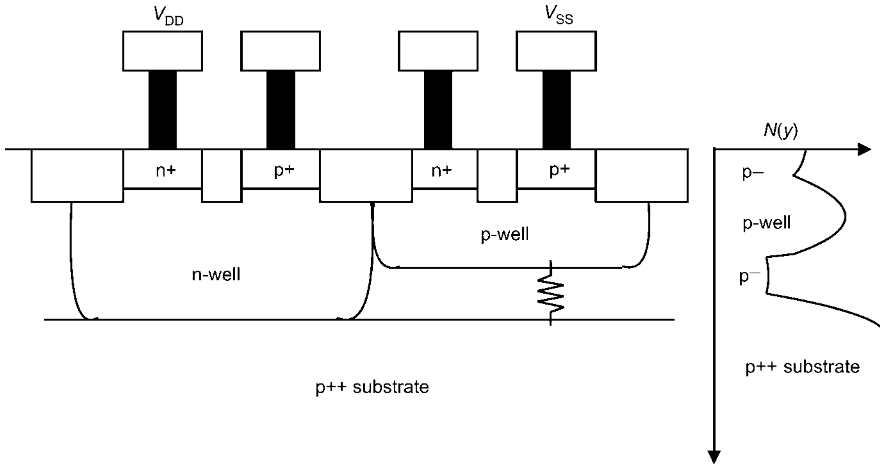



**Figure 5.8** Cross section region highlighting the vertical resistive region between the p-well and the p++ substrate.

Table 5.7 shows the parasitic $\beta_{npn}$ as a function of the epitaxial buried implant (EBI) dose (for a p+/n+ space of 1.2 µm ). In the case of no buried layer implant, the parasitic bipolar current gain was $\beta_{npn} = 9.1$ @ 140 °C. With a $1.5 \times 10^{13}\,\mathrm{cm}^{-2}$ implant, $\beta_{npn}$ was reduced to 4.3. As the implant was increased to $6.0 \times 10^{13}\,\mathrm{cm}^{-2}$, $\beta_{npn} = 2.75$ @ 140 °C. Four-point resistance test structure measurements also showed a resistance reduction from 2.25 to 1.18 Ω. The connecting implant served as a low-resistance shunt termination.

Table 5.8 shows the undershoot current $I_{under}$ (required to initiate latchup) as a function of the EBI dose (e.g. p+/n+ space of 1.2 µm). In the case of no buried layer implant, the undershoot current was $I_{under} = 0.77$ mA @ 140 °C. With a $1.5 \times 10^{13}\,\mathrm{cm}^{-2}$ implant, $I_{under} = 3.1$ mA. As the implant dose was further increased to $6.0 \times 10^{13}\,\mathrm{cm}^{-2}$, $I_{under} = 3.35$ mA @ $T = 140$ °C.

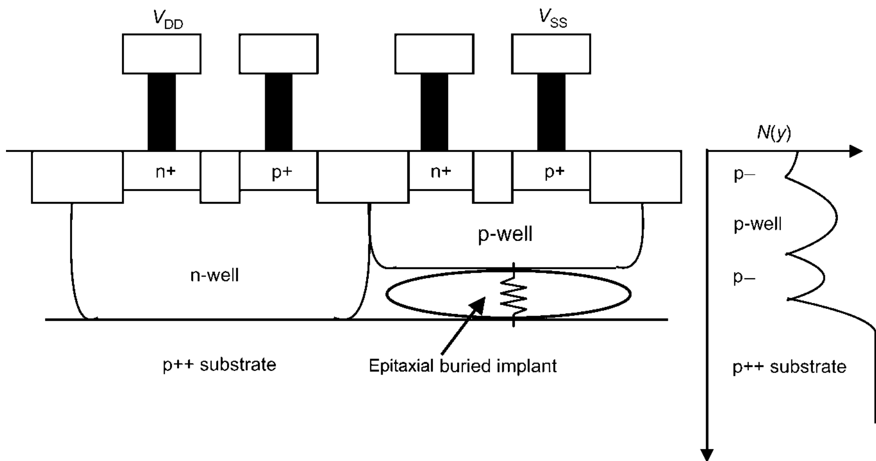



**Figure 5.9** Semiconductor technology with CMOS connecting implant layer (epitaxial buried implant).

**Table 5.7** Parasitic npn bipolar current gain $\beta_{npn}$ as a function of the EBI dose. Reproduced by permission of IBM.

| Epitaxial buried layer implant dose ($cm^{-2}$) | npn bipolar current gain (@ 140 °C), $\beta_{npn}$ |
|---|---|
| No epitaxial buried layer | 9.1 |
| $1.5 \times 10^{13}$ | 4.3 |
| $3.0 \times 10^{13}$ | 3.3 |
| $4.5 \times 10^{13}$ | 2.9 |
| $6.0 \times 10^{13}$ | 2.75 |

**Table 5.8** Undershoot current as a function of the EBI dose. Reproduced by permission of IBM.

| Epitaxial buried layer implant dose ($cm^{-2}$) | Undershoot current (@ 140 °C) (mA) |
|---|---|
| No epitaxial buried layer | 0.77 |
| 1.5 | 3.1 |
| 3.0 | 2.95 |
| 4.5 | 3.1 |
| 6.0 | 3.35 |

Table 5.9 shows the holding voltage $V_H$ as a function of the deep EBI layer for a 1.2-μm p+/n+ space (@ 140°C). This study was completed with a 10-Ω external $R_{SX}$ and a 0-Ω external $R_W$. As the connecting implant dose increases, the holding voltage, $V_H$ increased from 3.25 to 4.3 V. As a result, the connecting implant increased $V_H$ above the 3.3 V external power supply value and allowed for margin between the holding voltage state and the power supply voltage state. (Note: Valuable for a 2.5-V technology with a 3.3-V interface.)

A CMOS semiconductor process design practice for CMOS latchup is as follows:

- Vertical resistance regions in a p+/p−/p++ profile can exist leading to a higher shunt resistance.
- 'Connection implants' (e.g. also known as epitaxial buried implants) are added to establish a lower resistance path between the p+ and p++ implants for a lower substrate shunt resistance (the connecting implant serves as a 'termination shunt' in a resistive transmission line representation of the substrate).
- Lateral npn bipolar current gain is significantly reduced with the introduction of the epitaxial buried layer between the p-well and the p++ substrate.

**Table 5.9** Holding voltage $V_H$ as a function of the EBI dose. Reproduced by permission of IBM.

| Epitaxial buried layer implant dose ($cm^{-2}$) | Holding voltage $V_H$ (@ 140°C) |
|---|---|
| No epitaxial buried layer | 3.25 |
| $1.5 \times 10^{13}$ | 3.6 |
| $3.0 \times 10^{13}$ | 4.2 |
| $4.5 \times 10^{13}$ | 4.25 |
| $6.0 \times 10^{13}$ | 4.3 |

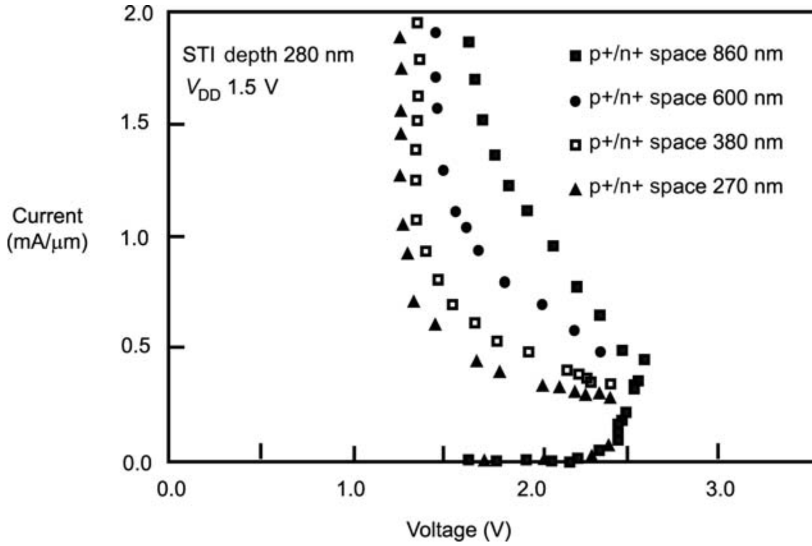



**Figure 5.10** Latchup simulation as a function of p+/n+ spacing (e.g. STI depth of 280 nm).

## 5.5 P+/N+ SCALING

With the scaling of the p+/n+ dimension, the holding voltage $V_H$ is reduced. Figure 5.10 shows latchup simulation as a function of p+/n+ spacing (e.g. STI depth of 280 nm). Table 5.10 contains trigger and holding voltage simulation results as a function of p+/n+ spacing [43]. Figure 5.11 shows the latchup simulation results as a function of p+/n+ spacing (@ $V_{DD} = 1.2$ V and STI depth of 200 nm).

## 5.6 ISOLATION AND LATCHUP

Isolation technology can have a strong influence on the CMOS latchup robustness of a technology. The isolation structure influences latchup in the following ways:

- lateral transport to vertical current transport ratio;
- effective geometric base width;

**Table 5.10** Trigger current and holding voltage simulation results as a function of p+/n+ spacing.

| p+/n+ spacing (nm) | $I_{TR}$ (µA/µm) | $V_H$ (V) |
|---|---|---|
| 860 | 445 | 1.6 |
| 600 | 400 | 1.45 |
| 380 | 330 | 1.35 |
| 270 | 280 | 1.29 |

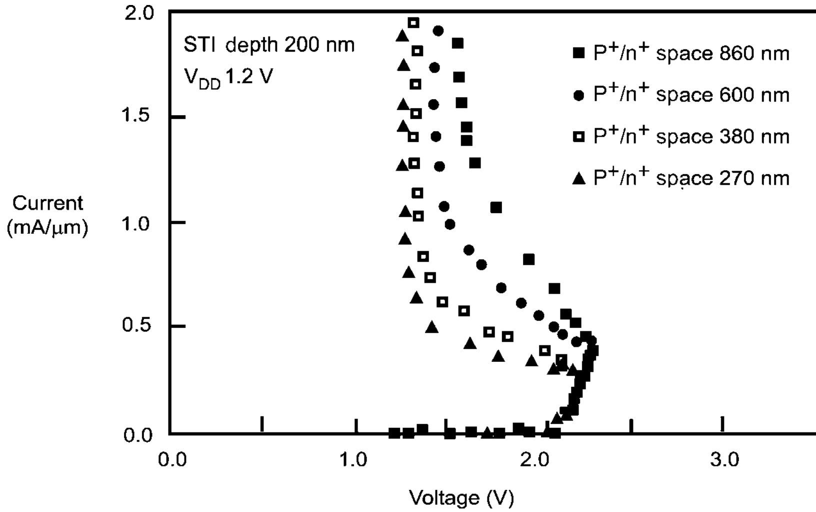



**Figure 5.11** CMOS latchup simulation results as a function of p+/n+ spacing (@ $V_{DD} = 1.2\,V$ and STI depth of 200 nm).

- bulk recombination;
- surface recombination;
- scaling impact of p+/n+ spacing.

## 5.6.1 Recess Oxide (ROX) and LOCOS Isolation

ROX and LOCOS isolation processes were early NMOS, PMOS and CMOS isolation strategies for MOSFET development. LOCOS isolation structures were the dominant isolation strategy during the 1980s to early 1990s (e.g. 1.0–0.25-μm CMOS technology). In ROX and LOCOS isolation concepts, half the isolation structure extended above the silicon surface and the other half formed below the silicon surface. At the edges of the ROX and LOCOS structures, a 'bird's beak' effect was formed at the silicon–silicon dioxide ($Si–SiO_2$) interface. In these technology generations, the semiconductor CMOS MOSFET source and drain junctions extended below the isolation depth. The key features that influence latchup in LOCOS isolation defined CMOS technology are as follows:

- Lateral pnp transistors were formed where the sidewall of the p-channel MOSFET source/drain junction, below the LOCOS isolation depth, defined the lateral transistor emitter area.
- Lateral npn transistors were formed where the sidewall of the n-channel MOSFET source/drain junction, below the LOCOS isolation depth, defined the lateral transistor emitter area.
- n+ guard rings contained within the n-well, between the p-channel MOSFET and the n-well edge, reduced the lateral current transport by increasing the surface recombination velocity and bulk recombination.

- p+ substrate guard rings contained within the p− substrate, between the n-channel MOSFET and the n-well edge, reduced the lateral current transport by increasing the surface recombination velocity and bulk recombination.

Today, LOCOS isolation technology is utilized in low-cost CMOS technology, high-voltage CMOS technology, to smart power technology. LOCOS isolation allows for lateral heat transfer reducing the surface heating in CMOS and power technologies. In these technologies, drain-extended MOS (DeMOS) and lateral diffused MOS (LDMOS) power transistors, and additional n-body and n-tub deeper implants are used below the LOCOS isolation leading to increased latchup issues.

## 5.6.2 Shallow Trench Isolation (STI)

Shallow trench isolation was integrated into CMOS technology to eliminate the negative attributes of LOCOS isolation. LOCOS isolation introduced lateral 'bird's beak' effect; this issue impacted MOSFET channel width control [36]. STI structures offers an abrupt 'lateral dielectric barrier'. Minority carriers that are injected from the emitter diffuse vertically as a result of the dielectric abruptness; this forces the minority carriers to move across potential gradients in the vicinity of the STI structure. In the presence of surface recombination and the lower recombination times with well depth (e.g. doping increases with depth due to retrograde wells), the recombination rate increases as the carrier diffuses toward the STI bottom region. In LOCOS isolation, there was little dielectric abruptness to impede lateral diffusion, and hence did not interfere with minority carriers moving across potential gradients. In 1991, STI technology was first integrated into a 0.5-μm CMOS DRAM and logic technology.

Voldman showed that for a 0.25-μm STI-defined CMOS technology pnp parasitic $\beta_{pnp}$ is not a strong function of the p+/n+ space; results showed that the n-well doping concentration had a significantly stronger effect [36]. With scaling, STI dimensions are scaled laterally and vertically to maintain width-to-depth aspect ratio. Consequently, the p- and n-well depths are scaled to maintain the same doping concentration under the STI and to reduce lateral implant scattering.

Brown first showed the STI scaling implications on latchup [42]. In Table 5.11, the parasitic $\beta_{npn}$ versus n+ to n-well space as a function of STI depth is shown (e.g. for a n+ to n-well space of 0.6–1.1 μm and $T = 140°C$). For the 0.6-μm n+ to n-well space, the parasitic $\beta_{npn}$ increased from 2.1 to 2.35 as the STI depth scaled from 0.44 to 0.35 μm. For the 1.1-μm n+ to n-well space, the parasitic $\beta_{npn}$ increased from 1.5 to 1.75. For all physical n− to n-well dimensions, there was a shift in $\beta_{npn}$ on the order of $\Delta\beta_{npn} = 0.25$. For the lateral pnp transistor, for a 0.3-μm p+ to n-well space, the parasitic $\beta_{pnp}$ increased from 1.7 to 2.25 as the STI depth scaled from 0.44 to 0.35 μm. For the 0.7-μm p+ to n-well space, the parasitic $\beta_{npn}$ increased from 1.6 to 2.1. For all physical p+ to n-well dimensions, there was a shift in $\beta_{pnp}$ on the order of $\Delta\beta_{pnp} = 0.4$ [42].

**Table 5.11** Latchup structure npn and pnp bipolar current gains versus STI depth.

| STI depth (μm) | npn bipolar current gain, 0.6 μm n+/nw (@ 140°C), $\beta_{npn}$ | pnp bipolar current gain, 0.3 μm p+/nw (@ 140°C), $\beta_{pnp}$ |
|---|---|---|
| 0.44 | 2.1 | 1.7 |
| 0.35 | 2.35 | 2.25 |

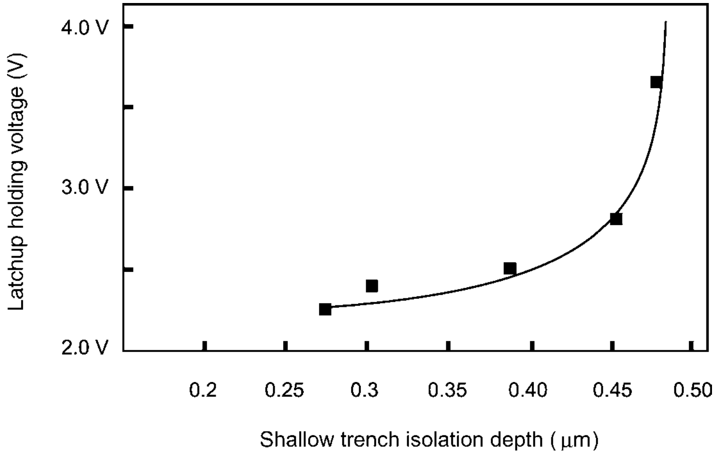



**Figure 5.12** CMOS holding voltage $V_H$ as a function of STI depth.

STI structures also have an influence on the holding voltage $V_H$ [43]. Figure 5.12 shows a plot of the latchup holding voltage versus STI depth. As the STI depth is reduced, the holding voltage decreases.

STI depth lowers the bipolar current gain because of an increase in both the effective base width and the recombination rate; hence, both the 'base defect' term and the 'emitter defect' term are influenced by the STI structure. As a result, it is anticipated that the holding voltage $V_H$ will increase with the lowering of the bipolar current gains. Figure 5.13 shows the holding voltage $V_H$ as a function of the bipolar current gain $\beta$ results. The holding voltage $V_H$ can also be shown to be a function of the bipolar

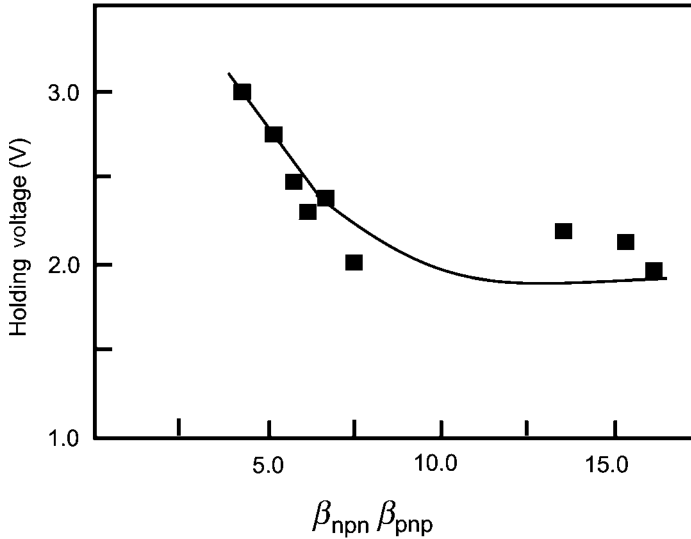



**Figure 5.13** CMOS holding voltage $V_H$ as a function of the bipolar 'beta product' ($\beta_{pnp}\beta_{npn}$).

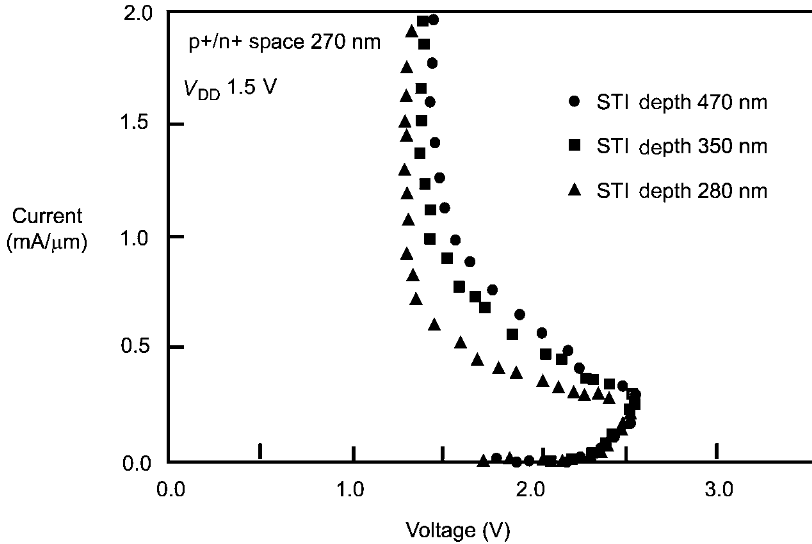



**Figure 5.14** Simulation of CMOS latchup I–V characteristic versus STI depth.

current gain product $\beta_{pnp}\beta_{npn}$. Figure 5.14 shows simulation as a function of three different STI depths (Figure 5.14). The deeper STI depth has spoiled the lateral $\beta_{npn}$, which results in an increase in $V_H$. The deeper STI structure slightly improves the trigger current $I_{TR}$; however, the largest effect is on the holding voltage $V_H$.

## 5.7 SILICIDE

Self-aligned silicide, known as salicide, is used to provide low MOSFET source and drain resistance. MOSFET source regions are the emitters of the parasitic transistor elements that are part of the parasitic pnpn structure. During the formation of the salicide, silicon is consumed, leading to the refractory metal penetration below the wafer surface [48]. There are two effects associated with the salicide formation. First, the lower emitter resistance influences the latchup criterion for CMOS latchup, as discussed in Chapter 3. As a result of the penetration into the MOSFET source and drain regions, the emitter width decreases. Second, the bipolar current gain (for both npn and pnp transistors) changes as a result of the effective emitter width change. The emitter width decreases, leading to a higher reverse current injection, and modifying the 'emitter defect' term in the bipolar current gain equation.

## 5.8 TRIPLE WELL

In advanced CMOS, RF CMOS and RF BiCMOS SiGe technologies, noise isolation and back-bias requirements for MOSFETs in an isolated epitaxial or well region that can be separated from the physical p− substrate are desirable. Structures that allow the separation of the p-well from the low-doped p− substrate to form an 'isolated MOSFET' are advantageous; this technology is also referred to as 'triple-well' technology.

### 5.8.1 Triple Well–Separate Wells

In the initial objective of ‘triple-well’ technology, it was believed that CMOS latchup would no longer be a technology issue because of the ability to separate the p-channel MOSFET structure from the n-channel MOSFET structure, avoiding formation of the parasitic pnpn device. In early concepts, the elimination of the parasitic silicon-controlled rectifier (SCR) was to free ‘triple-well’ technology from latchup concerns. As practiced today, circuit designers desire to remap dual-well structures to triple-well implementations without a change in the on-chip design, ground rules or p+/n+ spacing rules. As a result, ‘triple well’ is being practiced not as isolated regions but ‘merged triple well’, where the n-well and associated isolating buried layers are integrated. In the following sections, we will distinguish between dual well CMOS, triple well CMOS and merged triple well CMOS.

### 5.8.2 Decoupling of pnp and npn by Spatial Separation

With mixed signal, for wired and wireless applications, noise separation has led to an increased interest in triple-well technology. RF technology must also be placed in p− substrate to provide noise isolation between digital, analog and RF circuits. Additionally, the ability to bias the MOSFET devices at a different potential for application needs, reliability and latchup has lead to an increased interest in ‘isolated MOSFETs’ or triple-well technology. With the rapid remapping of digital and analog applications, circuit designers want to remap from ‘dual well’ to ‘triple well’ without the redesign and readjustment of designs and spacings.

In this section, latchup measurements of a standard ‘dual-well’ implementation to a ‘merged triple-well’ implementation in a high-resistivity p− substrate wafer will be compared. Figure 5.15 shows an example of STI-defined dual-well CMOS technology. The n-well and p-well implant energies and doses are established so that the junction between the two implants remains centered under the STI structure. A lateral parasitic pnp bipolar transistor is formed between the p+ diffusion, the n-well and the p− substrate region. Additionally, a parasitic npn bipolar transistor is formed between the n+ diffusion, the p-well and the n-well region. These cross-coupled pnp and npn form a parasitic pnpn element that can undergo latchup.

Figure 5.16 shows a triple-well structure, where the p-channel MOSFET is placed in an n-well and the n-channel MOSFET is placed in an isolated epitaxial region surrounded by an n-type structure. The

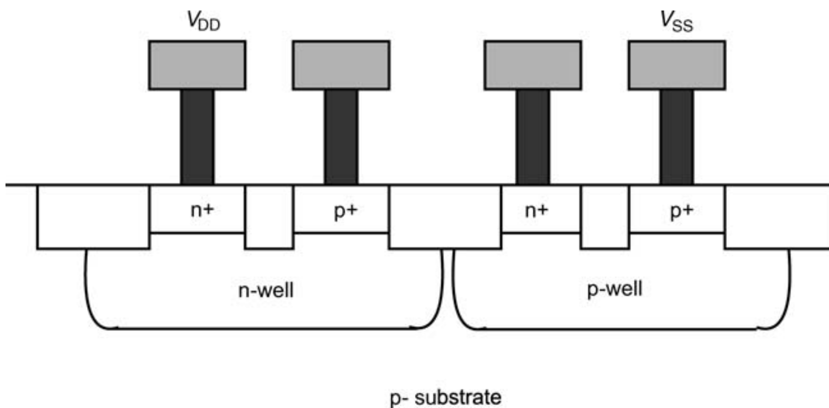



**Figure 5.15** Dual-well CMOS cross section.

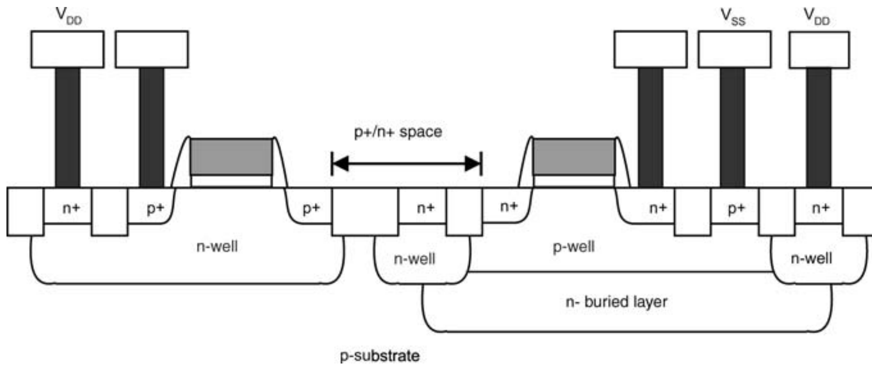



**Figure 5.16** Triple-well CMOS with p-well region separated from n-well region.

p-isolating structure consists of an n− implant underneath and bordered by an n-well ring structure. In this structure, the n-isolation structure is electrically connected to a power supply domain. With the placement of the n-isolation region, the parasitic pnp and npn transistors are decoupled preventing latchup between the p-channel and n-channel MOSFET structures.

## 5.8.3 Merged Triple-Well CMOS

With the rapid remapping of digital and analog applications, circuit designers want to remap from 'dual well' to 'triple well' without the redesign and readjustment of designs and spaces. Designers desire to remap dual-well structures to triple-well implementations without a change in the on-chip design, ground rules or p+/n+ spacing rules. As a result, 'triple well' is being practiced not as isolated regions but 'merged triple well', where the n-well and associated isolating buried layers are integrated. Hence, a key objective is to form a 'merged triple-well' implementation without the need for changing the p+/n+ space, additional guard rings or latchup rules. This is achieved by placing a n-type implant under the p-well region and using the n-well region to isolate the upper region. The n-well must form a ring structure enclosing the p-well and merge with the n-type isolating region below, while maintaining the dual-well p+/n+ space (Figure 5.17).

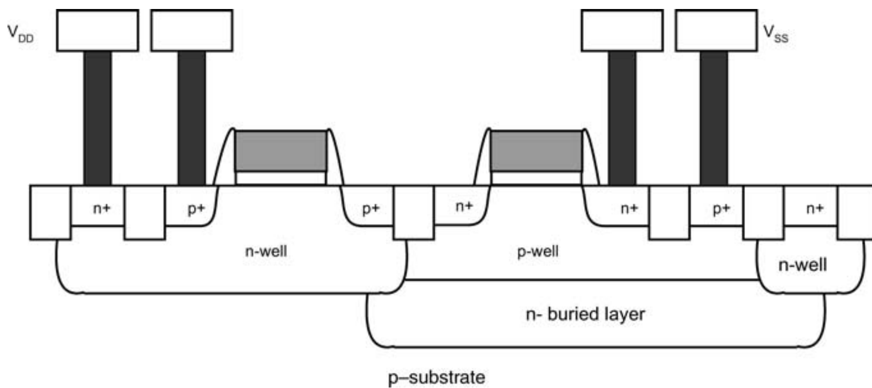



**Figure 5.17** Merged triple-well CMOS structure.

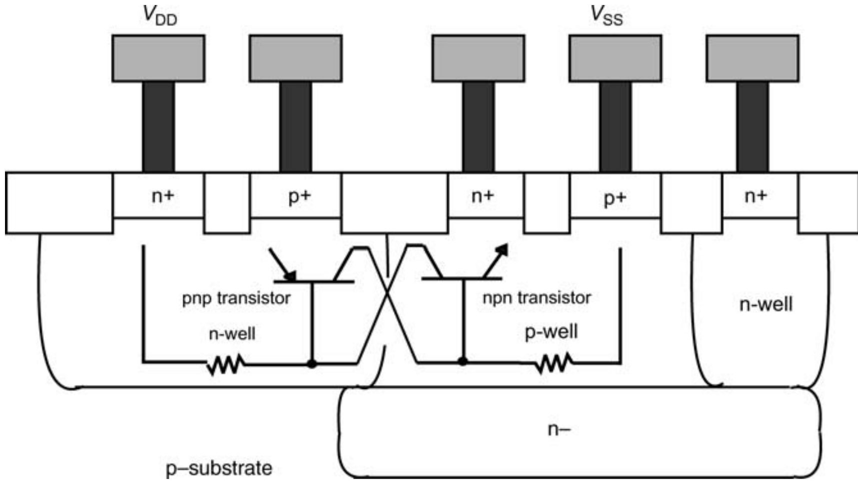



**Figure 5.18** Merged triple-well CMOS structure highlighting the parasitic lateral pnpn.

In the process of isolating the p-well regions, critical parameters that influence latchup robustness are modified. For the parasitic bipolar elements, the placement of an n-buried layer region under the n-channel MOSFET introduces a vertical npn bipolar transistor. The parasitic pnp transistor base region is also modified by the n-layer (Figure 5.18). Additionally, the 'isolated' epitaxial region resistance is modified by the separation from the p-type substrate region. To form the merged triple-well latchup test structure, the four-stripe structure is modified to mimic the ground rules of merged triple well (Figure 5.19). In the structure, the length and width of the anode, cathode, well contact and isolated epitaxial contact are 1 and 25 μm, respectively.

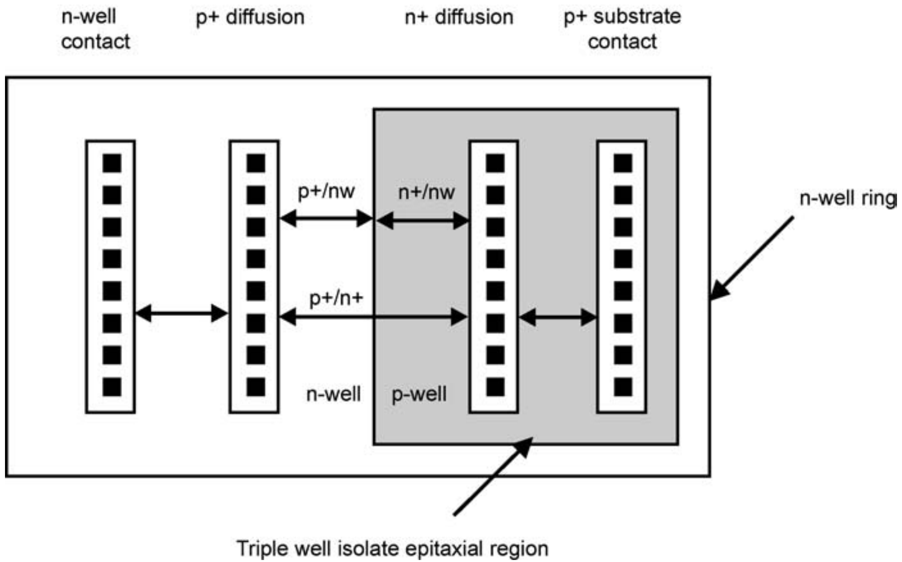



**Figure 5.19** Merged triple-well CMOS latchup test structure.

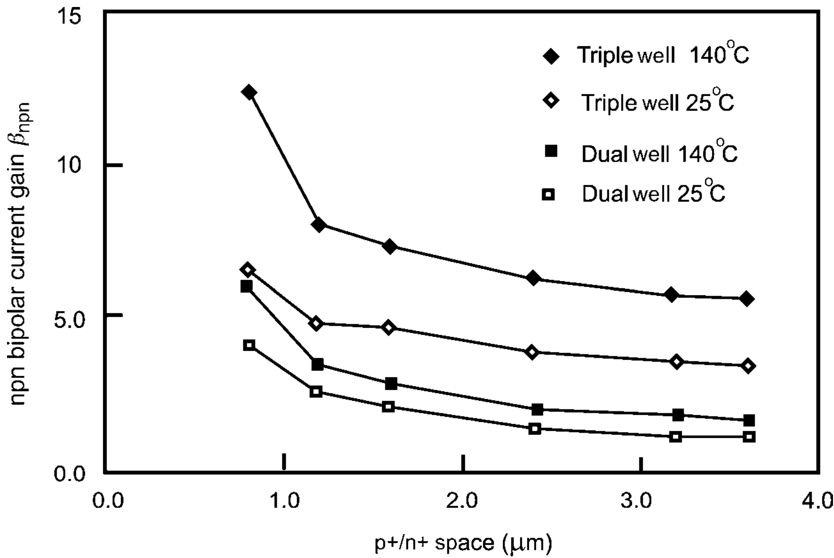



**Figure 5.20** Parasitic npn bipolar current gain as a function of p+/n+ spacing for dual-well and merged triple-well CMOS (@ 25 and 140 °C).

To provide a better understanding of the effect of the buried n+ layer on the latchup characteristics, it is important to analyze the npn and pnp transistors independently. Figure 5.20 shows the parasitic npn bipolar current gain $\beta_{npn}$ of a dual-well and a merged triple-well structure. In both cases, as the p+/n+ spacing increases, the lateral parasitic npn bipolar current gain $\beta_{npn}$ decreases. As a result of the isolating buried layer, the vertical npn bipolar current gain $\beta_{npn}$ of the merged triple-well structure is two to three times higher in magnitude compared to the dual-well structure. As the n-well to n+ spacing increases, the triple-well $\beta_{npn}$ is dominated by the vertical npn (formed from the n+ emitter, the p-well region and the n+ buried layer). As the p+/n+ space decreases, the lateral device effect in both the dual well and the triple well increases, with the triple-well curve shifted toward higher values.

Using a bipolar curve tracer, the parasitic forward and reverse $\beta_{npn}$, as well as the avalanche breakdown voltage $V_{AV}$, were measured (e.g., $T = 25\,°C$) (Table 5.12 and 5.13). The ratio of forward and reverse bipolar current gains is approximately 4–5. In the forward npn mode of operation, the avalanche breakdown is 11.8 V for all p+/n+ spaces; for the reverse npn mode of operation, $V_{AV}$ increases with p+/n+ spaces.

**Table 5.12** Dual-well CMOS parasitic npn forward and reverse bipolar current gains (@ 25 °C).

| p+/n+ spacing (μm) | Forward npn bipolar current gain $\beta_F$ | Reverse npn bipolar current gain $\beta_R$ | $V_{AV\ (REV)}$(V) |
|---|---|---|---|
| 0.8 | 3 | 0.6 | 9.2 |
| 1.2 | 2.2 | 0.55 | 10.2 |
| 1.6 | 1.8 | 0.40 | 11.8 |
| 2.4 | 1.5 | 0.35 | 13.0 |
| 3.2 | 1.2 | 0.29 | 14.0 |
| 3.6 | 1.18 | 0.20 | 14.2 |

**Table 5.13** Merged triple-well CMOS npn forward and reverse bipolar current gains (@ 25 °C).

| p+/n+ spacing (μm) | $\beta_F$ | $\beta_R$ | $V_{AV\ (REV)}$ (V) |
|---|---|---|---|
| 0.8 | 4.7 | 1.2 | 8.2 |
| 1.2 | 3.6 | 0.8 | 9 |
| 1.6 | 3.5 | 0.7 | 10.5 |
| 2.4 | 3.1 | 0.65 | 11 |
| 3.2 | 3.0 | 0.60 | 11 |
| 3.6 | 2.8 | 0.58 | 12 |

For the merged triple-well structures, the identical study was completed in evaluating the forward and reverse bipolar npn current gains (Table 5.13). From these measurements, the triple-well forward $\beta_{npn}$ is approximately three times higher at large p+/n+ spaces. In the forward npn mode of operation, the bipolar avalanche voltage was $V_{AV} = 8$–$8.2$ V. For the reverse bipolar characteristic, the bipolar avalanche breakdown increases from $V_{AV} = 8.2$ to 12 V; the reverse bipolar avalanche voltage is lower for the merged triple-well structure.

For evaluation of the lateral $\beta_{pnp}$, the pnp transistor formed between the p+ emitter, the n-well and the p-well region was measured for the dual well and the merged triple well. For the dual-well case, the collector is integrated with the p− substrate; for the merged triple-well case, the p− substrate is isolated from the interaction. For the parasitic $\beta_{pnp}$, the opposite effect occurs where the merged triple-well structure results are lower than those of the dual-well structure. In the case of the pnp, the triple-well bipolar current gain $\beta_{pnp}$ is reduced by approximately five times (Figure 5.21). In the case of the triple-well structure, $\beta_{pnp}$ is reduced well below unity and demonstrates a weaker dependence on the p+/n+ space.

To further evaluate the transistors, measurements were taken on a Tektronix 576 bipolar curve tracer. Table 5.14 contains the dual-well $\beta_{pnp}$ for the forward and reverse pnp parasitic BJT. As the p+/n+ space

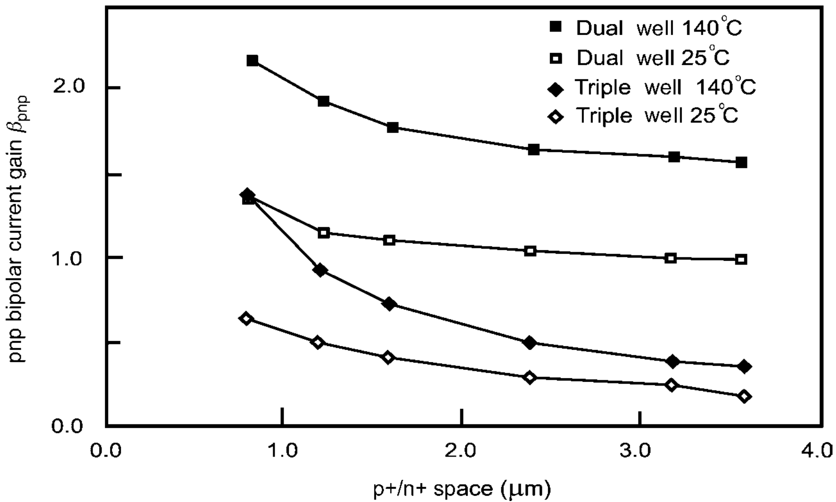



**Figure 5.21** Parasitic pnp bipolar current gain as a function of p+/n+ spacing for dual-well and merged triple-well CMOS (@ $T = 25\,°C$ and $T = 140\,°C$).

**Table 5.14** Dual-well CMOS parasitic pnp forward and reverse bipolar current gains (@ $T = 25\,^{\circ}\mathrm{C}$).

| p+/n+ spacing (μm) | Forward npn bipolar current gain $\beta_F$ | Reverse npn bipolar current gain $\beta_R$ |
|---|---|---|
| 0.8 | 1.2 | 0.16 |
| 1.2 | 1.1 | 0.16 |
| 1.6 | 1.05 | 0.09 |
| 2.4 | 0.9 | 0.08 |
| 3.2 | 0.9 | 0.08 |
| 3.6 | 0.9 | 0.08 |

increases, the forward and the reverse $\beta_{pnp}$ flatten to a constant value. Also note that there is a 10-fold difference between the forward and the reverse $\beta_{pnp}$ values.

For the case of the parasitic pnp, in the merged triple well, there exist separable lateral and vertical pnp elements that can be independently evaluated. In the automated measurements, the lateral device was addressed for the pnpn measurements and the pnp. Figure 5.22 highlights the lateral pnp element formed between the p+ emitter, the n-well and the p+ contact to the isolated p-well.

Figure 5.23 shows the vertical parasitic pnp BJT in the triple-well structure. In this case, the vertical pnp BJT does not participate in the latchup characteristics but does inject current into the p− substrate. By separating the p− substrate contact and the local p-well isolated contact, independent measurements can be taken.

Table 5.15 shows the forward and reverse pnp bipolar BJT for the lateral and the vertical BJT parasitics. For small p+/n+ spacings, the pnp lateral and vertical components are comparable in value. As the p+/n+ space increases, the forward vertical $\beta_{pnp}$ increases, and then saturates at $\beta_{pnp} = 0.8$. For the lateral forward $\beta_{pnp}$, the lateral component continues to decrease with the p+/n+ space from $\beta_{pnp} = 0.65$ to $\beta_{pnp} = 0.18$. Note for the largest p+/n+ spaces, the vertical-to-lateral $\beta$ component ratio is approximately 4. For the reverse bipolar characteristics, both the vertical and the lateral $\beta_{pnp}$ decrease with p+/n+ spaces; both decrease by approximately 4×.

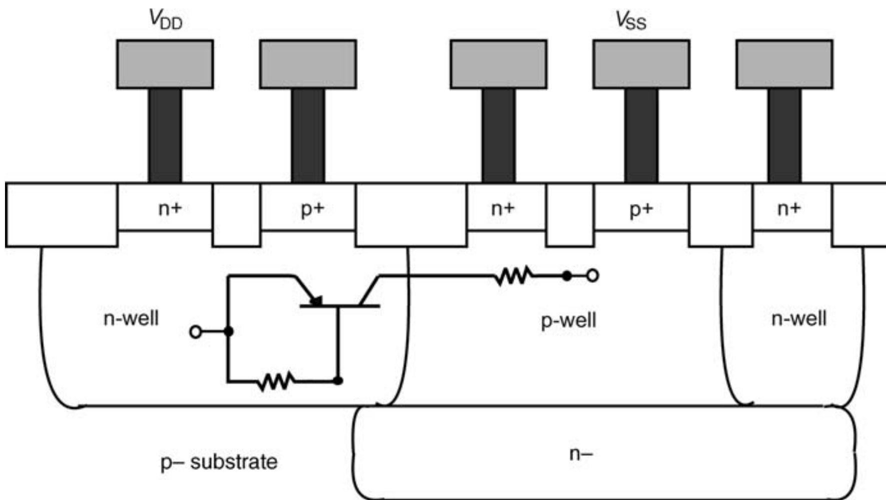



**Figure 5.22** Merged triple-well structure highlighting a lateral parasitic pnp bipolar transistor.

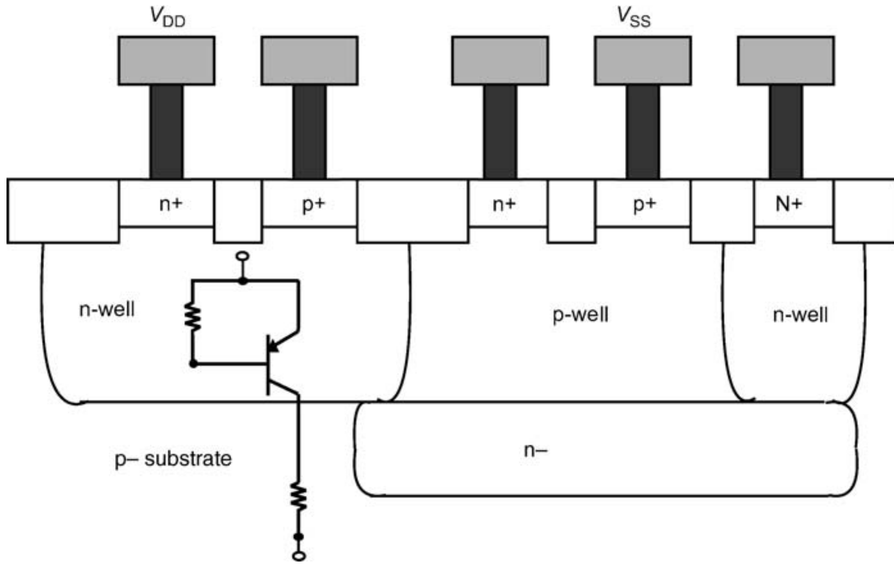



**Figure 5.23** Merged triple-well structure highlighting a vertical parasitic pnp bipolar transistor.

In Figure 5.24, the product of the two bipolar gain terms, $\beta_{pnp}\beta_{npn}$, is shown. The interesting results from this study show that the 'beta product' results are comparable in temperature and spacing relations (except minimum spacing). Hence, to fully appreciate the difference, the results must be deconvolved into the undershoot and overshoot sensitivities.

Figure 5.25 shows the undershoot sensitivity analysis for dual- and triple-well CMOS. As the temperature increases, the current needed to initiate undershoot for a given p+/n+ space decreases because of the $\beta$-enhancement with temperature. Note that for minimum structure size both structure types have comparable results. As the p+/n+ space increases, the merged triple-well structure has a higher sensitivity to undershoot because of the vertical npn bipolar transistor. Figure 5.26 shows the overshoot analysis. Because of the significant reduction of the pnp bipolar current gain $\beta_{pnp}$, the merged triple-well structure is less sensitive to the overshoot phenomenon.

The first key point is that although it was believed that triple-well CMOS would eliminate latchup concerns by isolating n-channel MOSFETs from p-channel MOSFETs, as it is being practiced today, in the form of a merged triple well, CMOS latchup has not been 'cured' and still remains a concern.

**Table 5.15** Merged triple-well parasitic pnp forward and reverse bipolar current gains for lateral and vertical structures (@ $T = 25\,^\circ$C).

| p+/n+ spacing (μm) | Lateral pnp bipolar forward $(\beta_F)_{LAT}$ | Vertical pnp bipolar forward $(\beta_F)_V$ | Lateral pnp bipolar reverse $(\beta_R)_{LAT}$ | Vertical pnp bipolar reverse $(\beta_R)_V$ |
|---|---|---|---|---|
| 0.8 | 0.65 | 0.8 | 0.17 | 0.040 |
| 1.2 | 0.5 | 0.88 | 0.13 | 0.017 |
| 1.6 | 0.5 | 0.85 | 0.118 | 0.017 |
| 2.4 | 0.28 | 0.82 | 0.085 | 0.015 |
| 3.2 | 0.22 | 0.80 | 0.068 | 0.014 |
| 3.6 | 0.18 | 0.80 | 0.041 | 0.013 |

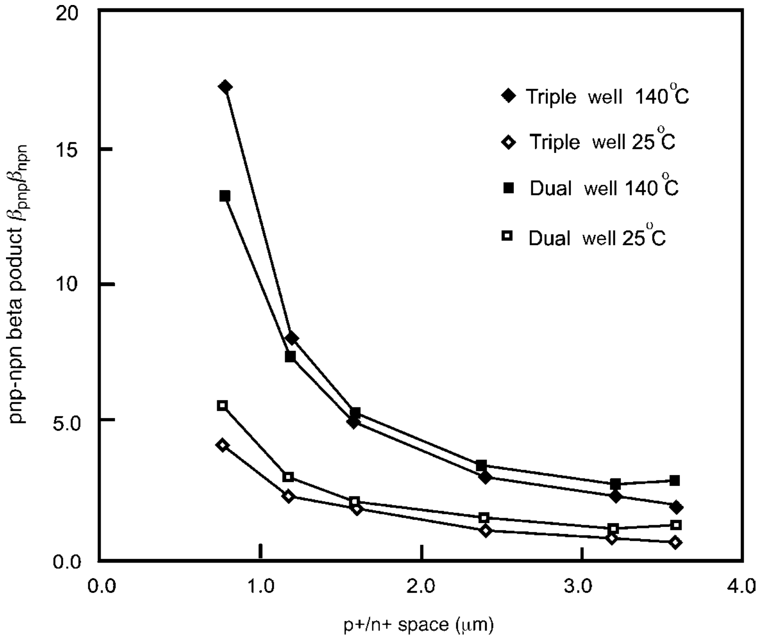



**Figure 5.24** Beta product of parasitic pnp and npn bipolar current gains versus p+/n+ space for dual-well and merge triple-well CMOS (@ $T = 25\,°C$ and $T = 140\,°C$).

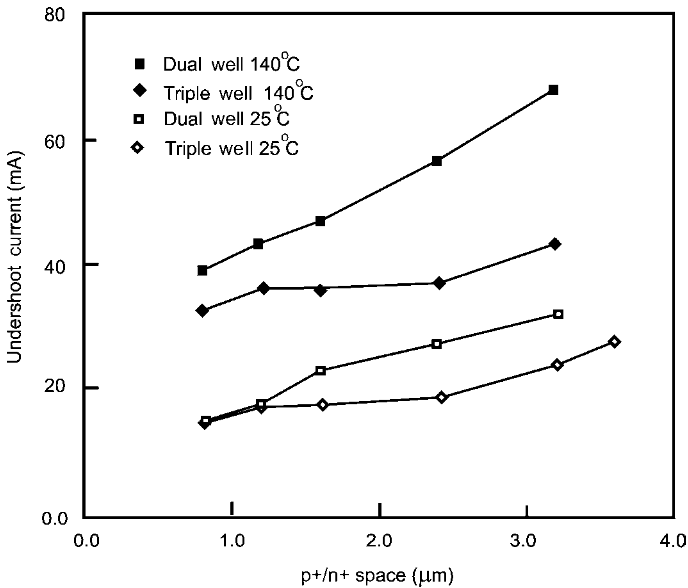



**Figure 5.25** Latchup undershoot versus p+/n+ space for dual-well CMOS and merged triple-well CMOS pnpn structures (@ $T = 25\,°C$ and $T = 140\,°C$).

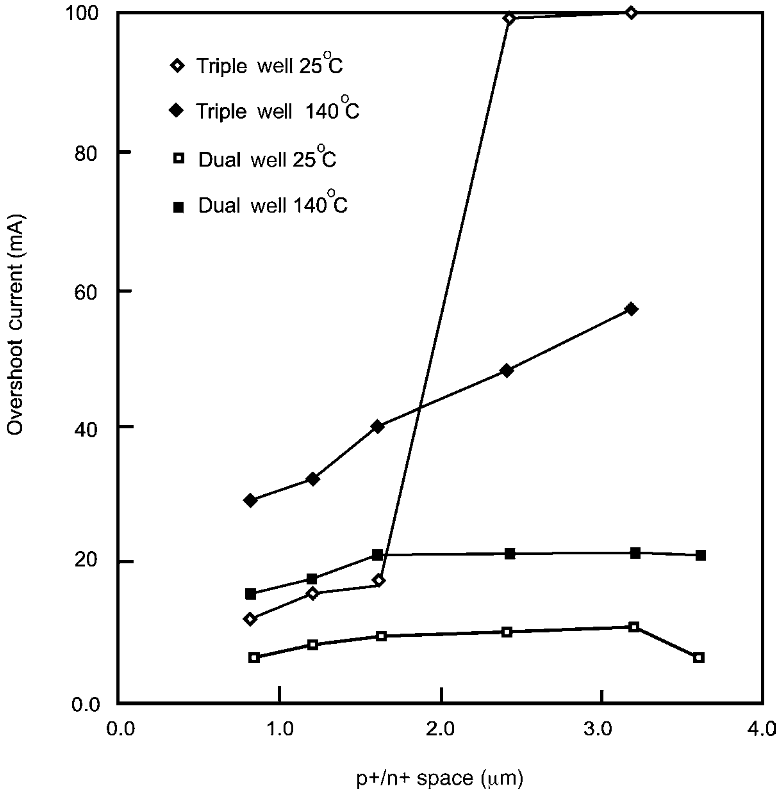



**Figure 5.26** Latchup overshoot versus p+/n+ space for dual-well CMOS and merged triple-well CMOS pnpn structures (@ $T = 25\,°\mathrm{C}$ and $T = 140\,°\mathrm{C}$).

The second key point is that the CMOS latchup in a merged triple-well structure has a different response from a dual-well CMOS. In our case, $\beta_{\mathrm{pnp}}$ was significantly reduced below the dual-well CMOS levels; additionally, the addition of the n-type buried layer increased $\beta_{\mathrm{npn}}$. The $\beta_{\mathrm{npn}}$ reduction will be a function of the buried layer design, but the $\beta_{\mathrm{npn}}$ increase is inevitable because of the n− region placement under the n-channel MOSFET structure. Additionally, the reduction of the first bipolar current gain and the enhancement of the second bipolar current gain changes the relationships, as well as the sensitivity of the undershoot and the overshoot currents.

Latchup occurs in CMOS technology when a semiconductor pnpn network undergoes regenerative feedback between the parasitic pnp and npn bipolar current gains. The first criterion is that the pnp and npn bipolar transistors must be in forward active mode. The second criterion for CMOS latchup is when the following inequality is satisfied:

$$\beta_{\mathrm{n}}\beta_{\mathrm{p}} \geq \frac{I + I_{\mathrm{s}x}\beta_{\mathrm{n}}}{\left[I - I_{\mathrm{w}}\frac{(\beta_{\mathrm{n}}+1)}{\beta_{\mathrm{n}}} - I_{\mathrm{s}x}\right]},$$

where

$$I_{\mathrm{w}} = \frac{(V_{\mathrm{BE}})_{\mathrm{pnp}}}{R_{\mathrm{w}}} = \frac{V_0}{R_{\mathrm{w}}} \ln\left[\frac{I - I_{\mathrm{w}}}{(I_o)_{\mathrm{p}}}\right]$$

and

$$I_{\mathrm{s}x} = \frac{(V_{\mathrm{BE}})_{\mathrm{npn}}}{R_{\mathrm{s}x}} = \frac{V_0}{R_{\mathrm{s}x}} \ln\left[\frac{I - I_{\mathrm{s}x}}{(I_o)_{\mathrm{n}}}\right].$$

When latchup is initiated, the holding current condition can be expressed as

$$I_{\mathrm{H}} = \frac{\beta_{\mathrm{p}}(\beta_{\mathrm{n}} + 1)I\mathrm{w} + (\beta_{\mathrm{p}} + 1)\beta_{\mathrm{n}} I_{\mathrm{s}x}}{\beta_{\mathrm{p}}\beta_{\mathrm{n}} - 1}.$$

In the case where the npn current gain is significantly increased,

$$I_{\mathrm{H}}\big|_{\beta_{\mathrm{n}} \gg 1} \cong I_{\mathrm{w}} + \left(\frac{\beta_{\mathrm{p}} + 1}{\beta_{\mathrm{p}}}\right) I_{\mathrm{s}x}$$

and the complementary case is

$$I_{\mathrm{H}}\big|_{\beta_{\mathrm{p}} \gg 1} \cong I_{\mathrm{s}x} + \left(\frac{\beta_{\mathrm{n}} + 1}{\beta_{\mathrm{n}}}\right) I_{\mathrm{w}}.$$

For our case, the pnp bipolar current gain is significantly reduced below unity, the npn bipolar current gain is greater than unity and the product of the bipolar current gain is greater than 1:

$$I_{\mathrm{H}} \approx \left(\frac{1}{\beta_{\mathrm{p}}}\right) I_{\mathrm{s}x} + \left(\frac{\beta_{\mathrm{n}} + 1}{\beta_{\mathrm{n}}}\right) I_{\mathrm{w}}.$$

Hence, the symmetry of the product of the bipolar current gains, $\beta_{\mathrm{npn}}\beta_{\mathrm{pnp}}$, in the triple-well structure is distinct from the dual-well bipolar current gain symmetry. In advanced CMOS, typically the npn bipolar current gain is higher than the pnp bipolar current gain; in merged triple-well structures this asymmetry of bipolar gains is further enhanced with a greater difference between the npn and the pnp bipolar transistors. As a result, the triple-well process may be more sensitive to the negative undershoot phenomenon, but less sensitive to the positive overshoot phenomenon.

A very important advantage of the merged triple-well structure over dual-well CMOS is that the latchup phenomenon is not a strong function of the substrate doping concentration but a function of the p-well and the n-type isolation layer design choices. Additionally, the electrical isolation of the n-channel MOSFETs from the substrate can also reduce the effects from SEL from ionizing radiation sources and cable discharge events (CDEs).

### 5.8.4 Merged Triple Well with Blanket Implant

In the migration from dual-well to triple-well CMOS, many product development teams want to 'map' the layout design without significant redesign of a product. In the migration of a layout design within the same technology generation, it is desired that the dimensional features of physical designs remain the same, and in the case of a design migration to a future technology generation, it is an expectation that the dimensional features will be scaled.

In the case of integration of the triple-well region that is placed selectively, there are dimensional issues. First, there must be overlap of the n-well edge and the triple-well implant edge to guarantee electrical isolation of the p-well epitaxial region. Second, implant scattering issues of the buried implant require a minimum space between the implant mask edge and n-channel and p-channel transistors. As a result of alignment issues and implant scattering issues, the ability to maintain the p+/n+ spacings or scale them to smaller spaces is more difficult when the edge of the triple-well implant is required to intersect the n-well to p-well region. Hence, it is desirable for technologies to allow either the buried layer extend under the n-well region or blanket buried layer implants (Figure 5.27). The extension of the n-doped triple-well implant under the n-well region leads to the following advantages:

- larger vertical pnp base width;
- formation of a high-low or low-high n−/n+ step transition;
- lower shunt resistance in series with the pnp transistor;
- electrical isolation from the p− substrate;
- higher capacitance in the latch network.

First, the vertical pnp bipolar current gain $\beta_{pnp}$ is significantly reduced with the buried layer under the n-well structure. Hence, the entire length of a p-channel MOSFET will have low vertical $\beta_{pnp}$ current gain. Second, a step transition is formed; depending on the doping concentration of the well and

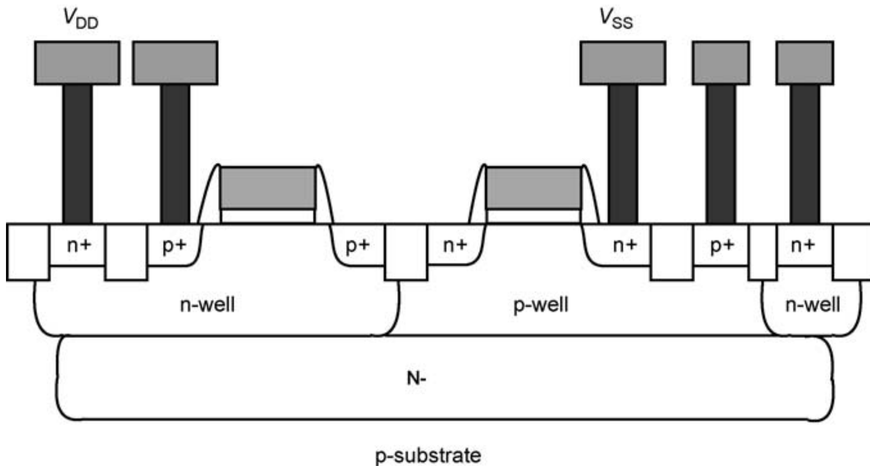



**Figure 5.27** Cross section with a 'blanket triple-well implant'.

the buried layer, this can lead to either a lower or a higher lateral parasitic pnp bipolar current gain. As noted in Chapter 3, the doping step transition of the vertical bipolar device influences the lateral bipolar device current gain. Third, the combined structure of the n-well and the triple-well implant leads to a lower 'n-well shunt resistance'. The fourth issue is that the injection into the p− substrate does not affect the local pnpn structure, which is electrically isolated.

The fifth issue is that the area of the buried layer is larger than the n-well region, leading to a significantly larger total capacitance to the chip substrate. As a result, the transient latchup response and external latchup will be altered. In the case of transient latchup, the capacitance of the pnpn latch network has been altered changing the ramp rate – capacitance (e.g. $KC$ term, where $K = \mathrm{d}V_{DD}/\mathrm{d}t$ from Chapter 3). Second, injection in the substrate will be collected by the blanket n-buried layer to p− substrate junction.

A latchup triple-well design practice is as follows:

- Extension of the triple-well layer under the n-well leads to lower vertical pnp bipolar current gain.
- Extension of the triple-well layer under the n-well lowers the n-well shunt resistance.
- Injection into the p− substrate from the vertical pnp transistor can be isolated from the lateral pnpn latchup structure.
- Large area blanket buried layer implants can lower the transient response of the pnpn latchup influencing the $KC$ product term for switching the latchup pnpn network.
- Large area blanket implant can minimize the issue of external latchup by collection of minority carriers in the large buried layer to p− substrate region.
- p− substrate injection sources can be eliminated with the blanket buried layer minimizing external latchup concerns.

## 5.9 HIGH-DOSE BURIED LAYER

A concept that has been proposed for mainstream CMOS is the high-dose buried layer (e.g. also referred to as the heavily doped buried layer) whose dopant type is the same as the substrate wafer. HDBL research focused on the influence of the MeV implanted layers on semiconductor devices. Zappe and Hu [49], Lin and Ting [50], Kuroi et al. [51], Jacobson et al. [52], Borland [53, 54] and Morris and Rubin [55–59] pursued the semiconductor research and integration issue to integrate HDBL into semiconductor technology. At that time, the concept was replaced with p++ substrate wafers. With the migration of low-doped substrate wafers, the issue of the use of the HDBL for advanced mainstream CMOS has a renewed interest and is valuable as technology integration continues to sub-45-nm technologies.

### 5.9.1 Buried Implanted Layer for Lateral Isolation (BILLI) Structure

Borland developed a structure known as the 'buried implanted layer for lateral isolation' for the purpose of latchup improvement and the elimination of epitaxial growth steps on a p++ substrate wafer [53, 54]. In this process, a blanket p++ implant was implanted across the entire wafer, where the implant impinged on a thick mask level. In this BILLI concept, the MeV implant penetrated through the mask and all sections of the semiconductor device forming a blanket layer with a transition step at

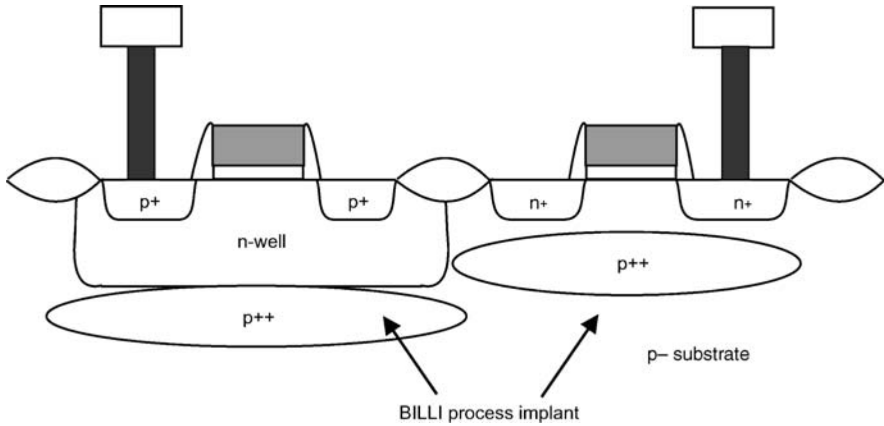



**Figure 5.28** Cross section of the BILLI structure.

the mask edge. In the area formed by the n-well, the p++ implant was below the n-well and abutted the n-well to p− substrate junction. In the masked region, the implant formed an 'upper layer' that was a continuous boron layer closer to the substrate surface (Figure 5.28). The objective was to surround the n-well with a p+ implant on the bottom and the edges to 'spoil' the vertical and lateral bipolar elements. In this concept, this was for full integration with retrograde n-well and p-well regions. In this BILLI concept, there are practical integration issues. First, the ability to provide a good 'transition region' between the two 'layers' of the p+ implant that guaranteed good latchup results was key to the success of the concept. Second, the ability to integrate the n-well design point, the p-well design point, isolation and the two BILLI layers is difficult. Third, the implantation through a mask leads to photoresist and mask hardening issues.

## 5.9.2 Continuous High-Dose Buried Layer

Morris and Rubin focused on the second concept for CMOS latchup improvement, which is the usage of a continuous buried layer that is not masked but uniform across a semiconductor chip [55–59]. Using a continuous p++ implanted buried layer in a p− substrate wafer solves implanting through photoresist masks and integration into the n-well, p-well and isolation design point. By not constraining the CMOS n-well/p-well/isolation integration, the implant can be integrated without significant alteration of the design ground rules or semiconductor product design. The latchup semiconductor design practice of HDBL demonstrated the following:

- Lower substrate shunt resistance is achieved allowing for improved latchup or design relaxation of the p+ contact to n-channel MOSFET space.
- Shorter minority carrier recombination time (e.g. electron recombination time) is achieved in a p− substrate improving immunity to external latchup.

Figures 5.29 and 5.30 show an example of a vertical doping profile of a dual-well process (e.g. under the n-well and p-well, respectively) with a boron HDBL in a boron p− substrate. The HDBL is a boron

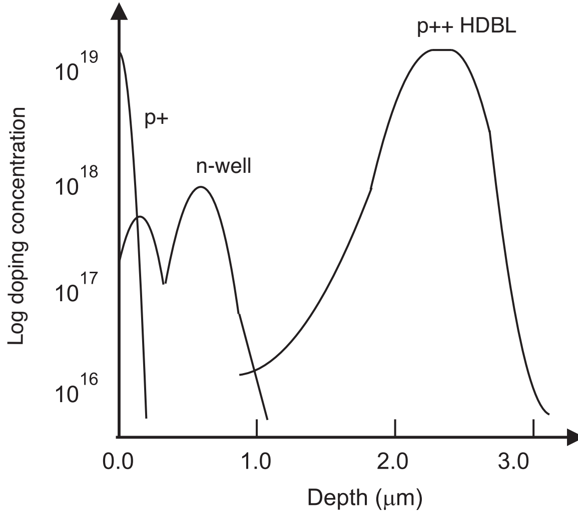



**Figure 5.29** CMOS vertical doping profile highlighting n-well and high-dose buried layer.

implant with an implant energy of 1.7 MeV and a dose of $1 \times 10^{15}\,\text{cm}^{-2}$. The deep p++ HDBL provides a low substrate shunt resistance. Figure 5.31 and Table 5.16 show that the HDBL has more than doubled the trigger current $I_{TR}$, as well as improved the holding voltage $V_H$. Since the bipolar pnp collector current will predominately flow in the lowest resistive path to the $V_{SS}$ terminal, the low-resistivity shunt layer is much more effective in sustaining the local ground potential at the n+/p− diode; this results in the much improved trigger current. As the substrate shunt resistance is reduced, a much larger pnp collector current is required to produce the *IR* drop within the p-well (e.g. necessary to forward bias the n+/p− diode and initiate latchup).

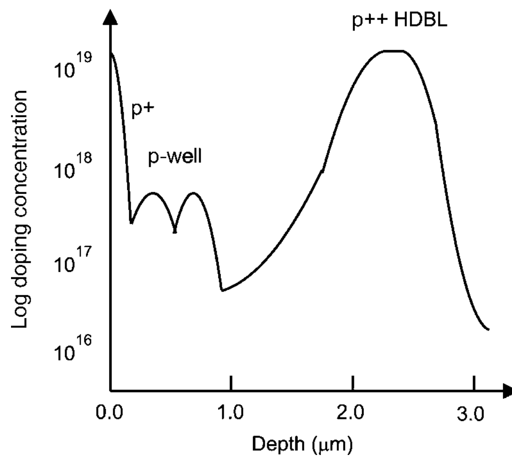



**Figure 5.30** CMOS vertical doping profile highlighting p-well and high-dose buried layer.

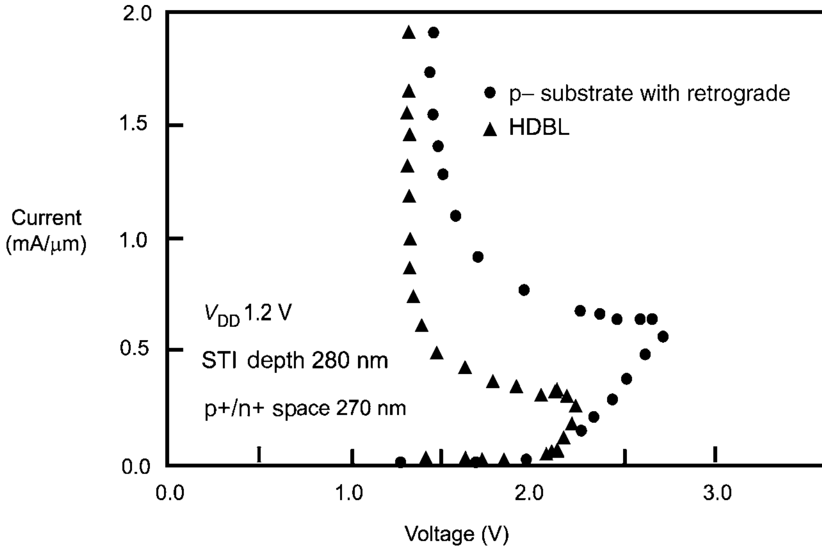



**Figure 5.31** Simulation of the CMOS latchup *I–V* characteristic with and without HDBL.

## 5.9.3 HDBL Characterization

Experiments have shown that the implanted HDBL is a more effective p+ shunt than the p−/p+ epi substrate, resulting in a more robust latchup isolation [43]. Even though the HDBL is much thinner than a p+ substrate, the peak doping is more than twice as high as the nominal p+ substrate doping, leading to increased trigger currents. Figure 5.32 and Table 5.17 illustrate the effect of dose on the trigger current $I_{TR}$, holding voltage $V_H$ and the holding current $I_H$. As can be seen, as the HDBL dose increases, the latchup parameters all improve. Figure 5.33 shows that with an increase in the STI depth as well as the HDBL, the CMOS latchup parameters improve.

## 5.9.4 HDBL Recombination Time and Resistance Measurements

With the introduction of the HDBL, as the doping concentration of the buried layer increases the minority carrier recombination time decreases. Figure 5.34 shows a plot of the minority carrier lifetime as a function of the boron dose. It can be observed that as the HDBL dose rate increases, the minority carrier lifetime is reduced. Morris *et al.* showed that in a p− wafer the minority carrier recombination time of a p− wafer was $\tau_e = 100\,\mu s$ and the diffusion length was 500−600 μm. With the introduction

**Table 5.16** Experimental results of CMOS latchup structures with and without HDBL.

| Parameters | Bulk Si | Bulk with HDBL | Percent improvement with HDBL (%) |
|---|---|---|---|
| $V_H$ (V) | 1.25 | 1.38 | 10.5 |
| Trigger voltage (V) | 2.2 | 2.69 | 22 |
| Trigger current (μA/μm) | 280 | 600 | 114 |

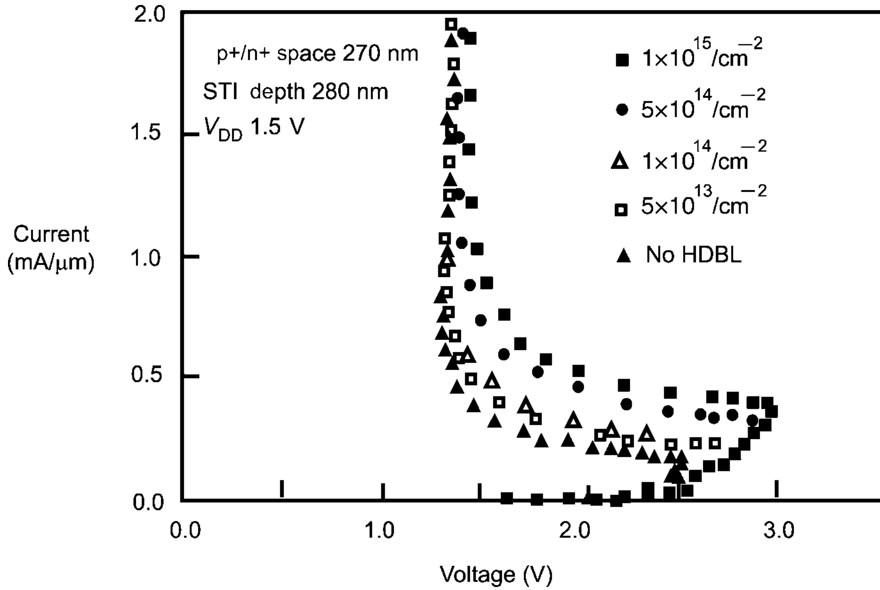



**Figure 5.32** Simulation of the latchup *I–V* characteristic versus HDBL dose.

of a p++ high-dose buried layer, the minority carrier recombination time was $\tau_e = 100\,\mu s$ and the diffusion length was 170–200 μm [59]. With the HDBL, series resistance was also reduced approximately three to five times. A latchup discipline design practice is as follows:

- HDBL can lead to a reduction in the minority carrier lifetime and diffusion length for improved CMOS latchup robustness.
- p++ HDBL provide a lower substrate shunt resistance.

## 5.9.5 HDBL and Reachthrough Implant

The HDBL concept can achieve a larger improvement with the addition of a ‘reachthrough’ implant. With the addition of a ‘reachthrough’ implant, there is an improvement in the dc and transient response. Morris showed that the addition of the reachthrough implant to the HDBL structure improves both the

**Table 5.17** Parametric data of the latchup *I–V* characteristic versus HDBL dose.

| HDBL dose ($cm^{-2}$) | $V_H$ (V) | $I_{TR}$ (μA/μm) |
|---|---|---|
| $1 \times 10^{15}$ | 1.43 | 600 |
| $5 \times 10^{14}$ | 1.39 | 490 |
| $1 \times 10^{14}$ | 1.33 | 345 |
| $5 \times 10^{13}$ | 1.31 | 320 |
| None | 1.28 | 280 |

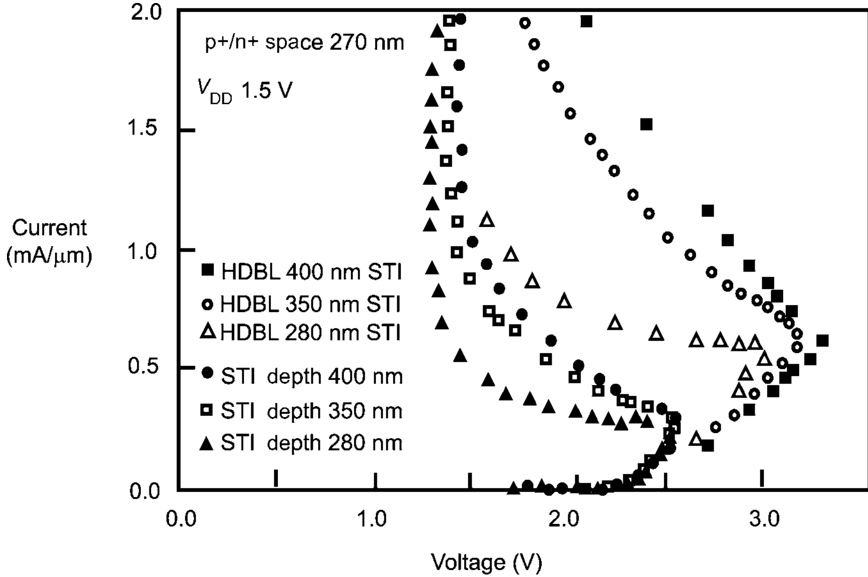



**Figure 5.33** Simulation of the latchup *I–V* characteristic with and without HDBL as a function of STI depth.

dc and the transient latchup tolerance [43]. With the lowering of the HDBL shunt resistance, the latchup trigger current will improve. Morris showed that the improved contact to the HDBL shunt layer is extremely effective in preventing latchup triggering up to a vertical pnp collector current as high as 1.6 mA/μm (with a $V_{DD}$ of 1.5 V and an STI depth of 280 nm) [43]. Note that this type of structure clearly offers a most robust solution compared to the HDBL without the reachthrough implant. Figure 5.35 shows simulated latchup *I–V* characteristics of different processes with and without the reachthrough implant. With this structure, the parasitic pnpn has not triggered at high injection current and is kept in the desired 'off' blocking state (Figure 5.35).

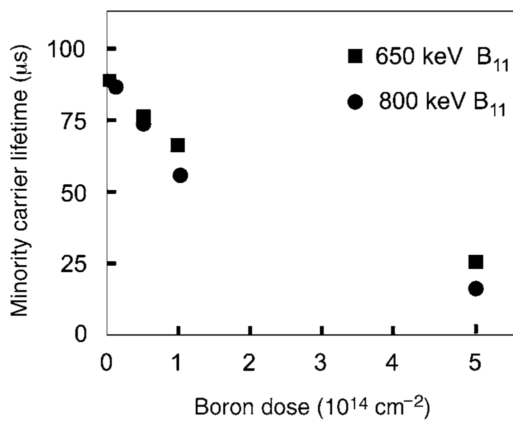



**Figure 5.34** Minority carrier lifetime as a function of the boron concentration.

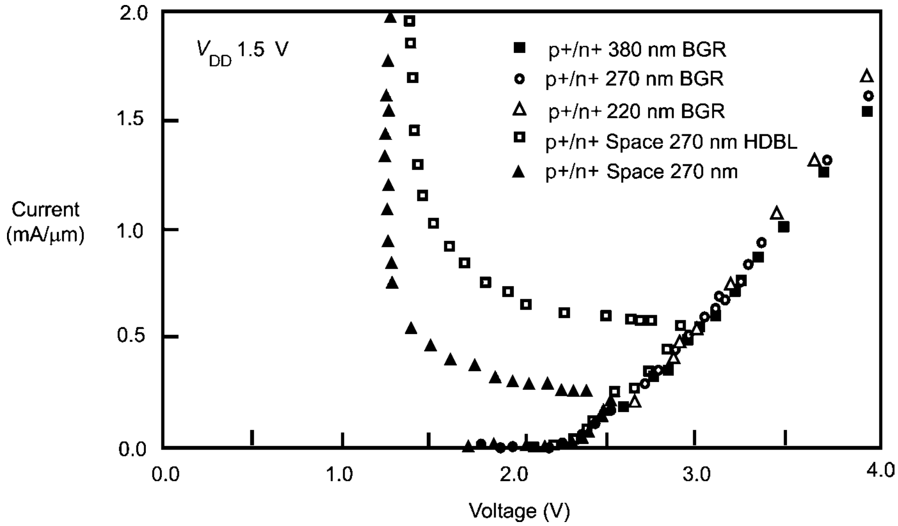



**Figure 5.35** Simulation of latchup *I–V* characteristics of different semiconductor processes.

## 5.10 FUTURE CONCEPTS

With the continued scaling of CMOS technology, alternative concepts may be needed to address CMOS latchup in the future. These concepts can include the following:

- dual depth shallow trench isolation [60];
- dual depth shallow trench isolation and edge implant [61];
- dual tone resist and edge implants [62];
- buried grid of opposite doping;
- deep trench [63–65];
- BiCMOS implants (e.g. subcollectors and varactor implants) [65];
- fully isolated triple well (isolated n-channel and p-channel regions);
- silicon on insulator.

Although today these concepts are not implemented into mainstream CMOS, it is possible to integrate them with today's semiconductor tools.

## 5.11 SUMMARY AND CLOSING COMMENTS

This is a critical chapter in the understanding of the influence of semiconductor revolutionary and evolutionary semiconductor processes and structures on CMOS latchup. Many of the transitions were part of the natural evolution of CMOS technology as it was scaled from 1-μm to 45-nm technology

generations. As technology migrated from single well, to dual well, to present day triple well, the latchup robustness also altered.

In the next chapter, semiconductor process solutions associated with BiCMOS technology will be discussed. In Chapter 6, the influence of semiconductor features associated with bipolar transistors will be utilized on the CMOS sector of the semiconductor chip. The structures of interest include epitaxial formed subcollectors, implanted subcollectors, polysilicon-filled deep trench, biased deep trench and a new isolation known as trench isolation. These will be discussed in a dual-well and a triple-well CMOS environment. In this discussion, different structures will look at both independent and combined effects on latchup.

## PROBLEMS

1. Substrate wafers have evolved from p− substrate to p++ substrates, and then back to p− wafers. As the substrate doping concentration is decreased, project how the substrate doping concentration will influence the p− substrate contact (tap) to n-channel MOSFET scaling.
2. Given that the spacing between the substrate contact (tap) and the n-channel MOSFET must be reduced, what is the influence on the noise as the density of substrate contacts increases in magnitude and periodicity?
3. Given a 'constant latchup robustness' with MOSFET technology scaling, how should the p-well doping concentration scale with dimensional scaling for the cases of the lateral dimensions: the case of lateral dimensions and substrate scaling and the case of lateral and vertical dimensional scaling.
4. Experimental results showed that with an increase in the n-well doping concentration, the ESD robustness of a p+/n-well diode improved. How did the latchup robustness improve with the ESD robustness?
5. With technology scaling, the depth of STI is reduced with the lateral dimensions to maintain a constant aspect ratio. How does the latchup robustness of a technology change with the decrease of the STI width (with technology scaling)?
6. Assume that the lateral parasitic bipolar base width is one half of the width of the STI isolation between the p+ and n+ diffusion. Assuming STI is scaled, show how the bipolar current gain should scale for 180, 130, 65, 40 and 22 nm. Assume the doping concentration is a constant.

## REFERENCES


1. D.B. Estreich. *The physics and modeling of latch-up and CMOS integrated circuits.* Technical Report No. G-201-9, Integrated Circuits Laboratory, Stanford Electronic Laboratories, Stanford University, Stanford, CA, November 1980.
2. R. Troutman. *Latchup in CMOS Technology: The Problem and the Cure.* Boston, MA: Kluwer Academic Publishers; 1986.
3. D.B. Estreich, A. Ochoa Jr. and R.W. Dutton. An analysis of latchup prevention in CMOS IC's using an epitaxial-buried layer process. *International Electron Device Meeting (IEDM) Technical Digest*, December 1978. p. 230–234.
4. R.R. Troutman. Epitaxial layer enhancement of n-well guard rings for CMOS circuits. *IEEE Electron Device Letters*, **4**, 1983, 438–440.

5. P. Cottrell, S. Warley, S. Voldman, W. Leipold and C. Long. N-well design for trench DRAM arrays. *International Electron Device Meeting (IEDM) Technical Digest*, December 1988. p. 584–587.
6. S. Voldman, M. Marceau, A. Baker, E. Adler, S. Geissler, J. Slinkman, J. Johnson and M. Paggi. Retrograde well and epitaxial thickness optimization for shallow- and deep-trench collar merged isolation and node trench (MINT) SPT cell and CMOS logic technology. *International Electron Device Meeting (IEDM) Technical Digest*, 1992. p. 811–815.
7. E. Adler, J.K. DeBrosse, S.F. Geissler, S.J. Holmes, M.D. Jaffe, J.B. Johnson, C.W. Koburger, J.B. Lasky, B. Lloyd, G.L. Miles, J.S. Nakos, W.P. Noble Jr., S. Voldman, M. Armacost and R. Ferguson. The evolution of IBM CMOS DRAM technology. *IBM Journal of Research and Development*, **39**(1/2), 1995, 167–188.
8. C.W. Koburger, W.F. Clark, J.W. Adkisson, E. Adler, P.E. Bakeman, A.S. Bergendahl, A.B. Botula, W. Chang, B. Davari, J.H. Givens, H.H. Hansen, S.J. Holmes, D.V. Horak, C.H. Lam, J.B. Lasky, S.E. Luce, R.W. Mann, G.L. Miles, J.S. Nakos, E.J. Nowak, G. Shahidi, Y. Taur, F.R. White and M.R. Wordeman. A half-micron CMOS logic generation. *IBM Journal of Research and Development*, **39**(1/2), 1995, 215–228.
9. Y. Taur, Y.J. Mii, D.J. Frank, H.S. Wong, D.A. Buchanan, S.J. Wind, S.A. Rishton, G.A. Sai-Halasz and E.J. Nowak. CMOS scaling into the 21st century: 0.1 μm and beyond. *IBM Journal of Research and Development*, **39**(1/2), 1995, 245–260.
10. S. Voldman, E. Gebreselasie, X.F. Liu, D. Coolbaugh and A. Joseph. The influence of high resistivity substrates on CMOS latchup robustness. *Proceedings of the Electrical Overstress/Electrostatic Discharge (EOS/ESD) Symposium*, 2005. p. 90–99.
11. G. Boselli, V. Reddy and C. Duvvury. Latch-up in 65 nm CMOS technology: a scaling perspective. *Proceeding of the International Reliability Physics Symposium (IRPS)*, 2004. p. 137–144.
12. S. Voldman. ESD and latchup in advanced technologies. Tutorial, *Tutorial Notes of the International Reliability Physics Symposium (IRPS)*, April 25, 2004.
13. S. Voldman. Latchup physics and design. Tutorial Notes, *ESD Tutorials of the Electrical Overstress/Electrostatic Discharge (EOS/ESD) Symposium*, September 20, 2004.
14. S. Voldman. CMOS latchup. *Tutorial Notes of the International Reliability Physics Symposium (IRPS)*, April 17, 2005.
15. S. Voldman. Latchup – it's back! *Electrostatic Discharge (ESD) Association Threshold Magazine*, **19**(5), 2003, 7–8.
16. S. Voldman. Latchup: Part II – Why is it still here? *Electrostatic Discharge (ESD) Association Threshold Magazine*, **19**(6), 2003, 9–10.
17. J. Quinke. Novel test structures for the investigation of the efficiency of guard rings used for I/O latchup prevention, *Proceedings on the International Conference on Microelectronic Test Structures (ICMTS)*, 1990. p. 35–40.
18. G.G.E. Low. Carrier concentration disturbances in semiconductors. *Proceedings of the Physical Society B*, **68**, 1955, 310–314.
19. J.B. Arthur, A.F. Gibson and J.B. Gunn. Current gain at L-H junctions in germanium. *Proceedings of the Physical Society B*, **69**, 1956, 705–711.
20. J.B. Gunn. On carrier accumulation, and the properties of certain semiconductor junctions, *Journal of Electronics and Control*, **4**, 1958, 17–50.
21. R.W. Lade and A.G. Jordan. On the static characteristics of high-low junction devices, *Journal of Electronics and Control*, **13**, 1962, 23–31.
22. D.L. Scharfetter. Minority carrier injection and charge storage in epitaxial Schottky barrier diodes. *Solid State Electronics*, **8**, 1965, 299–311.
23. R.W. Dutton and R.J. Whittier. Forward current-voltage and switching characteristics of $p^{+}$-n-$n^{+}$ (epitaxial) diodes. *IEEE Transactions on Electronic Devices*, **16**, 1969, 458–467.
24. S. Dabral, R. Aslett and T. Maloney. Designing on-chip power supply coupling diodes for ESD protection and noise immunity. *Proceedings of the Electrical Overstress/Electrostatic Discharge (EOS/ESD) Symposium*, 1993. p. 239–249.
25. S. Dabral and T. Maloney. *Basic ESD and I/O Design*. New York: John Wiley & Sons, Inc.; 1995.
26. S. Voldman and G. Gerosa. Mixed voltage interface ESD protection circuits for advanced microprocessors in shallow trench and LOCOS isolation CMOS technology. *International Electron Device Meeting (IEDM) Technical Digest*, December 1994. p. 811–815.

27. S. Voldman, G. Gerosa, V. Gross, N. Dickson, S. Furkay and J. Slinkman. Analysis of snubber clamped diode string mixed voltage interface ESD protection networks for advanced microprocessors. *Proceedings of the Electrical Overstress/Electrostatic Discharge (EOS/ESD) Symposium*, 1995, p. 43–61.
28. S. Voldman, G. Gerosa, V. Gross, N. Dickson, S. Furkay and J. Slinkman. Analysis of snubber clamped diode string mixed voltage interface ESD protection networks for advanced microprocessors. *Journal of Electrostatics*, 1996, **38**(1–2), 3–32.
29. S. Voldman, S. Furkay and J. Slinkman. Three dimensional transient electro-thermal simulation of electrostatic discharge protection networks. *Proceedings of the Electrical Overstress/Electrostatic Discharge (EOS/ESD)*, 1994. p. 246–257.
30. S. Voldman. *ESD: Physics and Devices*. Chichester, UK: John Wiley & Sons, Ltd; 2004.
31. S. Voldman. ESD: *Circuits and Devices*. Chichester, UK: John Wiley & Sons, Ltd; 2005.
32. D. Takacs, J. Harter, E.P. Jacobs, C. Werner and U. Schwabe. Comparison of latch-up in p− and n-well CMOS circuits. *International Electron Device Meeting (IEDM) Technical Digest*, December 1983. p. 159–163.
33. J.Y.-T. Chen. An n-well CMOS with self-aligned channel stops. *International Electron Device Meeting (IEDM) Technical Digest*, 1983. p. 526–529.
34. S. Voldman. Retrograde well implants and ESD. *SEMATECH Vertical Modulated Well PTAB*, November 1994.
35. S. Voldman. Optimization of MeV retrograde wells for advanced logic and microprocessor/PowerPC and electrostatic discharge (ESD). *Smart and Economic Device and Process Designs for ULSI Using MeV Implant Technology Seminar: SEMICON West, SEMICON West GENUS Seminar*, San Francisco, 1994.
36. S. Voldman. MeV implants boost device design. *IEEE Circuits and Devices*, **11**(6), 1995; 8–17.
37. S. Voldman, M. Marceau, A. Baker, E. Adler, S. Geissler, J. Slinkman, J. Johnson and M. Paggi, Shallow trench isolation (STI) double-diode electrostatic discharge (ESD) circuit and interaction with DRAM circuitry. *Proceedings of the Electrical Overstress/Electrostatic Discharge (EOS/ESD) Symposium*, 1992. p. 277–288.
38. S. Voldman and V. Gross. Scaling, optimization, and design considerations of electrostatic discharge protection circuits in CMOS technology. *Proceedings of the Electrical Overstress/Electrostatic Discharge (EOS/ESD) Symposium*, 1993. p. 251–260.
39. S. Voldman and V. Gross. Scaling, optimization, and design considerations of electrostatic discharge protection circuits in CMOS technology. *Journal of Electrostatics*, **33**(3), 1994, 327–357.
40. S. Voldman. Electrostatic discharge protection, scaling, and ion implantation in advanced semiconductor technologies. Process Integration Issues/Technical Trends Session, *Proceedings of the Ion Implantation Conference ($I^2CON$)*, Napa, CA, 1999.
41. S. Voldman. The impact of MOSFET technology evolution and scaling on electrostatic discharge protection. *Microelectronics Reliability*, **38**, 1998, 1649–1668.
42. M. Hargrove, S. Voldman, J. Brown, K. Duncan and W. Craig. Latchup in CMOS. *Proceedings of the International Reliability Physics Symposium (IRPS)*, 1998. p. 269–278.
43. W. Morris. CMOS latchup. *Proceedings of the International Reliability Physics Symposium (IRPS)*, 2003. p. 76–84.
44. L. Parillo, R.S. Payne, R.E. Davis, G.W. Reulinger and R.L. Field. Twin-tub CMOS – a technology for VLSI circuits. *International Electron Device Meeting (IEDM) Technical Digest*, 1980. p. 752–755.
45. G. Goto, H. Takahashi and T. Nakamura. Latchup immunity against noise pulses in a CMOS double well structure. *International Electron Device Meeting (IEDM) Technical Digest*, 1983. p. 168–171.
46. G. Coquin, W.T. Lynch and L.C. Parillo. Methods for fabricating latchup-preventing CMOS devices. US Patent No. 4,766,090 (August 23, 1988).
47. J. Brown and S. Voldman. Semiconductor device fabrication method and apparatus using connecting implants. US Patent No. 6,097,068 (August 1, 2000).
48. S. Murarka. *Silicides for VLSI Applications*. Orlando, FL: Academic Press; 1983.
49. H.P. Zappe and C. Hu. Device characteristics of MOSFETs in MeV implanted substrates. *Nuclear Instrumentation Methods Physical Review B*, **21**, 1987, 163–167.
50. H.Y. Lin and C.H. Ting. Improvements of CMOS latchup using a high energy buried layer. *Nuclear Instrumentation Methods Physical Review B*, **38–39**, 1989, 960–964.
51. T. Kuroi, S. Komori, H. Miyatake and K. Tsukamoto. Self-gettering and proximity gettering for buried layer formation by MeV ion implantation. *International Electron Device Meeting (IEDM) Technical Digest*, 1990. p. 261–264.

52. D.C. Jacobson, A. Kamgar, D.J. Eaglesham, E.J. Lloyd, S.J. Hillenius and J.M. Poate. High energy implantation for profiled tub formation and impurity gettering in deep submicron CMOS technology. *Nuclear Instrumentation Methods Physical Review B*, **96**, 1995, 416–419.
53. J.O. Borland. Method of constructing CMOS vertically modulated wells (VMW) by clustered MeV BILLI (buried implanted layer for lateral isolation) implantation. US Patent No. 5,501,993 (March 26, 1996).
54. J.O. Borland. Method for CMOS latch-up improvement by MeV BILLI (buried implanted layer for lateral isolation) plus buried layer implantation. US Patent No. 5,821,589 (October 13, 1998).
55. K.C. Leong, P.C. Liu, W. Morris and L. Rubin. Superior latchup resistance of high dose energy implanted p+ buried layers. *Proceedings of the XII International Conference on Ion Implantation Technology*, Kyoto, Japan, 1998. p. 99–103.
56. K.K. Bourdelle, D.J. Eaglesham, D.C. Jacobson and J. Poate. The effect of As-implanted damage on the microstructure of threading dislocations in MeV implanted silicon. *Journal of Applied Physics*, **86**, 1999, 1221–1225.
57. K.K. Bourdelle, Y. Chen, R. Ashton, L. Rubin, A. Agarwal and W. Morris. Epi-replacement in CMOS technology by high dose, high energy boron implantation into Cz substrates. *International Electron Device Meeting (IEDM) Technical Digest*, 2000. p. 312–316.
58. K.K. Bourdelle, Y. Chen, R.A. Ashton, L.M. Rubin, A. Agarwal and W. Morris. Evaluation of high dose, high energy boron implantation into Cz substrates for epi-replacement in CMOS technology. *IEEE Transactions on Electron Devices*, **48**, 2001, 2043–2049.
59. S. Voldman, L. Lanzerotti, W. Morris and L. Rubin. The influence of heavily doped buried layer implants on electrostatic discharge (ESD), latchup, and a silicon germanium heterojunction bipolar transistor in a BiCMOS SiGe technology. *Proceeding of the International Reliability Physics Symposium (IRPS)*, 2004. p. 143–151.
60. M. Bohr. Isolation structure formation for semiconductor circuit fabrication. US Patent No. 5,536,675 (July 16, 1996).
61. J. Brown and S. Voldman. Structure and method for improved latch-up using dual depth STI with impurity implant. US Patent No. 6,144,086 (November 7, 2000).
62. F.D. Baker, J.S. Brown, S.J. Holmes, R.K. Leidy, E.J. Nowak and S.H. Voldman. Method and structure to reduce latch-up using edge implants. US Patent No. 6,232,639 (May 15, 2001).
63. R.D. Rung, H. Momose and Y. Nagabuko. Deep trench isolated CMOS devices. *International Electron Device Meeting (IEDM) Technical Digest*, 1982. p. 237–240.
64. T. Yamaguchi. S. Morimoto, G.H. Kawamoto, H.K. Park and G.C. Eiden. High-speed latchup-free 0.5-μm channel CMOS using self-aligned $TiSi_2$ and deep trench isolation technologies. *International Electron Device Meeting (IEDM) Technical Digest*, 1983. p. 522–525.
65. S. Voldman. A review of CMOS latchup and electrostatic discharge (ESD) in bipolar complimentary MOSFET (BiCMOS) silicon germanium technologies: Part II–Latchup. *Microelectronics and Reliability*, **45**, 2005, 437–455.

# 6 CMOS Latchup Process Features and Solutions – Bipolar and BiCMOS Technology

## 6.1 CMOS LATCHUP IN BIPOLAR AND RF BiCMOS TECHNOLOGY

In mainstream CMOS technology, the technology goals are focused on circuit density, chip performance and cost. Derivative technologies, such as RF CMOS, bipolar, RF BiCMOS and power technologies, have a significant larger set of semiconductor devices that contain many additional masks, implants and structures. Many of these features can be used to improve the latchup robustness of the CMOS circuitry within the derivative technology, or can be added to a base CMOS technology. In this chapter, the focus will be on utilization of these semiconductor features and structures for latchup prevention. Today, many of these features are not present in mainstream CMOS or RF CMOS; but, these processes and structures can be added in future sub-0.1 μm and sub-65 nm CMOS technologies.

## 6.2 SUBSTRATES – HIGH-RESISTANCE SUBSTRATES

Today, high-resistance substrates are needed to avoid noise coupling between circuits. In a mixed signal chip, a concern exists between the digital CMOS circuits, analog and radio frequency (RF) circuitry. High-resistance substrates will be needed in mixed signal CMOS (MS-CMOS), radio frequency CMOS (RF CMOS), BiCMOS silicon germanium (SiGe) and gallium arsenide (GaAs) technologies [1–5]. As technologies are scaled and as the application frequency increases, noise coupling, crosstalk and noise rejection methods will be needed. Hence, a key issue in the CMOS latchup robustness is the role of the substrate resistance. Figure 6.1 shows a plot of the trend of substrate resistance in the future.

*Latchup* Steven H. Voldman
© 2007 John Wiley & Sons, Ltd

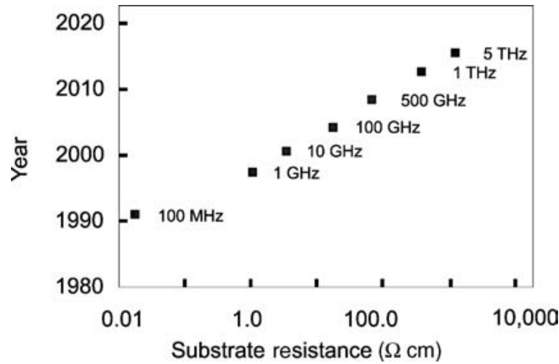



**Figure 6.1** Substrate resistance projection as a function of year and application frequency.

The trend in the industry is the migration from 1 to 50 Ω cm [4,5]. RF BiCMOS typically migrates to higher resistance substrates prior to advanced digital CMOS or RF CMOS due to the needs of the bipolar transistor in the MS applications. Additionally, BiCMOS technologies provide bipolar transistors that are typically one to three generations ahead of the RF performance of a RF CMOS technology. For example, BiCMOS migration to 50 Ω cm substrates will occur at 0.13-μm technology generation, whereas Boselli *et al.* showed that 50 Ω cm substrates occur later at a 65-nm CMOS technology node [4]. The starting wafer substrate resistance has an affect on the substrate spreading resistance term. As a result, this influences both ESD and latchup [4–7]. For latchup, the spreading resistance has a role in the cathode-to-substrate contact spacing. As the distance increases, the influence is more significant. Additionally, the spreading resistance does not saturate but continues to increase in latchup analysis [4]. In order to initiate latchup, the product of the npn and the pnp bipolar current gain needs to satisfy the following inequality,

$$\beta_{\mathrm{n}}\beta_{\mathrm{p}} \geq \frac{1+\left(\dfrac{I_{\mathrm{s}x}}{I}\right)\beta_{\mathrm{n}}}{\left[1-\left(\dfrac{I_{\mathrm{w}}}{I}\right)\left(\dfrac{\beta_{\mathrm{n}}+1}{\beta_{\mathrm{n}}}\right)-\left(\dfrac{I_{\mathrm{s}x}}{I}\right)\right]},$$

where

$$I_{\mathrm{w}} = \frac{(V_{\mathrm{be}})_{\mathrm{pnp}}}{R_{\mathrm{nw}}} = \frac{V_0}{R_{\mathrm{nw}}}\ln\left[\frac{I-I_{\mathrm{w}}}{(I_0)_{\mathrm{p}}}\right],$$

and where the substrate current expression can be modified as follows [2]

$$I_{\mathrm{s}x} = \frac{(V_{\mathrm{be}})_{\mathrm{npn}}}{R_{\mathrm{pw}} \parallel R_{\mathrm{s}x}} = \frac{V_0}{R_{\mathrm{pw}} \parallel R_{\mathrm{s}x}}\ln\left[\frac{I-I_{\mathrm{s}x}}{(I_0)_{\mathrm{n}}}\right],$$

where the effective substrate resistance comprises both the p-well resistance, $R_{\mathrm{pw}}$, and the base wafer resistance, $R_{\mathrm{s}x}$, spreading resistance terms,

$$\frac{1}{R_{\mathrm{pw}} \parallel R_{\mathrm{s}x}} = \frac{1}{R_{\mathrm{pw}}} + \frac{1}{R_{\mathrm{s}x}}.$$

In the generalized tetrode formulation, the latchup condition can be expressed as

$$\alpha^*_{\mathrm{fns}} + \alpha^*_{\mathrm{fps}} \geq 1,$$

where

$$\alpha^*_{\mathrm{fns}} = \frac{\alpha_{\mathrm{fns}}}{1 + \dfrac{r_{\mathrm{en}}}{R_{\mathrm{s}x} \parallel R_{\mathrm{pw}}}}$$

and

$$\alpha^*_{\mathrm{fps}} = \frac{\alpha_{\mathrm{fps}}}{1 + \dfrac{r_{\mathrm{ep}}}{R_{\mathrm{nw}}}},$$

where the substrate resistance is explicitly shown as the parallel resistance of the p-well resistance and the substrate base wafer resistance. As the substrate wafer resistance increases, it influences the generalized npn transport factor. As the substrate resistance becomes large compared to the p-well resistance, the affect on the differential generalized stability criteria is minimal. Note that the parallel resistance of the substrate and the well is approximately the p-well resistance,

$$\frac{1}{R_{\mathrm{pw}} \parallel R_{\mathrm{s}x}} \approx \frac{1}{R_{\mathrm{pw}}} \quad R_{\mathrm{s}x} \gg R_{\mathrm{pw}}.$$

From the $\beta_{\mathrm{pnp}}\beta_{\mathrm{npn}}$ criteria, the solution of what is the substrate resistance value that initiates latchup can be derived. Given,

$$\beta_{\mathrm{n}}\beta_{\mathrm{p}} \geq \frac{1 + \left(\dfrac{I_{\mathrm{s}x}}{I}\right)\beta_{\mathrm{n}}}{\left[1 - \left(\dfrac{I_{\mathrm{w}}}{I}\right)\left(\dfrac{\beta_{\mathrm{n}}+1}{\beta_{\mathrm{n}}}\right) - \left(\dfrac{I_{\mathrm{s}x}}{I}\right)\right]},$$

the expression for the substrate current can be factored and expressed as follows,

$$\frac{I_{\mathrm{s}x}}{I} \leq \frac{\beta_{\mathrm{n}}\beta_{\mathrm{p}}\left[1 - \left(\dfrac{I_{\mathrm{w}}}{I}\right)\left(\dfrac{\beta_{\mathrm{n}}+1}{\beta_{\mathrm{n}}}\right)\right]}{\beta_{\mathrm{n}}(1+\beta_{\mathrm{p}})},$$

and

$$I_{\mathrm{s}x} \leq I\left\{\frac{\beta_{\mathrm{n}}\beta_{\mathrm{p}}\left[1 - \left(\dfrac{I_{\mathrm{w}}}{I}\right)\left(\dfrac{\beta_{\mathrm{n}}+1}{\beta_{\mathrm{n}}}\right)\right]}{\beta_{\mathrm{n}}(1+\beta_{\mathrm{p}})}\right\}.$$

From this expression, the relationship of the forward bias of the npn and the resistance can be substituted into the equation. The effective substrate resistance can be expressed as the parallel product of the p-well and the substrate resistances. The effective substrate resistance needed in order to initiate latchup can be expressed as

$$R_{\mathrm{s}x} \parallel R_{\mathrm{pw}} \geq \frac{(V_{\mathrm{be}})_{\mathrm{npn}}}{I}\left\{\frac{\beta_{\mathrm{n}}(1+\beta_{\mathrm{p}})}{\beta_{\mathrm{n}}\beta_{\mathrm{p}}\left[1 - \dfrac{I_{\mathrm{w}}}{I}\left(\dfrac{\beta_{\mathrm{n}}+1}{\beta_{\mathrm{n}}}\right)\right]}\right\}.$$

Hence, the 'effective substrate resistance condition' to initiate latchup can be expressed as

$$(R_{sx})_{\text{eff}} \geq \frac{V_0 \ln\left(\frac{I - I_{sx}}{I_0}\right)}{I} \left\{ \frac{\beta_n(1+\beta_p)}{\beta_n\beta_p\left[1 - \frac{I_w}{I}\left(\frac{\beta_n + 1}{\beta_n}\right)\right]} \right\}.$$

In future applications, high-resistance substrates will be sensitive to cable discharge events (CDE) and injection sources. Hence, the condition for initiation of latchup, in the 'alpha formulation' is when [5,8,9]

$$(\alpha_p + \alpha_n) = 1 + \alpha_p \frac{(I_w)}{I} + \alpha_n \frac{(I_{sx})}{I} - \frac{I^*}{I},$$

or the current for latchup initiation for an external current source occurs when

$$I^* = I\left\{1 + \alpha_p\left[\frac{(I_w)}{I} - 1\right] + \alpha_n\left[\frac{(I_{sx})}{I} - 1\right]\right\},$$

where the well and substrate current expressions can be shown to be (respectively),

$$I_w = \frac{(V_{be})_{pnp}}{R_w} = \frac{V_0}{R_w}\ln\left[\frac{I - I_w}{(I_0)_p}\right]$$

and

$$I_{sx} = \frac{(V_{be})_{npn}}{R_{sx}} = \frac{V_0}{R_{sx}}\ln\left[\frac{I - I_{sx}}{(I_0)_n}\right].$$

Hence, the condition for latchup from an external source can also be expressed as

$$I^* = I\left\{1 + \frac{\beta_p}{\beta_p + 1}\left[\frac{I_w}{I} - 1\right] + \frac{\beta_n}{\beta_n + 1}\left[\frac{I_{sx}}{I} - 1\right]\right\},$$

or

$$I_{inj} = I\left\{1 + \frac{\beta_p}{\beta_p + 1}\left[\frac{I_w}{I} - 1\right] + \frac{\beta_n}{\beta_n + 1}\left[\frac{I_{sx}}{I} - 1\right]\right\} - I_{cn0} - I_{cp0}.$$

For the same circuit, this can be expressed as a function of the bipolar current gains. The '$\beta$ product' relation for undergoing latchup in the environment of an external injection can be written in the form,

$$\beta_n\beta_p = \frac{1 + \beta_p\left[\frac{I_w - I^*}{I}\right] + \beta_n\left[\frac{I_{sx} - I^*}{I}\right] - \left[\frac{I^*}{I}\right]}{\left[1 - \frac{I_w + I_{sx} - I^*}{I}\right]},$$

where

$$\left.\begin{aligned} I^* &= I_{\text{cp0}} + I_{\text{cn0}} + I_{\text{inj}}(x,t), \\ I_{\text{w}} &= \frac{(V_{\text{be}})_{\text{pnp}}}{R_{\text{w}}} = \frac{V_0}{R_{\text{w}}} \ln\left[\frac{I - I_{\text{w}}}{(I_0)_{\text{p}}}\right], \\ I_{\text{s}x} &= \frac{(V_{\text{be}})_{\text{npn}}}{R_{\text{s}x}} = \frac{V_0}{R_{\text{s}x}} \ln\left[\frac{I - I_{\text{s}x}}{(I_0)_{\text{n}}}\right]. \end{aligned}\right\}$$

In an alpha representation,

$$(\alpha_{\text{p}} + \alpha_{\text{n}}) = 1 + \alpha_{\text{p}} \frac{(I_{\text{w}})}{I} + \alpha_{\text{n}} \frac{(I_{\text{s}x})}{I} - \frac{I^*}{I}.$$

Solving for the substrate current needed to initiate latchup,

$$I_{\text{s}x} = I\left[1 + \left(\frac{\alpha_{\text{p}}}{\alpha_{\text{n}}}\right)\left[1 - \frac{I_{\text{w}}}{I}\right] + \left(\frac{1}{\alpha_{\text{n}}}\right)\left[\frac{I^*}{I} - 1\right]\right]$$

and

$$I_{\text{s}x} = \frac{(V_{\text{be}})_{\text{npn}}}{(R_{\text{s}x})_{\text{eff}}} = \frac{V_0}{(R_{\text{s}x})_{\text{eff}}} \ln\left[\frac{I - I_{\text{s}x}}{(I_0)_{\text{n}}}\right].$$

Hence, the substrate resistance condition needed to initiate external latchup under a condition of an external injection source can be quantified,

$$(R_{\text{s}x})_{\text{eff}} \geq \frac{(V_{\text{be}})_{\text{npn}}}{I\left[1 + \left(\frac{\alpha_{\text{p}}}{\alpha_{\text{n}}}\right)\left[1 - \frac{I_{\text{w}}}{I}\right] + \left(\frac{1}{\alpha_{\text{n}}}\right)\left[\frac{I^*}{I} - 1\right]\right]}$$

or expressed as follows,

$$(R_{\text{s}x})_{\text{eff}} \geq \frac{V_0 \ln\left[\frac{I - I_{\text{s}x}}{(I_0)_{\text{n}}}\right]}{I\left[1 + \left(\frac{\alpha_{\text{p}}}{\alpha_{\text{n}}}\right)\left[1 - \frac{I_{\text{w}}}{I}\right] + \left(\frac{1}{\alpha_{\text{n}}}\right)\left[\frac{I^*}{I} - 1\right]\right]}.$$

From these formulations, we get the substrate resistance condition for the amount of resistance in the substrate in order to initiate latchup for the case of external injection. As the external injection increases, the denominator of the expression increases in magnitude, where the effective resistance needed to undergo latchup is lower. Hence, as the external source increases, the substrate resistance requirement to undergo latchup decreases. In the case that the substrate resistance continues to increase with scaling, circuits will become more sensitive to the external sources. High-resistance substrates will influence the latchup design practice as follows:

- The impact of substrate resistance on the local pnpn latchup substrate shunt resistance will lessen as the substrate resistance is significantly less than the p-well resistance.
- A high–low p+/p− step junction transition will occur as the p− substrate doping concentration is lower than the p-well doping concentration forming a n++/p+/p− diode structure; this will increase the vertical diode injection to the substrate.

- The substrate thermal impedance will increase, leading to a higher self-heating within the pnpn structure.
- Minority carrier recombination time will increase, leading to larger diffusion lengths and longer propagation distances of minority carriers.

First, since the shunt resistance is the parallel resistance of the p-well and the p− substrate, the p-well will play a greater role as the p− substrate resistance is significantly increased. Second, with the transition from a first substrate resistance to a higher second substrate resistance, the change will be less than anticipated. For the second item, as the p− substrate decreases below the p-well, a p+/p− high–low junction is formed; this increases the vertical injection of carriers. As a result, the lateral npn current gain should be reduced based on the early discussion in Chapter 2. Third, as the conductivity of the substrate decreases, the thermal conductivity also decreases; this leads to a higher thermal impedance. Hence, latchup events that introduce self-heating will have a higher junction temperature for the same amount of power dissipated. As a result, the rise in temperature and the change in parameters with the temperature will be more significant for high-resistance substrate wafers. Fourth it is anticipated that transient and external latchup considerations will be worse. Minority carrier recombination time will increase, leading to longer diffusion lengths and longer distances that carriers traverse the substrate. Additionally, the ability to shunt the excess carriers is increased, making products more vulnerable to single-event latchup to transient phenomenon.

The majority of reported latchup measurements were taken on CMOS wafers with substrate resistivity below 1 Ω cm wafer. The first measurements on 50 Ω cm were taken by Boselli *et al.* [4] and Voldman *et al.* [5]. As an example of the latchup characterization and the influence of higher substrate resistance, latchup measurements were evaluated for 10 and 50 Ω cm substrate wafer. Experimental pnpn latchup test structures were constructed in a 0.13-μm CMOS within a BiCMOS technology. The latchup experimental work was completed using a low resistivity 10 Ω cm and a high resistivity 50 Ω cm wafer.

### 6.2.1 Fifty Ohm Centimeter Substrate Resistance

Figure 6.2 is a plot of $\beta_{npn}$ versus the p+/n+ space as a function of the 10 and 50 Ω cm substrate wafers. Measurements of $\beta_{npn}$ show that as the p+/n+ space increases, $\beta_{npn}$ decreases. At a small

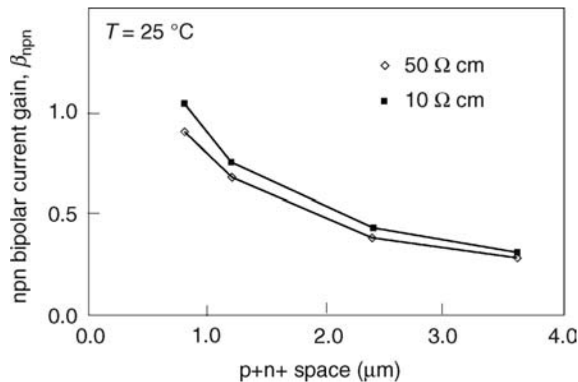



**Figure 6.2** Parasitic npn bipolar current gain, $\beta_{npn}$, versus p+/n+ spacing as a function of substrate resistivity.

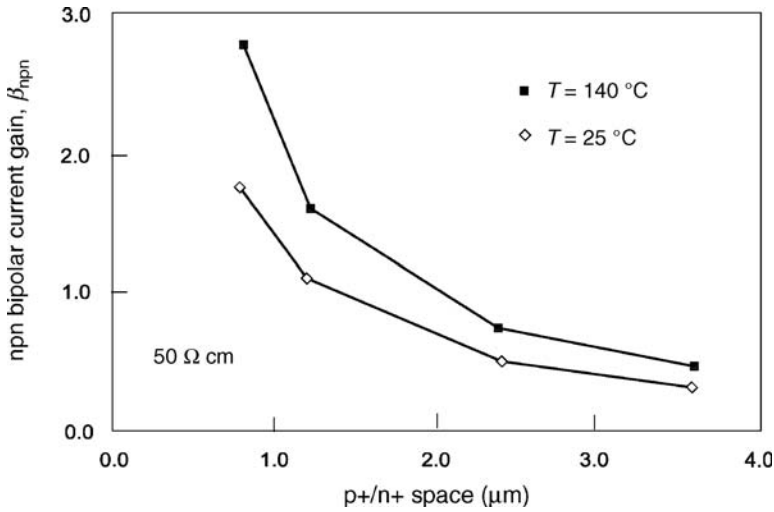



**Figure 6.3** npn bipolar current gain as a function of p+/n+ spacing in a 50 Ω cm substrate wafer (at $T = 25$ °C and $T = 140$ °C).

n+ diffusion to p− substrate contact space and the presence of p-well implants, only a small change in the extracted npn $\beta_{npn}$ is evident (e.g. $\beta_{npn}$ extraction is at a 1 mA current level) [5].

Figure 6.3 shows the npn transistor bipolar current gain, $\beta_{npn}$, versus the p+/n+ space for the 50 Ω cm substrate wafer for both ambient and elevated temperatures [5].

Figure 6.4 is a plot of the pnp bipolar current gain, $\beta_{pnp}$, versus the p+/n+ space as a function of the 10 and 50 Ω cm substrate wafers. Experimental measurements of $\beta_{pnp}$ show a weak sensitivity to the p+/n+ space; for both substrate wafers, $\beta_{pnp}$ exceeds unity.

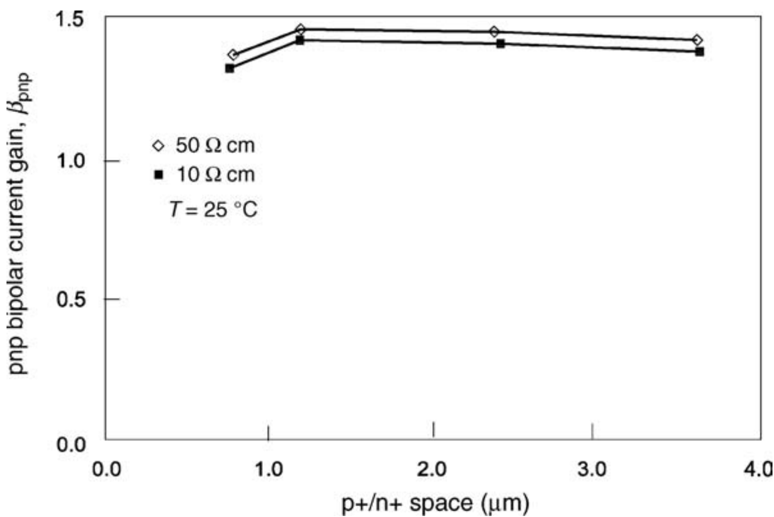



**Figure 6.4** Parasitic pnp bipolar current gain, $\beta_{pnp}$, versus p+/n+ spacing for 10 and 50 Ω cm substrate wafers.

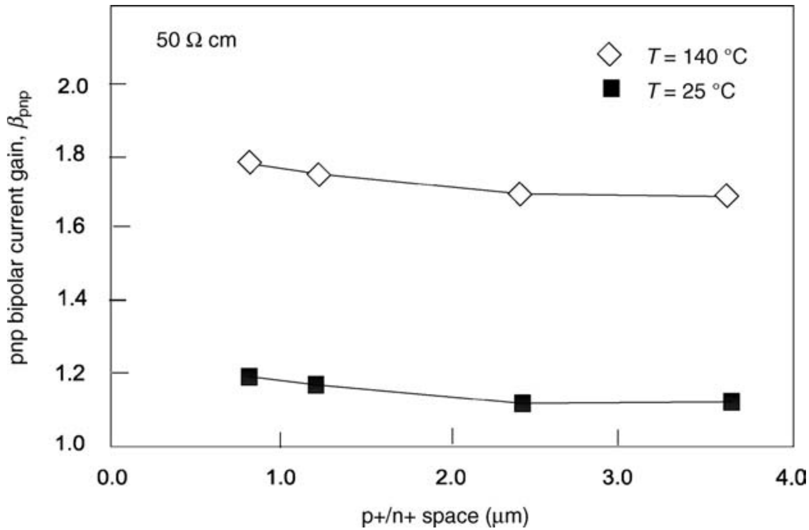



**Figure 6.5** pnp bipolar current gain, $\beta_{pnp}$, versus p+/n+ spacing for a 50 Ω cm substrate (at $T = 25$ °C and $T = 140$ °C).

Figure 6.5 shows the parasitic $\beta_{pnp}$ as a function of p+/n+ space for ambient and elevated temperatures in the 50 Ω cm substrate wafer; $\beta_{pnp}$ is near unity at room temperature but increases with both decreased p+/n+ space and increased temperature.

Figure 6.6 is the evaluation of $\beta_{pnp}\beta_{npn}$ as a function of large substrate spacing for 10 and 50 Ω cm substrate wafers. A CMOS design practice for the evaluation of large-scale space effects is that in a test site structure, a p+ substrate contact from the next pnpn test structure can be utilized; in that fashion, a p+ substrate contact effect can be evaluated over thousands of microns within a family of test structures (e.g. a 1 × 25 macro). As the space increases, $\beta_{pnp}\beta_{npn}$ approaches unity for both wafers [5].

The latchup overshoot results are shown in Figure 6.7. In the case of the 50 Ω cm wafer, the overshoot voltage decreased, as expected due to the higher substrate shunt resistance. At small substrate spacings, the overshoot voltage is comparable; but as the spacing increases, the overshoot voltage is reduced by approximately 0.2 V [5].

Undershoot current versus p+/n+ spacing for the 50 Ω cm wafer is shown in Figure 6.8. As the p+/n+ space is decreased, the current required to initiate latchup decreases. As the bipolar current gain increases, the current needed to latchup at a minimum structure spacing decreases from 25 to 10 mA (for a 20-μm wide latchup test structure).

The undershoot voltage for the 50 Ω cm wafer as a function of p+/n+ space is shown in Figure 6.9. With the elevated temperature and smaller p+/n+ spacing, the voltage required to initiate latchup decreases. For increased temperature and small p+/n+ space, the undershoot voltage needed to initiate latchup decreases toward −1 V. Hence, latchup is initiated in these structures for voltage levels 0.3 V beyond the diode turn-on voltage. For −1 V undershoots, the 0.13-μm technology is vulnerable to potential latchup conditions.

Figure 6.10 is a plot of the current required to initiate latchup due to an overshoot phenomenon. The overshoot current is shown as a function of the p+/n+ space and temperature (e.g. $T = 25$ and 140 °C); results show that the temperature dependence leads to approximately 2× reduction in the overshoot current.

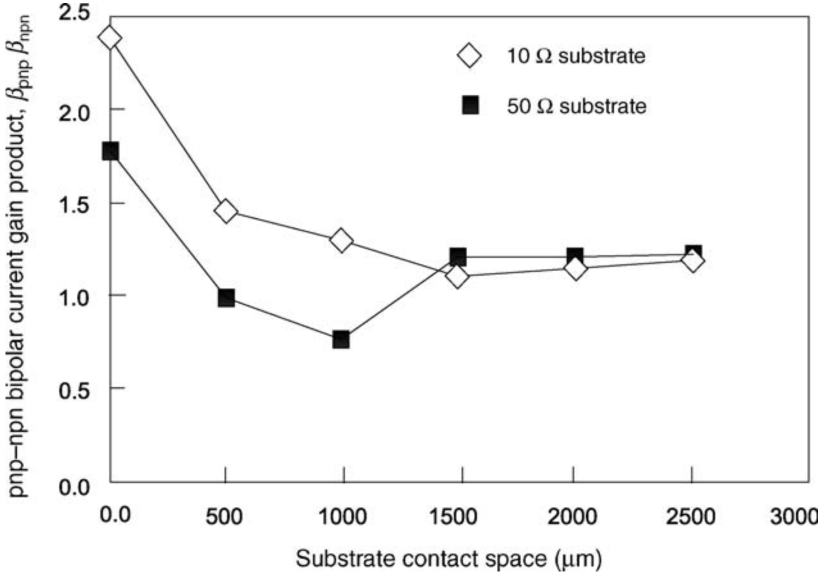



**Figure 6.6** $\beta$ Product, $\beta_{pnp}\beta_{npn}$, as a function of large substrate contact spacing for 10 and 50 Ω cm substrate wafers.

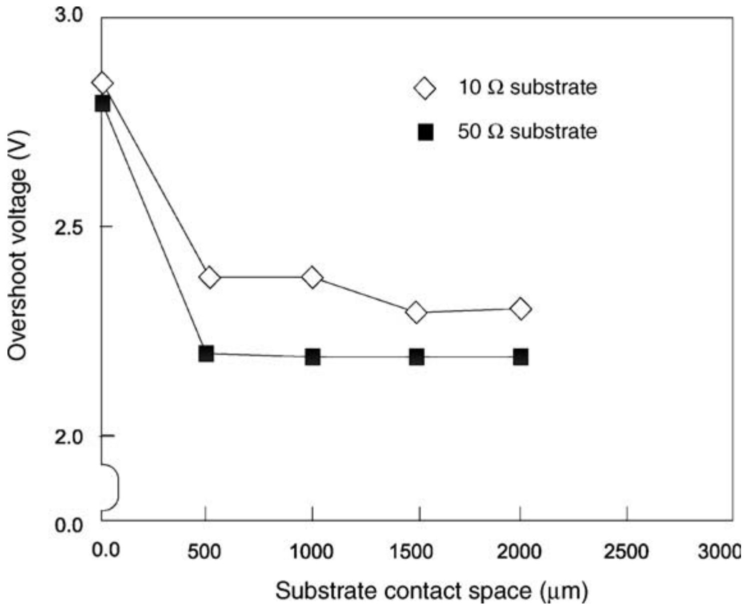



**Figure 6.7** Overshoot voltage as a function of substrate spacing for 10 and 50 Ω cm wafers (for the case of p+/n+ of 1.2 μm).

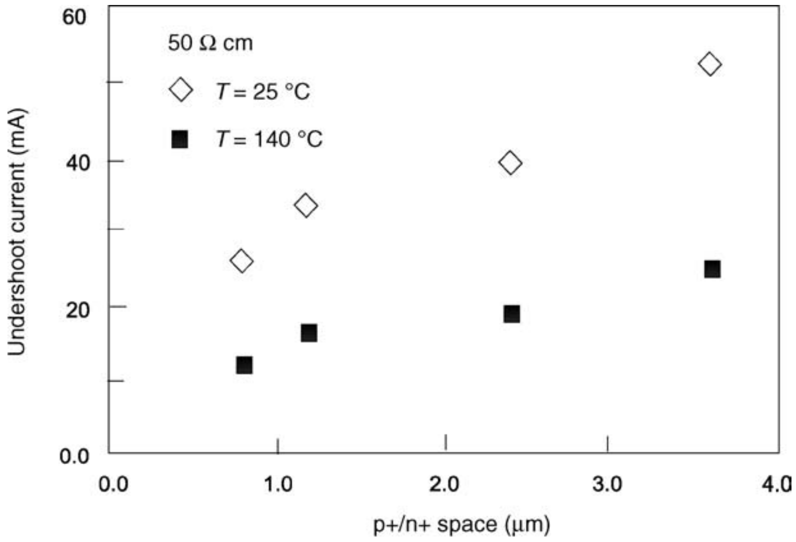



**Figure 6.8** Undershoot current as a function of p+/n+ spacing for 50 Ω cm substrate for ambient and elevated temperatures.

## 6.2.2 Ultrahigh Substrate Resistance

Presently, and in the future, there is, and will be, a high level of focus on integration for analog and mixed signal, as well as system-on-chip (SOC) and network-on-chip (NOC) semiconductor chips [1,2,5]. The integration of different circuit functions, performance objectives and noise isolation will require even lower substrate doping concentrations. Hence, the question is what will happen when the

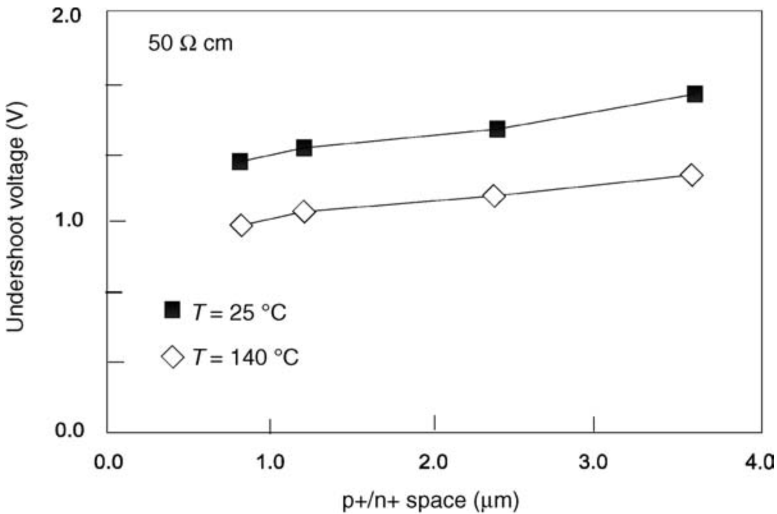



**Figure 6.9** Undershoot voltage as a function of p+/n+ spacing for 50 Ω cm substrate (for ambient and elevated temperatures).

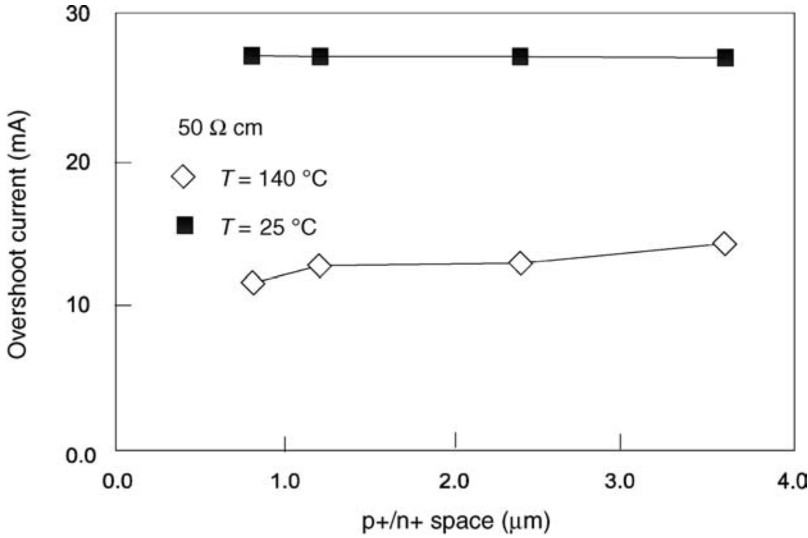



**Figure 6.10** Overshoot current as a function of p+/n+ spacing for 50 Ω cm substrate (for ambient and elevated temperatures).

substrate wafer concentrations are extended logarithmically from 50 to 100, 200 and 500 Ω cm substrate wafers? What will results be with 1 and 5 kΩ cm substrates? What will happen to CMOS latchup as technologies progress to ultrahigh resistance?

The first measurements of CMOS latchup structures in this ultrahigh resistance regime have been measured recently [5]. At first glance, it would seem that as substrate wafers approach 1, 2 and 5 kΩ cm resistance values, it would lead to forward-bias voltage of the diffusions at milliampere (mA) substrate current levels (e.g. simplistically, $V_{be} = I_{sx}R_{sx}$). But, as the substrate resistance becomes significantly larger than the p-well resistance, the influence of low-doped substrate wafers will decrease. The local nature of the spreading resistance, as well as the p-well resistance itself will dominate the parallel resistance condition. Yet, at the same time, as the substrate resistance increases, injected current will lead to less minority carrier recombination and longer diffusion lengths. But, as the substrate resistance increases to very low doping concentrations, the differences also will become smaller; as the substrate doping concentration decreases significantly, the influence on the diffusion coefficient will be weaker, and the recombination will depend only on Shockley–Read–Hall (SRH) and be inversely proportional to the doping of the substrate. With the relative doping differential between the p− substrate and the p-well, the vertical built-in field will also serve to reduce the net collection near the surface.

Table 6.1 contains latchup results of a dual-well CMOS process in standard and ultrahigh substrate resistivity wafers. The parameters were measured on a curve tracer to evaluate the latchup 'turn-on'

**Table 6.1** Experimental results of latchup parametrics for ultrahigh resistance substrates for dual-well CMOS structures.

| Substrate resistivity | $V_{ON}$ (V) | $V_{TR}$ (V) | $I_{TR}$ (mA) | $V_H$ (V) | $I_H$ (mA) |
|---|---|---|---|---|---|
| 10 Ω cm | 11 | 12.1 | 26 | 1.4 | 12 |
| 1 kΩ cm | 11 | 12.2 | 25 | 1.8 | 15 |
| 5 kΩ cm | 11 | 12.2 | 24 | 1.8 | 15 |

**Table 6.2** Experimental results of latchup parametrics for high-resistance substrates for triple-well latchup structures.

| Substrate resistivity | $V_{ON}$ (V) | $V_{TR}$ (V) | $I_{TR}$ (mA) | $V_H$ (V) | $I_H$ (mA) |
|---|---|---|---|---|---|
| 10 Ω cm | 10 | 11 | 12.5 | 1.5 | 3 |
| 1 kΩ cm | 10 | 11 | 13 | 1.2 | 3 |
| 5 kΩ cm | 10 | 11 | 12 | 1.2 | 4 |

voltage as well as the trigger and holding states. The pnpn structure is a 0.6-μm p+/n+ space, with a 50-μm p+ to n-well contact space and a 50-μm n+ to p+ substrate contact space [5].

Latchup measurements clearly show that the latchup parameters are weakly influenced by the high substrate resistance wafers. Note that the 'turn-on' voltage is not influenced by the substrate resistance. Secondly, there is a slight increase in $V_{TR}$ and a slight decrease in $I_{TR}$. The holding points, $V_H$ and $I_H$, show a slight increase with the ultrahigh resistivity wafers.

Table 6.2 contains latchup results of a triple-well process in standard and ultrahigh substrate resistance wafers. The test structure is a 0.6-μm p+/n+ space, a 50-μm p+ to n-well contact space and a 50-μm n+ to p+ substrate contact space. It is clear that the substrate resistance has a small role in the pnpn latchup parameters; comparing the results of Table 6.1 with table of Table 6.2, the difference between the dual-well and a merged triple-well structure has more influence than the presence of high-resistance substrates [5].

The considerations for a latchup design practice are as follows:

- Latchup sensitivity to high-resistance wafers is lowered by the presence of low-resistance p-well region.
- Latchup shunt resistance is a function of the parallel resistance of the p-well and the p− substrate region.
- Compensate the p− substrate latchup scaling impact by reducing the p-well resistance.

## 6.3 SUBCOLLECTORS

In CMOS processes, retrograde n-well design was shown to have a significant impact on latchup [10–16]. Bipolar and BiCMOS technologies have used subcollectors in both homojunction and heterojunction (HBT) bipolar transistors for transistor performance. Subcollector has a number of advantages for bipolar transistors which can be utilized for latchup [17]. First, they provide a low-resistance collector. Second, they minimize the Kirk effect at high currents. Third, they establish a low subcollector/substrate junction capacitance. Fourth, they minimize minority carrier injection into the substrate. These fundamental characteristics of a subcollector have natural advantages for latchup prevention.

There are two methods of forming a subcollector region – epitaxial and nonepitaxial. In the first process, subcollectors are formed prior to an epitaxial growth step. A substrate wafer is implanted through a masked area, followed by an epitaxial growth process; in this process, the region can be placed deep into the substrate wafer with a very high doping concentration. The dose level of these subcollectors is typically $10^{16}\,cm^{-2}$. This doping concentration can be at a silicon saturation level (e.g. $10^{19}$ to $10^{21}\,cm^{-2}$) providing very low sheet resistance (e.g. 1–10 Ω/square sheet resistance). These subcollectors provide a deep subcollector to p− substrate metallurgical junction on the order of 3–5 μm below the wafer surface. Additionally, the doping concentration is so significant that the

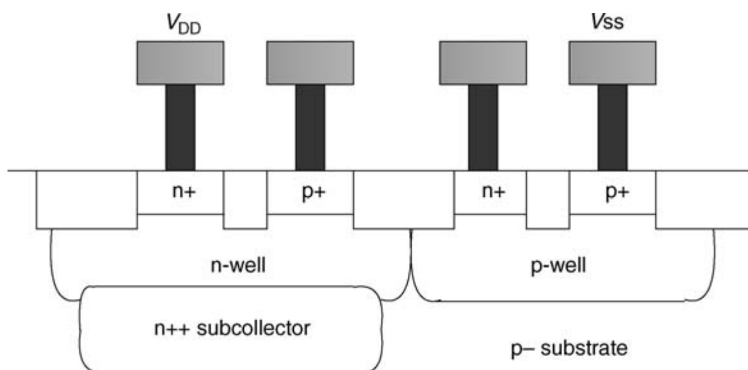



**Figure 6.11** Latchup structure for dual-well CMOS with implanted subcollector.

minority carrier recombination time is Auger recombination dominant. In the second process, the subcollectors are formed in a nonepitaxial process using a high-energy MeV implant. In this case, the implant depth is shallower, limiting the dose, concentration and sheet resistance. From a CMOS latchup design practice, the addition of a subcollector (added under CMOS devices) provides the following advantages:

- wider pnp base width of vertical parasitic pnp transistor;
- higher dopants in the base region of the vertical parasitic pnp transistor;
- lower 'n-well' shunt resistance between the n-well contact and the p-channel MOSFET device.

Experimental work demonstrates CMOS latchup trade-offs exist between the enhancement of $\beta_{\mathrm{npn}}$ and reduction of $\beta_{\mathrm{pnp}}$.

Figure 6.11 shows the dual-well CMOS latchup structure incorporating the subcollector implant, where the subcollector implant was placed under the retrograde n-well region. The implanted subcollector design point for the subcollector was defined to be suitable for a low-cost bipolar transistor technology for a 0.13-μm CMOS base technology [17].

## 6.3.1 Subcollector – npn and pnp Bipolar Current Gain

Figure 6.12 shows $\beta_{\mathrm{npn}}$ as a function of the p+/n+ space for the case of standard CMOS (e.g. without subcollector) and with the additional subcollector implant. The first important result is that the addition of a subcollector increases the bipolar npn current gain, $\beta_{\mathrm{npn}}$ (contrary to the desired result). At large p+/n+ space, the two cases converge to a common result – but the case of the dual-well CMOS with an additional subcollector is worse; the addition of a subcollector increases the bipolar npn current gain, $\beta_{\mathrm{npn}}$, approximately 2× (e.g. 0.76–1.7). This latchup degradation effect can be associated with three possible issues: (a) an increase in the effective 'collector area,' (b) subcollector lateral out-diffusion leading to a smaller base width and (c) an improved electron transport with deep collector.

There are three factors that lead to an increased bipolar current gain. First, the increased collector area (due to the subcollector depth) increases the bipolar current gain. Second, the out-diffusion of the subcollector decreases the base width. Third, the deeper structure leads to a change in the minority

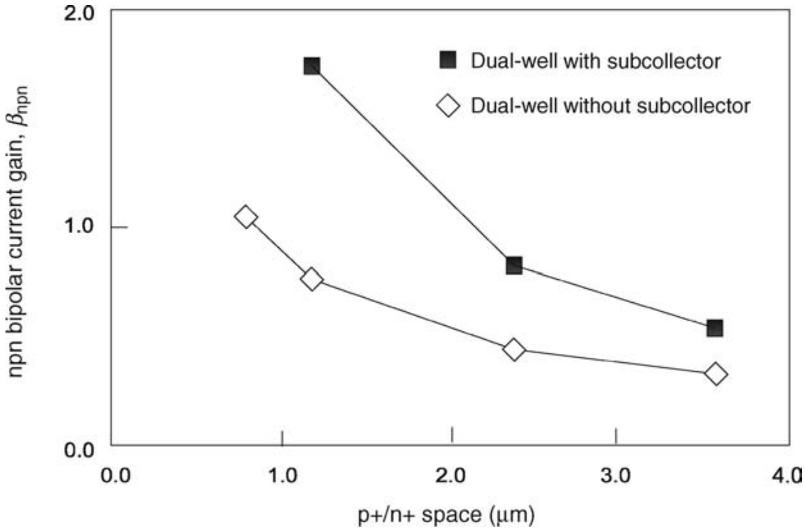



**Figure 6.12** npn bipolar current gain for dual-well CMOS structure with and without subcollector implant.

carrier trajectory. In the case of a shallower well region, the electron trajectory is of a more two-dimensional nature. However, as the emitter or the collector extends below, the worst-case trajectory path is decreased, leading to a shorter effective base width and higher bipolar gain. Figure 6.13 shows $\beta_{pnp}$ versus p+/n+ space for the case of standard dual-well CMOS (e.g. without subcollector) and CMOS with the additional subcollector implant [17].

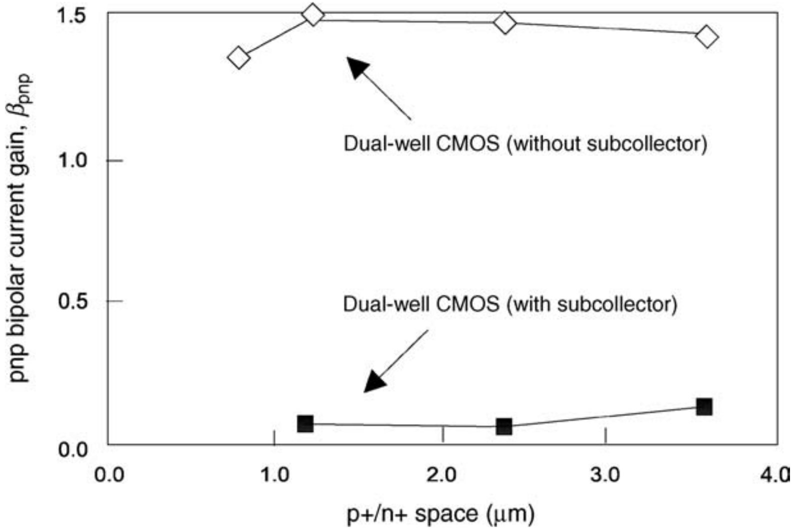



**Figure 6.13** pnp bipolar current gain, $\beta_{pnp}$, for dual-well CMOS structure with and without subcollector implant.

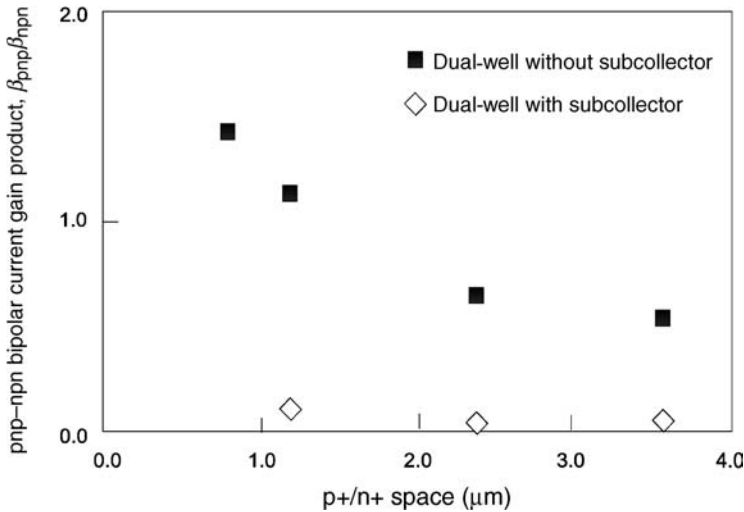



**Figure 6.14** Product of the pnp and npn bipolar current gain, $\beta_{pnp}\beta_{npn}$; versus p+/n+ space for dual-well with/without subcollector implant.

The subcollector implant provides a significant reduction in the bipolar pnp current gain, $\beta_{pnp}$. Note that the $\beta_{pnp}$ for dual-well CMOS with and without subcollector shows a weak sensitivity to the p+/n+ space.

## 6.3.2 Subcollector – $\beta$ Product $\beta_{pnp}\beta_{npn}$

Figure 6.14 plots $\beta_{pnp}\beta_{npn}$ with and without subcollector. First, the dual-well CMOS $\beta_{pnp}\beta_{npn}$ is significantly higher. Second, as the p+/n+ space is reduced, the $\beta_{pnp}\beta_{npn}$ increases significantly for the case of dual-well CMOS technology. An interesting result is that the subcollector depth increases the total area, leading to an increase in $\beta_{npn}$by approximately 2×, but at the same time leading to a 10× decrease in $\beta_{pnp}$. But, $\beta_{pnp}$ reduction has a greater effect of lowering the '$\beta$ product' term. Additionally, one can note that in the dual-well CMOS case, $\beta_{pnp}\beta_{npn}$ increases above unity as the p+/n+ space decreases; yet, in the case of the n-well and subcollector implant, $\beta_{pnp}\beta_{npn}$ remains well below unity.

## 6.3.3 Subcollector – Overshoot and Undershoot Currents

Figure 6.15 shows overshoot current, $I_{OVER}$, as a function of the p+/n+ space. For dual-well CMOS, the $I_{OVER}$ is 23 mA (for a 25 μm long structure). As the p+/n+ space decreases, $I_{OVER}$ decreased slightly from 25 to 23 mA. With subcollector, the overshoot current exceeded the tester limit of 100 mA (e.g. at least a 4× increase was evident for all p+/n+ space) [17].

Figure 6.16 shows the latchup undershoot results. As the $\beta_{npn}$ and $\beta_{pnp}$ increases, it is more easier to initiate latchup; as the p+/n+ space is decreased to a minimum space of 0.8 μm, the undershoot current required to initiate latchup decreases to 27 mA (e.g. a 2× decrease with the spacing). When the subcollector is added to the n-well structure, the undershoot current increases over

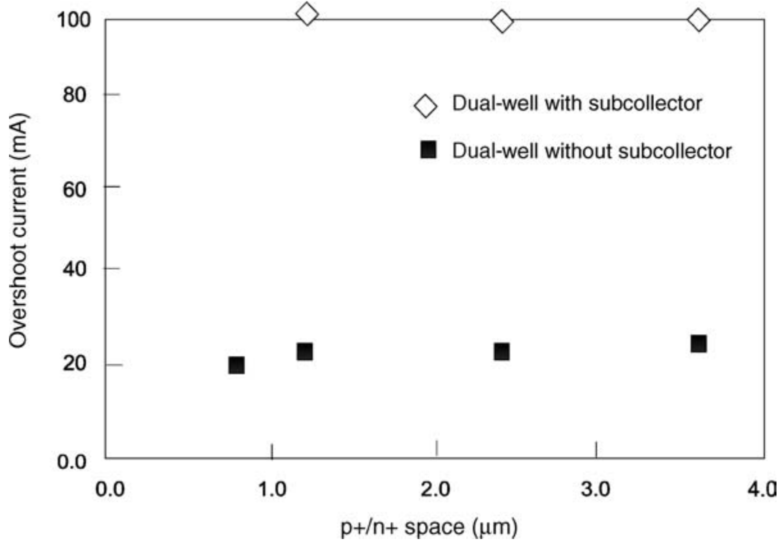



**Figure 6.15** Overshoot current versus p+/n+ spacing for dual-well CMOS for the cases of with and without implanted subcollector.

100 mA (tester maximum condition). Hence, the addition of the subcollector provides *at least* a 4× improvement (at minimum p+/n+ spacing).

From a latchup design practice, the addition of subcollectors provides the following advantages:

- significantly lower n-well shunt resistance is achieved (e.g. 10–100 ×);
- lower $\beta_{pnp}$ is achieved with subcollector;

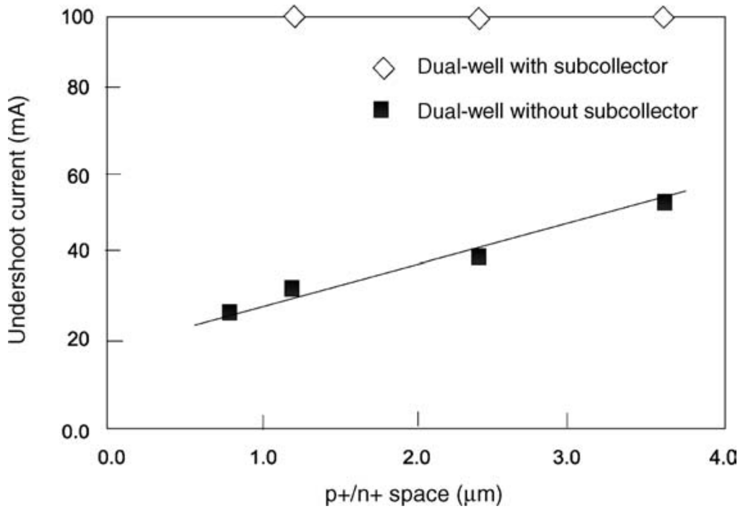



**Figure 6.16** Undershoot current versus p+/n+ spacing for dual-well CMOS with and without implanted subcollector.

- lower $\beta_{pnp}\beta_{npn}$ is achieved providing improved latchup robustness;
- Higher undershoot and overshoot currents are required to achieve latchup.

## 6.4 ALTERNATIVE ISOLATION CONCEPTS

With technology scaling, each technology generation is reducing physical dimensions; as a result, the p+/n+ spacing is being reduced. As the p+/n+ space is reduced, the STI width and depth are also being scaled. A solution to minimize the scaling impact to CMOS latchup is the usage of two isolation depths in the physical structure between p-channel and n-channel MOSFETs. This was first proposed by Rung *et al.* [18], Yamaguchi [19] and more recently Bohr [20]. Bohr proposed a 'dual depth isolation structure' that has a scalable shallow depth and a deeper second depth [20]. By providing alternate process solutions and new isolation structures, the latchup robustness of CMOS circuits can be improved. There are second isolation structures that are normally intended for the bipolar transistor that can be utilized for this purpose.

## 6.5 TRENCH ISOLATION (TI)

Figure 6.17 shows a low-cost TI structure that is utilized today in a low-cost BiCMOS technology [21]. The TI structure is an isolation structure with no polysilicon-filled regions [21]. The TI structure penetrates through the standard CMOS STI. The final structure formed is similar to a proposed 'dual-depth' STI, but is dissimilar in that the integration is not completed as a dual-damascene process. As a result, the structure is constrained neither to the STI process optimization (e.g. density, polishing requirements) nor to the process flow placement. As an independent technology step, it can be independently optimized for width, depth and density requirements. The usage of the TI structure was intended to provide a low-capacitance subcollector-to-substrate junction region and an improved density (e.g. improved spacing relative to adjacent wells and other structures). The TI structure also has advantages for ESD protection elements [21–23].

**Figure 6.17** CMOS pnpn structure with TI.

### 6.5.1 Trench Isolation and Latchup

For a CMOS latchup design practice, the TI structure can be integrated into a mainstream base CMOS technology. The TI structure can be used as independent guard ring structures or integrated into the n-well and p-well regions. It is formed as a ring, making it a natural process for bordering the n-well region. Placement of the TI structure on the n-well edge makes n-well breakdown voltage a function of neither the p-well nor the adjacent structures. From a latchup perspective, this improves the latchup trigger voltage, $V_{TR}$, for positive ramp test modes of the n-well. Additionally, the lateral parasitic pnp transistor gain will be significantly reduced because of the inability for the holes to flow from the p+ emitter to the p-well collector region. As a result, the lateral parasitic pnp bipolar device is eliminated, leaving only a vertical parasitic pnp bipolar device. An advantage of this feature is that as the technology is scaled, there will be no p+/n+ space sensitivity. Additionally, the placement of the TI-bordered n-well also impacts the lateral parasitic npn bipolar. The minority carrier electron flow from an npn emitter structure to the trench-bound n-well collector structure is inhibited by the TI structure; this decreases the npn bipolar current gain. Additionally, the dependence on spacing will also be weakened, since it will be less dependent on lateral current flow and more dependent on the flow from the emitter to the collector under the isolation structure.

Figure 6.18 shows the TI-defined isolation structure test site elements. The latchup structures consisted of a four-stripe pnpn structure with the TI integrated along the n-well border.

Figure 6.19 shows the npn bipolar current gain as a function of the p+/n+ space for the cases of standard CMOS with and without TI structures [21].

At large p+/n+ space, $\beta_{npn}$ of the no-TI and the TI-bound structures are low. For the case of the standard CMOS, $\beta_{npn}$ is 1.2 and 1.69 at ambient and accelerated temperatures, respectively (e.g. space of p+/n+ of 3.2 μm). In the case of the TI-bound structures, $\beta_{npn} = 0.68$ and 0.91 at ambient and accelerated temperatures, respectively (e.g. p+/n+ space of 3.2 μm). Notice that $\beta_{npn}$ is below unity in both the cases. As the p+/n+ spacing is decreased, $\beta_{npn}$ increases for the standard CMOS case. In the case of the TI-bound structure, a key experimental result actually shows a *decrease* in the result with *decreased* spacing. Hence, the no-TI and TI-bound structures results show an opposite CMOS latchup scaling characteristic response for the parasitic npn bipolar transistor [21]. Figure 6.20 shows the measurement results for the lateral pnp structure. For the case of the standard CMOS, $\beta_{pnp} = 0.99$ and 1.53 at ambient and accelerated temperatures, respectively (e.g. p+/n+ space of 3.2 μm). At $T = 25$ °C, the pnp bipolar current gain was unity in the ambient case and above unity in accelerated temperature

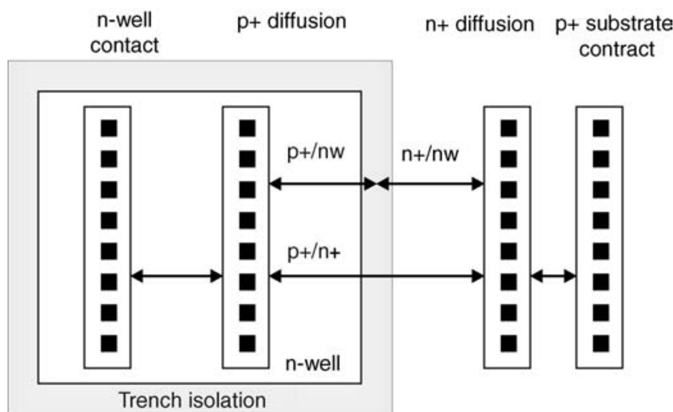



**Figure 6.18** CMOS pnpn latchup structure with TI bordered n-well region.

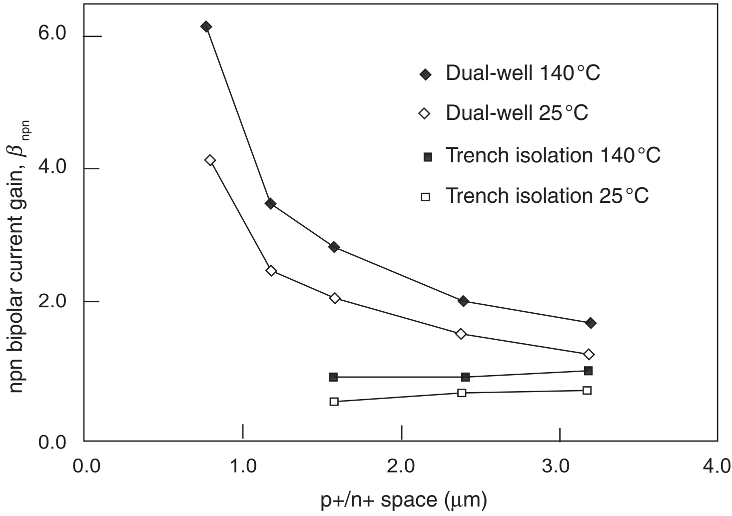



**Figure 6.19** npn bipolar current gain versus p+/n+ spacing with and without TI.

case. As a result, for all p+/n+ spacings, $\beta_{pnp}\beta_{npn}$ will be higher than unity for all CMOS cases for both room and accelerated temperatures. For the TI-bound structures, the pnp bipolar current gain, $\beta_{pnp}$, is 0.96 and 1.45 at ambient and accelerated temperatures, respectively (e.g. p+/n+ space of 3.2 μm). As the p+/n+ spacing decreases, the non-TI structure $\beta_{pnp}$ increases, while the TI-bound structure $\beta_{pnp}$ decreases. From these results, again, the p+/n+ scaling impact on CMOS latchup is *opposite* in nature [21].

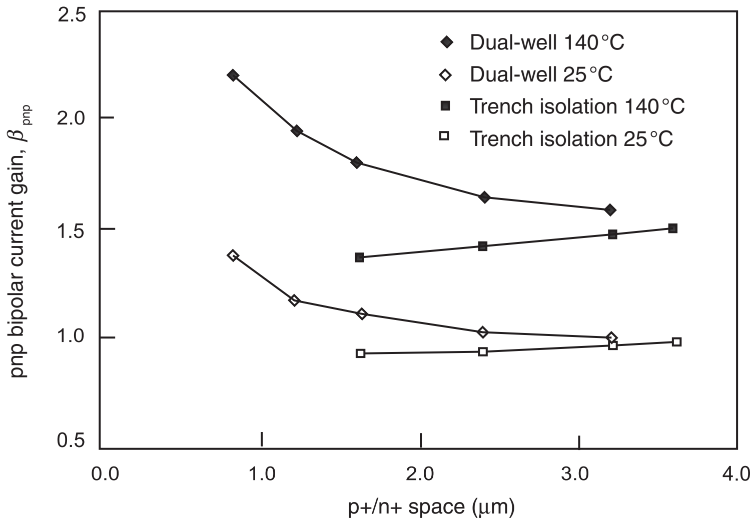



**Figure 6.20** pnp bipolar current gain, $\beta_{pnp}$, versus p+/n+ space with and without TI.

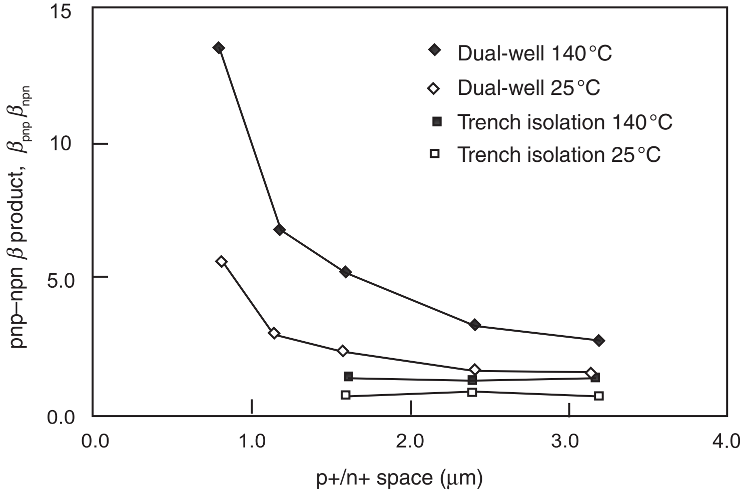



**Figure 6.21** Bipolar pnp–npn current gain product versus p+/n+ spacing with and without TI.

To evaluate the forward and reverse $\beta_{pnp}$, the measurements were taken on a bipolar curve tracer. The forward $\beta_{pnp}$ uses the p+ diffusion as the emitter and the local substrate contact as the collector; for the reverse bipolar current gain, the electrodes are switched. Bipolar current gain measurements were extracted for $I_B = 1$ mA and $V_{CE} = 5.0$ V at T = 25 °C. In the experimental results on the bipolar curve tracer, the forward and reverse pnp bipolar current gains in a p+/n+ space is $\beta_{pnp} = 0.95$ and 0.055, respectively. Notice that the forward bipolar current gain is also significantly greater than the reverse bipolar current gain (e.g. on the order of 20×). As the p+/n+ space increased, the reverse pnp bipolar current gain decreased from $\beta_{pnp} = 0.055$ to $\beta_{pnp} = 0.047$ (as the spacing increased from 1.6 to 3.6 μm). Figure 6.21 shows the latchup results for the bipolar current gain product. As the p+/n+ spacing is reduced, the $\beta_{pnp}\beta_{npn}$ *increases* from 2.69 to 13 for the non-TI case, whereas $\beta_{pnp}\beta_{npn}$ *decreases* from 1.33 to 1.197 for the TI case [21].

Figure 6.22 shows the undershoot current, $I_{UNDER}$, for the pnpn structures. For the no-TI case, as the p+/n+ spacing decreases, $I_{UNDER}$ needed to initiate CMOS latchup decreases. As the temperature increases, the amount of undershoot current, $I_{UNDER}$, required to initiate latchup further decreases because of the enhanced bipolar current gain. With TI, at room temperature, the amount of $I_{UNDER}$ needed to initiate latchup *exceeded* the JEDEC latchup tester capability of 100 mA. Second, at 140 °C, $I_{UNDER}$ decreases to a level that is still above the base technology. Third, from these results, it is evident that $I_{UNDER}$ required for latchup *increases* with decreased p+/n+ spacings; based on the improvement in the parasitic bipolar current gain with decreased p+/n+ space, these results are not surprising. Although these increases are small, what is important is the trend in the results that the TI latchup robustness improves with decrease in p+/n+ space while the non-TI latchup robustness continues to degrade.

Figure 6.23 shows the latchup overshoot study. For the case of no TI, $I_{OVER}$ required to initiate latchup decreases from 25 to 15.85 mA at $T = 25$ °C. At accelerated temperature, the overshoot current decreases from $I_{OVER} = 10.9$ mA to $I_{OVER} = 6.31$ mA.

However, for the case of trench isolation, the overshoot current remains close to 100 mA levels at room temperature. In the case of large p+/n+ space above 2.4 μm, the elevated temperature results show a degraded $I_{OVER}$ level to 17 mA. As the p+/n+ space decreases, the overshoot current improves. Hence, the scaling of the p+/n+ spacing leads to an improved overshoot condition.

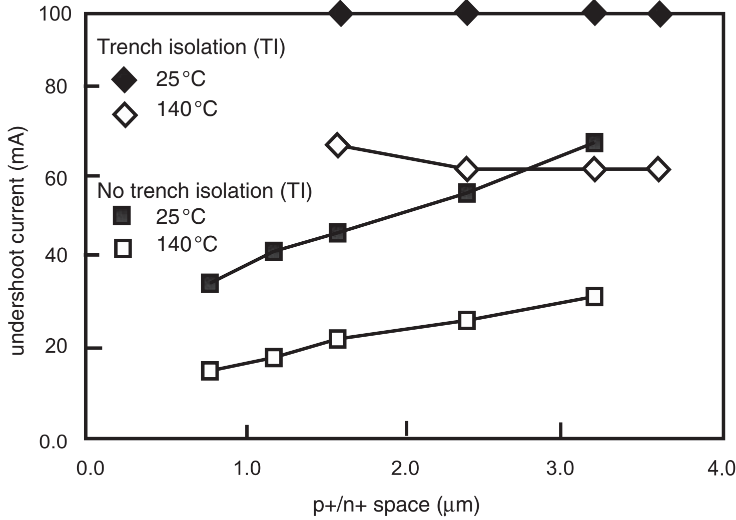



**Figure 6.22** Undershoot current versus p+/n+ spacing with and without TI.

An additional issue is the voltage triggering conditions. Prior to latchup initiation, current begins to flow while the CMOS pnpn is in a high-voltage/low-current state. The onset of the triggering is preceded by avalanche breakdown at the n-well-to-substrate voltage. As avalanche multiplication increases, current flow (e.g. in the milliampere (mA) level) begins in the pre-triggering state. We refer to this point as the 'turn-on' voltage, $V_{ON}$. To evaluate the onset of current flow, the 'turn-on'

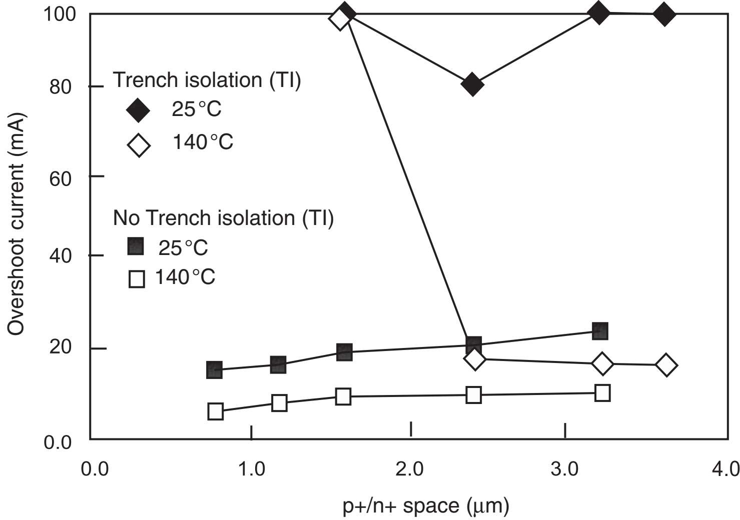



**Figure 6.23** Overshoot current versus p+/n+ spacing with and without TI.

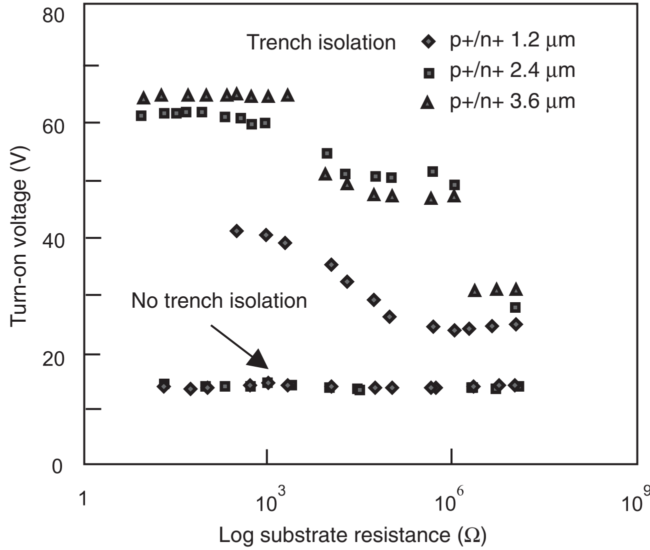



**Figure 6.24** Turn-on voltage, $V_{ON}$, versus external substrate resistance as a function of with and without trench isolation (for p+/n+ spacings of 1.2, 2.4 and 3.6 μm).

voltage, $V_{ON}$, is measured with and without TI versus substrate resistance (Figure 6.24). In many cases, $V_{ON}$ and $V_{TR}$ are equal. In a 0.22-μm technology, with no additional external substrate resistance, $V_T = 13.7$ V without TI-bordered n-wells. Without a TI border, the 'no-TI' case shows no sensitivity to the external substrate resistance or the p+/n+ spacings of 1.2, 2.4 and 3.6 μm. With a TI-bordered n-well, the turn-on voltage increases to over 60 V. The increase in the 'turn-on' voltage has increased over 43 V for all p+/n+ spacings evaluated. With a turn-on voltage of $V_{ON} = 60$ V, the latchup trigger voltage was measured at levels of $V_{TR} = 90$ V conditions. Hence, in order to initiate latchup, a trigger state of over 90 V applied to the p+ anode must be applied in order to initiate latchup. As the p+/n+ spacing decreases and the external resistance increases, these results are reduced toward the no-TI case. Trigger voltages increase from $V_{TR} = 25$ V to $V_{TR} = 90$ V for no-TI and TI, respectively [21].

In the 0.18-μm technology, as the turn-on voltage increases, the trigger voltage also increases. Figure 6.25 shows the trigger voltage as a function of p+/n+ spacings for the case of dual-well technology with and without TI. The $V_{TR} = 13$ V for dual-well technology with no TI and $V_{TR} = 70$ V for that with TI.

A latchup design practice is as follows:

- a TI perimeter to an n-well region lowers the lateral $\beta_{npn}$ scaling with p+/n+ space;
- a TI perimeter structure on a CMOS n-well region lowers the lateral $\beta_{pnp}$ with p+/n+ scaling;
- a TI perimeter structure to a CMOS n-well region lowers the $\beta_{pnp}\beta_{npn}$ scaling with p+/n+ space;
- a TI structure demonstrates an improvement in all latchup metrics (e.g. bipolar current gain, bipolar current gain product, overshoot, undershoot, turn-on voltage and trigger voltage).

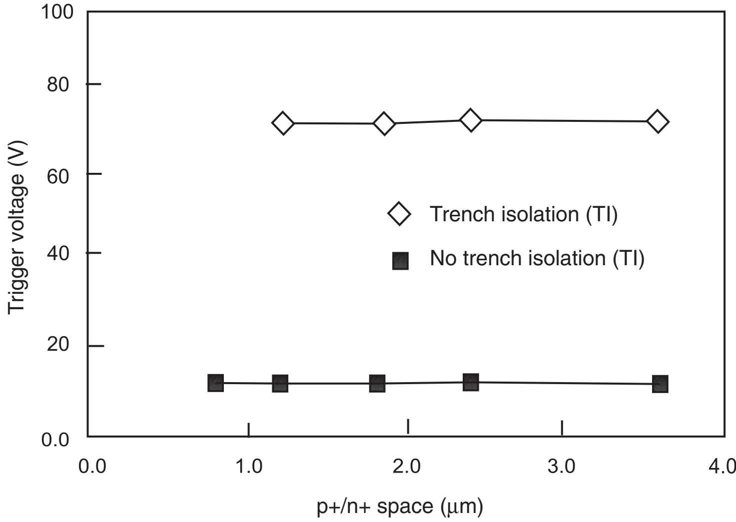



**Figure 6.25** Trigger voltage, $V_{TR}$, versus p+/n+ space for the case of with and without TI.

## 6.5.2 Trench Isolation and Subcollector

Latchup robustness can be further improved by integrating both the subcollector implants and TI. The addition of implants to the well structure can improve the latchup by decreasing the pnp bipolar current gain and reducing the n-well shunt resistance. In this study, an implanted subcollector and the TI are integrated [21].

The development of implanted subcollectors avoids the usage of epitaxial-formed subcollector semiconductor processes. With the scaling of BiCMOS SiGe HBT devices, the vertical profile is reduced allowing for the integration with lower energy implanted profiles. The implants available for the SiGe HBT devices can be used for improving latchup robustness in the digital circuits, analog circuits, I/O circuits and ESD elements. These structures, both the implanted subcollector and the trench isolation structure, can also be integrated into RF CMOS and mainstream sub-0.1 μm scaled CMOS technology. Figure 6.26 shows the integration of the subcollector implant and TI structure. The subcollector implant is placed below the CMOS n-well implant. The subcollector sidewall capacitance is also reduced using the TI structure. The TI depth and the subcollector are chosen to optimize the subcollector capacitance and the other device parameters.

Figure 6.27 shows the lateral $\beta_{npn}$ versus p+/n+ space with subcollector and TI for ambient and elevated temperatures, as well as the standard base process. With the addition of subcollector and TI, the lateral $\beta_{npn} = 1.02$ at large p+/n+ space and $\beta_{npn} = 0.99$ at the smallest space (at $T = 140\,°C$). Again, in contrast to the standard CMOS process, the lateral $\beta_{npn}$ decreases with p+/n+ space (e.g. instead of an increasing nature in the base CMOS technology). Additionally, the temperature sensitivity is weak in the case of the TI case, remaining close to unity even at elevated temperatures.

In Figure 6.28, the $\beta_{pnp}$ is shown for the case of (a) no TI, (b) TI and (c) TI and subcollector implant. Evaluation of the parasitic lateral pnp current gain for case (a) is that as the p+/n+ spacing decreases, the bipolar pnp current gain increases. With TI, there is a decrease in the lateral $\beta_{pnp}$ with a decrease in the p+/n+ spacing; as the p+ diffusion approaches the isolation sidewall, the vertical current

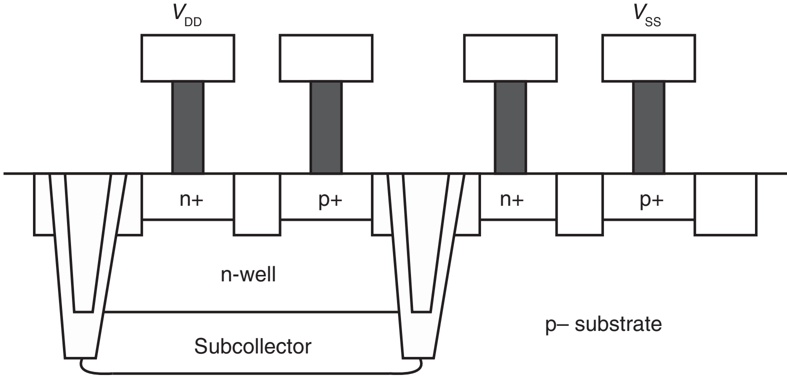



**Figure 6.26** CMOS pnpn latchup structure with trench isolation and implanted subcollector.

decreases, providing a small decrease in the current gain. With the addition of an implanted subcollector, the parasitic bipolar current gain significantly decreases well below unity.

In Figure 6.29, the effect of the subcollector on CMOS latchup is highlighted. At the smallest p+/n+ space and the highest temperature, the case of TI has a pnp bipolar current gain of $\beta_{pnp} = 1.38$ and the case of the TI and sub-collector has the current gain of $\beta_{pnp} = 0.41$; hence, the presence of the subcollector provides an additional 3× reduction in the pnp bipolar current gain, $\beta_{pnp}$, at worst-case conditions [21].

To evaluate the forward and reverse $\beta_{pnp}$, measurements were take on a bipolar curve tracer for improved accuracy (Table 6.3). For the forward bipolar current gain, the p+ diffusion is the emitter and the local substrate contact is the collector; for the reverse bipolar current gain, the electrodes are

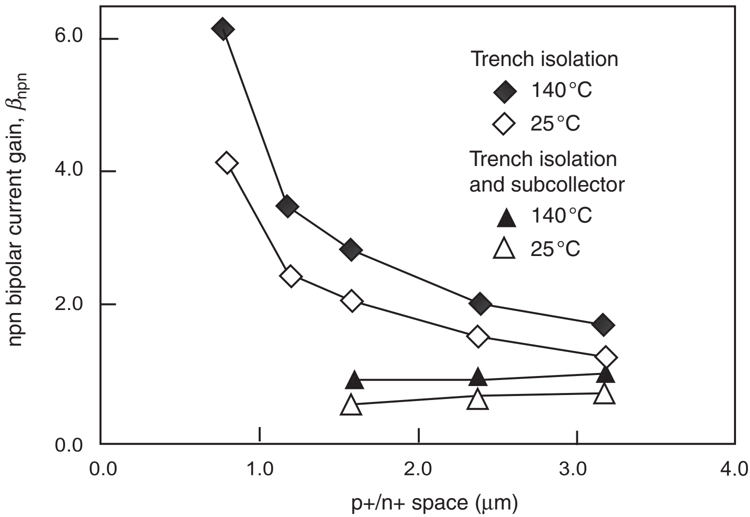



**Figure 6.27** Parasitic npn bipolar current gain with p+/n+ space with and without the addition of trench isolation and subcollector implants.

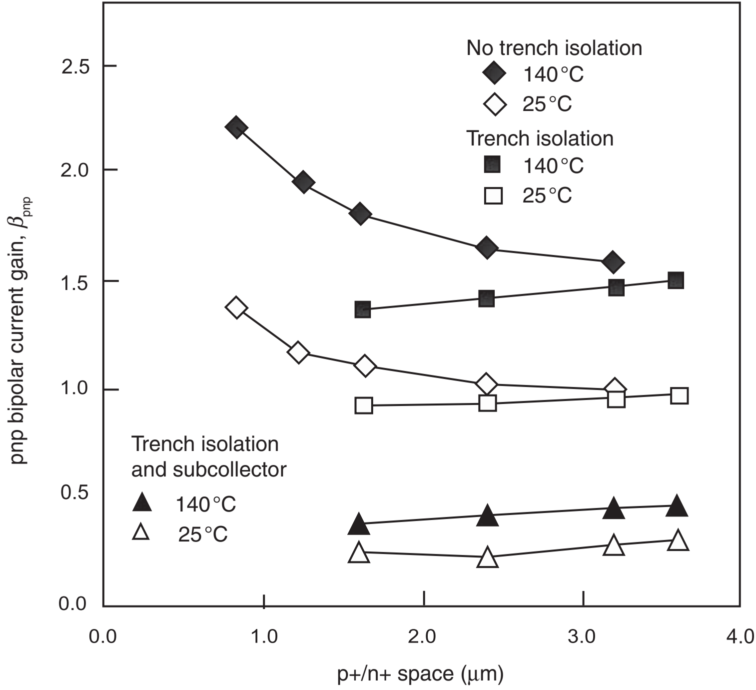



**Figure 6.28** Parasitic pnp bipolar current gain, $\beta_{pnp}$, with p+/n+ space with and without the addition of TI and subcollector implants.

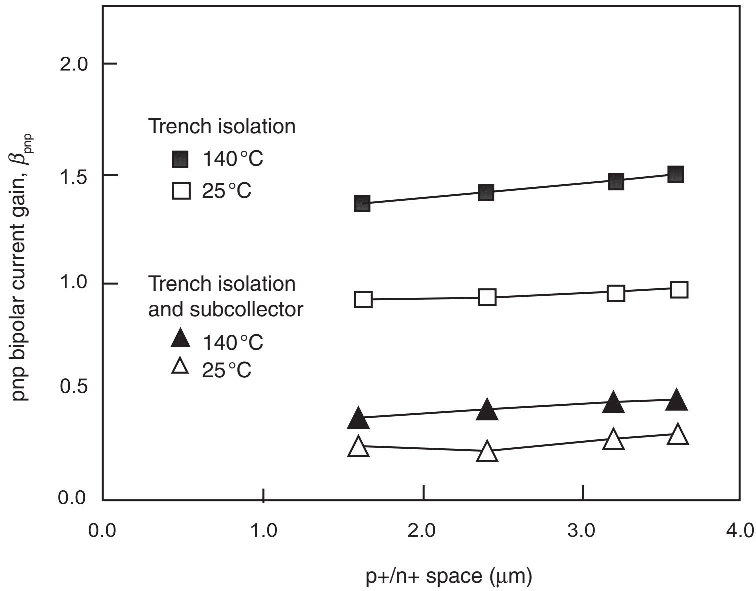



**Figure 6.29** Trench isolation parasitic pnp bipolar current gain, $\beta_{pnp}$, with p+/n+ space with and without subcollector implants.

**Table 6.3** Bipolar parameters versus p+/n+ space for CMOS pnpn structures with TI and subcollector.

| p+/n+ spacings | Forward pnp bipolar current gain, $\beta_F$ | Reverse pnp bipolar current gain, $\beta_R$ | Avalanche breakdown voltage, $V_{AV}$ (V) |
|---|---|---|---|
| 1.6 | 0.21 | 0.009 | 55 |
| 2.4 | 0.22 | 0.009 | 55 |
| 3.2 | 0.22 | 0.008 | 55 |
| 3.6 | 0.24 | 0.007 | 55 |

switched. Bipolar current gain measurements were extracted for $I_B = 1$ mA and $V_{CE} = 5$ V at ambient temperature. In the results, as the spacing of the p+/n+ decreases, $\beta_{pnp}$ decreases instead of increasing. Notice that this is not true for the reverse bipolar current gain. Notice that the forward $\beta_{pnp}$ is also significantly greater than the reverse $\beta_{pnp}$. Additionally, the avalanche voltage, $V_{AV}$, occurs at 55 V (for the case of zero base current).

In Table 6.4, the parameters are shown for three different processes for the forward and reverse $\beta_{pnp}$ for the p+/n+ spacing of 1.6 µm. The results show that the 'subcollector-only' case has a stronger influence on the pnp bipolar current gain (e.g. 3 × lower), whereas $\beta_{pnp}$ is reduced at the addition of TI 30 %.

Figure 6.30 shows the impact of the subcollector and TI on the npn and pnp bipolar transistors. The experimental results show that $\beta_{pnp}\beta_{npn}$ with only TI is below unity at ambient temperature, but above unity at elevated temperatures, for all p+/n+ spacings. Measurements show that $\beta_{pnp}\beta_{npn}$ decreases with the smaller p+/n+ spacings. For the case of the additional subcollector, the latchup $\beta_{pnp}\beta_{npn}$ results remain below unity for all structure sizes tested and maintaining a 3 × improvement (over the TI case only).

For the undershoot analysis, it was shown that with TI, the latchup undershoot parametric does not decrease with spacing. With TI, but without the subcollector implant, the undershoot current drops to $I_{UNDER} = 70$ mA (e.g. for a 25 µm wide pnpn structure) at elevated temperatures. Figure 6.31 shows the undershoot with both TI and subcollector (*note*: with the subcollector and TI, the undershoot current level, $I_{UNDER}$, is pinned at the maximum tester level of 100 mA).

This same effect is evident with the $I_{OVER}$ analysis of the pnpn structures. Figure 6.32 shows the experimental results of a latchup overshoot condition. With elevated temperature, the wider spacings show evidence of $I_{OVER}$ degradation in the TI case. With the addition of the implanted subcollector, the maximum test level of 100 mA overshoot (e.g. 4 mA/µm) is preserved at 140 °C.

It was also discovered that the advantages of TI structure and subcollector implants provide excellent low capacitance ESD network elements [23]. ESD elements that utilized both trench isolation and subcollector achieved the best HBM and MM experimental results [23].

The introduction of TI has significant implications for the future of prevention of latchup in advanced semiconductors for RF BiCMOS, as well as power, automobile and space applications. In

**Table 6.4** Bipolar parameters for CMOS pnpn structures with combinations of TI and implanted subcollector.

| Process | p+/n+ spacing (µm) | Forward pnp bipolar current gain, $\beta_F$ | Reverse pnp bipolar current gain, $\beta_R$ |
|---|---|---|---|
| Trench isolation | 1.6 | 0.95 | 0.055 |
| Subcollector | 1.6 | 0.29 | 0.019 |
| Trench isolation and subcollector | 1.6 | 0.21 | 0.009 |

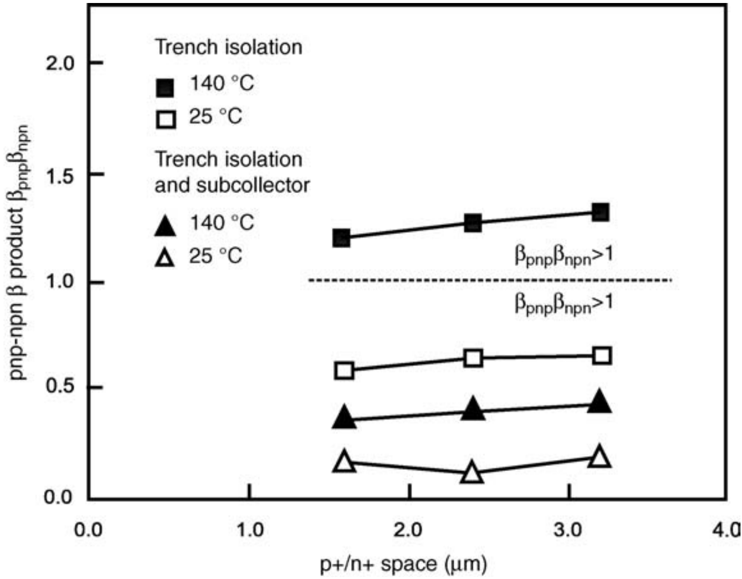



**Figure 6.30** *β* Product versus p+/n+ space of CMOS pnpn with TI for the cases of with and without subcollector implants.

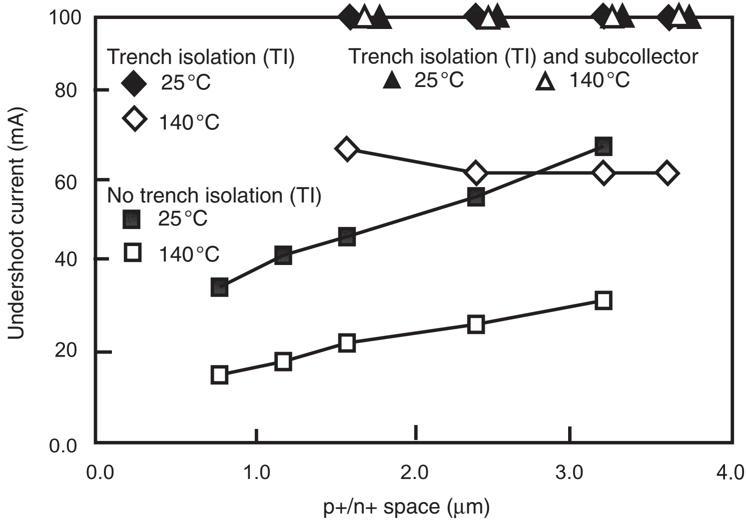



**Figure 6.31** Undershoot current versus p+/n+ space for three technology cases (no TI, TI, and TI and subcollector).

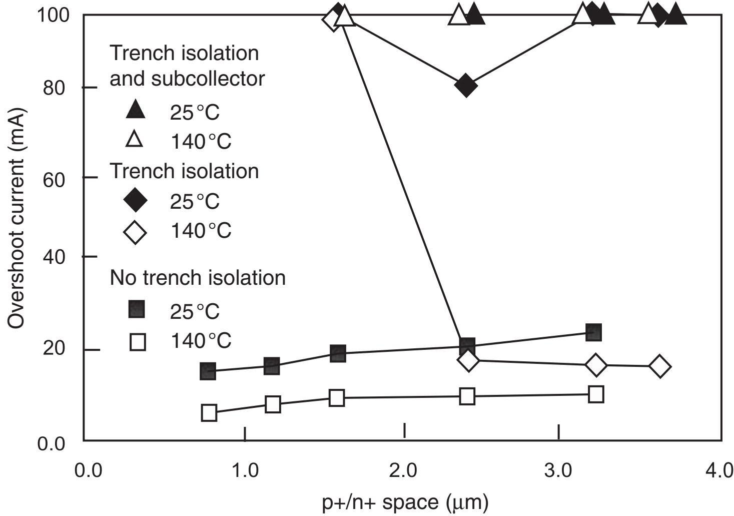



**Figure 6.32** Overshoot current, $I_{OVER}$, versus p+/n+ space for three technology cases (case A: no TI; case B: TI; and case C: TI and subcollector).

essence, the integration of TI structure with the STI structure is structurally equivalent to a 'dual depth' isolation structure that has been proposed for mainstream CMOS technology, but has never been implemented. In the 1980s, Rung *et al.* [18] and the team of Yamaguchi *et al.* Eiden [19] proposed the concept for mainstream CMOS for sub-1 μm technology generations. More recently, in the 1990s, to address STI scaling, the concept of a dual depth isolation structure was proposed by Bohr [20]. From a scaling perspective, the TI dimensions can be scaled with the scaling of the subcollector implant. The width scaling of the TI structure will allow smaller p+/n+ space, as this structure is utilized for latchup solutions. Hence, the scaling of the structure is natural to the objectives of CMOS MOSFET scaling and BiCMOS SiGe HBT device scaling.

A CMOS latchup design practice, combining both trench isolation and subcollector, is as follows:

- A TI perimeter to an n-well region with a subcollector lowers the lateral npn current gain scaling and eliminates any subcollector npn enhancement concerns.
- A TI perimeter structure and subcollector implant on a CMOS n-well region significantly lowers the lateral pnp current gain.
- A TI structure and subcollector demonstrates an improvement in all latchup metrics (e.g. bipolar current gain, bipolar current gain product, overshoot, undershoot, turn-on voltage and trigger voltage) and eliminates the dimensional scaling impacts observed in CMOS technology.

## 6.6 DEEP TRENCH

Deep trench (DT) structures have been utilized for early bipolar development for the high-performance bipolar memory. In the early development, the 'fill' region changed from fully silicon dioxide, to an interlevel dielectric polyimide to finally a polysilicon fill material. Today, high-performance BiCMOS

applications utilize the polysilicon-filled deep trench structure for definition of the subcollector. In a high-performance BiCMOS silicon germanium carbon (SiGeC) technology, DT structures are used to provide both performance and density. For RF bipolar applications, DT isolation reduces subcollector-to-substrate capacitance by elimination of the subcollector sidewall capacitance.

DT structures are also utilized in trench DRAM capacitor elements and embedded DRAM technology for SOC applications. In DT capacitor structures, thin dielectric exists to provide a high-capacitance element. For smart power applications, standard CMOS technology is extended to integrate high-voltage power devices. Drain-extended NMOS (DeNMOS), lateral diffused MOS (LDMOS) and insulated gate bipolar trench devices (IGBT) are used in high-voltage CMOS to provide 20–120 V CMOS applications. Trench structures exist in power electronics to provide sidewall vertical MOSFETs and other novel structures. With the high-voltage CMOS applications, there is a motivation to provide latchup robustness at levels significantly above the native power supply voltage of advanced CMOS. In automotive applications, there are states of mis-installation of electronics leading to reverse pin conditions (e.g. negative instead of positive voltage) to mis-installation of batteries. In addition, there are environments with high inductive load dumps and switching conditions that lead to high transient states. Hence, there is motivation to provide high latchup immunity. For space applications, it is desirable to provide technology that is immune to SEL events. Hence, process solutions that can provide a high trigger voltage, $V_{TR}$, or high holding voltage, $V_H$, are desirable. As discussed in Chapters 1 and 3, SEL can occur in space applications from heavy-ion particles to cosmic ray events.

Deep trench isolation can be utilized in multiple ways to provide improved CMOS latchup robustness:

- Stand-alone DT guard rings structures can be used to isolate injection sources from adjacent circuitry in conjunction with other independent guard ring structures [24–26].
- DT guard rings can be utilized on the perimeter of n-well regions, such as the n-well containing p-channel MOSFETs [27–32,34].

## 6.6.1 Deep Trench as a Guard Ring

DT structures can be used as a stand-alone guard ring for CMOS latchup prevention, structural isolation and noise isolation. DT guard rings provide barriers for lateral parasitic npn transistors, as well as it can be integrated into both passive and active guard ring concepts (see Chapter 7). Electrons injected into a p− substrate must diffuse below the DT structure in order to be transported from one n− diffusion region to another. In the minority carrier diffusion process, bulk minority carrier recombination can occur along the path. Additionally, surface recombination can also occur on the sidewall of the DT structure. Given the injection source exists within the guard ring, the probability to escape is the probability that an electron collects or recombines outside the DT guard ring [24–26].

## 6.6.2 Deep Trench Guard Ring Characterization – Guard Ring Efficiency Methodology

To characterize the DT guard ring efficiency, as discussed in Chapter 4, the factors $F$, $1/F$ and $1-1/F$ can be evaluated. An injection source is used within the DT ring structure and an n-well is used outside

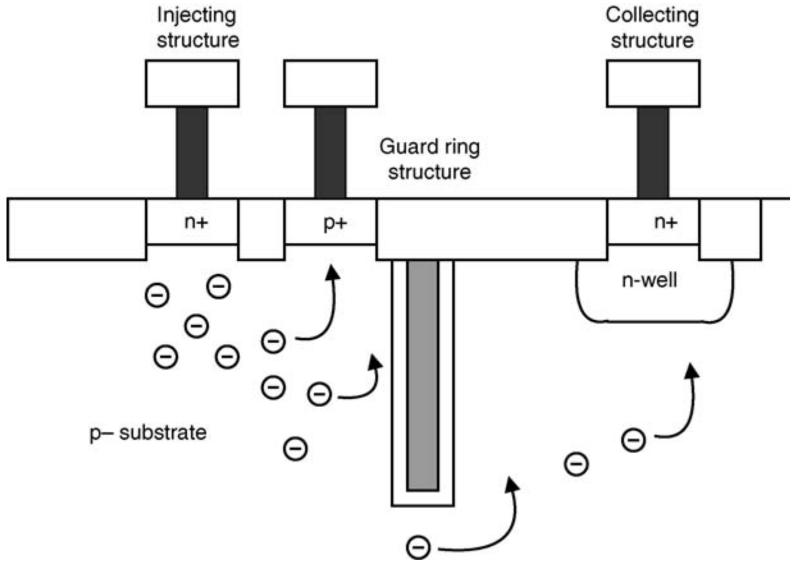



**Figure 6.33** Characterization structure highlighting the n+ injection source, an internal p+ guard ring, a deep trench guard ring structure and the n-well collecting structure.

the trench to collect the current, as shown in Figure 6.33. In CMOS, injection sources may be n-diffusions, n-well regions, buried layers or triple-well n-diffusion regions. In BiCMOS, the injection source may be a trench-defined or nontrench-defined subcollector region; and in smart power applications, the injection source can be a LDMOS n-tub region. From Chapter 4, metric $F$ is defined as

$$F = I(injector)/I(collector).$$

Table 6.5 shows the collection efficiency results as a function of the trench depth for a DT guard ring. The inner injector is a STI-bound diode structure and the outer collector is an n-well ring to measure the collected current that escapes the guard ring.

As the injection current is ramped, the collected current is measured and the ratio factor $F$ is plotted. In the study, the bias condition on the collecting structure is increased, leading to an improved collection with the increasing depletion region of the collection structure. In our study, the value of $F$ is of interest when the injection current is 100 mA. As part of the CMOS design practice, injection

**Table 6.5** Deep trench depth versus guard ring metrics (F, $1/F$, $1-1/F$ and guard ring efficiency).

| Experimental split trench depth (μm) | Actual trench depth (μm) | $F$ | $1/F$ | $1-1/F$ | Guard ring efficiency (%) [$100 \times (1-1/F)$] |
|---|---|---|---|---|---|
| 1 | 1.6 | 4.64 | 0.2155 | 0.7845 | 78.45 |
| 2 | 2.46 | 6.29 | 0.1590 | 0.8410 | 84.10 |
| 3 | 3.38 | 8.71 | 0.1148 | 0.8852 | 88.52 |
| 4 | 4.14 | 10.6 | 0.094 | 0.9060 | 90.60 |
| 6 | 6.25 | 17.1 | 0.05848 | 0.9415 | 94.15 |

ratio plots are produced for all physical structures. In our experimental work, the procedure was as follows [26]:

- Choose an injection source dimension. This remains ‘fixed’.
- Choose a collection region design/dimension. This remains ‘fixed’.
- Choose a guard ring structure.
- Sweep a forced current in the injection structure (with the collecting structure having a fixed d c voltage bias state).
- Repeat the forced current in the injection structure, with a new fixed d c voltage bias state (e.g. step different voltage states after each sweep of the forced current).
- Plot $F$ versus injection current for each fixed d c voltage bias step of the collection structure.
- Plot $1/F$ versus injection current for each fixed d c voltage bias step of the collection structure.
- Extract the value of $F$ and $1/F$ at 100 mA current level.

Figure 6.34 shows an injection ratio plot. The $x$-axis is the injection current, the left-hand side $y$-axis is the $F$ parameter and the right-hand side $y$-axis is the collected current. As the current increases, the values become well controlled and weak function of the bias conditions on the collecting structure, and the derivative of $F$ as a function of injection current approaches a constant.

From a semiconductor CMOS latchup design perspective, the slope of the metric $F$ (e.g. equivalently $1/F$) is a measure of the guard ring effectiveness as a function of current. Hence, the derivative of the guard ring efficiency metric is of significant interest. Notice that in Figure 6.34, the slope of the derivative approaches a linear value, or

$$\frac{\partial F(V)}{\partial I_{\text{inj}}} \approx K.$$

Plotting the results, Figure 6.35 shows the injection ratio $F$ as a function of DT depth (*note*: value extracted at the 100 mA current injection level). At the extracted value of 100 mA, the metric $F$ shows a

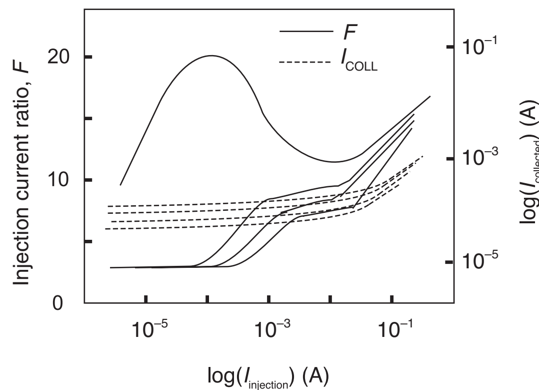



**Figure 6.34** Guard ring injection ratio plot $F$ versus forced injection current.

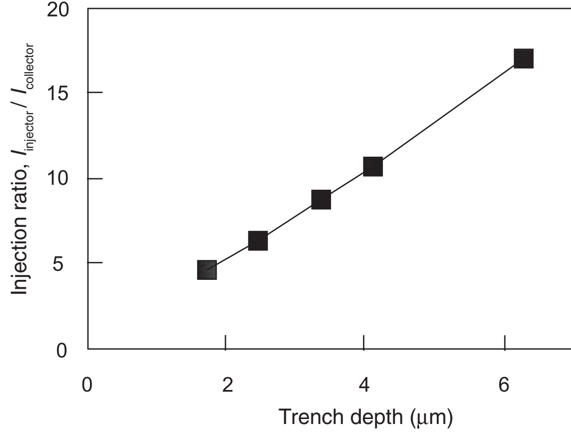



**Figure 6.35** Guard ring injection ratio plot $F$ versus deep trench (DT) depth.

linear increase with trench depth. Evaluation of the DT depth in the form of the $F$ factor shows that the expression becomes linear as a function of the DT depth. In linear form, we can assume a form $F = AL_{\mathrm{DT}} + B$, where $A$ and $B$ are constants. Figure 6.36 shows $F$ as a function of trench depth, leading to the fitting parameters [24],

$$F = 2.7087L_{\mathrm{DT}} - 0.19.$$

Inverting the factor, we can introduce the $1/F$ factor that is more physically associated with the probability of escape and the guard ring efficiency. Assuming a functional and power relationship [24]

$$\frac{1}{F} = \frac{A}{(L_{\mathrm{DT}})^B},$$

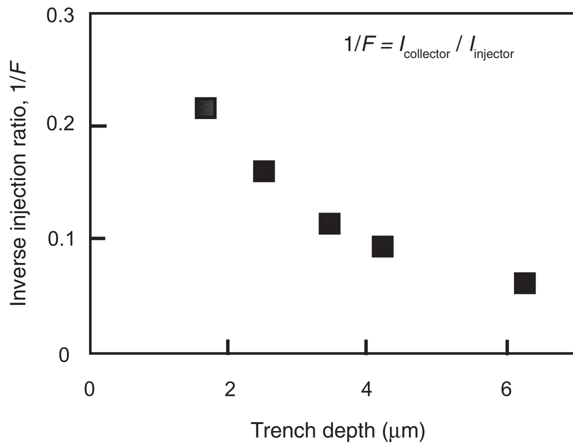



**Figure 6.36** Guard ring injection ratio plot $1/F$ versus DT depth.

where $L_{DT}$ is the DT depth (in μm) and $A$ and $B$ are constants, the data fit for $A = 0.3575$ and $B = 0.9568$ (e.g. unity). From this definition, it can be shown that the guard ring efficiency can be expressed in a power form,

$$\Psi_{GRE} = 1 - \frac{1}{F} = 1 - \frac{A}{(L_{DT})^B}.$$

In this form, it can be observed that as the factor $1/F$ approaches zero, the guard ring efficiency approaches unity (e.g. or in percentage form, 100 %).

## 6.6.3 Deep Trench Guard Ring Characterization – Bipolar Transistor Methodology

A second characterization method viewed the structure as a parasitic bipolar transistor and the 'beta spoiling' nature of the DT within the base region. From the electrical characterization of the guard ring structure, the n+ injection structure, the p− substrate and the n-well collecting structure form a parasitic bipolar transistor. From this structure, the forward and reverse bipolar current gains of the structure as a function of the trench depth can also be characterized. In this measurement technique, the substrate was used as a bipolar base region and $I_C$ versus $V_{CE}$ characteristics were obtained to evaluate the injection characteristics with the substrate grounded [24–26].

As the DT depth increases, the ability to transport the electrons from the injecting structure to the collecting structure decreases. This is measured by the decrease in the forward bipolar current gain. Figure 6.37 shows the forward and reverse bipolar current gains as a function of the DT depth using the injection–collection structure as a parasitic bipolar transistor element. The measurements were taken by obtaining the collector current versus the collector–injector voltage as a function of the base current. The parasitic bipolar current gain was then extracted from the $I_C$ versus $V_{CE}$ characteristics. As the trench depth increases, the forward bipolar current gain decreases. This is anticipated on the basis that the trench structure prevents the lateral diffusion of the electrons from the injector to the collector decreasing the parasitic bipolar current gain and the transport efficiency. This 'bipolar method' uses the bipolar characteristics as a measure of the 'beta spoiling' within the structure, and is related to the

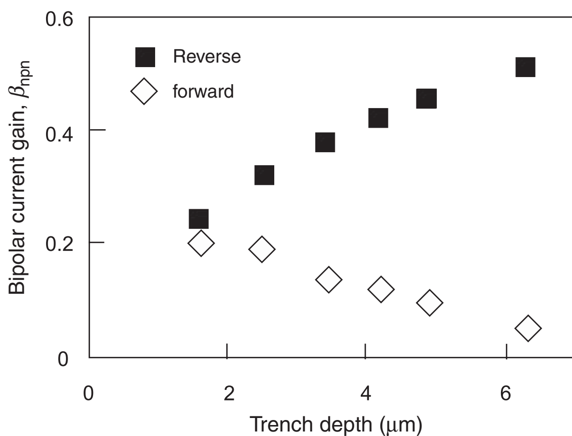



**Figure 6.37** Forward and reverse bipolar current gains versus DT depth.

**Table 6.6** Trench depth study and the 'bipolar' and guard ring efficiency methodologies.

| Experimental split trench depth (μm) | Actual trench depth (μm) | Forward bipolar current gain, $\beta_F$ | Reverse bipolar current gain, $\beta_R$ | $F$ | $1/F$ |
|---|---|---|---|---|---|
| 1 | 1.6 | 0.205 | 0.244 | 4.64 | 0.2155 |
| 2 | 2.46 | 0.191 | 0.323 | 6.29 | 0.1590 |
| 3 | 3.38 | 1.36 | 0.381 | 8.71 | 0.1148 |
| 4 | 4.14 | 1.20 | 0.424 | 10.6 | 0.094 |
| 6 | 6.25 | 0.050 | 0.513 | 17.1 | 0.05848 |

guard ring efficiency metrics. Table 6.6 shows the bipolar methodology and the guard ring efficiency methodology results. Notice that the forward bipolar current gain tracks with the $1/F$ metric [24].

Hence, the usage of DT-independent guard rings lowers the probability of electron escape and improves the latchup margin to the 'external latchup' problem observed in integrated electronics.

These results demonstrate that the usage of DT structures as independent guard rings in a p− substrate wafer significantly improves the problem of 'external latchup' where an injection source leads to the triggering of latchup in adjacent circuitry. Electrons injected from the forward biasing of an n− or n-well region in a p− substrate can initiate latchup in an adjacent structure. From this experimental work, the deeper the trench, the lesser the probability of an electron to escape outside the trench guard ring and get collected by an adjacent well structure.

## 6.6.4 Deep Trench Structure Within the pnpn Structure

Trench structures can be used to improve the latchup robustness of a technology by reducing the lateral bipolar current gain of parasitic elements. To evaluate the DT structure, the standard latchup structure was modified to allow placement of the DT structure within the pnpn structure (Figures 6.38 and 6.39). The DT structure was centrally placed on the perimeter of the n-well region that contains the p+ diffusion. The p+/n+ spacings were not altered to allow the placement of the DT structure [28–32].

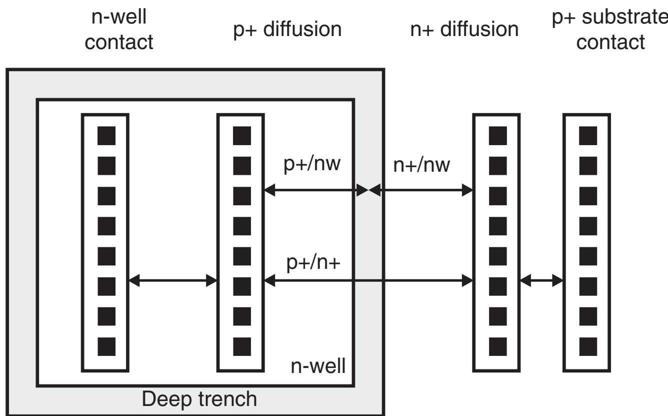



**Figure 6.38** DT pnpn latchup structure.

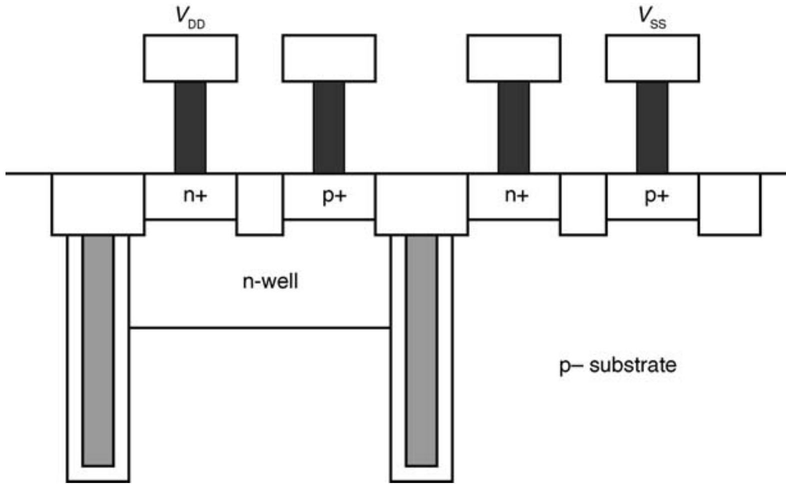



**Figure 6.39** DT pnpn latchup structure cross section.

The lateral $\beta_{pnp}$ can be reduced by placement of a trench region on the sidewall of the n-well region. Forward biasing of the p-diffusion in the n-well leads to minority carrier holes in the n-well region that diffuse to the collector region. The minority carrier holes flow from the low-doped well region to the heavily doped retrograde well region. The built-in electric field established by the mega electron volt retrograde well implant creates a drift component opposite to the flow of the holes to the collector region. Holes will flow laterally out of the well region in the case where no trench structure is present. Holes that flow from the p− diffusion to the p−substrate contribute to the forward pnp bipolar transport factor, $\alpha_{fpnp}$. Placement of the trench region also reduces the lateral parasitic npn formed between an n-diffusion outside the n-well and the n-well itself. Minority carrier electrons that are transported from the n+ diffusion region (emitter) to the n-well (collector) contribute to the forward lateral npn bipolar transport factor, $\alpha_{npn}$.

The latchup trigger voltage is shown as a function of the DT depth (Figure 6.40). In the case of no DT, the latchup trigger voltage is approximately 10 V. When DT depth is below 2 μm, the trench depth does not penetrat through the retrograde n-well region. As the DT structure extends below the retrograde n-well region, $V_{TR}$ increases 2×. As the DT depth increases, $V_{TR}$ continues to increase. With a 6-μm DT structure, the trigger condition increases 5×. For a 2.5 V technology, this establishes a latchup trigger condition over 25× the $V_{DD}$ power supply value. For applications, such as 10 V $V_{DD}$ conditions, this provides a 4× margin above $V_{DD}$. For automotive or space applications, this provides an increased assurance of high-voltage spikes initiating latchup.

Figure 6.41 shows $V_H$ versus DT depth results [28–32]. These data are associated with the aforementioned trigger voltage conditions in the prior figure. Experimental results show that $V_H$ is approximately 2.5 V without a DT structure. As the DT depth is increased, $V_H$ increases above the 2.5 V $V_{DD}$ levels. As the DT depth increases to more than 3 μm, $V_H$ increases over 10 V.

This provides a 4× margin to a 2.5 V $V_{DD}$ power supply application. With the full DT depth of a SiGe HBT device, $V_H$ increases to approximately 20 V. To fall into this state, a voltage impulse must exceed 55 V and then be able to sustain 20 V condition to remain in this latched state. The trench structure provides a latchup robustness of $7 \times V_{DD}$ for this technology generation. Addressing the trigger voltage condition, we can further explore its sensitivity in its design space. Trigger and holding voltage conditions form loci of points in its design space which satisfy the latchup stability criterion.

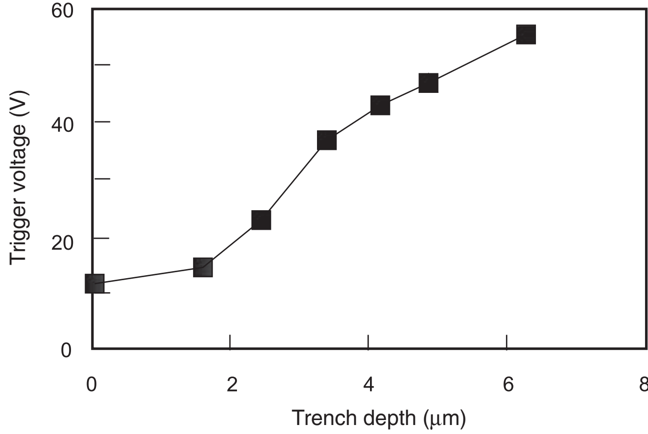



**Figure 6.40** CMOS latchup trigger voltage, $V_{TR}$, results of the DT pnpn latchup structure as a function of deep trench depth.

There are four parameters that strongly influence the latchup conditions from the design perspective. These parameters include the npn, and the pnp bipolar current gain, the n-well contact to p+ diffusion space and the substrate contact to n-well space. The bipolar current gain terms in the plane can be modulated by varying the p+/n+ space between the two physical emitters. Well and substrate bypass resistors can be used in parallel with the emitter–base junctions in order to evaluate the stability in resistance space. In a two-dimensional (2-D) plot of $\log(R_{sx})$ and $\log(R_{\text{well}})$, the loci of points for the holding and the trigger voltage can be plotted to demonstrate the values of the resistance of the well and the substrate that provide the latchup state for a given p+/n+ spacing. This two-dimensional $\log(R)$ space can be viewed as a three-dimensional (3-D) space by adding the DT parameter as an additional axis in the design space. In our experiment, it was found that the results were a weak function in

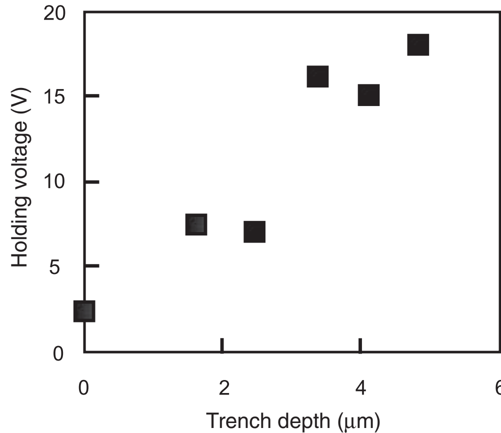



**Figure 6.41** CMOS latchup holding voltage, $V_H$, results of the DT pnpn latchup structure as a function of deep trench depth.

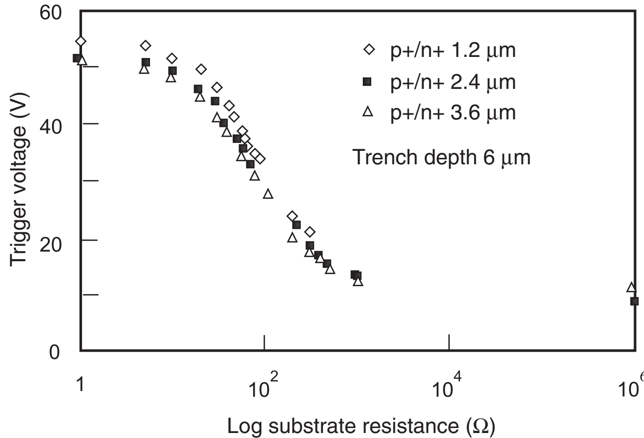



**Figure 6.42** CMOS latchup trigger voltage results versus the logarithm of the external substrate resistance for a 6 μm deep trench (as a function of p+/n+ spacing).

$\log(R_{\text{well}})$ space and a very strong function of the $\log(R_{sx})$ space. For easier viewing and understanding, we will focus on the two dimensions of DT and $\log(R_{sx})$ space cuts in the 3-D design space.

Figure 6.42 shows an example of the latchup trigger voltage as a function of external substrate resistance on a trigger voltage–$\log(R_{sx})$ plot. External substrate and well resistance boxes were used to modulate the well and substrate shunt resistances. A bipolar curve tracer was used to observe the latchup turn-on mechanism, the trigger regime, the negative resistance regime and the holding condition.

Significant 'structure' was evident in the latchup electrical measurements in the turn-on and instability regimes. As the external resistance was increased, the trigger condition was modulated and the observations in trigger and holding conditions could be dynamically recorded for a given structure. In Figure 6.42, the p+/n+ spacings are shown for the cases of p+/n+ space of 1.2, 2.4 and 3.6 μm. In these cases, the npn and pnp spacings are symmetrical and equal. Experimental latchup results show that the trigger condition is a weak function of the p+/n+ space and very strong function of the substrate resistance value. As the substrate resistance increases, the trigger value decreases [27–32].

Hence, it is clear that using minimum spacings does not significantly degrade the latchup robustness in the presence of a DT structure and a low-doped substrate wafer. As the substrate resistance increases, the trigger voltage decreases significantly with the increase in the external substrate resistance. For small resistance values from 1 to 10 Ω, there is little reduction in the latchup trigger voltage. As the resistance increases from 10 to 300 Ω, the trigger voltage rapidly decreases from 55 to 15 V. For significantly higher resistance values, the trigger voltage is asymptotic. Hence, our experimental results show that for maintaining high trigger voltage conditions, the external resistance must be kept below 100 Ω. This sets the design box for the substrate contact spacing that is required to preserve the latchup stability.

In Figure 6.43, $V_{\text{TR}}$ versus $\log(R_{sx})$ is shown as a function of the DT depth for a fixed p+/n+ space. From the plot, it is evident that the trigger condition is a strong function of DT and $\log(R_{sx})$ up to an external resistance below 300 Ω. The trigger voltage as a function of DT depth is a function of the substrate external resistance value. As the DT depth decreases, the trigger voltage decreases and each successive contour has a more gradual dependence on the external substrate resistance condition [27]. In this plot, there exist multiple solutions for a given trigger voltage condition. There is a set of points

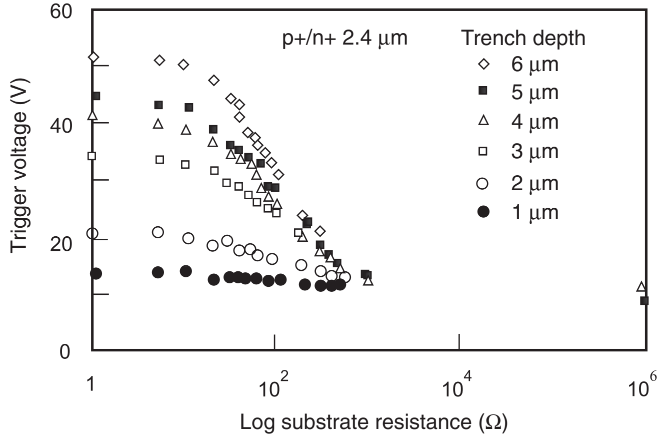



**Figure 6.43** CMOS latchup trigger voltage results versus the logarithm of the external substrate resistance as a function of DT depth (for a fixed p+/n+ spacing of 2.4 μm).

(DT depth, $R_{sx}$) that will provide a given trigger condition. Hence, re-plotting this into a new (DT depth, log($R_{sx}$)) design space, the loci of points form a contour in this space for each trigger condition.

Trigger condition loci can be created in log($R_{sx}$)–DT depth space to optimize the trigger condition for the technology (Figure 6.44). In log($R_{sx}$)–DT depth space, trigger contours can be constructed from the results in Figure 6.43. These loci of points can be used as a design curve to trade-off the DT and p− substrate contact distance for a fixed n-well to p+ space and fixed minimum p+/n+ space [27].

DT structures when bordering a n-well structure have the advantage of reducing the lateral bipolar gain of both the lateral pnp transistor and the lateral npn transistor. The differential tetrode condition of

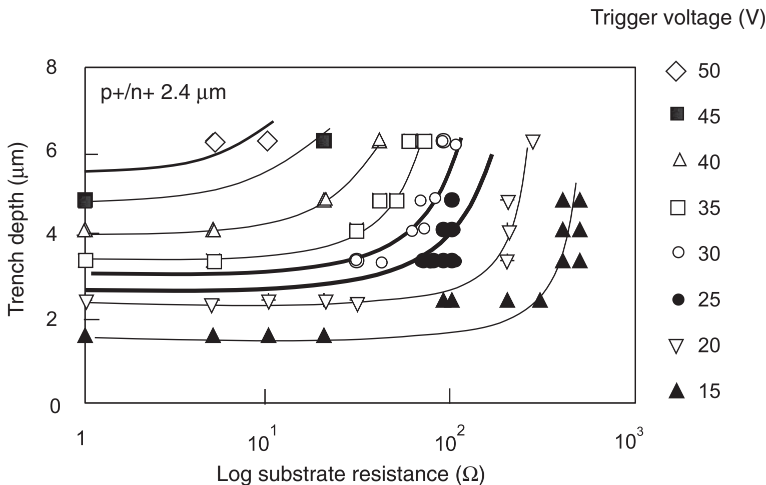



**Figure 6.44** CMOS latchup trigger voltage ($V_{TR}$) contours in DT–log($R_{SX}$) space for a fixed p+/n+ space (e.g. p+/n+ spacing of 2.4 μm).

inequality for initiation of latchup can be expressed as the sum of the generalized effective transistor gain transport factors exceeding or equaling unity,

$$\alpha^*_{\mathrm{fns}}(L_{\mathrm{DT}}) + \alpha^*_{\mathrm{fps}}(L_{\mathrm{DT}}) \geq 1,$$

where the generalized transport terms can be stated as a function of the trench depth, DT,

$$\alpha^*_{\mathrm{fns}} = \frac{\alpha_{\mathrm{fns}}(L_{\mathrm{DT}})}{1 + \frac{r_{\mathrm{en}}}{R_{\mathrm{s}}}} \qquad \alpha^*_{\mathrm{fps}} = \frac{\alpha_{\mathrm{fps}}(L_{\mathrm{DT}})}{1 + \frac{r_{\mathrm{ep}}}{R_{\mathrm{w}}}}.$$

Also the resistance term can be represented as the derivative of the emitter–base resistance with respect to the emitter current for the npn and pnp transistors, respectively,

$$r_{\mathrm{en}} = \frac{(\mathrm{d}V_{\mathrm{be}})_{\mathrm{n}}}{(\mathrm{d}I_{\mathrm{e}})_{\mathrm{n}}} \qquad r_{\mathrm{ep}} = \frac{(\mathrm{d}V_{\mathrm{be}})_{\mathrm{p}}}{(\mathrm{d}I_{\mathrm{e}})_{\mathrm{p}}}.$$

From this general inequality expression, the effective transistor gain factors are a function of the transistor gain and the shunt resistance values that establish an alternative current path. In the generalized transport term, the numerator is influenced by the transport within the transistor structure and the denominator is influenced by the substrate and the well resistance terms. Hence, the choice of substrate and well doping concentrations influence the denominator of the generalized transport terms.

In the case of DT structure, the numerators of both the lateral pnp and npn transistors are influenced by the trench depth. Hence, in the generalized alpha space, as the trench depth increases, the transport terms decrease for both transistors. Assuming that the initial state was in the unstable region of alpha space (e.g. latchup instability is satisfied and outside the SAFE regime), as the trench depth is increased, the condition follows a characteristic from outside the SAFE region to inside the triangular SAFE region. The characteristic will not be parallel to the axis of the SAFE space since both transport factors are being reduced with increasing trench depth. This is a double advantage of the DT structure in that it addresses the reduction in the current gains of both the pnp and the npn transistor concurrently (Figure 6.45).

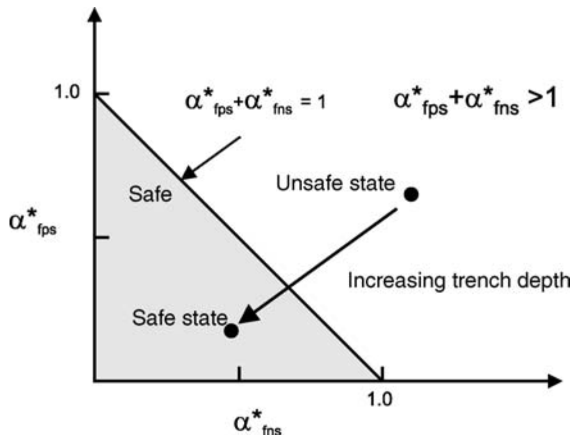



**Figure 6.45** DT characteristic in generalized 'alpha' space ($\alpha^*_{\mathrm{fps}}$–$\alpha^*_{\mathrm{fns}}$). A characteristic path is shown highlighting the path of the technology with increasing DT depth.

The utilization of the DT structure for CMOS peripheral circuits such as I/O output driver networks to isolate the p-channel MOSFET pull-up, the n-channel MOSFET pull-down, resistor elements and the ESD elements can provide a considerable latchup robustness improvement. The stand-alone guard ring structures, as discussed in Section 6.1, demonstrate the improvement in the probability of escape for the trench guard rings. The usage of DT guard rings around n-channel networks improves the latchup robustness for negative undershoot phenomenon. The usage of DT guard rings abutting the n-well structure of the p-channel pull-up transistor demonstrates latchup robustness improvement from positive overshoot events. Additionally, the isolation of the p+/n-well diode ESD element and the p-channel MOSFET from the n-channel MOSFET provides improvement in the latchup of the p-doped elements in the well with the n-doped elements in the substrate. Placement of the DT guard ring around I/O output driver cells also prevents the interaction between two I/O cells when placed adjacent to each other; this will reduce I/O to I/O latchup events. In 'array I/O', I/O elements may be adjacent to logic circuitry or gate array elements. Placement of the guard rings about an array I/O book will reduce the likeliness of intralatchup events between the I/O and adjacent circuitry.

The DT structure can also be used for bordering CMOS elements such as ESD elements and circuits. Minority carrier injection into the substrate can lead to the initiation of latchup in adjacent circuits or in conjunction with the injecting element. Voldman demonstrated the use of the DT and subcollector structures on STI-bound p+/n-well diodes. The new ESD STI-bound p+/n-well diode elements proposed demonstrate that excellent ESD results can be achieved with the presence of DT in these elements in a forward bias mode of operation [23]. In this fashion, the subcollector region and the DT region act as an isolation region. In this fashion, the DT region serves as an isolation region and the subcollector region separates the p-well region from the p−substrate. Latchup was also minimized in this structure by the formation of a 'triple-well' type structure. Using the DT structure as a guard ring, a DT-bound n-well region or a DT-bound triple-well isolation structure, the latchup robustness can be significantly improved.

## 6.6.5 Deep Trench Structure and Subcollector

The DT structure can also be integrated with a subcollector implant [28,29]. The first advantage of the subcollector is the lowering of the well shunt resistance. Lowering of the well resistance modulates the denominator of the generalized transport term for the pnp transistor improving the stability of the pnpn network. Hence, the combination of the DT and the subcollector leads to variation of three variables in the generalized tetrode relationship, leading to an improved stability. The second advantage is when using the DT structure integrated with the subcollector, there are no additional cost or density implication concerns since the DT structure limits the out-diffusion of the subcollector structure. Figure 6.46 shows the CMOS latchup structure with a DT and subcollector implant.

Figure 6.47 shows the $V_{ON}$ with and without subcollector. For a 1-μm DT depth, with or without subcollector implant, there was no improvement in the turn-on mechanism. As DT depth increased (with subcollector), the turn-on voltage, $V_{ON}$, increased as the trench penetrated through the subcollector region. This effect was first discovered experimentally by Watson and Voldman [28,29]. When the DT depth exceeded the subcollector depth, $V_{ON}$ remained constant; this experimental discovery shows that at the pre-trigger voltage condition, the 'turn-on' mechanism can be eliminated when the trench penetrates through the subcollector.

To evaluate the nature of the turn-on mechanism, the latchup structure was biased as an npn transistor between the n-well/subcollector structure, the substrate and the n-diffusion. The avalanche

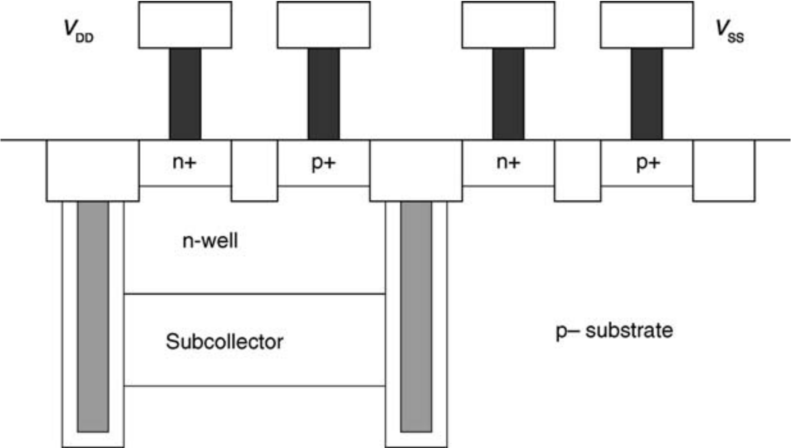



**Figure 6.46** Latchup structure with deep trench and subcollector.

breakdown voltage was extracted from the $I_C$–$V_{CE}$ npn characteristics (Figure 6.48). As the DT depth increased, the reverse $\beta_{npn}$ decreased and $V_{AV}$ increased. From these experimental results, it is evident that the turn-on voltage condition tracks with this condition [27].

Figure 6.49 shows the trigger voltage condition, $V_{TR}$ versus DT depth, with and without the subcollector implant in the n-well. A key experimental discovery was the improvement of $V_{TR}$ from 50 to 100 V. Hence, the combined effect of the trench structure and subcollector allows for a significant latchup robustness improvement by an increase in the trigger voltage of the technology [28,29].

DT structures have the advantage of reducing the lateral bipolar current gain of both the pnp and npn transistors. The addition of the subcollector leads to the elimination of the sidewall trench mechanism, lowering of $\beta_{pnp}$ and lowering of internal well shunt resistance. Lowering of the well resistance modulates the denominator of the generalized transport term for the pnp transistor,

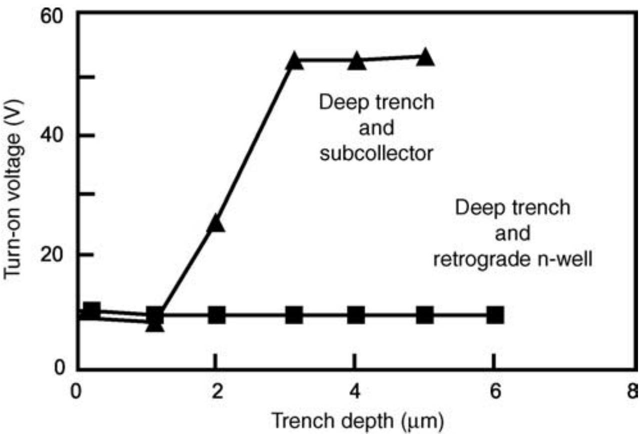



**Figure 6.47** Latchup 'turn-on' voltage versus trench depth for the case of with and without subcollector implant.

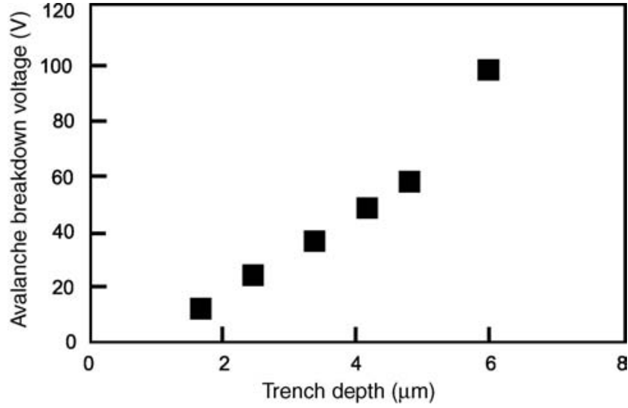



**Figure 6.48** Lateral parasitic npn bipolar avalanche breakdown voltage versus trench depth (in a pnpn structure with deep trench and subcollector).

improving the stability of the pnpn network. Hence, the combination of the DT and the subcollector leads to variation of three variables in the generalized tetrode relationship, leading to an improved stability. Another advantage is that there are no additional cost or density implications of using the DT structure integrated with the subcollector, since the DT structure limits the subcollector out-diffusion.

From the SAFE condition, as the trench depth increases, the trajectory along the vertical line moves away from the east corner of $\alpha^*_{\mathrm{fns}} = 1$ as $\alpha_{\mathrm{fns}}$ decreases until it intersects the hypotenuse. In the case of the addition of the subcollector, the trajectory moves away from the northern most corner of the SAFE triangular region to a lower $\alpha^*_{\mathrm{fps}}$ away from the condition of $\alpha^*_{\mathrm{fps}} = 1$. As the trench depth increases and the subcollector is added, the trajectories for given $R_{sx}$ and $R_{\mathrm{well}}$ are changed within the SAFE regime widening the possible resistance values in resistance space (e.g. $\log(R_{sx})$ versus $\log(R_{\mathrm{w}})$).

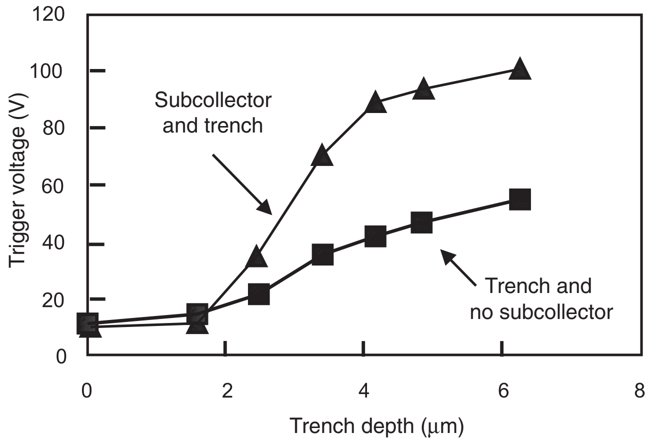



**Figure 6.49** Latchup pnpn trigger voltage versus trench depth (for the cases of with and without heavily doped subcollector).

## 6.6.6 Electrically Connected Deep Trench Structure

In the evaluation and comparison of a polysilicon-filled DT isolation and TI it was noted that the latchup robustness of the TI structure for a given depth was higher than anticipated (based on the projection from the deep trench structure). From the DT measurements, it would be anticipated that for a 2 μm trench isolation structure, the trigger voltages would increase to 20 or 30 V [22].

One of the primary reasons of the paradox is the fact that in the case of DT structures, the polysilicon-doped region is 'floating'. As the n-well region is biased to high voltages, the polysilicon region within the DT capacitively couples to a positive voltage; a capacitor divider is formed between the well-to-polysilicon DT fill region and the polysilicon DT fill region to the substrate (e.g. p-well) [32]. Assuming the trench polysilicon region is a capacitor node, it is clear that the potential of the physical node will be between the voltage of the n-well region and the substrate during CMOS latchup testing. Figure 6.50 shows the DT structure highlighting the capacitor elements of the DT structure [32].

During latchup testing, the voltage state of the DT polysilicon region is associated with the capacitor divider formed between the n-well to trench capacitor, $C_{\mathrm{NW-TR}}$, and the trench capacitor-to-substrate capacitor, $C_{\mathrm{TR}-SX}$, where

$$V_{\mathrm{TR}} = \frac{V_{\mathrm{NW}} - V_{SX}}{\left[1 + \dfrac{C_{\mathrm{TR-SX}}}{C_{\mathrm{NW-TR}}}\right]}.$$

Assuming an equal bias condition, as the trench region is raised to voltage $V^*$, the polysilicon trench region will be biased at a potential between the value $V^*$ to the p-well/p− substrate potential. The voltage potential is a function of the size of the capacitor formed between the n-well and trench sidewall, $C_{\mathrm{TR-NW}}$, and that formed between the trench and substrate capacitor, $C_{\mathrm{TR}-SX}$. Notice that the

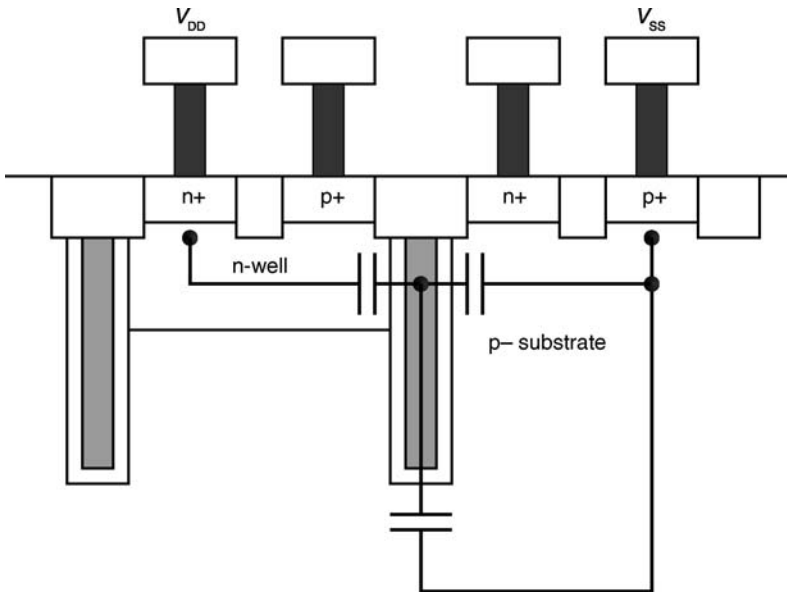



**Figure 6.50** Latchup pnpn test structure highlighting the capacitance components of the deep trench structure.

area of the trench in the substrate region, $A_{TR-SX}$, is larger than the area of the trench in the n-well region, $A_{TR-NW}$.

As the DT depth is increased, the capacitance voltage divider ratio also changed with a larger coupling to the substrate node. Additionally, as the n-well doping concentration increases, the size of the n-well to trench capacitor increases (e.g. adding the subcollector implant). Additionally, adding the subcollector region also increases the area of the sidewall capacitor element. Also notice that as the doping concentration on the side of the n-well increases, the threshold voltage of the sidewall vertical p-channel MOSFET increases.

In all the earlier DT studies, the polysilicon-filled DT structure is capped with a STI; as a result, the state of the polysilicon-filled region is floating and unbiased. Hence, in the utilization of the STI-capped polysilicon-filled DT structure, the polysilicon-filled region would capacitively couple between the n-well and the substrate voltage. A first capacitor is formed between the n-well tub and polysilicon-filled DT trench region ($C_{NW-TR}$), and a second capacitor is formed between the polysilicon-filled DT and the p-well/p− substrate region ($C_{TR-SX}$).

A new structure was constructed that allowed contaction of the polysilicon DT structure; a series of masks were modified to allow the contact of the deep trench. An opening was formed in the STI mask. The STI mask was opened to allow the physical connection to the polysilicon trench fill. The STI etch formed a trough over the polysilicon and sidewall of the trench structure. The STI trough was filled and polished, exposing the polysilicon region of the DT structure to the surface. This step was followed by the UHV/CVD deposition of the silicon germanium (SiGe) epitaxial film, depositing the SiGe base film on the DT structure. In this fashion, the SiGe film is utilized as an electrical connecting structure. This is followed by silicidation of the SiGe film, interlevel dielectric films and the tungsten contact. Figure 6.51 shows a cross section of the pnpn structure with a contact placed in the opening of the STI to contact the polysilicon DT ring (e.g. surrounding the n-well region).

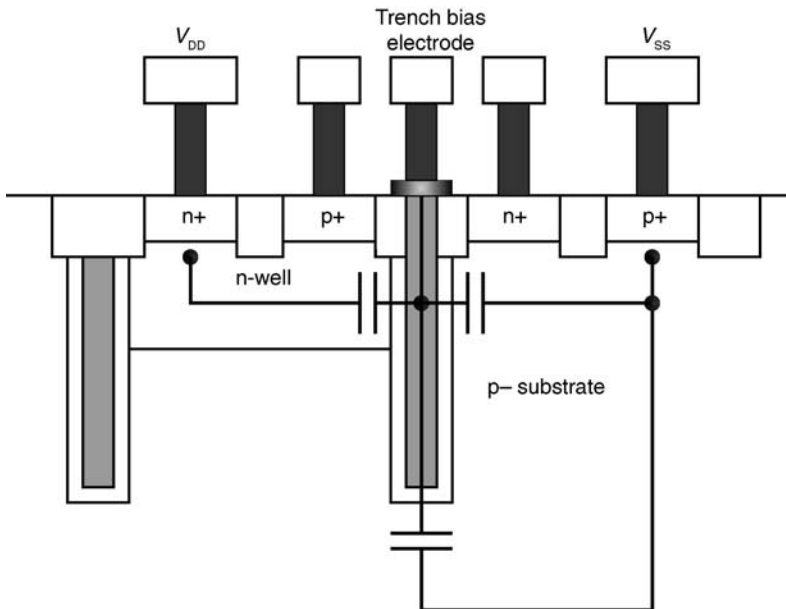



**Figure 6.51** Latchup pnpn test structure with contacted trench structure with the capacitance components highlighted.

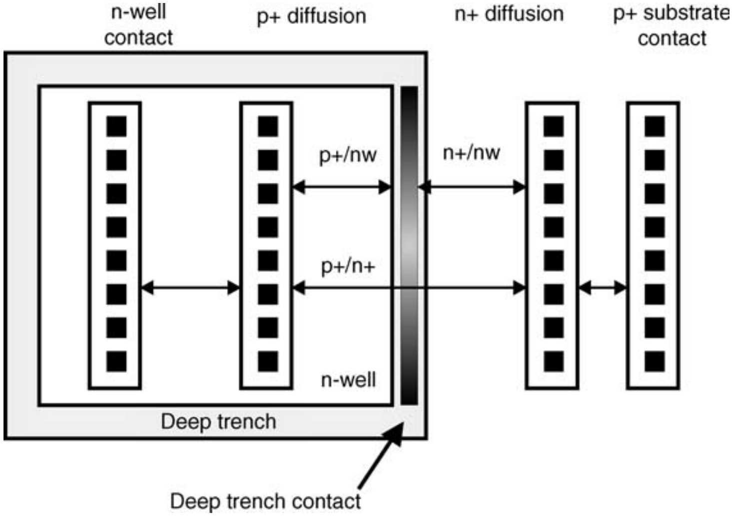



**Figure 6.52** Latchup pnpn test structure with contacted trench structure.

Figure 6.52 shows the electrical contact with the electrical connection within the p+/n+ space. In this structure, five electrical contacts were used, adding one more additional signal pad to the DT bias contact region [32].

Experimental work was completed on a Tektronix 576 bipolar curve tracer to be able to observe the latchup characteristics and modify the states of the experiment. In a first condition, the DT structure was biased to a ground potential (e.g. $V_{SS}$). For the case of the DT structure biased at a ground potential, the maximum voltage condition occurs across the n-well to trench capacitor (e.g. and no voltage drop between the trench and substrate potential). In this state, the latchup trigger voltage was 40 V. Figure 6.50 shows a latchup *I–V* characteristics highlighting the trench grounded state.

In a second condition, the DT was electrically connected to the n-well as the n-well was ramped; in this case, the latchup trigger voltage, $V_{TR}$, increased to 80 V (Figure 6.53).

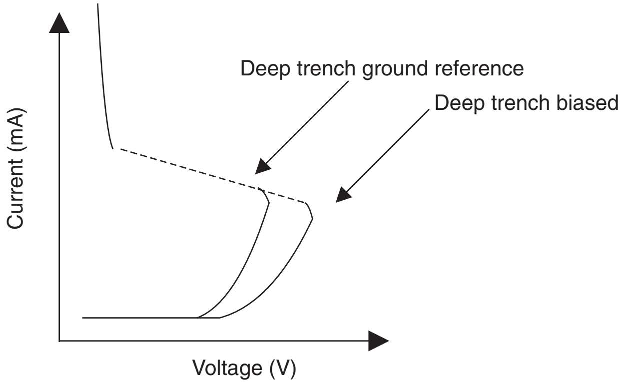



**Figure 6.53** Latchup pnpn *I–V* characteristics with (a) trench-biased ramped with n-well and (b) trench grounded at substrate.

**Table 6.7** Latchup pnpn metrics as a function of the DT bias state.

| Trench state | $V_{ON}$ (V) | $V_{TR_1}$ (V) | $I_{TR_1}$ (mA) | $V_H$ (V) | $I_H$ (mA) |
|---|---|---|---|---|---|
| Trench grounded to $V_{SS}$ | 12 | 40 | 8 | 25 | 45 |
| Floating trench | 12 | 58 | 8 | 22 | 38 |
| Trench biased at nw | 12 | 80 | 3.5 | N/A | N/A |

In a third condition, the DT structure was electrically disconnected and allowed to 'float' during the sweep characteristics of the bipolar curve tracer. Table 6.7 summarizes the latchup metrics results with the electrically contacted deep trench structure as a function of the trench bias state. In all cases, the turn-on voltage was the same. But, $V_{TR}$ and the holding conditions were modified by the trench bias state [32].

In a second study, accidentally, an additional condition was discovered. As the DT structure was biased with the n-well, the CMOS latchup structure achieved the high trigger condition. In anticipation of performing the low-state test, the electrical connection to the n-well was removed. While the electrical connection was removed, the SCR did not undergo a low-state transition, but remained in the high $V_{TR}$ state as if biased. It was determined that because of charge storage in the DT structure, the trigger state remained at the higher trigger state; this state remained without decay. The charge remained in the polysilicon DT capacitor structure and modulated the latchup state. As a second step, the electrical connection was then manually shorted to ground (e.g. the oscilloscope ground electrode) and the latchup characteristic switched to a low trigger state.

In a third study, the DT electrode was connected to a step current generator (e.g. base drive of a bipolar curve tracer). In this condition, while the voltage was applied, a series of *I–V* characteristics were generated between the 40 and 80 V states. In this condition, the *I-V* characteristic was a function of the step generator number of steps and current injected into the DT structure. In this condition, the latchup *I–V* characteristic was a function of the step generator number of steps and current injected into the DT structure. Hence, by charge pumping the current into the deep trench structure, intermediate trigger states were established between the highest and lowest states. In a fourth study, adding current to the trench structure varied the voltage trigger to 58 V at 10 mA, 52 V at 1 mA and, for negative current, 38 V at $-1$ mA and 32 V at $-10$ mA [32].

Experimental results show that in the case of floating polysilicon DT structure, the bias state and stored charged state of the polysilicon region modulate the latchup trigger voltages. Additionally, this has significant implications on charging phenomenon and on the high-voltage, power electronics, automotive, and space applications.

## 6.7 TRIPLE-WELL AND BiCMOS PROCESSES INTEGRATION

As practiced today, circuit designers prefer to re-map dual-well structures into triple-well implementations without changing the on-chip design, ground rules or p+/n+ spacing rules. As a result, 'triple well' is being practiced not as isolated regions but as 'merged triple well' where the n-well and associated isolating buried layers are integrated. With the introduction of merged triple well in advanced CMOS technologies, the vertical bipolar transistor is enhanced well above the dual-well CMOS structures. In advanced BiCMOS technologies, for merged triple-well CMOS, there are many options [33,34]:

- merged triple well with deep trench;
- merged triple well with deep trench, with triple-well implant only under p-well region;
- merged triple well with deep trench, with triple-well implant under both n-well and p-well;

- merged triple well with deep trench, with triple-well implant under both n-well and p-well, with additional BiCMOS implants (e.g. subcollector).

### 6.7.1 Triple Well – Deep Trench and Triple Well Integration

This experimental work is completed in a high resistivity 50 Ω cm p− substrate wafer in a 0.13-μm BiCMOS technology to highlight the enhancement of the npn bipolar current gain; these results will be followed by the first measurements demonstrating integration of DT structures and subcollector implants in a triple-well environment [34].

In order to reduce the lateral $\beta_{npn}$, the DT structures can be used to separate the triple-well buried layer from the n-well region. Because of scattering phenomenon influencing MOSFET threshold implants, the spacing between the implanted buried layer edge and the n-channel MOSFET and the p-channel MOSFET must be large to avoid p-channel MOSFET and n-channel MOSFET threshold modulations. Hence, in some technologies, adequate spacing must be established between the edge of the buried layer implant, the n-channel MOSFET and the p-channel MOSFET; this provides an opportunity to place deep trench structures between the p−and n-channel MOSFETs in merged triple-well technology (Figure 6.54).

To avoid the issue of the buried layer implant edges, some circuit designers would like to convert the technology from dual-well CMOS to triple well using a blanket n− buried layer under all structures [34]. Using a 'blanket' buried layer (e.g. unmasked), this can impact devices and circuit elements. Additionally, using a continuous n− buried layer under the p+/n+ spacing allows for avoidance of mask edge scattering phenomenon and provides improved latchup. Figure 6.55 shows the triple-well structure with DT isolation and a continuous n-buried layer under both the n-well and the p-well region.

In this structure, the role of the n-buried layer under the n-well provides latchup advantages:

- Low resistance shunt for the pnp transistor.
- Lower pnp bipolar gain due to increase base width and larger Gummel number (e.g. total implant dose integrated over the base region). The placement of the n-layer under the

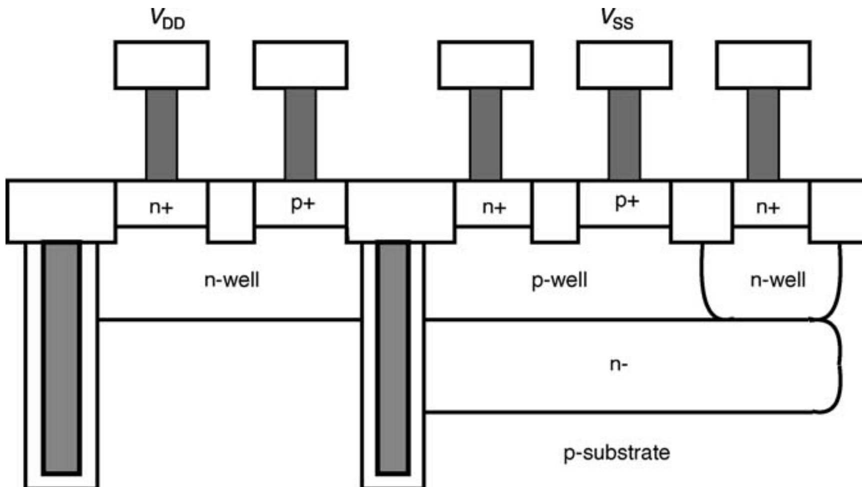



**Figure 6.54** Triple-well structure with DT and triple well under p-well region only.

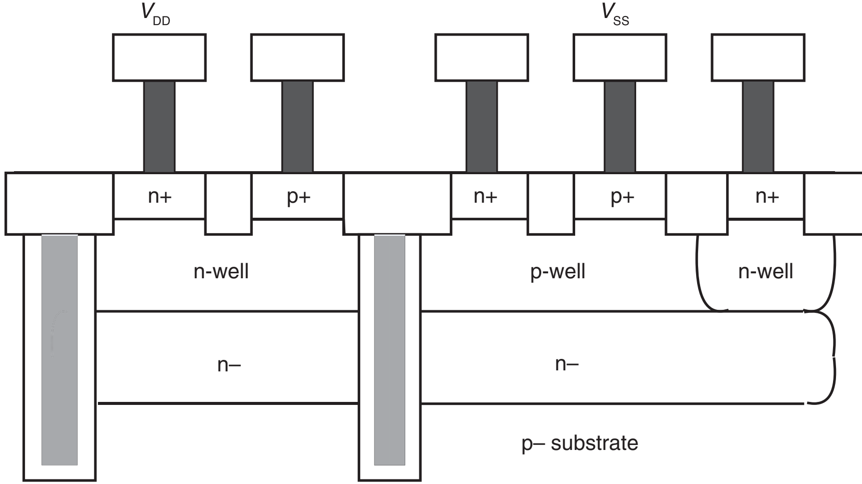



**Figure 6.55** Triple-well structure with DT and continuous triple-well implant.

n-well prevents scattering mask phenomenon and improves the latchup robustness of the structure.

Figure 6.56 shows an example of the parasitic bipolar current gain for three cases: (1) the triple well under the p-well region (e.g. p− epi) only, (2) a continuous n-buried layer under the p− epi and n-well and (3) the n-buried layer and the DT structure. From the experimental results, with a continuous n-buried layer (e.g. case II), the npn bipolar current gain is the highest for smallest spacings; the second highest is the case of the n-buried layer only under the p-well. With the presence of the DT structure, the lateral npn current gain is significantly reduced for small spacings [34].

## 6.7.2 Triple Well – Deep Trench and Subcollector

With the introduction of a subcollector, the pnp bipolar current gain can be reduced due to the added dopant to the vertical parasitic pnp base width (Figure 6.57). Typically, the bipolar subcollector doping concentration is significantly higher than the n− buried layer.

Figure 6.58 shows the results of the three cases of merged triple well, where the subcollector is added in the n-well region in the third case. In the first case, the scaling of the npn base shows significant increase in the npn bipolar current gain. In the case of continuous triple well, the npn current gain also increases to very levels for small p+/n+ spacings. With continuous triple well implant, the high level of npn current gain is evident. With the introduction of the DT structure and the subcollector implant, the npn current gain is significantly lowered [34].

Focusing on the parasitic pnp bipolar current gain, the dual-well CMOS with DT can be compared to the triple well (Figure 6.59). An interesting issue is the pnp bipolar current gain is higher than with dual-well CMOS with DT compared to the merged triple-well structure. These two converge as the p+/n+ spacing is reduced. But, the dual-well CMOS with DT and subcollector is superior to the merged triple-well only structure [34].

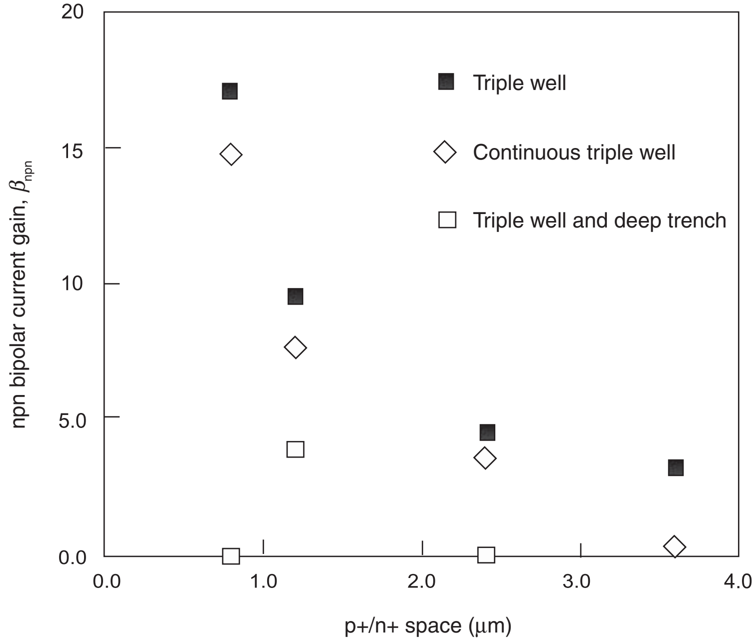



**Figure 6.56** Parasitic npn bipolar current gain with (a) triple well, (b) triple well with continuous buried layer and (c) triple well with DT.

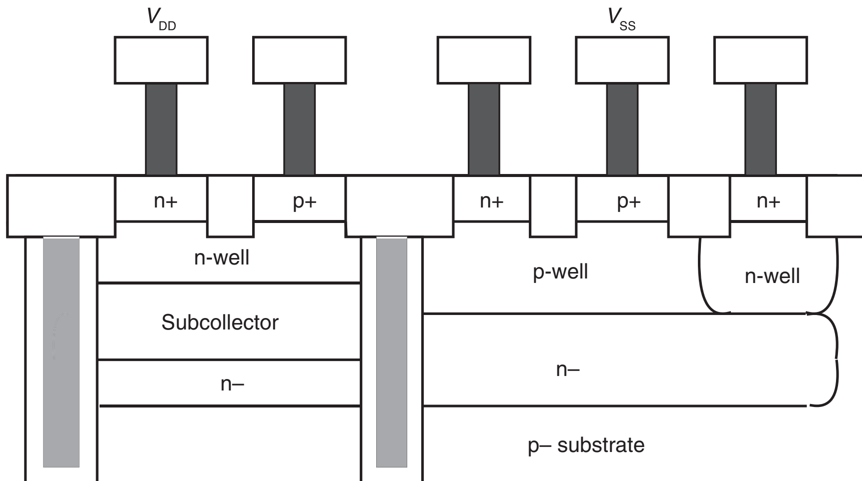



**Figure 6.57** Continuous triple well, deep trench and subcollector.

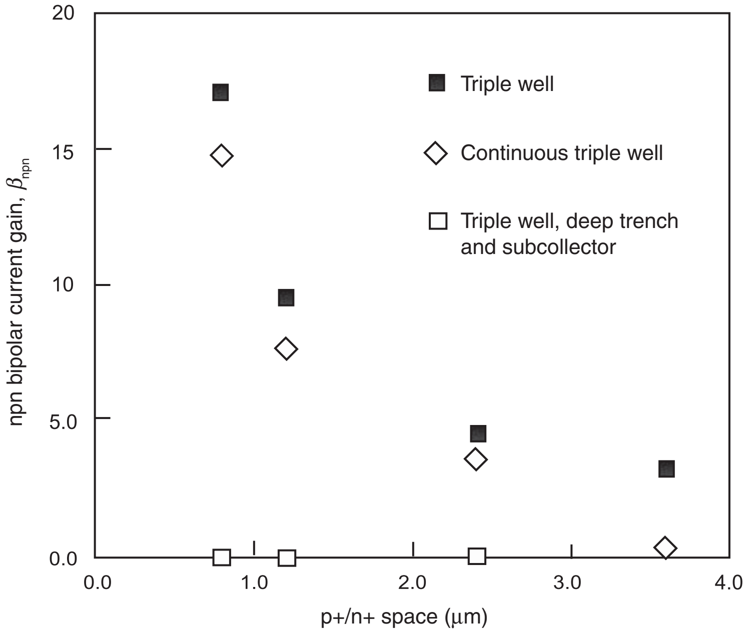



**Figure 6.58** Parasitic npn bipolar current gain with (a) triple well, (b) triple well with continuous buried layer and (c) triple well with DT and subcollector.

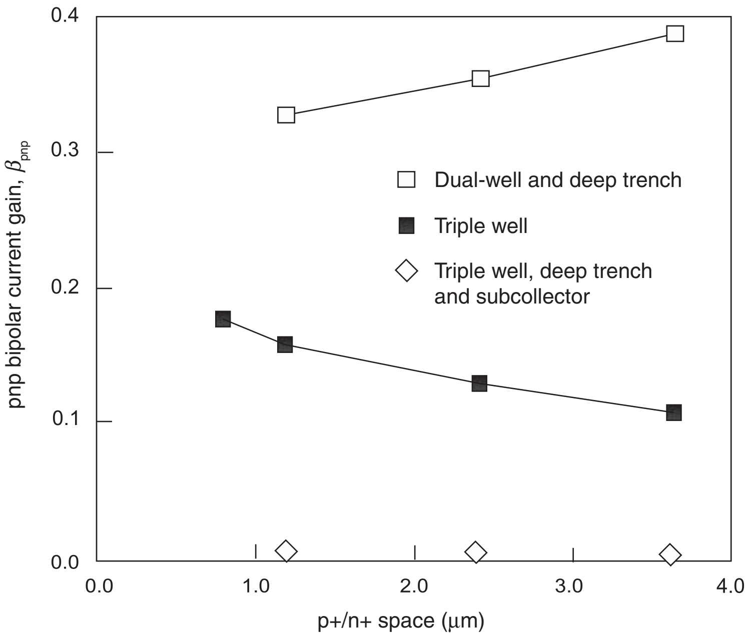



**Figure 6.59** Parasitic pnp bipolar current gain with (a) triple well, (b) dual well and deep trench and (c) dual well, DT and subcollector.

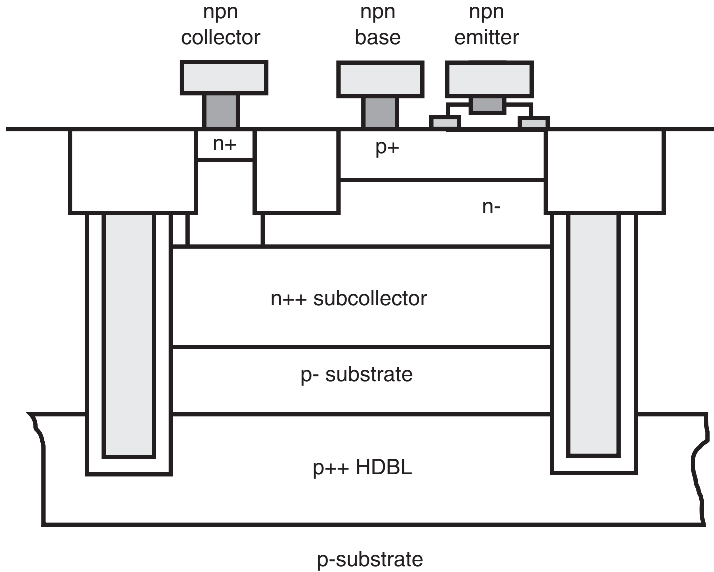



**Figure 6.60** Integration of a HDBL structure into a BiCMOS technology.

## 6.8 HEAVILY DOPED BURIED LAYER IMPLANT AND BiCMOS TECHNOLOGY

High dose buried layers (HDBL) can also be integrated into CMOS [35–39] or BiCMOS technology [40]. Integration of HDBL structures into a BiCMOS process requires the ability to place the implant deep below the subcollector through the subcollector or selectively implant regions outside the subcollector region [40]. With epitaxially formed subcollectors, the subcollectors are implanted prior to the epitaxial deposition. With the long hot process cycle and high doping concentrations used in traditional high-performance subcollectors, the depth of the subcollector region can be significantly greater than that of n-well regions used in CMOS base processes. As a result, the implant energy of the heavily doped buried layer must be above 1 MeV and extend to 2.0 MeV or higher [40] (Figure 6.60).

HDBL implants were performed in p− substrate wafers prior to processing of the silicon germanium process steps. The implant energy and dose were chosen to provide a deep implant that would minimize impact on the semiconductor devices and yet provide a low-resistance shunt for latchup improvement. The HDBL implants were placed in a DT-defined SiGe HBT technology. The placement of the HDBL implant can influence the collector-to-emitter junction that can influence functional characteristics. Optimization of the HDBL implant may be critical for some RF applications. Table 6.8 shows the collector-to-substrate breakdown voltage of the SiGe HBT device as a function of the HDBL implant dose and energy. Results show that the high-energy implant has the least affect on the breakdown voltage [40].

With the integration of HDBL implants into BiCMOS technology, the integration may require masking to avoid impact to the bipolar transistor. Using a blanket HDBL, the trench structure can be used to segment chip sectors for noise isolation and guard ring structures [40].

**Table 6.8** DT bound collector-to-substrate breakdown voltage as a function of HDBL energy and dose.

| Split | Implant energy (MeV) | Implant dose (no. #cm$^{-2}$) | $V_{BR}$(V) |
|---|---|---|---|
| No HDBL | N/A | N/A | 20.7 |
| HDBL | 2.0 | $5 \times 10^{14}$ | 19.0 |
| | 2.0 | $1 \times 10^{15}$ | 16.0 |
| | 1.6 | $5 \times 10^{14}$ | 11.0 |
| | 1.6 | $1 \times 10^{15}$ | 9.0 |
| | 1.2 | $5 \times 10^{14}$ | 7.0 |

## 6.9 SUMMARY AND CLOSING COMMENTS

This chapter focused on semiconductor process features of a BiCMOS process for improving CMOS latchup in a mixed signal chip environment. In this chapter, the topics systematically evaluated different process features and then demonstrated the cosynthesis of these features. In this chapter, at the same time, the analysis of the features taught new CMOS design practices; new characterization methods and design plots were highlighted as a way to analyze the CMOS latchup results. These CMOS design methods can be applied to other technology features. In this chapter, it is also noted that although these features today are integrated in BiCMOS technology, all of these features can be integrated into mainstream CMOS technology that was discussed in Chapter 5.

Chapter 7 will discuss circuit solutions to address CMOS Latchup. First, the discussion will address intradevice, intracircuit, intercircuit and interfunction latchup in integrated chip design. Chip-level and system-level latchup issues and solutions will be discussed. As part of the CMOS latchup design discipline, circuit solutions exist to address power-up to circuits that determine the onset of CMOS latchup. The chapter will discuss CMOS, BiCMOS to smart power issues. A key area of interest is sequence-independent circuits and networks that allow decoupling from power supply rails. The chapter will close with examples of ‘anti-overshoot and anti-undershoot’ circuits, such as active clamp networks, to active guard ring networks. In these examples, circuit concepts will be demonstrated that are utilized in the CMOS latchup circuit design discipline.

## PROBLEMS

1. Draw a latchup pnpn test structure with a subcollector implant where the subcollector is under the n-well region. Draw a second latchup pnpn test structure with a subcollector under both the n-well and p-well, where the subcollector does not isolate the p-well from the p− substrate. Draw a third latchup pnpn test structure where an n-well encloses the p-well (isolates the p-well from the p− substrate).

2. As in Problem 1, create a circuit schematic of the pnpn structure for each of the three cases. Label the resistance terms and identify the resistances in the test structure cross section and in the circuit schematic.

3. A trench isolation is formed after the STI isolation, leading to a ‘dual depth’ isolation structure. What is the depth required for maintaining latchup robustness?

4. Draw a pnpn test structure with shallow trench isolation and the TI structure. Assume three cases: (1) the trench isolation is shallower than the n-well region, (2) the trench isolation is equal to the n-well depth and (3) the trench isolation is deeper than the n-well depth. Explain the lateral and vertical parasitic transistors in the three cases.

5. Derive a parasitic transistor model for trench isolation, where (1) the trench isolation is shallower than the n-well region, (2) the trench isolation is equal to the n-well depth and (3) the trench isolation is deeper than the n-well depth.

6. Draw a pnpn test structure with shallow trench isolation and the DT structure, where the DT is polysilicon-filled and borders the n-well. Assume that the wafer is a p−/p++ substrate wafer. Assume the following cases: (1) the deep trench is shallower than the n-well depth, (2) the deep trench is equal to the n-well depth, and (3) the deep trench is deeper than the n-well depth but shallower than the p−/p++ interface, (4) the deep trench is deeper than the n-well depth but abuts the p−/p++ interface and (5) the deep trench is deeper than the n-well and extends beyond the p−/p++ interface. Explain the lateral and vertical parasitic transistors in the three cases.

7. For Problem 6, create a lateral bipolar model for all the above cases.

8. A series of deep trench rings are placed around an n+ injection source. Show the cases of the minority carrier transport for the following structure where the listed guard rings are in sequence:

   a. p+ diffusion, n-well;

   b. p+ diffusion, n-well, deep trench;

   c. p+ diffusion, n-well, deep trench, n-well;

   d. p+ diffusion, n-well, deep trench, n-well, deep trench;

9. Explain the voltage conditions of the polysilicon region in a deep trench structure used as a border on the edge of an n-well region in a pnpn. What is the potential of the polysilicon?

10. Show the test structure for integration of deep trench, subcollector and triple-well implant under both the n-well and p-well regions. Describe the lateral and vertical transports for the parasitic pnp and npn structures.

## REFERENCES


1. S. Voldman. Latchup – it's back! *Electrostatic Discharge* (*ESD*) *Association Threshold Magazine*, **19**(5), 2003. p. 7–8.
2. S. Voldman. Latchup: Part II – Why is it still here? *Electrostatic Discharge* (*ESD*) *Association Threshold Magazine*, **19**(6), 2003. pp. 9–10.
3. W. Morris. CMOS latchup. *Proceedings of the International Reliability Physics Symposium* (*IRPS*), May 2003. p. 86–92.
4. G. Boselli, V. Reddy and C. Duvvury. Latch-up in 65 nm CMOS technology: A scaling perspective. *Proceeding of the International Reliability Physics Symposium* (*IRPS*), 2004. p. 137–144.
5. S. Voldman, E. Gebreselasie, X.F. Liu, D. Coolbaugh and A. Joseph. The influence of high resistivity substrates on CMOS latchup robustness. *Proceedings of the Electrical Overstress/Electrostatic Discharge* (*EOS/ESD*) *Symposium*, 2005. p. 90–99.
6. S. Voldman. A review of CMOS latchup and electrostatic discharge (ESD) in bipolar complimentary MOSFET (BiCMOS) silicon germanium technologies: Part I – ESD. *Microelectronics and Reliability*, **45**, 2005, 323–340.
7. S. Voldman. A review of CMOS latchup and electrostatic discharge (ESD) in bipolar complimentary MOSFET (BiCMOS) silicon germanium technologies: Part II – Latchup. *Microelectronics and Reliability*, **45**, 2005, 437–455.
8. S. Voldman. Latchup and the domino effect. *Proceedings of the International Reliability Physics Symposium* (*IRPS*), 2005. p. 145–156.

9. S. Voldman. Cable discharge event and CMOS latchup. *Proceedings of the Taiwan Electrostatic Discharge Conference* (*T-ESDC*), 2005. p. 23–28.
10. P. Cottrell, S. Warley, S. Voldman, W. Leipold and C. Long. N-well design for trench DRAM arrays. *International Electron Device Meeting* (*IEDM*) *Technical Digest*; December 1988. p. 584–587.
11. S. Voldman, M. Marceau, A. Baker, E. Adler, S. Geissler, J. Slinkman, J. Johnson and M. Paggi. Retrograde well and epitaxial thickness optimization for shallow- and deep-trench collar merged isolation and node trench (MINT) SPT cell and CMOS logic technology. *International Electron Device Meeting* (*IEDM*) *Technical Digest*; 1992. p. 811–815.
12. S. Voldman. Optimization of MeV retrograde wells for advanced logic and microprocessor/PowerPC and electrostatic discharge. *Smart and Economic Device and Process Designs for ULSI Using MeV Implant Technology: Semicon West GENUS Seminar*, San Francisco, 1994.
13. S. Voldman. MeV implants boost device design. *IEEE Circuits and Devices*; **11**(6), 1995, 8–16.
14. S. Voldman. ESD and latchup in advanced technologies, *Tutorial Notes of the International Reliability Physics Symposium* (*IRPS*), April 25, 2004.
15. S. Voldman. Latchup physics and design, *ESD Tutorials of the Electrical Overstress/Electrostatic Discharge* (*EOS/ESD*) *Symposium*, September 20, 2004.
16. S. Voldman. CMOS latchup, *Tutorial Notes of the International Reliability Physics Symposium* (*IRPS*), April 17, 2005.
17. S. Voldman and E.G. Gebreselasie. The influence of implanted sub-collector on CMOS latchup. *Proceedings of the Electrical Overstress/Electrostatic Discharge* (*EOS/ESD*) *Symposium*, 2005. p. 108–117.
18. R.D. Rung, H. Momose and Y. Nagabuko. Deep trench isolated CMOS devices. *International Electron Device Meeting* (*IEDM*) *Technical Digest*, 1982. p. 237–240.
19. T. Yamaguchi. S. Morimoto, G.H. Kawamoto, H.K. Park and G.C. Eiden. High-speed latchup-free 0.5-μm channel CMOS using self-aligned $TiSi_2$ and deep trench isolation technologies. *International Electron Device Meeting* (*IEDM*) *Technical Digest*; 1983. p. 522–525.
20. M. Bohr. Isolation structure formation for semiconductor circuit fabrication. U.S. Patent No. 5,536,675, July 16, 1996.
21. S. Voldman, E. G. Gebreselasie, L. W. Lanzerotti, N. B. Feilchenfeld, S. A. St. Onge, A. Joseph and J. Dunn. The influence of silicon dioxide-filled trench isolation (TI) structure and implanted sub-collector on latchup robustness. *Proceedings of the International Reliability Physics Symposium* (*IRPS*), 2005. p. 112–120.
22. S. Voldman, E.G. Gebreselasie and A. Watson. Comparison of CMOS latchup with trench isolation (TI) and polysilicon-filled deep trench (DT) isolation structures for CMOS and BiCMOS technology. *Proceedings of the Taiwan Electrostatic Discharge Conference* (*T-ESDC*), 2005. p. 23–28.
23. S. Voldman. The effect of deep trench isolation, trench isolation, and sub-collector on the electrostatic discharge (ESD) robustness of radio frequency (RF) ESD STI-bound p+/n-well diodes in a BiCMOS silicon germanium technology. *Proceedings of the Electrical Overstress/Electrostatic Discharge* (*EOS/ESD*) *Symposium*, 2003. p. 214–223.
24. A. Watson, S. Voldman and T. Larsen. Deep trench guard ring structures and evaluation of the probability of minority carrier escape for ESD and latchup in advanced BiCMOS SiGe technology, *Proceedings of the Taiwan Electrostatic Discharge Conference*, National Chiao-Tung University, Hsinchu City, Taiwan, November 12–13, 2003. p. 97–103.
25. S. Voldman, C.N. Perez and A. Watson. Guard rings: theory, experimental quantification, and design. *Proceedings of the Electrical Overstress/Electrostatic Discharge* (*EOS/ESD*) *Symposium*, 2005. p. 100–107.
26. S. Voldman, C.N. Perez and A. Watson. Guard rings: structures, design methodology, integration, experimental results and analysis for RF CMOS and RF mixed signal BiCMOS silicon germanium technology. *Journal of Electrostatics*, **64**, 2006, 730–743.
27. S. Voldman and A. Watson. The influence of deep trench and substrate resistance on the latchup robustness in a BiCMOS silicon germanium technology. *Proceedings of the International Reliability Physics Symposium* (*IRPS*), April 25–27, 2004. p. 135–142.
28. A. Watson and S. Voldman. The effect of deep trench and sub-collector on the latchup robustness in BiCMOS silicon germanium technology. *Proceedings of the Bipolar Circuit Technology Meeting* (*BCTM*), Montreal, Canada, September 12–14, 2004. p. 172–175.

29. S. Voldman and A. Watson. The influence of polysilicon-filled deep trench and sub-collector implants on latchup robustness in RF CMOS and BiCMOS SiGe technology. *Proceedings of the Taiwan Electrostatic Discharge Conference* (*T-ESDC*), 2004. p. 15–19.
30. S. Voldman. ESD and latchup in advanced technologies, *Tutorial Notes of the International Reliability Physics Symposium* (*IRPS*), April 25, 2004.
31. S. Voldman. Latchup physics and design, *ESD Tutorials of the Electrical Overstress/Electrostatic Discharge* (*EOS/ESD*) *Symposium*, September 20, 2004.
32. S. Voldman. The influence of a novel contacted poly-silicon filled deep trench (DT) biased structure and its voltage state on CMOS latchup. *Proceedings of the International Reliability Physics Symposium* (*IRPS*), 2006. p. 151–158.
33. S. Voldman, E.G. Gebreselasie, M. Zierak, D. Hershberger and D.S. Collins. Latchup in merged triple well structure. *Proceedings of the International Reliability Physics Symposium* (*IRPS*), 2005. p. 129–136.
34. S. Voldman and E.G. Gebreselasie. The influence of merged triple well, deep trench and subcollector on CMOS latchup. *Proceedings of the Taiwan Electrostatic Discharge Conference* (*T-ESDC*), 2006. p. 49–52.
35. H.Y. Lin and C.H. Ting. Improvements of CMOS latchup using a high energy buried layer. *Nuclear Instrumentation Methods Physics Review*, **B38/39**, 1989, 960–964.
36. K.C. Leong, P.C. Liu, W. Morris and L. Rubin. Superior latchup resistance of high dose energy implanted p+ buried layers. *Proceedings of the XII International Conference on Ion Implantation Technology*, Kyoto, Japan, 1998. p. 99–108.
37. T. Kuroi, S. Komori, H. Miyatake and K. Tsukamoto. Self gettering and proximity gettering for buried layer formation by MeV ion implantation. *International Electron Device Meeting* (*IEDM*) *Technical Digest*, December 1990. p. 261–264.
38. K.K. Bourdelle, D.J. Eaglesham, D.C. Jacobson and J. Poate. The effect of as-implanted damage on the microstructure of threading dislocations in MeV implanted silicon. *Journal of Applied Physics*, **86**, 1999, 1221–1225.
39. K.K Bourdelle, Y. Chen, R. Ashton, L. Rubin, A. Agarwal and W. Morris. Epi-replacement in CMOS technology by high dose, high energy boron implantation into Cz substrates. *Proceedings of the International Conference on Ion Implantation*, 2000. p. 312–315.
40. S. Voldman, L. Lanzerotti, W. Morris and L. Rubin. The influence of heavily doped buried layer implants on electrostatic discharge (ESD), latchup, and a silicon germanium heterojunction bipolar transistor in a BiCMOS SiGe technology. *Proceeding of the International Reliability Physics Symposium* (*IRPS*), 2004. p. 143–151.

# 7 CMOS Latchup – Circuits

Latchup is highly influenced by the types of circuits and how these circuits interact during design integration and synthesis. In this chapter, we will focus on the types of circuits that cause latchup, the integration of these networks and the interactions that can occur. We will close the discussion by providing examples of circuit solutions, novel circuits, active clamps and active guard rings to address latchup.

## 7.1 TABLE OF CIRCUIT INTERACTIONS

Latchup occurs in semiconductor chips as a result of interaction within a circuit, between circuits and between chip subfunctions. Table 7.1 shows examples of CMOS latchup events within a given circuit. As a first example, latchup occurred between an ESD 'dual diode' (DD) network and an OCD p-channel MOSFET pull-up device for a negative undershoot phenomenon; improper sequencing leads to a latchup issue. In this case, $V_{DD}$ was established on power-up, but due to a poor ground connection, $V_{SS}$ ground was not well defined. Negative signals were applied to the input signal prior to the establishment of the ground potential. In this case, an n-well to substrate diode forward biased, and the current flowed to the p-channel MOSFET of the OCD network. This case study is interesting in that all the proper solutions and design layout issues were addressed, yet latchup occurred due to improper power sequencing on the system level.

In the second case study, latchup occurred in a *RC*-triggered MOSFET ESD power clamp. In this implementation, a p-channel MOSFET clamp device is present between the $V_{DD}$ and $V_{SS}$ power supplies. For this circuit, an *RC* discriminator circuit responds to an ESD event for initiating the p-channel MOSFET clamp device. An n-diffusion resistor was electrically connected to the $V_{SS}$ power rail. Latchup occurred between the p-channel MOSFET, which is electrically connected to $V_{DD}$, and the n-type resistor electrically connected to $V_{SS}$. In this case study, this interaction was unanticipated, so no spatial spacings and physical size of the single p-channel MOSFET were considered.

Table 7.2 provides examples of intercircuit latchup events. Intercircuit latchup events occur due to unanticipated interactions between circuits. Hsu *et al*. [1] quantified interaction between a dual-diode ESD input circuit and an ESD power clamp. Latchup can occur between ESD input circuits and ESD power clamps. In the second case study, latchup can occur in identical circuits that are adjacent but in

*Latchup* Steven H. Voldman
© 2007 John Wiley & Sons, Ltd

**Table 7.1** Internal circuit CMOS latchup.

| Internal circuit | | | Solution |
|---|---|---|---|
| ESD dual diode and off-chip driver | p-channel MOSFET pull-up and n-well diode | W. Noble | Power sequencing |
| *RC*-triggered MOSFET ESD power clamp | n-resistor and p-channel MOSFET | M.D. Ker | Spatial separation |

opposite states. Salcedo-Suner *et al*. [2] reported a case study of latchup between two adjacent I/O networks. Given two adjacent off-chip driver networks, both contain a p-channel MOSFET pull-up and an n-channel MOSFET pull-down. As a result, there is a possibility of interaction of a p-channel MOSFET pull-up of one I/O network with the n-channel MOSFET pull-down of the second network. Given the state that one p-channel MOSFET has a signal that leads to overshoot, and at the same time, an adjacent n-channel MOSFET pull-down has a signal that leads to undershoot, the voltage difference is such that latchup may occur between these two elements. Salcedo-Suner also reported interaction within the ESD network that consisted of a diode string serving as a 'substrate pump' for a grounded ESD n-channel device. Inadequate guard rings between the diode string and the MOSFET led to latchup.

Latchup can occur between circuits of chip segments or subfunctions. Latchup can occur between I/O and core memory circuits, I/O and core CMOS logic, digital and analog segments, and digital and RF segments. In chip design synthesis, many segments are designed separately, and not integrated until the final product chip. For example, the I/O peripheral circuit designs are optimized, tested and qualified for ESD and latchup independent of the chip cores and other sectors. Hence, interactions between the chip sectors are evaluated as a second step after all the subfunctions are defined, and then placed. As a result, the interactions on the subfunction 'boundaries' are not understood until afterward. Additionally, latchup events that spatially propagate may not be understood when only local interaction is addressed. Transient events and external source issues can bypass today's latchup ground rules. Examples are shown in Table 7.3.

As a first case study, Huh *et al.* [3, 4] demonstrated that latchup can occur between two power rails, where a parasitic npn was formed. Given the n-well contains a p-type device, a parasitic pnpn is created between two chip subfunctions. As a second example, Huh *et al*. demonstrated latchup between an OCD in the I/O sector and another chip sector that contained decoupling capacitors [3, 4].

As a third example, Weger *et al*. demonstrated interaction between 'array I/O' and surrounding CMOS logic [5]. Minority carrier injection from negative undershoot in an ESD DD network can initiate latchup in the surrounding core logic. Brennan also demonstrated that this can occur using GGNMOS ESD networks [6].

**Table 7.2** Intercircuit CMOS latchup.

| Circuit-to-circuit | | | Solution |
|---|---|---|---|
| Dual-diode ESD and field oxide device ESD power clamp | Dual-diode p+/n-well diode and field oxide device ESD power clamp | C.T. Hsu | Physical separation of input ESD and FOD ESD power clamp |
| Off-chip driver to Off-chip driver | p-channel MOSFET to n-channel MOSFET | J. Salcedo-Suner | Improved guard rings within ESD circuit |

Table 7.3 Subfunction to subfunction CMOS latchup.

| Subfunction to subfunction | Interaction | | Solution |
|---|---|---|---|
| $V_{DD}$ to $V_{DD}$ | Parasitic npn formation between two n-wells | Y. Huh | Spatial separation of power domains |
| OCD-to-decoupling capacitors | p-channel MOSFET pull-up and n-type decoupling capacitor | Y. Huh | Ground rule development between power domains |
| Dual-diode ESD and CMOS logic | n-well diode and internal logic | A. Weger | n-well guard ring |
| GGNMOS ESD and CMOS logic | GGNMOS ESD and CMOS logic | C. Brennan | Improved guard ring design |

## 7.2 INTRABOOK LATCHUP MECHANISMS

Latchup can involve ESD networks since these networks exist at the periphery of the semiconductor chip. ESD networks can contain p+/n-well diodes, p-channel MOSFETs and p-diffusion resistors that can serve as parasitic pnp elements. ESD networks may also contain n+/p− substrate diode elements, n-channel MOSFETs and n-type resistors that can serve as parasitic npn elements. These elements can be contained within the same circuit or between circuits. CMOS latchup can occur from the following cases:

- within an ESD input network;
- within an ESD power clamp;
- between an ESD input network and an I/O circuit;
- between an ESD power clamp and an I/O circuit;
- between the ESD input network and the ESD power clamp.

### 7.2.1 CMOS Latchup Within an ESD Input Network

For latchup to occur within an ESD input circuit, the ESD network must contain both parasitic pnp and npn elements, which are electrically configured to form a pnpn network. Since the parasitic elements are optimized for ESD, the injection characteristics of the pnp and the npn are designed to allow the forward active state condition of the silicon-controlled rectifier (SCR) to occur. Parasitic pnpns can also form between the desired SCR and unanticipated parasitic SCRs in parallel with the desired element. It is in these cases that latchup can occur between the input node and the power supply (e.g. interaction with guard rings and power supplies). ESD input circuits that only contain n-channel MOSFET device can not undergo latchup. In a grounded gate NMOS (GGNMOS) ESD network, typically there is no parasitic pnpn element; hence, ESD input networks that only contain GGNMOS devices are not latchup prone.

ESD DD networks can consist of a mix of both p-type and n-type elements used for the positive and negative ESD events [7–9]; this network contains both a pnp and an npn parasitic element. But the

electrical configuration is such that latchup does not occur within the two elements since they are electrically connected to the same electrical input node. Also note that ESD input networks are encompassed by a guard ring structure. Latchup can occur between the pnp element and the guard ring (e.g. n-diffusion or n-well guard ring) under the condition of current injection into the p+/n-well diode with the n-well guard ring at the $V_{DD}$ potential. Whereas this is possible, in the case of a p+/n-well diode which is designed with low resistance, the current is shunted through the diode to the $V_{DD}$ power supply without latchup. The electrical potential across the diode is a function of the diode series resistance. Hence, a latchup condition exists where as the resistance increases, the likelihood of this to occur is possible. The condition for latchup is that the forward bias voltage of the first diode and the voltage drop across the diode must exceed the forward active condition of the parasitic pnpn element. Hence, one of the conditions for this to occur is as follows:

$$(V_{be})_{pnp} + IR \geq (V_{be})_{pnp} + (V_{be})_{npn},$$

with $I$ the diode current and $R$ the diode resistance.

Hence, this can be simplified to

$$IR \geq (V_{be})_{npn}.$$

In the case of a mixed voltage diode string ESD network, the electrical potential between the input node and $V_{DD}$ increases with the diode resistance and the number of diodes. Additionally, in a ‘diode string ESD network’ the n-well region is ‘floating’ and not directly coupled to the $V_{DD}$ power supply. As the series resistance of the diode elements increases, as well as the number of diodes, a parasitic pnpn can form between the p+/n-well diode and the n-well guard ring structure (electrically connected to $V_{DD}$). Hence, the condition of latchup is when the diode string turn-on condition exceeds the forward active state of the parasitic pnpn. A precondition for latchup is as follows [8, 9]:

$$V_T = N(V_{be})_{pnp} + \frac{N(N-1)}{2}\left(\frac{kT}{q}\right)\ln[\beta_{pnp} + 1] \geq (V_{be})_{pnp} + (V_{be})_{npn}.$$

This can be simplified to

$$[N-1](V_{be})_{pnp} + \frac{N(N-1)}{2}\left(\frac{kT}{q}\right)\ln[\beta_{pnp} + 1] \geq (V_{be})_{npn}.$$

Note that the inclusion of the voltage drop through the diodes increases the likelihood of latchup between the first stage and an adjacent guard ring structure. A key latchup design practice is as follows:

- Latchup can occur within an ESD network when a parasitic pnpn element is formed that is electrically configured such that a forward active state can occur during chip operation.
- ESD input circuits can form parasitic pnpns between the elements and its own electrical guard rings when p-type elements exist within the ESD network.
- Latchup can be eliminated with low-resistance ESD elements and adequate guard ring structures to avoid the forward active condition for CMOS latchup.

### 7.2.2 Intrabook Latchup Mechanisms – ESD Input Circuit and I/O Circuit Latchup

Latchup can involve ESD networks since these networks exist at the periphery of the semiconductor chip. As a result, intrabook latchup can occur from the following cases:

- within an ESD input network;
- within an ESD power clamp;
- between an ESD input network and an I/O circuit;
- between an ESD power clamp and an I/O circuit;
- between the ESD input network and the ESD power clamp.

Within a peripheral circuit 'book', a semiconductor chip may contain an I/O OCD circuit, a predrive network and an ESD input node device. An example of a latchup event that occurred within a peripheral book can occur between an ESD input node device and the CMOS off-chip driver.

In this case, a 'dual-diode' network was placed on an input node (Figure 7.1) [8]. The DD network consisted of a LOCOS-defined p+ diffusion within a retrograde n-well CMOS technology. The technology was a 0.8-μm CMOS technology on a p++ substrate and a p− epitaxial region. In addition, there was an n-well diode electrically connected to the input node as the diode element for negative discharge to the $V_{SS}$ power rail.

In the system, the semiconductor chip sequencing of the power rails and the input signals was not adequately defined. The input pins biasing was established prior to the ground power rail $V_{SS}$. In this case, a negative signal was established that led to a negative discharge through the ESD n-well diode to the p++ substrate. The ESD network current was injected into the substrate and into the n-well of the LOCOS-defined p-channel MOSFET. The OCD p-channel and n-channel MOSFETs established a parasitic pnpn that underwent latchup. In this case, the failure mechanism leads to the destruction of the semiconductor chip and package module. A key issue in this condition was that in this incident the

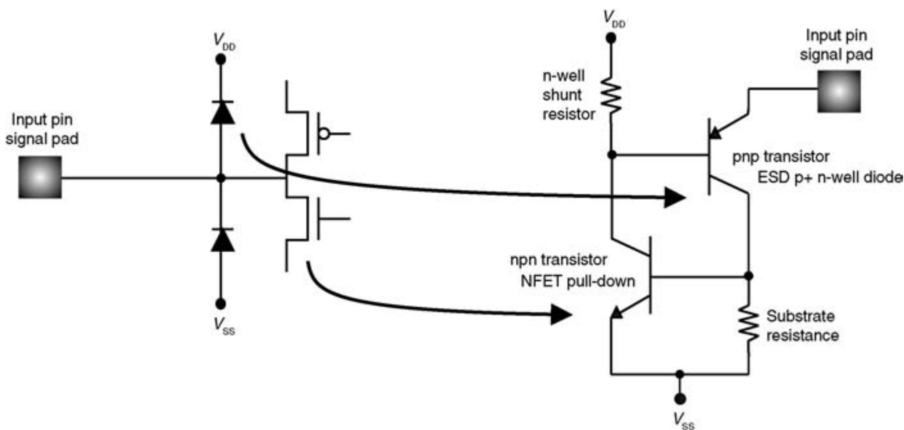



**Figure 7.1** Intrabook ESD dual-diode input network and OCD latchup.

latchup robustness of the base technology was significant with adequate process solutions, good design and established design rules, yet with the improper sequencing condition, latchup still occurred.

A key latchup design practice is as follows:

- Sequence-dependent ESD input networks can lead to injection into the $V_{DD}$ or $V_{SS}$ rail.
- Latchup can occur in latchup robust technologies given improper sequencing of the power rails, ground power rails and input pins.
- ESD input circuits electrically connected to $V_{DD}$ and $V_{SS}$ can lead to external injection into the input circuitry leading to an intrabook latchup.

## 7.3 INTERBOOK LATCHUP MECHANISMS

Interbook latchup events can occur between different electrical circuits or subfunctions. When the analysis of the circuits is evaluated independently, these events will not be observed until full-chip integration. In this section, interbook latchup events will be discussed in detail, as shown in Tables 7.1–7.3.

### 7.3.1 Interbook Latchup Mechanisms – ESD Input Circuit and Power Clamp Latchup

Latchup can involve ESD networks since these networks exist at the periphery of the semiconductor chip. Within ESD networks are semiconductor devices as well as parasitic devices that can contribute to latchup. Interbook latchup can occur from the following cases:

- between an ESD input network and an I/O circuit;
- between an ESD input network and core logic circuits;
- between an I/O and core logic circuitry;
- between an ESD power clamp and an I/O circuit;
- between the ESD input network and the ESD power clamp.

### 7.3.2 Interbook Latchup Mechanisms – ESD Input Circuit and Power Clamp Latchup (Hsu *et al.*)

Hsu *et al.* demonstrated a case study where latchup occurred between an input node ESD network and an ESD power clamp [1]. In this case, an ESD DD network was placed on an input node (Figure 7.2). The dual-diode network consisted of a LOCOS-defined p+ diffusion within an n-well structure in a single-well CMOS technology. In addition, an n-type field oxide device (FOD) was used as a local ESD power clamp. The n-type ESD power clamp consisted of a MOSFET gate on a LOCOS region to form the n-type field oxide device (e.g. also known as a LOCOS thick oxide MOSFET structure). The placement of the FOD ESD power clamp was adjacent to the signal pad (Figure 7.2) A p+ substrate guard ring was also placed between the ESD DD input and power clamp circuits. With the local placement of the ESD DD input network near the thick oxide MOSFET ESD power clamp, a parasitic

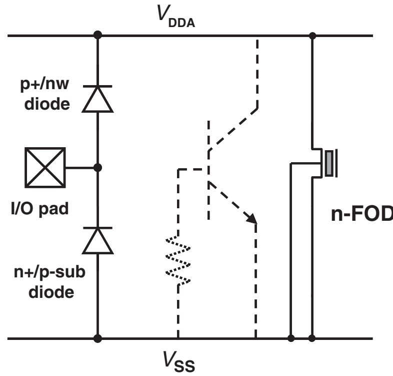



**Figure 7.2** ESD DD and n-type FOD circuit highlighting the parasitic npn bipolar transistor of the ESD power clamp. *Source*: [1].

pnpn structure was formed between the input node and the ground node of the power rail. Hsu *et al*. stated that two parasitic npn BJTs were involved in the latchup triggering [1]; the first npn was formed between the ESD input network n-well and the source of the thick oxide NFET ESD power clamp. The second npn also exists between NFET ESD power clamp drain and the NFET ESD power clamp source. The first npn triggering is associated with an anticipated pnpn structure. In the case of the second npn interaction, the discharge of the current to the n-well contact is then electrically connected to the $V_{DDA}$ power supply; this power supply is electrically connected to the ESD power clamp thick oxide drain, which also leads to a parasitic npn bipolar turn-on initiating latchup.

Figure 7.3 shows an example of a positive latchup pulse on the $V_{DDA}$ power rail. From the EMMI tool emission study, it can be observed that the latchup trigger pulse initiated turn-on of the FOD power clamp. The FOD structure can undergo MOSFET snapback without involvement of the ESD input network.

Figure 7.4 shows the latchup interaction between the ESD DD network and the FOD ESD power clamp. From the photograph, photon emissions can be observed between the ESD networks and the FOD region. In this case, current flowed from the input node through the p+/n-well diode. In the operation of

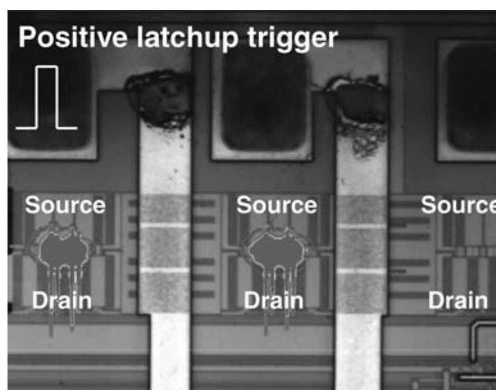



**Figure 7.3** Emission microscope (EMMI) tool results highlighting FOD drain-to-source current and the metallization due to positive trigger latchup pulse. *Source*: [1].

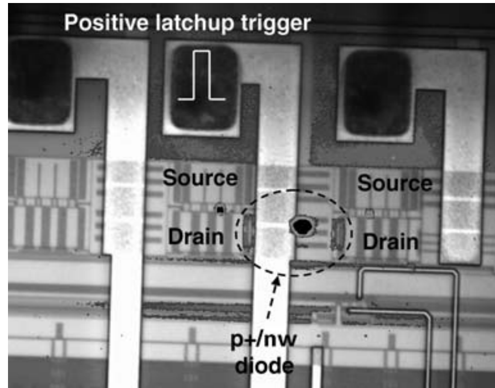



**Figure 7.4** A photograph highlighting the latchup failure associated with the interaction of the ESD dual-diode input circuit and the FOD ESD power clamp. *Source*: [1].

this network, some of the current flowed through the vertical parasitic pnp transistor to the substrate, whereas some of the current flowed to the $V_{DDA}$ power supply and through the FOD power clamp.

Figure 7.5 shows the CMOS latchup failure mechanism in the p+/n-well diode structure. The p+ anode is shown with the n-well region for the contact. From the failure analysis image, it can be observed that failure damage is observable at the p− substrate region.

From the latchup event, the FOD drain structure also showed evidence of silicon damage at the drain region (Figure 7.6).

A key latchup design practice is as follows:

- Placement of ESD input networks near the ESD power clamps can lead to the formation of parasitic pnpn elements between them.
- ESD input circuits electrically connected to $V_{DD}$ can lead to electrical coupling to an ESD power clamp leading to latchup through the power supply rail.
- ESD CAD systems should address the parasitic pnpn formed between ESD input and ESD power clamp networks.
- Latchup can be eliminated with adequate guard rings around the ESD input circuits to prevent latchup pnpn parasitic interaction with ESD power clamps.

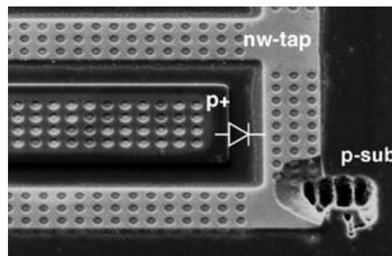



**Figure 7.5** ESD input device p+/n-well diode failure due to CMOS latchup. *Source*: [1].

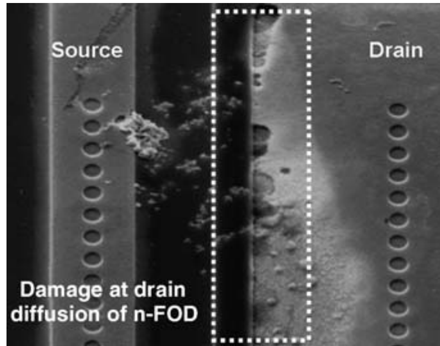



**Figure 7.6** CMOS latchup damage at the drain diffusion of the n-type FOD. *Source*: [1].

### 7.3.3 Interbook Latchup Mechanisms – I/O to ESD Power Clamps (Huh *et al.*)

Since I/O circuits are exposed to voltage overshoot and undershoot phenomena caused by signal reflection, I/O latchup design layout rules are more conservative compared to internal logic layout rules. Unfortunately, latchup design layout guidelines do not typically address the parasitic pnpn formed between I/O devices. Huh noted that in a semiconductor chip latchup can occur through adjacent I/O cell to I/O cell interactions [3, 4].

In Figure 7.7, the emission pattern is shown in a 5-V PCI buffer immediately next to ESD power clamp cells. The p-channel MOSFET OCD pull-up and a $V_{SS}$ n+ junction in a neighboring ESD power clamp cell form an unexpected latchup path. In this interaction, the diffusion area of the p-channel OCD pull-up element and the n-diffusion of an ESD power clamp are both large area regions; this forms a large parasitic pnpn. Since I/O latchup rules are checked in unit I/O cells, it is hard to detect this type of I/O to I/O cell interaction-induced latchup path between the first I/O and the second adjacent region. In most design environments, conformance to the latchup rules is verified within the I/O cell, but not between an I/O cell and an ESD power clamp in an adjacent cell. A key latchup design practice is as follows:

- Placement of OCD networks near the ESD power clamps can lead to the formation of parasitic pnpn elements between the OCD input network and the ESD power clamps.
- OCD p-channel pull-up MOSFET can lead to electrical coupling to an n-diffusion within an ESD power clamp (whose n-diffusion is at a ground potential) leading to latchup.

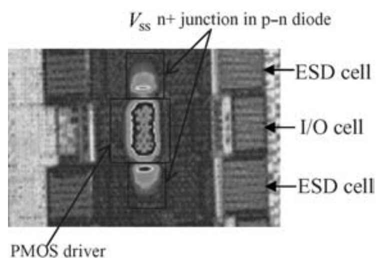



**Figure 7.7** I/O to I/O latchup of adjacent signal pads. Reproduced by permission of ESD Association.

- ESD CAD systems should address the parasitic pnpn formed between OCD input networks and adjacent ESD power clamp networks.
- Latchup can be eliminated with adequate guard rings around the OCD circuits to prevent latchup pnpn parasitic interaction with ESD power clamps.

### 7.3.4 Interbook Latchup Mechanisms – Power Supply to Power Supply (Huh)

Today, most semiconductor chips with different power supply voltages exist in a common system. Some semiconductor chips have multiple voltage levels contained within the same chip. Typically, the peripheral I/O voltage and the core circuit blocks will be at different voltage levels. Unfortunately, latchup design layout guidelines do not typically address the parasitic pnpn formed between (1) power supply to power supply (e.g. I/O $V_{DD}$ to core $V_{DD}$) and (2) between I/O devices and core devices, and other interchip subfunctions (e.g. I/O to I/O, I/O to memory). Huh *et al.* demonstrated that the aforementioned cases of 'I/O $V_{DD}$' to 'core $V_{DD}$' latchup and other cross-'book' latchup can occur in semiconductor products [3,4]. For instance, 3.3- and 1.2-V I/Os can be placed adjacent to each other with different cell heights in an advanced IC design. In core-limited design floor plans, I/O cells can be occupied by decoupling capacitors to reduce core power supply noise. Thus, a 3.3-V I/O cell can be physically placed adjacent to a 1.2-V decoupling capacitor. Since latchup design layout guidelines for I/O cell design do not typically address latchup between I/O power (3.3 or 2.5 V) and core power (1.2 V) supplies, latchup occurred between the two different $V_{DD}$ power supplies. Latchup failure was observed in a mixed signal design that had a decoupling capacitor powered by 1.2 V in the I/O area.

Failure analysis and design layout review revealed that the OCD p-channel transistor pull-ups in the 3.3-V I/O cell and p-channel MOSFET decoupling capacitor cell powered by 1.2 V form a parasitic pnpn latchup structure. To verify this failure mechanism, latchup testing with varying core $V_{DD}$ voltage was performed at an ambient temperature with an injection current of 100 mA [3,4].

Huh *et al.* [3,4] noted that since there were no $V_{DD}$ to $V_{DD}$ latchup rules, the 3.3-V n-well to 1.2-V n-well space did not have any guard ring structures electrically separating the physical domains. An indication of latchup was that core $I_{DD}$ current was reduced by approximately 100 mA once the device under test (DUT) went into latchup mode; this implies that a latchup path exists between $V_{DD}$ (I/O) and $V_{DD}$ (core) through the neighboring decoupling capacitor cell placed near the I/O region. Latchup simulation was performed to confirm the trigger mechanism. The simulation structure results show that the vertical pnp device turns on by injecting current through the I/O pad and its base begins to push out; the npn transistor also turns on, injecting carriers into the base, leading to regenerative feedback and initiating of the pnpn 'turn-on' (note that the current flow is deep below the n-well). The device stays 'latched' even after the injection pulse is removed. A latchup design solution for this failure was as follows:

- Increase the pnpn emitter resistance by removal of n+ contacts (e.g. $V_{DD}$ n+ contacts).

By removing $V_{DD}$ n+ contacts in decoupling capacitor cells close to an abutting I/O cell, the emitter resistance is increased; consequently, the lateral npn transistor does not supply enough trigger current to couple to the vertical pnp transistor since current flow through the npn is limited by the increased n-well resistance. Huh *et al.* [3,4] also noted that $V_{DD}$ to $V_{DD}$ latchup can also occur for the case where a high-voltage $V_{DD}$ (I/O) is abutting a low-voltage $V_{DD}$ (I/O) or ESD power clamp cells.

A proposed concept in these environments is the introduction of a 'transition cell'. With the introduction of a transition cell with p+, n+ and n-well guard rings, the following objectives would be achieved:

- Transition cells can be integrated into a design methodology.
- Transition cell can provide a method for layout design checking.
- Transition cell can provide a method for design verification.

A transition cell concept would be a good solution to prevent undesirable $V_{DD}$ to $V_{DD}$ latchup since it will provide enough space between two neighboring I/O cells powered by different $V_{DD}$ power supplies, can be automatically placed by using a CAD tool, as well as provide conformance to CMOS latchup design rules.

### 7.3.5 Interbook Latchup Mechanisms – I/O to I/O Networks (Salcedo-Suner *et al.*)

I/O to I/O latchup can occur due to interaction of adjacent circuits. In these cases, the I/O to I/O latchup can occur between adjacent OCDs, ESD to OCD or ESD to ESD. A unique characteristic of this failure type is that the parasitic pnpn anode is the signal pin, instead of the $V_{DD}$ power supply. Parasitic pnpn elements can exist within an ESD network itself, wherein injection from other I/O cells can influence its trigger state.

Salcedo-Suner *et al.* reported latchup in a voltage-tolerant I/O cell [2]. The low-impedance path is formed between a signal I/O pin and a ground pin while triggering an adjacent I/O pin with negative current pulse. In this case study, the ESD circuit consisted of a series cascode n-channel MOSFET network between the signal pin and the ground rail $V_{SS}$ (Figure 7.8). In addition, there exists a series diode string to the $V_{DD}$ power supply (forming a set of common collector parasitic pnp structures). With the network containing both p+/n-well diode and MOSFET elements, a parasitic pnpn exists between the parasitic pnp of the p+/n-well diode and the parasitic npn of the n-well and n-channel MOSFET source. Since a pnpn structure with floating n-well is contained within the ESD protection device, the ESD network is vulnerable to latchup. In this case, the parasitic pnpn structure easily turned on when a negative current pulse was injected through the neighboring I/O cell.

When a negative pulse was applied to a neighboring input pin, the injection current would influence the parasitic pnpn of the ESD network. Table 7.4 shows the relationship between the physical spacing of the I/O cells and the current needed to initiate latchup. As the spacing between the I/Os decreases, the current needed to initiate latchup decreases. As the injecting source moves farther from the I/O parasitic pnpn, it requires more current to lead to latchup [2].

Salcedo-Suner *et al*. provided solutions to address latchup [2]. In this implementation, there was a trade-off between the 'substrate pump' feature of the injection utilized for improved ESD and the latchup robustness. Salcedo-Suner *et al*. noted that with the introduction of guard ring to improve the latchup robustness, the low substrate resistance impacted the 'substrate pump' feature of the circuit, and hence the ESD results. Second, separating the diode string from the cascode elements to improve the latchup robustness also impacted ESD robustness since more current was needed to influence the potential under the MOSFET elements.

The following solutions were provided:

- reduce the 'substrate pump' feature of the ESD network;

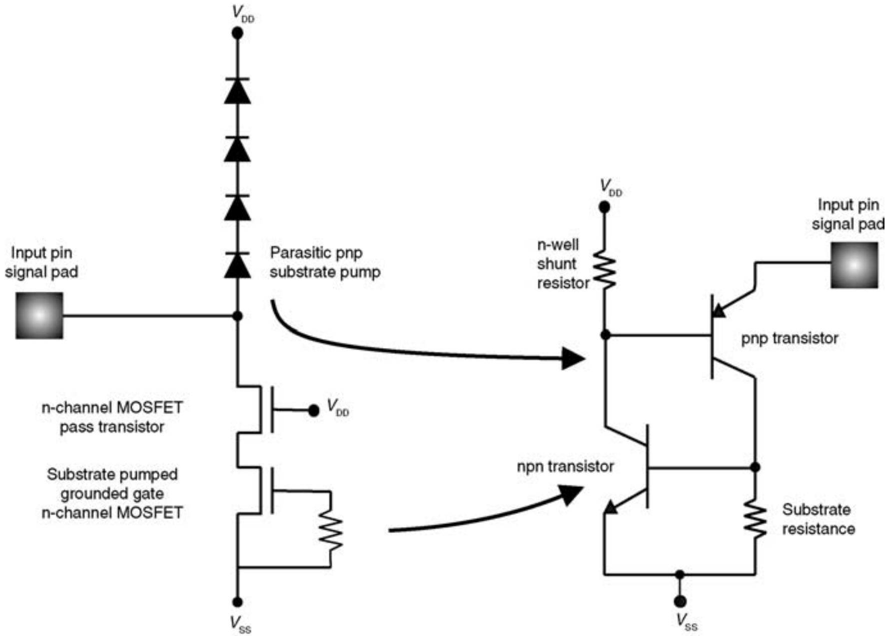



**Figure 7.8** Schematic of the I/O signal pin ESD network.

- add p+ substrate contacts between each p+/n-well diode region;
- add n-well guard ring around first diode and electrically connect to the first signal pad;
- add n-well guard ring around all diodes after the first diode element and electrically connect to the $V_{DD}$ power supply.

## 7.3.6 System-Level Latchup

System-level latchup failure can occur due to many interactions. Typical failures in systems can occur due to the following issues:

- system component power-up missequencing;

**Table 7.4** CMOS latchup injection current as a function of I/O to I/O space (e.g. injection source to pnpn parasitic spacing).

| I/O to I/O space (μm) | Latchup initiation current (mA) |
|---|---|
| 64 | −30 |
| 90 | −80 |
| 150 | −250 |

**Table 7.5** System-level chip-to-chip latchup.

| Chip-to-chip system | Interaction | Solution |
|---|---|---|
| Wire-bond failure in parallel multichip system | Wire-bond breakage leads to subsystem failure, leading to nonuniform current distribution in system | None |
| Wire-bond inductor mismatch in parallel multichip system | Voltage differential occurs due to inductive mismatch, leading to nonuniform current distribution in system | None |

- system ground not well established prior to signal pin power-up;
- reverse polarity initiated on a signal pin;
- cable discharge event (CDE) into signal pin;
- reverse battery installation.

In addition to these system-level events, multichip interactions can occur that can lead to the system-level latchup. Table 7.5 shows chip-to-chip interactions in a multichip system where system-level anomalies can occur.

System-level latchup can be addressed by the introduction of latchup prevention specific semiconductor components. Today, specific control chips can be integrated into systems to avoid latchup. For example, spacecraft system electronics have incorporated techniques into systems to provide single event latchup (SEL) immunity to events. The latchup protection circuitry is integrated on the same physical package as the latchup-sensitive device. These latchup prevention networks can set a 'threshold' based on a known sensitivity level, provide a means of detection of particle event, provide a current limitation of the device that is undergoing latchup, can shut down the system, and then reset to its original state. These are techniques used today to address latchup to avoid system-level latchup concerns.

## 7.4 CIRCUIT SOLUTIONS – INPUT CIRCUIT

### 7.4.1 Latchup Prevention Circuit – Dual-Well ESD Networks

Depending on the circuit and the physical design, ESD networks can either be a latchup preventive network or cause latchup. In this section, the preventive nature of ESD networks will be discussed. In dual-well CMOS networks, ESD input networks consist of the following classes [7–9]:

- dual-diode ESD networks;
- diode string ESD network;
- grounded gate n-channel MOSFET (e.g. GGNMOS) ESD networks;
- SCR ESD networks.

Dual-diode networks consist of a p-diffusion in an n-well and an n-well diode in the p− substrate. Dual-diode networks can lead to CMOS latchup depending on the semiconductor process and physical design layout. In the physical design, both internal and external guard ring structures can be used to minimize the impact of CMOS latchup. In the physical design of the p-diffusion, placing an n+ well contact on all sides of the p-diffusion reduces the risk of lateral injection and minimizes the well series resistance. Using a heavily doped retrograde n-well minimizes the vertical injection [7,8]. In the physical design of the n-well diode to substrate, placement of an n-well guard ring around the n-well to p− substrate diode minimizes injection to the p− substrate contact and injection outside of the n-well guard ring. External latchup will occur given that this is not adequately designed. A latchup circuit practice to decrease the risk of latchup in diode networks is as follows:

- utilize an n-diffusion to p− substrate diode;
- use two p-diffusion/n-well diodes for the dual-diode network;
- provide ESD networks with low well series resistances.

In the first case, since the n+ diffusion is not as deep as the n-well layer, the injection characteristics are lower. With a smaller area injector region and shallower structure, the 'emitter' injection characteristics are decreased using an n-diffusion. A lateral bipolar transistor is formed between the n-well and the n-well guard ring structure [7]. When the n-diffusion is used, the effective emitter area is decreased.

In the second case, two p+/n-well diodes are used, as shown in Figure 7.9 [9]. In this case, the 'down diode' electrically connects the n-well to the input circuit and the p+ diffusion to the $V_{SS}$ ground rail. An advantage of this network is that the current is not directly injected into the substrate region, but discharged to the power rail; this reduces the chip substrate injection and lowers the risk of external latchup. Additionally, in this concept, the diode series resistance is a function of the n-well resistance, not the p− substrate.

In diode string networks, there are a plurality of p-diffusions in an n-well, which are interconnected in series [7,8]. In these networks, one disadvantage is that the n-well regions of the successive diodes are not

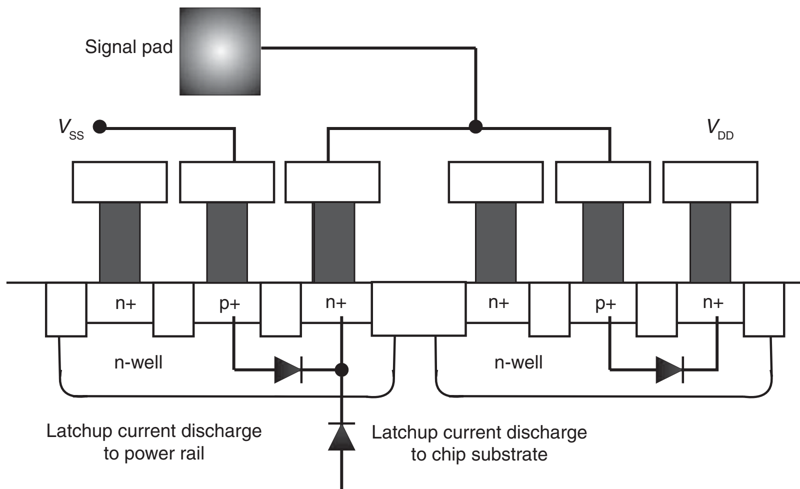



**Figure 7.9** Dual-diode ESD network using two p+/n-well diodes.

electrically connected to the power supply. As a result, the diode potential can be raised, leading to substrate injection. It has been observed that latchup can occur between the first input diode and the adjacent guard ring; a parasitic pnpn element can be formed between the first p+/n-well diode and the n-well guard ring. Latchup can be minimized by (1) reducing the number of diodes in series, (2) increasing the spacing between the guard ring and the first diode element, and (3) low series resistance ESD elements.

Grounded gate n-channel MOSFET ESD devices can minimize latchup by electrically limiting the voltage level observed on input pads and discharging current to the $V_{SS}$ power rail, instead of substrate injection. As MOSFETs are scaled to lower voltages, the MOSFET snapback voltage reduces. As a result, the GGNMOS ESD network limits the input signal voltage levels and will voltage clamp the input pads. Note, in this fashion, the p-channel MOSFET pull-up of an off-chip driver circuit will only discharge current into the substrate to the GGNMOS turn-on. But, for negative pulse events, the GGNMOS will serve as an injection source into the substrate. In this case, it can serve as a source of current injection into the substrate [7].

For silicon-controlled rectifiers, or pnpn networks, these ESD elements operate on the same principles as latchup; as a result, they are vulnerable to CMOS latchup. Many solutions to these networks are used to turn 'off' these ESD SCR elements during chip operation and turn 'on' when the chip is unpowered. One advantage of ESD SCR elements is that low-voltage triggers are integrated to minimize the turn-on voltage. This is achieved using n-channel MOSFET elements. Low-voltage trigger SCR (e.g. LVTSCR) ESD networks will reduce the voltage level of the signal pad and reduce the risk of the MOSFET OCD from undergoing CMOS latchup. LVTSCR ESD network trigger voltage scales with the MOSFET channel length [7].

## 7.4.2 Latchup Prevention Circuit – Triple-Well ESD Networks

Triple-well technology can provide opportunities for the prevention of latchup failures. With integration of triple-well ESD networks, isolation from the physical substrate can lead to the minimization of injection phenomenon. Triple-well ESD networks can include latchup solutions to avoid latchup. Depending on the circuit and the physical design, triple-well ESD networks can either be a latchup preventive network or cause latchup. In triple-well networks, ESD input networks consist of the following classes:

- triple-well grounded gate n-channel MOSFET (e.g. GGNMOS) ESD networks;
- triple-well diode string ESD networks.

In dual-well CMOS, GGNMOSs are not physically isolated from the semiconductor chip substrate, leading to avalanche current injection and interaction with other device elements. With the physical isolation of the p− epitaxial region from the p− substrate, injection phenomenon can initiate latchup such as adjacent circuit or single events. Figure 7.10 shows an example of an 'isolated MOSFET' structure [10]. Minority carrier injection from adjacent structures is removed by the triple-well n+ buried layer, preventing 'external latchup'. In addition, 'substrate pump' phenomenon will also not influence the MOSFET device. As a result, a solution to avoid latchup is the utilization of the triple-well isolating region for electrical decoupling of the substrate region.

In dual-well CMOS, ESD networks are not isolated from the semiconductor chip substrate. Triple-well ESD diode strings can physically isolate these elements from the chip substrate. Figure 7.11 shows an example of a triple-well diode string developed for triple-well technology by Pequignot *et al.* [11]. In this case, to avoid injection into the substrate, the triple-well region consists of a buried layer. The buried layer prevents the lateral shallow trench isolation p+/n+ diode from injecting into the substrate

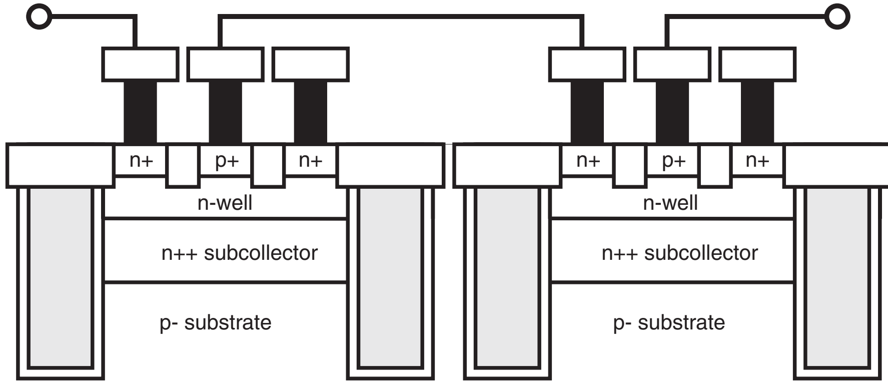



**Figure 7.10** Triple-well MOSFET structure isolated using deep trench and subcollector.

region. In addition, the triple-well 'tub' isolates the ESD network from external injection phenomenon. In this network, further 'power supply decoupling' can be instituted with a 'triple-well tub bias control network' [11]. A latchup circuit practice to decrease the risk of latchup is as follows:

- utilize 'isolated MOSFET' for ESD networks;
- utilize 'triple-well diodes' for ESD networks.

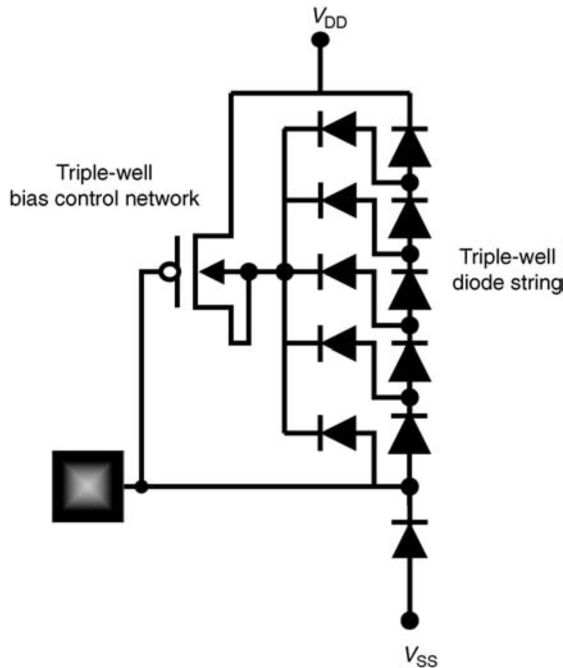



**Figure 7.11** Triple-well ESD device with bias control network.

### 7.4.3 Sequence-Independent Input Networks

In some system applications, it is desirable to not require a sequencing condition between the signal pins and the power supply. In this case, the sequencing of the power-up or power-down of $V_{DD}$ does not need to be established prior to the states of the signal pins; this is a 'sequence-independent' condition between the signal pins and the power supply. This condition can also be stated as a 'power boundary condition' where no power flows in or out of the signal pins independent of the power supply state. In this condition, signal pins can be voltage activated without current flowing out of the signal pins into a sector of a system that is 'shut off' or below the signal pins state. The sequence-independent condition applied to signal pins requires a certain class of circuitry that can support this condition [12–20]. For input pin sequence independence, the OCD, the receiver and the ESD input protection networks must not forward bias or lead to the flow of dc constant current either into the system or out of the system. Hence, there are classes of circuits that can provide this capability.

### 7.4.4 Sequence-Independent Input Networks: Off-Chip Drivers

The sequence-independent condition applied to signal pins requires a certain class of circuitry used in OCD networks [12–17]. For input pin sequence independence, the OCD must not forward bias or lead to the flow of dc constant current either into the system or out of the system. An OCD network that does not provide this feature is a simple p-channel MOSFET pull-up whose n-well is electrically connected to the power supply. A similar condition also exists for many mixed voltage interface (MVI) circuits. Mixed voltage interface circuits have many classes of networks, where some MVI circuits are sequence dependent and other classes are sequence independent [12–17]. For mixed voltage interfaces, OCDs and receivers must be able to transmit or receive voltage level conditions above its native power supply voltage. In the circuit design practices provided for MVI compatibility, the circuitry must be able to support the voltage state without reliability. In addition, the circuits must not forward bias when the power supply of one of the two chips is 'off' or not at its final voltage condition. Hence, MVI OCD circuits must address the 'power boundary condition' that continuous current flow into or out of the interface circuit exists as a result of the power supply states.

From a latchup perspective, a forward bias condition of the p-type device (e.g. contained within an n-well) within receivers, OCDs or ESD networks can initiate latchup. Therefore, the class of input signal sequence-independent networks that prevent forward biasing helps reduce the risk of latchup. Input signal sequence-independent circuits eliminate continuous forward bias injection into the substrate. In addition, many of these networks allow the circuit to 'decouple' the circuit from the power supply state.

Off-chip driver circuits were developed to address MVI conditions by Lundberg [12], Adams *et al.* [13, 14], Austin *et al.* [15], Hoffman *et al.* [16] and Dobberpuhl [17]. An OCD circuit design technique is the use of an 'n-well bias control network' to prevent forward biasing of the p-channel MOSFET. This class of circuits is also referred to as 'floating well' OCD [17]. Flaker instituted the idea of an n-well bias control switch where when a signal pad is low, the 'switch' is turned on, allowing biasing of the n-well to the $V_{DD}$ power supply. When the signal pad is 'high', the 'switch' is turned off, decoupling the n-well from the state of the power supply voltage [13, 14]. In this fashion, given the signal pad is above the power supply voltage state, the p-channel MOSFET does not continuously forward bias leading to a constant current flow from the signal pad into the semiconductor chip.

Figure 7.12 shows the circuit of Austin *et al.* [15]. In the circuit network, various elements exist to prevent forward biasing and electrical overvoltage. The 'n-well bias control network' consists of a

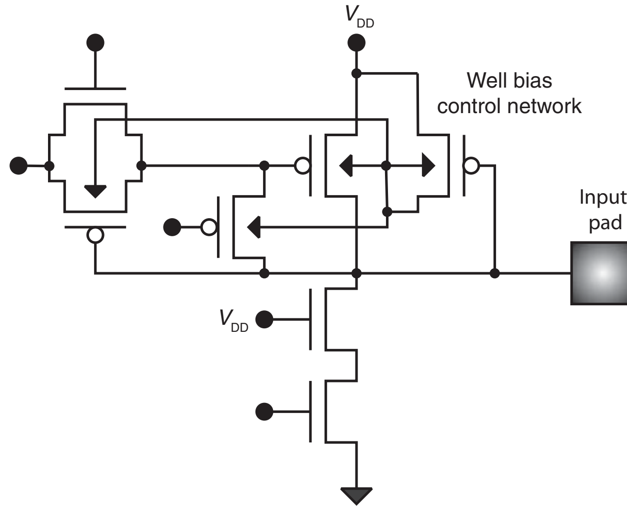



**Figure 7.12** Sequence-independent input off-chip driver circuit.

single p-channel MOSFET device whose MOSFET gate is electrically connected to the input node, whose source is connected to the $V_{DD}$ power supply rail and whose MOSFET drain is connected to its own n-well. All OCD p-channel MOSFETs are contained within the same n-well so that none of the supporting control elements forward bias. Dobberpuhl also developed OCD networks that improved the voltage bias state of the n-well during chip operation [17]. Hoffman *et al.* developed a circuit, allowing the ability to shut down redundant networks to reduce power dissipation and provide a ‘fault-tolerant’ environment [16]. The ability to shut down redundant network elements for space applications is known as ‘cold sparing’. This technique allows for redundancy without power loss. Hence, a latchup circuit design practice is as follows:

- Sequence-independent input circuit techniques provide a solution to avoid substrate injection, as well as avoid a forward bias state in parasitic pnp elements.
- CMOS peripheral OCD circuit techniques exist that avoid forward biasing of the p-channel MOSFET transistor using an ‘n-well bias control network’.
- n-well bias control networks ‘decouple’ the n-well tubs from the $V_{DD}$ power supply to avoid continuous dc current flow and injection into the p− substrate.

## 7.4.5 Sequence-Independent Input Networks: ESD Networks

The condition of ‘sequence independence’ between the signal pin and the power supply voltage state must be true for both the OCD and the ESD network. From a latchup perspective, a forward bias condition of the ESD networks can initiate latchup; therefore, the class of input signal sequence-independent ESD networks that prevent forward biasing helps reduce the risk of latchup. The concept used in the sequence-independent OCD can be applied to ESD networks. A

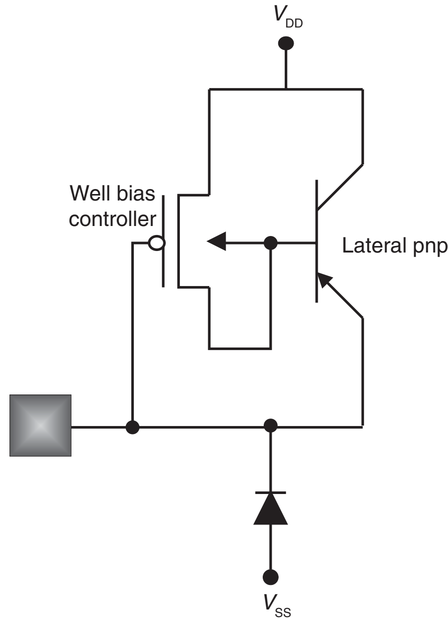



**Figure 7.13** Sequence-independent input ESD circuit.

sequence-independent ESD circuit can eliminate continuous forward bias injection into the substrate. In addition, ESD networks that allow the circuit to 'decouple' the circuit from the $V_{DD}$ power supply during power-up operation or power-down will also prevent injection into the semiconductor chip substrate [18,19]. Figure 7.13 shows an ESD network that uses an n-well bias control switch. For an n-well bias control switch, when a signal pad is low, the 'switch' is turned on, allowing biasing of the n-well to the $V_{DD}$ power supply. When the signal pad is 'high', the 'switch' is turned off, decoupling the n-well from the state of the power supply voltage [18, 19]. Hence, a latchup circuit design practice is as follows:

- use of an 'n-well bias control network' within a ESD network;
- using an n-well bias control network allows for the ESD network to 'decouple' from the power supply voltage and avoid continuous dc current flow and injection into the p− substrate.

## 7.5 POWER SUPPLY CONCEPTS

Chip design architecture also plays a role in latchup prevention. In latchup, the semiconductor chip can not latchup given that (1) the bipolar transistors do not become forward active, (2) the circuit does not reach the knee voltage (e.g. trigger voltage), or (3) the load line of the circuit cannot support the holding voltage or the holding current condition. In this discussion, it is inherent that the power supply voltage and the relationship of the circuit to the power supply influence all of the above conditions. The interrelation between the circuit subfunction and the power rails can influence latchup.

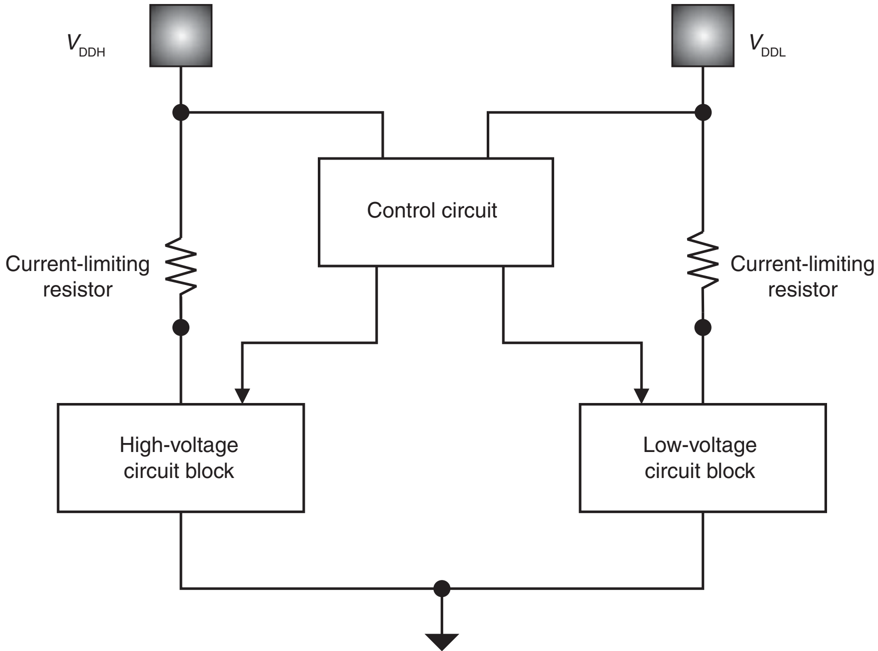



**Figure 7.14** Power supply current-limiting resistor element.

## 7.5.1 Power Supply Current Limit – Series Resistor

A latchup circuit design concept to avoid failure of semiconductor chips is to provide a circuit element in the power grid that limits the current flow to the circuitry. The first method is to place a resistor element between the power supply rail and the entire circuit set (Figure 7.14). This method is not a very elegant solution but is suitable for semiconductor chips that do not draw significant current into the functional circuit block. As the chip is in functional operation, the series introduces a small voltage drop during semiconductor chip operation. To avoid destruction of the system, the semiconductor chip and the internal metal power bus of the semiconductor chip, adding a resistance can introduce an *IR* drop at high current levels. The size of the resistance can be estimated by (1) allowable current through the chip without thermal failure and (2) an allowed current that is less than the holding current.

## 7.5.2 Power Supply Current Limit – Current Source

The second solution is the usage of a current source within a semiconductor chip (Figure 7.15). Using a current source, the amount of allowed current through the chip can be maintained as a circuit within the network attempts to undergo latchup. The choice of the current source can be chosen so that the allowed current magnitude is below $I_H$ of the circuits within the subfunction or chip. Examples of the current source can be a bipolar or a CMOS current mirror.

To avoid system-level latchup issues, integration of ‘on-chip’ or ‘off-chip’ latchup prevention networks also exists, which provides this current-limiting function. The role of the current-limiting

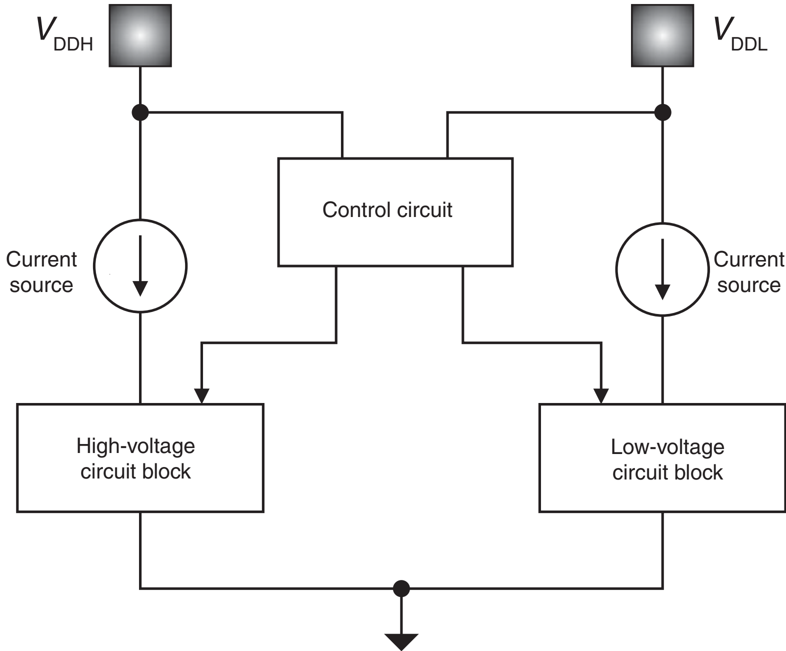



**Figure 7.15** Power supply current-limiting current source.

device is to sense the 'event' according to a preset 'threshold' level. The current-limiting network detects an increase in the current during an SEL event above the preset threshold level. When the current exceeds this threshold level, a 'shutdown' is initiated of the chip that is undergoing latchup. The 'shutdown' remains in a shutdown mode for a preset time interval. After a preset time, the chip supply voltage returns to its initial state. In ESD events, ESD circuitry is initiated by external ESD event itself (e.g. *RC*-triggered ESD power clamps); analogously, internal single events that can lead to SEL can be used to sense the current increase on the power rails, and latchup prevention networks can be initiated to 'shut down' the event, and then restore the chip to its initial state. A key difference between ESD event prevention networks and latchup prevention networks is 'powered' versus 'unpowered' condtions. A latchup circuit design practice is as follows:

- Introduction of a latchup prevention network to provide current limitation based on a preset magnitude.
- The current-limiting preset magnitude is set based on known system-level transients, space environment or known device sensitivity.
- Providing a current-limiting network to the device.
- Detection of system-level or single events from detection of current level increase.
- Providing a means of 'shutdown' of the chip undergoing latchup.
- Setting a preset time for 'shutdown' during the latchup event.
- Resetting the chip power after recovery from the transient event to the original power state.

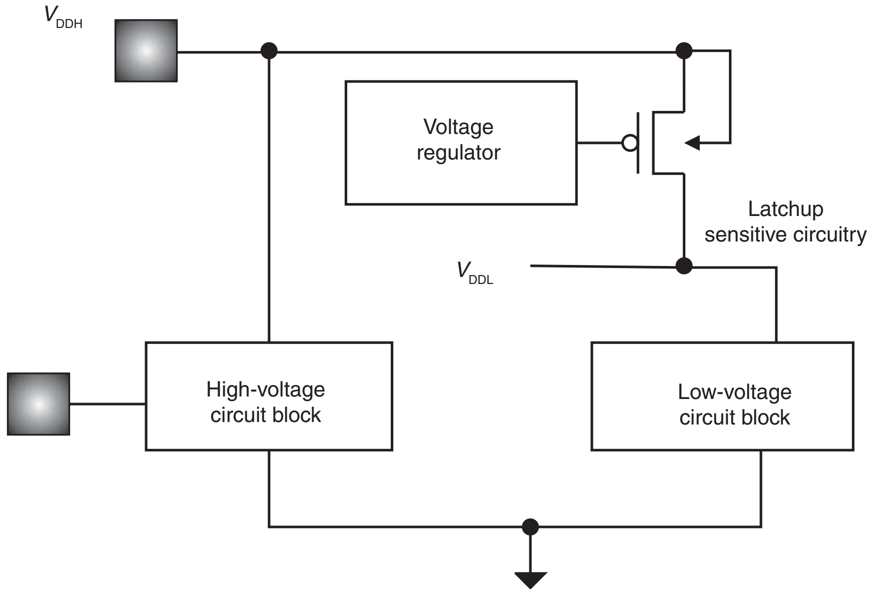



**Figure 7.16** Power supply current-limiting voltage regulator.

## 7.5.3 Power Supply Solutions – Voltage Regulator

Another method of controlling the voltage levels within a chip is the usage of a voltage regulator. Figure 7.16 shows an example of a regulator network. In many applications, the voltage regulator is placed between the external power rail and an internal power rail. Voltage regulators control the power supply internal to the chip from external noise and perturbations from outside sources.

A CMOS latchup circuit design practice is as follows:

- A current-limiting element or network is placed between the power supply and the circuit functional block wherein the element is chosen to guarantee the limiting of the power in the circuit.
- A current-limiting element or network is placed between the power supply and the circuit functional block wherein the element is chosen to guarantee the limiting of the current in the circuit such that it less than the holding current.
- A regulator is placed between the external power rail and the internal power rail to avoid overvoltage of the core circuitry.

## 7.5.4 Latchup Circuit Solutions – Power Supply Decoupling

Latchup is a concern in both internal circuits and peripheral circuitry. Latchup may also occur as a result of interaction between the ESD device, the I/O off-chip driver and adjacent circuitry initiated in the substrate from overshoot and undershoot phenomena. These may be generated by CMOS OCD

circuitry, receiver networks and ESD devices. In order for latchup to occur, the anode of the pnpn structure must be electrically connected to a power supply source (such as the $V_{DD}$ power supply) and the cathode of the pnpn structure must be electrically connected to a low potential (such as the $V_{SS}$ power supply). Hence, a latchup design discipline is the decoupling of the parasitic pnpn structure from the power supply voltages; this is achieved by the decoupling of the CMOS inverter circuits from the power supplies during overshoot and undershoot phenomena [21].

In CMOS I/O circuitry, undershoot and overshoot may lead to injection in the substrate. Hence, both a p-channel MOSFET and an n-channel MOSFET may lead to substrate injection. Simultaneous switching of circuitry, where overshoot or undershoot injection occurs, leads to injection into the substrate, which leads to both noise injection and latchup conditions. An interesting case study is the design methodology of the interaction of 'activated' and 'unactivated' elements in a gate array environment. In a gate array environment, a 'sea of gates' philosophy allows customization and personalization of circuit elements at a metallization level, where the silicon regions are predefined. Unused n-diffusion regions are grounded (e.g. $V_{SS}$ power supply) and unused p-diffusion regions are connected to the high power supply (e.g. $V_{DD}$ power supply). As the substrate voltage potential rises relative to the n-diffusion, all the n-diffusion elements of the gate arrays will tend toward the forward active state. As the substrate voltage potential lowers, the unused p-diffusion elements, the n-well and the substrate may activate the vertical pnp. This may occur as a result of minority carrier injection in wells and substrate regions.

In gate array design, latchup can occur from external latchup injection. An additional issue is the propagation of the CMOS latchup process. As an initial source injects electrons into the substrate, the first circuit element may latchup. The latchup of the first circuit leads to the turn-on of a pnp parasitic element, leading to more injection into the substrate. Hence, a circuit solution of detachment of the rails from a vulnerable pnpn parasitic structure is a viable solution to 'truncate the latchup propagation' through the semiconductor chip and array region [21]. In this case, the methodology may be addressed by certain functional blocks instead of spatial dependence. Hence, the methodology of detachment and connection to the latchup control networks may be according to the circuit type as well as physical localness (placement) to the injection source.

A latchup circuit design practice to eliminate latchup is the electrical decoupling of the power supplies during transient events. A latchup decoupling network consists of a latchup control isolation network electrically coupled to the substrate. In an ASIC environment, the latchup control isolation network purpose is to electrically isolates the 'sea of gates' (e.g. CMOS logic circuits) from the power rail. Figure 7.17 shows a latchup control isolation network integrated into a sea of gates environment. The circuit consists of the following [21]:

- A p-channel MOSFET array whose source is in series with a p-channel MOSFET 'switch' between the p-channel MOSFET array and the $V_{DD}$ power supply.
- An n-channel MOSFET array whose source is in series with an n-channel MOSFET 'switch' between the n-channel MOSFET array and the $V_{SS}$ power supply.
- The gate of the p-channel MOSFET 'switch' is electrically connected to the output of an inverter circuit, where the inverter circuit input is electrically connected to the n-well contact.
- The gate of the n-channel MOSFET 'switch' is electrically connected to output of an inverter circuit, where the inverter circuit input is electrically connected to the p-well (or p− substrate) contact.

In this circuit, perturbations to the well or substrate voltage potential are sensed by well and substrate contacts. This perturbation signal is then sensed on inverter circuit. When a perturbation signal is

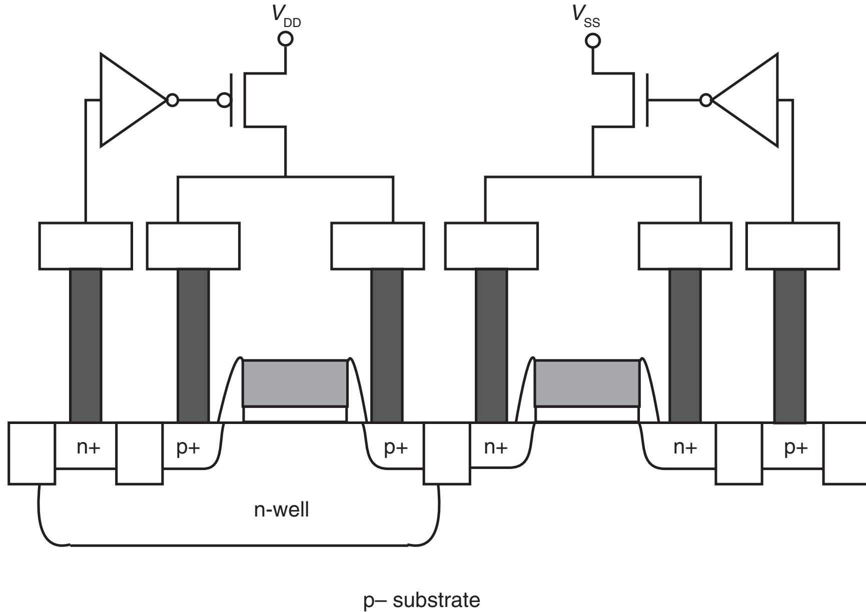



**Figure 7.17** CMOS latchup power supply decoupling network.

undesirable, leading to a latchup event, the 'switches' isolate the p-channel (or n-channel) transistor from its respective power supply rails.

Another implementation of this concept can utilize an 'active clamp' network [21] (Figure 7.18). Instead of inverter networks, active clamp networks generally have the advantage of being able to respond at excursions outside of the normal voltage operational regime. While diode-based implementations respond to excursions greater than or equal to $V_{DD} + V_{be}$, (e.g. $V_{be}$ being the forward bias voltage of a MOSFET junction), active clamp networks may respond to excursions greater than or equal to the $V_{DD}$ power supply voltage. Additionally, active clamp networks may also respond to voltage excursions below $V_{SS}$, instead of below $V_{SS} - V_{be}$ (e.g. as in diode-based scheme). An active clamp network is designed utilizing a reference control network. In this case, the active clamp network is electrically coupled to the n-well and a second active clamp is electrically connected to the p-well substrate contacts. As will be discussed in this chapter, an active clamp is turned on for excursions outside of normal voltage range (e.g. undershoot where $V^* < V_{SS}$ and overshoot where $V^* > V_{DD}$).

In an active clamp network, a reference control network is used where the reference potential is set to $V_{Tn}$ and a second reference control network is used whose reference potential is $V_{DD} - V_{Tp}$; these references may be established using a MOSFET whose gate is coupled to its own drain connection.

The network is established where the second NFET has its gate connected to reference $V_{Tn}$. The network is established where the second PFET has its gate connected to a reference $V_{DD} - V_{Tp}$; these control elements sense the $V_{DD}$ and $V_{SS}$ substrate local potentials. When the local substrate connection potential or the local well connection potential extends outside of the normal voltage range, these elements 'turn on'. The 'turn-on' of these elements leads to the 'turn-off' of a transistor element between the gate array diffusions and corresponding power supplies [21].

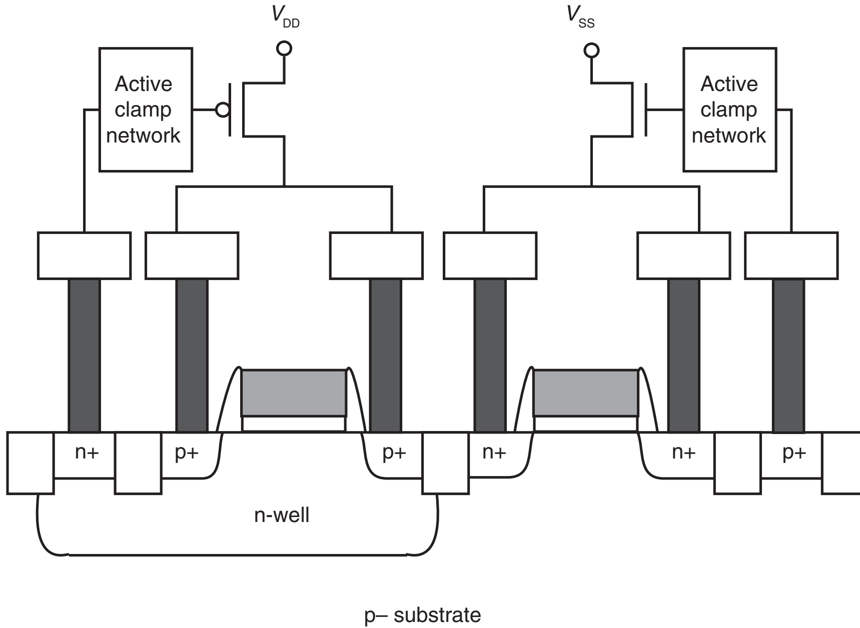



**Figure 7.18** Power supply decoupling network utilizing active clamp networks.

A latchup discipline circuit solution is as follows:

- A latchup decoupling network is established that senses perturbations outside of the normal operational range.
- The latchup network decouples the parasitic pnpn element from at least one power supply (e.g. $V_{DD}$ or $V_{SS}$) preventing CMOS latchup.
- The power supply decoupling networks can utilize inverter logic or active clamp networks that are electrically connected to a local well or substrate contact.

## 7.6 LATCHUP CIRCUIT SOLUTIONS – POWER SUPPLY TO POWER SUPPLY SEQUENCING CIRCUITRY

Latchup can occur as a result of the power supply sequencing of the different power supplies. In many semiconductor chips, the peripheral power rail is isolated from the internal power rail. For mixed voltage applications, the exterior power rail is at a higher voltage level to interface with chips from a prior technology generation. As a result, in many applications, two power supply voltages are within a single chip. In the case of a chip with an external peripheral and an internal core power rail, the capacitance of the internal power rail is significantly larger due to the number of circuits associated with the core chip (e.g. memory chips). As a result, the rate at which the exterior power rail and internal power rails charge is different in power-up and power-down; this can lead to forward bias states, as well as influence transient latchup. In the case where ESD circuits exist between the two power rails, such as series diode elements, these elements can forward bias leading to latchup when the rate of charging of the chip sectors is

different. In this example, the ESD network can forward bias leading to CMOS latchup. In addition, these networks can also lead to electrical overstress and metal failure during burn-in and other overvoltage conditions. Two methods can be provided to prevent CMOS latchup in this scenario:

- sequence-independent ESD networks;
- power supply to power supply switch networks.

## 7.6.1 Power Supply to Power Supply Sequence-Independent Networks

Figure 7.19 shows an example of a sequence-independent network used between the two power supply rails [20]. Voldman developed two versions of this network: the first network provides sequence independence between a signal pin and the power supply voltage, and the second version provides sequence independence between two power supplies. In the second version, the network contains a p-channel MOSFET contained in its own well, wherein the n-well is electrically connected to the p-channel MOSFET drain. Contained within the n-well is a lateral pnp element. The well node is self-biased leading to avoidance of forward biasing and latchup. This network prevents continuous forward biasing of the p+/n-well metallurgical junction during power-up and power-down states. When the p-channel MOSFET gate is 'low', the p-channel MOSFET is 'on', leading to the biasing of the n-well region to the higher power supply voltage and preventing forward biasing of the p-channel MOSFET junctions and the lateral pnp ESD element. When the MOSFET gate is 'high', the p-channel MOSFET is 'off', leading the n-well junction to 'float'. In this process, the power supplies are 'decoupled' preventing continuous current flow between the two power supplies providing 'sequence independence between power supplies'.

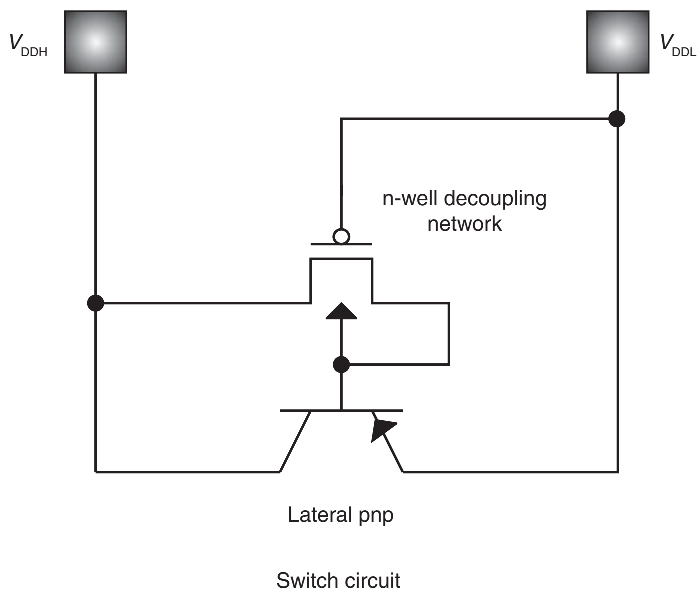



**Figure 7.19** Sequence-independent power supply to power supply network.

A CMOS latchup circuit design practice is as follows:

- use of networks that allow electrical decoupling to at least one power supply for the prevention of forward bias states between power supply rails;
- using an n-well bias control network allows to 'decouple' from the power supply voltage and avoid continuous dc current flow and injection between two power supplies.

## 7.6.2 Power Supply to Power Supply Control Circuits (Lin *et al.*)

Another control network methodology to prevent latchup between power supplies was developed by Lin *et al.* [22]. The latchup prevention network contained a 'control' circuit and a 'switch' circuit. The control circuit is electrically connected between the two power rails. Conceptually, the control network inputs are the two power supply voltage states. A control network detects a voltage difference between the first and the second power rail. The output of the control network produces two output signals that are connected to the 'switch network'. The 'switch circuit' senses the relative voltage. When the relative voltage is greater than a given preestablished value, the switch circuit (in response to the first control signal output) electrically connects the first power rail; when the relative voltage is smaller than the first preestablished value, the switch electrically disconnects the first power rail. Figure 7.20 shows a high-level diagram of the network.

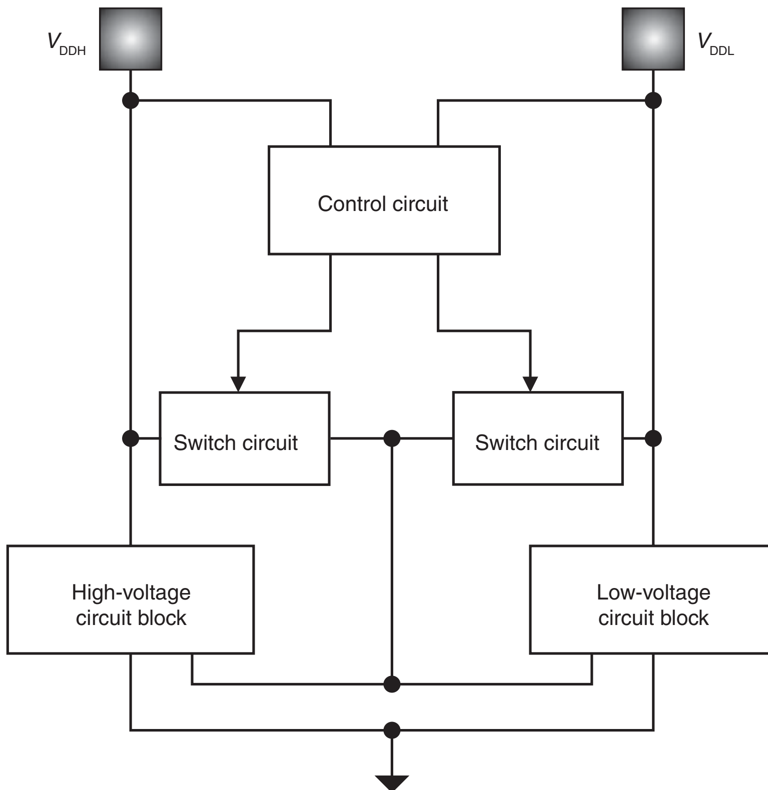



**Figure 7.20** Power supply to power supply network sequence network (higher level diagram).

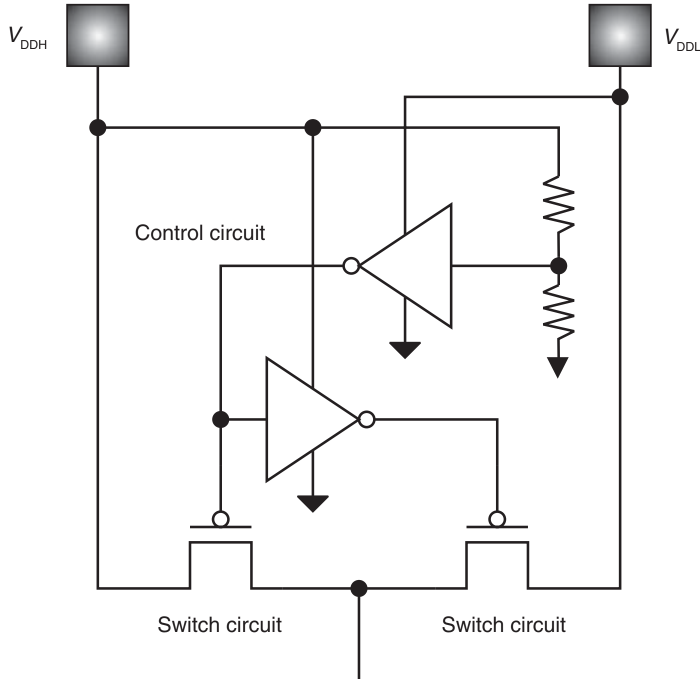



**Figure 7.21** Power supply to power supply network sequence network.

Figure 7.21 shows one of the networks proposed by Lin *et al.* [22]. In the control network there is a voltage divider, formed from two resistors, electrically connected to the input of an inverter circuit. The voltage divider is electrically connected between the higher power supply voltage and ground. When the higher voltage rail $V_{DDH}$ voltage is established across the two resistors, the input signal to the inverter leads the inverter to turn 'on'. The output of the inverter is electrically connected to a p-channel MOSFET 'switch' that electrically turns 'on' providing the electrical connectivity to the $V_{DDH}$ power rail. The output of the control inverter is also inverted by the second inverter control circuit. When the first inverter input is 'on' the output is 'low'. The first inverter 'low' output signal is passed into the input of the second inverter control circuit element, which provides a 'low' output. This signal is fed into the second p-channel MOSFET 'switch'; this p-channel MOSFET 'switch' is 'off' electrically decoupling the $V_{DD}$ power supply. The 'switch network' has two input states that are complementary from the two control network inverters; the output of the 'switch network' guarantees that only one of the power rails is electrically connected.

A CMOS latchup circuit design solution for sequencing is as follows:

- a control/switch network that provides electrical 'decoupling' from a power supply rail;
- a control/switch network that provides electrical 'decoupling' between two power supply rails;
- a control/switch network that prevents forward biasing independent of the states of the two power rails [22].

## 7.7 OVERSHOOT AND UNDERSHOOT CLAMP NETWORKS

Latchup events can be minimized with the utilization of a class of circuits that are used to minimize electrical overshoot and undershoot [23–32]. Given undershoot and overshoot can be minimized, forward biasing as well as substrate injection can be reduced.

Latchup occurs as a result of circuits that have the first node serving as an anode that is electrically connected to a high power supply (e.g. $V_{DD}$), and the second circuit whose cathode is electrically connected to a low power supply (e.g. $V_{SS}$). A p-channel MOSFET (or any p-diffusion) electrically connected to $V_{DD}$ serves as an anode of the pnpn and an n-channel MOSFET (or any n-diffusion) electrically connected to $V_{SS}$ serves as the cathode of the pnpn. Hence, a key concept of latchup is as follows:

- Circuits that have p-channel MOSFETs electrically connected to $V_{DD}$ and an n-channel electrically connected to $V_{SS}$ are prone to latchup from overshoot events on $V_{DD}$ and undershoot events on $V_{SS}$.

While this may obvious, the complement case is less obvious. If these circuits are latchup prone, then the inverse – are they latchup immune? or do they assist in lowering the latchup concern? There are circuits whose p-channel MOSFETs (or p-diffusions are electrically connected to $V_{SS}$) and whose n-channel MOSFETs (or n-diffusions) are electrically connected to $V_{DD}$. Examples of circuits that have n-channel MOSFETs (or n-diffusions) electrically connected to $V_{DD}$ are $V_{DD}$ to $V_{SS}$ ESD *RC*-triggered n-channel MOSFET power clamps, HSTL OCD (e.g. with an n-channel MOSFET pull-up transistor) and an ‘active clamp’ network; these networks prevent overshoot phenomenon either through clamping or through, parasitic bipolar turn-on. Hence, there are networks that provide ‘anti-overshoot’ and ‘anti-undershoot’ characteristics that assist in preventing latchup. Hence, a key concept of CMOS latchup is as follows:

- Circuits that have p-channel MOSFETs electrically connected to $V_{SS}$ and/or an n-channel electrically connected to $V_{DD}$ are not prone to latchup from overshoot events on $V_{DD}$ and undershoot events on $V_{SS}$ and provide ‘anti-latchup’ characteristics.

Therefore, there are circuit families or classes that can assist in the latchup prevention. In the following sections, we will learn about some interesting concepts, which provide ‘anti-overshoot’ and ‘anti-undershoot’ characteristics.

### 7.7.1 Passive Clamp Networks

Latchup events can be minimized with the utilization of a class of circuits that are used to minimize electrical overshoot and undershoot. This class of circuits are called ‘passive clamp networks’. ‘Passive’ clamping circuits are unable to effectively meet these opposing requirements for high-performance applications. Passive clamping circuits have been used in both bipolar and CMOS high-performance environments for 2- to 0.5-μm technology generations. These concepts were first utilized using bipolar transistors in high-performance environments. In CMOS technology, passive clamp networks have historically taken advantage of the parasitic bipolar transistors for overshoot and undershoot, directing the current between the MOSFET source and drain. The advantage of passive clamps is that the current is directed to the power rails instead of the injection into the substrate. Without active clamping, undershoot current is directed into the substrate, which can lead to injection, as discussed in Chapter 3.

Figure 7.22 is an example of a passive clamp network. For example, n-channel and p-channel MOSFET parasitic bipolar transistors were connected between the input and the power rails such that

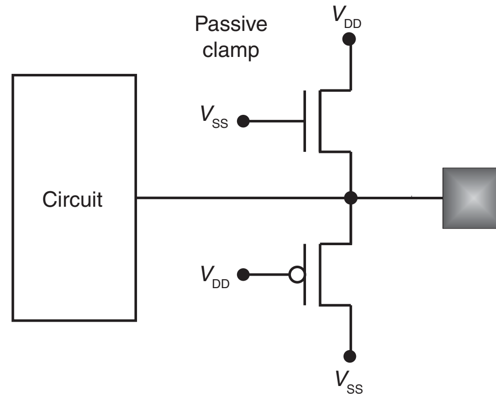



**Figure 7.22** Passive clamp network.

the MOSFET devices remained 'off' and the MOSFET parasitic bipolar devices would turn 'on' during overshoot and undershoot conditions. By connecting the n-channel MOSFET to $V_{DD}$ power rail and the p-channel MOSFET to $V_{SS}$ power rail, and disabling the MOSFET gate electrode, the overshoot and undershoot of the signal pad would activate the parasitic bipolar transistors. In these cases, the current will flow to the power grid $V_{DD}$ and $V_{SS}$ for overshoot and undershoot phenomena. For negative undershoot, the MOSFET parasitic npn bipolar transistor must become forward active, at $V_{BE} = 0.7\,V$, and discharge the current to the $V_{DD}$ power rail. For positive overshoot, the MOSFET pnp bipolar transistor must become forward active, at $V_{BE} = 0.7\,V$, and discharge the current to the $V_{SS}$ power rail.

In this practice, there is an interesting feature to these circuits. Note that this 'circuit feature' of 'sinking' a negative undershoot to the $V_{DD}$ power supply and a positive undershoot to the $V_{SS}$ power supply is the same 'utility' that a guard ring provides. For an n-well, guard ring serves as a 'sink' for a negative undershoot to the $V_{DD}$ power supply. For a p+ substrate, guard ring serves as a 'sink' for a positive overshoot to the $V_{SS}$ power supply. Hence, in essence, these passive MOSFET networks serve as a well-controlled scalable 'pseudo-guard ring' circuit. But the advantage is that these are scalable, self-enclosed designs, can be modeled and do not inject to the chip substrate. ESD networks, such as diodes, were also used as passive undershoot and overshoot clamps at 0.7 V noise levels. ESD diodes and MOSFET parasitic bipolar devices have served a useful role when the power supply voltage $V_{DD}$ was above 5 V.

The differences between the ESD diode networks and passive clamp networks are as follows:

- ESD diode networks inject current into the semiconductor chip substrate $V_{SS}$ for negative polarity events;
- passive clamps inject current into the $V_{DD}$ power rails for negative polarity events;
- passive clamps inject current into the $V_{SS}$ power rails for positive polarity events;
- ESD networks are designed to 'sink' high current levels well above the functional application (e.g. 1–10 A current levels);
- passive clamp networks are designed to 'sink' overshoot and undershoot current levels (e.g. 10–100 mA current levels).

Passive clamps have a functional limitation as the power supply voltage is scaled. A 'clamping figure of merit' is the ratio of the clamping voltage $\Delta V$ and power supply voltage $V_{DD}$:

$$\mathrm{FOM} = \frac{\Delta V}{V_{DD}}.$$

In order to damp out ringing and noise, a constant ratio must be maintained as the native power supply or peripheral supply voltage is scaled. A CMOS latchup circuit design practice is as follows:

- passive clamps are used to limit overshoot and undershoot on signal nodes;
- passive clamp networks 'sink' current to the power supply rails;
- passive clamp networks can 'sink' undershoot current to the $V_{DD}$ power supply, serving as a pseudo-'guard ring';
- passive clamp networks can 'sink' overshoot current to the $V_{SS}$ power supply, serving as a pseudo-'guard ring'.

## 7.7.2 Active Clamp Networks

Latchup events can be minimized with the utilization of a class of circuits called 'active clamp networks' [23–32]. Active clamp circuits are key to minimize electrical overshoot and undershoot, minimize reflected signals and achieve performance objectives and reliability requirements in high-performance circuits; these challenges include more significant challenge impedance matching conditions, MOSFET gate dielectric reliability, MOSFET off-current levels, low power, latchup and ESD protection. Active termination elements are of interest in the transfer of signals between receivers and transmitters [23–32]. Active termination elements exist in peripheral circuits; in the high-performance peripheral I/O environment, there are the receivers, OCDs, active clamp elements and ESD networks. DeClue and Muller [23], Davidson and Lane [24], Slaughter [25], Kosson [26], Petersen [27], Furman [28], Honningford [29], Voldman and Hui [30,31], and Mashak *et al.* [32] addressed the issues of voltage clamping, overshoot and undershoot, reflections and performance impacts.

In an ideal system, the input voltage switches instantaneously between the high-voltage state, 'digital 1', and the low-voltage state, 'digital 0', never exceeding the power supply and ground states and zero transition time. In a real environment, the signal has a finite transition time, overshoots and undershoots the power supply and ground voltages, oscillates and undergoes ringing. The ideal 'clamp' circuit eliminates ringing and noise such that the signal at the input remains at or near the two desired voltage states and switches between those states in the minimum time. The ideal 'clamp' must drain or supply current instantaneously to/from the network at the input to the circuit being clamped whenever the voltage at the input exceeds or falls below the desired voltage state. In order to achieve this, the clamping must have a low dynamic impedance (e.g. resistance) and a low reflection coefficient in the vicinity of the upper and lower voltage corresponding to the two digital logic states. On the contrary, in order to maximize switching speed between the two logic states, the impedance of the clamping circuit and the reflection coefficient should be very high during switching for a brief time when the input voltage is between the upper and lower digital voltages.

From a latchup circuit design practice, active clamps have significant benefits. Active clamp networks improve latchup tolerance of peripheral circuits by shunting the overshoot and undershoot signals to the power supply rails without forward biasing of the n- and p-channel MOSFET devices and

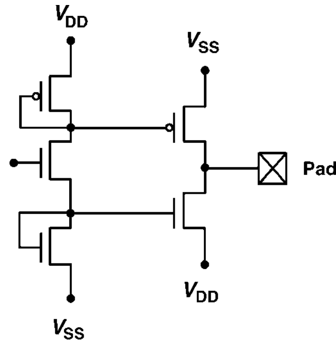



**Figure 7.23** CMOS active clamp network.

resistors in the peripheral circuits. Active clamp circuits also prevent electrical overstress (EOS) on the MOSFET gate dielectric in receiver networks and driver networks; this is very important with the rapid scaling of the MOSFET gate dielectric to avoid the electrical overstress.

Figure 7.23 shows an example of a CMOS active clamp circuit. In this circuit, an n-channel MOSFET source is connected to the pad node and the drain is connected to $V_{DD}$ power supply rail, and a p-channel MOSFET source is connected to the pad node and MOSFET drain is connected to the $V_{SS}$ power supply rail. The n-channel MOSFET gate is biased using a reference control element at the n-channel MOSFET threshold voltage $V_{Tn}$. The p-channel MOSFET gate is biased using the second reference control element at the voltage $V_{DD} - V_{Tp}$, where $V_{Tp}$ is the p-channel MOSFET threshold voltage. These clamping transistors are serially connected between the input pad and the power rails. The reference voltage is set so that when there is noise, ringing or undershoot event below the ground potential, the n-channel clamp element turns on. In this case, the input node is rapidly damped to the ground potential. Similarly, the p-channel reference circuit is set such that whenever there is an excursion above the $V_{DD}$ power supply, the p-channel MOSFET turns on and rapidly reduces the voltage to the power supply voltage. The center reference transistor controls current flow through the reference elements [30, 31].

Active clamps are 'pseudo-zero $V_T$' networks in that they turn on at the power supply rails. As a result, they 'turn on' at the power supply voltage rails. An 'ideal' active clamp would have no 'on resistance' providing an 'ideal switch' characteristic, which is 'off' for voltages between $V = V_{SS}$ and $V_{DD}$, and 'on' for all voltage excursions. Active clamp networks would also turn on prior to any ESD circuit network. For example, a dual-diode ESD network turn-on occurs at $V = -V_{BE}$ for undershoot and $V = V_{DD} + V_{BE}$ for overshoots, whereas the active clamp network turns on at $V = 0$ and $V = V_{DD}$.

Figure 7.24 shows the low-current $I$–$V$ characteristic of a clamp network for a 1.8-V power supply chip as a function of the input voltage $V_{in}$. When the input pad voltage is less than the $V_{SS}$ potential, the n-channel clamp element is 'on', and the current flows out of the clamping circuit. When the input voltage $V_{in}$ is between $V_{SS}$ and $V_{DD}$, no current flows through the active clamp network. When the input pad voltage $V_{in}$ exceeds the power supply voltage $V_{DD}$, the p-channel MOSFET clamp element is on and the current flows in the opposite direction from the input pad through the p-channel MOSFET to the $V_{SS}$ power rail. Using a MOSFET control switch serially connected between the two reference MOSFET circuits, the active clamp can be turned off by controlling the current through these two MOSFET control references [30, 31]. Note that in Figure 7.24 the magnitude of the current level for these elements is tens of milliamperes (e.g. on the order of overshoot and latchup events but well below the level of ESD events).

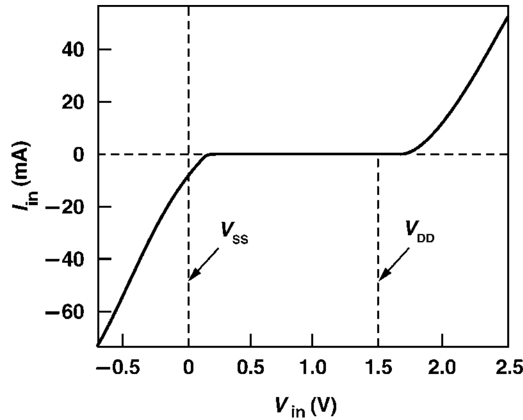



**Figure 7.24** *I–V* characteristics of an active clamp network.

A figure of merit of a clamp circuit is the reflection coefficient. The reflection coefficient is defined as the difference of the terminator (receiver) load $Z_1$ and the characteristic impedance of the transmission line, $Z_0$, over the sum of $Z_1$ and $Z_0$,

$$\Gamma = \frac{Z_1 - Z_0}{Z_1 + Z_0}.$$

For a perfect matched system, the reflection coefficient would be zero for $V_{in}$ below the substrate ground potential, unity for $V_{in}$ equal to the substrate potential and the native power supply voltage $V_{DD}$, and zero for $V_{in}$ greater than the native power supply voltage $V_{DD}$. For a perfectly matched active clamp network, the reflection coefficient should be $-1$ for $V_{in} < V_{SS}$, unity for $V_{SS} < V_{in} < V_{DD}$ and $-1$ for $V_{in} > V_{DD}$. Figure 7.25 shows the reflection coefficient of the circuit as a function of the input voltage.

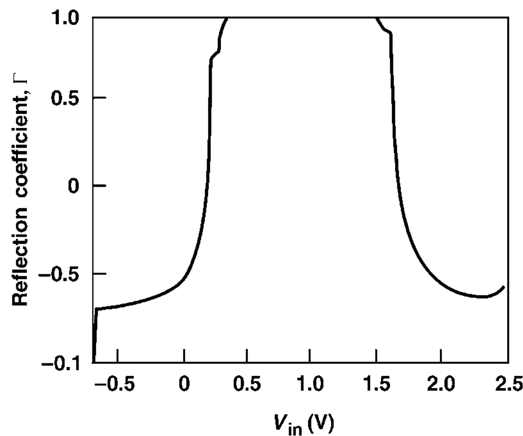



**Figure 7.25** Reflection coefficient of an active clamp network.

For the condition where $V_{in} < V_{SS}$, the reflection coefficient of the circuit is near ideal (e.g. $\Gamma = -1$). This corresponds to a low-impedance state and allows current to flow rapidly from the clamp circuit (e.g. p-channel MOSFET) to the input node to prevent undershoot. For $V_S < V_{in} < V_{DD}$, the reflection coefficient is near unity [30, 31].

A latchup circuit design practice is as follows:

- Active clamp networks can be utilized to reduce undershoot and overshoot on signal pads at the power supply rails prior to a forward biasing of any other circuit element.
- Active clamp networks can be used to 'sink' overshoot and undershoot voltage excursions outside of the $V_{SS}$ to $V_{DD}$ functional voltage range.
- Active clamp networks can be utilized to 'sink' overshoot and undershoot signal currents to the power supply rails (e.g. 10–100 mA levels).

## 7.7.3 Dynamic Threshold Triple-Well Passive and Active Clamp Networks

The third class of clamp networks that can be utilized in triple-well CMOS technology is the dynamic threshold triple-well clamp networks. Triple-well CMOS technology has an advantage over twin-well bulk CMOS in that the n-channel MOSFET substrate is not bound to the substrate potential. As a result, the triple-well CMOS can construct symmetric circuits and apply dynamic threshold techniques. As a result, active clamp networks can use complementary methods for the n-channel and p-channel transistors to address positive and negative voltage and current excursions [31].

Dynamic threshold voltage MOSFETs provide a higher transconductance ($g_{msat}$), a higher drain saturation current ($I_{dsat}$), a dynamic threshold voltage that can be used as a lower voltage triggering device, a high $I_{on}/I_{off}$ ratio and are scalable to sub-0.7 V power supply voltages. Figure 7.26 shows an example of a dynamic threshold active clamp. This network will limit voltage excursions in triple-well environments.

Figure 7.27 shows a modified version of the bulk CMOS active clamp network. In this case, the bodies of the network are electrically coupled to the input. In the case of triple-well technology, this leads to the current injection into the triple-well structures and not the chip substrate.

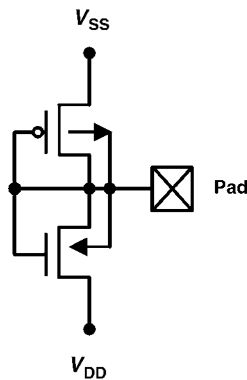



**Figure 7.26** Dynamic threshold triple-well body- and gate-coupled ESD network.

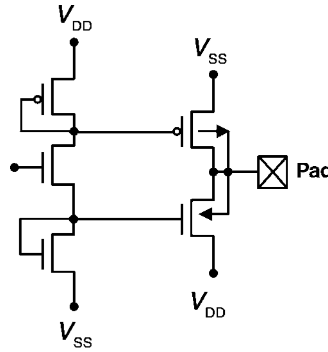



**Figure 7.27** Dynamic threshold triple-well active clamp network.

A latchup circuit design discipline incorporates the following:

- Anti-overshoot and anti-undershoot networks can minimize system-level transients and reflections.
- Anti-overshoot and anti-undershoot prevent the forward biasing of semiconductor devices used in semiconductor peripheral circuits.
- Active clamp networks provide a means of discharging the transient currents to the power rails and avoid substrate and well minority carrier injection.

## 7.8 PASSIVE AND ACTIVE GUARD RINGS

Latchup and noise are initiated in the substrate from overshoot and undershoot phenomena. These can be generated by CMOS OCD circuitry, receiver networks and ESD devices. Unfortunately, parasitic pnpns are often unrecognized or anticipated.

Today, system-on-chip (SOC) solutions have been used for solving the mixed signal and RF requirements. SOC applications have a wide range of power supply conditions, independent power domains and circuit performance objectives. Different power domains are established between digital, analog and RF functional blocks on an integrated chip. The integration of different circuits and system functions into a common chip has also resulted in solutions for ensuring that noise from one portion or circuit of the chip does not affect a different circuit within the chip [33–41]. With the chip integration issues, the need for better guard rings and alternative guard ring solutions has had increased interest [39–52]. In the mid-1990s, there was a focus in semiconductor chip design to achieve both noise isolation and ESD protection. This was achieved by establishing isolated power rails, which were then electrically reconnected with ESD solutions. Since 2000, there has been an increased focus on guard ring solutions that achieve the following objectives:

- solutions that achieve noise isolation, latchup robustness and ESD results;
- solutions that do not inject current back into the power grid.

With the growth of interaction between digital, analog and RF domains, guard ring concepts have increased in importance. In addition, with the growth of smart power technology, solutions are needed

for avoidance of interaction of the high-voltage CMOS (HVCMOS) chip sectors and the low-voltage sectors of a CMOS chip [39–59]. In smart power technology, the concern of latchup between the LDMOS power devices and low-voltage CMOS has been a focal point for the development of new guard ring concepts from 1990 to present day [39–59]. New guard ring solution for smart power and BiCMOS applications were proposed by the following teams: Bafluer *et al*. [41, 43], Peppiette [42], Winkler and Herzl [45], Gonnard and Charitat [46], Zhu *et al.* [47], Gonnard *et al.* [48], Schenkel *et al.* [49], Parthasarathy *et al.* [50, 51], Laine *et al.* [52], Khemka *et al.* [53], Horn [54], Voldman *et al.* [55], Gupta *et al.* [56], Stella *et al.* [57], Singh and Voldman [58, 59], and Zhu *et al.* [60]. In the next sections, examples of both the passive and active guard ring circuits and structures will be shown.

## 7.8.1 Passive Guard Ring Circuits and Structures

Passive guard ring structures can be used in an SOC application by the utilization of process features. Passive guard rings are valuable between digital, analog and RF domains to avoid latchup. In a BiCMOS technology with trench technology, passive guard ring elements are as follows:

- n-well rings;
- subcollector rings;
- n-well ring and deep trench (DT) ring;
- plurality of n-well and DT rings;
- plurality of deep trench rings and high-dose buried layer (HDBL).

With the integration of trench structures and implants, p−/p++ substrates and HDBLs can be used to influence the vertical and lateral transport between domains.

In smart power applications, technologies integrate bipolar, CMOS and DMOS devices (e.g. referred to as BCD technology). In BCD smart power applications that integrated high-voltage DMOS with low-voltage CMOS, passive guard ring solutions are utilized between the different power domains. These solutions are as follows:

- DeMOS implant guard rings;
- LDMOS n-well and n-body guard rings;
- LDMOS n-buried layer (e.g. n-tub) guard rings;
- p− epitaxy/p++ substrate wafers;
- heavily doped buried layer implants below the LDMOS n-tub regions.

## 7.8.2 Active Guard Ring Circuits and Structures

One of the problems with diffused junction ‘passive guard rings’ is that in order for the metallurgical junction based guard rings to improve CMOS latchup, current enters the semiconductor chip power grid. The guard ring improves the latchup tolerance; however, the overshoot noise that can initiate the latchup is injected into the $V_{SS}$ ground rail (through the vertical pnp), and possibly spreads to other

circuits. Solutions for improving latchup tolerance have been used; however, these circuits introduce noise into the power rails (e.g. $V_{DD}$ or $V_{SS}$). In digital, analog and RF chip applications, a circuit solution that improves latchup tolerance and at the same time limits the amount of the current injection or noise introduced into the power rails is valuable to achieve both CMOS latchup objectives and noise objectives.

In HVCMOS technology, a concern is the interaction of the LDMOS transistors and the adjacent low-voltage CMOS circuitry [39, 40]. In HVCMOS, inductive 'load dumps' initiate injection of minority carriers in the chip substrate. As a result of the physical size of the HVCMOS LDMOS devices, as well as the magnitude of the current injection, it is critical not to disturb the other chip functions on the smart power chips.

Different 'active' guard ring circuit concepts have been introduced for latchup improvement [41–52, 56–60]. In 'active' guard rings, the objective is not only to collect minority carriers, but to actively compensate the effect as well. The latchup circuit design discipline includes the following concepts:

- Electrically collecting minority carriers at a metallurgical junction that is electrically connected to the chip substrate to alter the substrate potential.
- Electrically collecting minority carriers at a metallurgical junction that is electrically connected to the chip substrate to alter the substrate potential, with the objective of reducing forward bias of the injection structure.
- Electrically collecting minority carriers at a metallurgical junction that is electrically connected to the chip substrate to alter the substrate potential, with the objective of introducing a lateral electrical field assist to reduce the lateral bipolar current gain.
- Electrically sensing the substrate potential drop and inverting the polarity of the potential drop using inverting amplifier networks.

Figure 7.28 is an example of an active guard ring. Typically, in a passive guard ring concept, a p+ substrate contact is electrically connected to a $V_{SS}$ power rail and an n-well ring is electrically connected to a $V_{DD}$ power rail. But in an active guard ring, an n-well region is not electrically

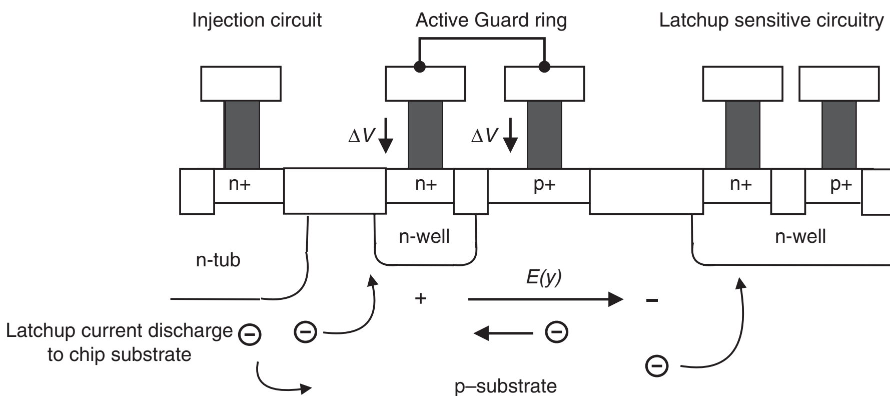



**Figure 7.28** Active guard ring with n-well junction electrically connected to a p+ substrate contact to establish lateral opposing electric field.

connected to a power rail. In an active guard ring, the n-well structure collects the minority carrier electrons in its metallurgical junction formed with the p− substrate region. The n-well ring is electrically connected to a 'soft grounded' p+ substrate contact. When the minority carrier electrons traverse the metallurgical junction, the electrical potential of the n-well region is reduced (e.g. denoted in the figure as $\Delta V$). By electrically connecting the n-well to the p+ substrate contact, the electrical potential of the substrate is also lowered by the same potential magnitude. In this case, the electrical potential of the region is lowered. The lowering of the substrate potential can be utilized by two means. First, given the p+ diffusion is near a forward biased structure (e.g. an injecting structure), the reduction of the potential can lower the forward bias state, turning off the injection process. Second, given another p+ substrate contact, a lateral electric field can be established that inhibits the flow of minority carriers. Given a parasitic npn bipolar transistor is formed between the injection source and a collecting victim circuit, if the lateral electrical field opposes the current transport, the lateral npn bipolar current gain is reduced. From Chapter 2, the derivation of 'electric field assist' gives the transport solution, wherein the electric field reduces the lateral bipolar current gain. In this methodology, the placement of the p+ region can be on the injection side or collection side of the n-well region. Gupta *et al.* [56] added an additional p++ substrate diffusion inside of the n-well ring/p+ substrate contact to establish a well-defined lateral electric field where a p+ region is flanking both sides of the n-well ring, with the outer p+ substrate contact electrically connected to the n-well ring. In the second implementation of Gupta *et al.* [56], two p+ diffusions are placed within the n-well ring, forming three p+ substrate rings and a single n-well ring. In this case, the two 'floating' p+ diffusions flanking the n-well are electrically connected to the n-well and the third p+ diffusion is shorted to a power rail $V_{SS}$.

Various implementations of guard rings are utilized, where a plurality of p+ substrate contacts and n-wells are integrated, mixing both the active and passive concepts, where some of the wells are 'floating' and some electrically connected to the power supplies. In these implementations, a plurality of trench structures can also be added to reduce the lateral bipolar current gain. In all cases, as the number of additional guard rings is increased, as well as the effective base width, the bipolar current gain decreases. In the work of Gupta *et al.*, it was shown that without a guard ring, the lateral bipolar current gain decreased from $2 \times 10^2$ to $1 \times 10^{-1}$ as the base width increased from 20 to 500 μm; with an unbiased guard ring, a bipolar current gain of $5 \times 10^{-3}$ was achieved at a spacing of 20 μm [56]. The measurement of Gupta *et al.* demonstrates that greater than a three order of magnitude of improvement was achieved with the active guard ring concept.

The concept of the active guard ring can also be integrated with passive guard ring solutions. Figure 7.29 shows the structure of Figure 7.28 with an additional n-well region. Electrons that bypass the active guard rings can be collected by the second n-well structure.

In order to improve the guard ring efficiency, additional solutions can be added to prevent the migration of injected carriers. Gupta *et al.* [56] and Stella *et al.* [57] added additional implant structures to the active guard ring concept to enhance the guard ring depth, as well as restrict the minority carriers' vertical transport [56, 57]. Figure 7.30 shows a guard ring concept that includes the n-well and an n-doped buried layer underneath the n-well to deepen the guard ring surface area. The n-doped buried layer can be a subcollector, an HDBL implant, or an LDMOS n-tub or deep n-well implant. In addition, a p−/p++ substrate wafer is used to restrict the vertical trajectory of the minority carrier. Hence, the following concepts are introduced in this implementation:

- n-buried layer under the n-well to provide an increase in the guard ring collection area;
- an n-well/n++ buried layer region electrically connected to a p++ substrate contact to modulate the substrate potential and induce an opposing lateral electric field;

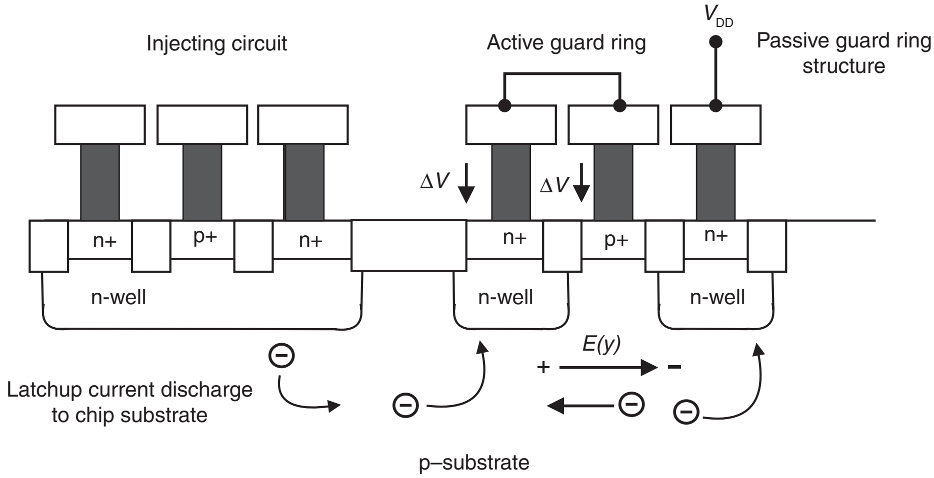



**Figure 7.29** Active guard ring with n-well junction electrically connected to a p+ substrate contact to establish lateral opposing electric field and passive n-well guard ring.

- a low–high (L–H) p−/p++ step junction to induce a vertical electric field to (a) decrease transport deep under the guard ring structure and (b) increase the current density in the region between the p− epitaxial region.

The first concept increases the guard ring efficiency with the deeper guard ring structure. The second concept provides a means of lowering the lateral transport with the introduction of the lateral electrical field. Third, the p−/p++ step junction reflecting boundary redirects the minority carrier toward the guard ring structure for collection. Fourth, the combination of the p−/p++ step junction and the guard ring

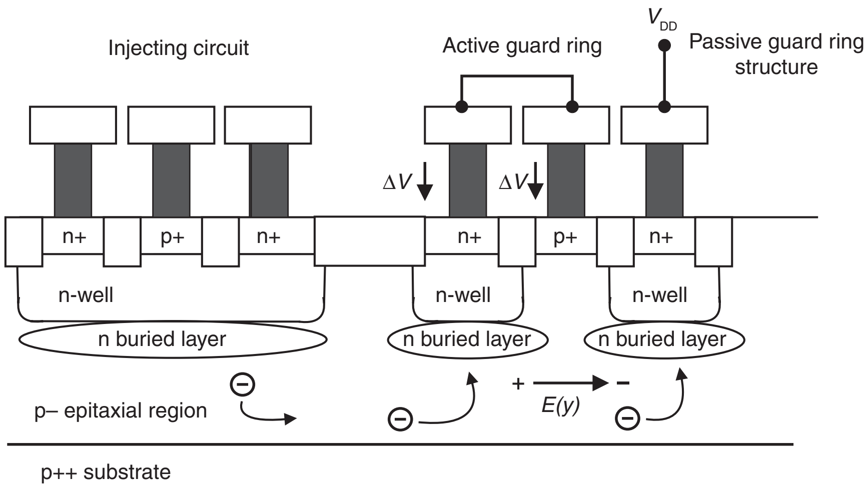



**Figure 7.30** Active guard ring with n-well, a n+ buried layer and a passive guard ring on a p−/p++ substrate wafer.

region restricts the carrier flow in the epitaxial region. At high currents, this lowers the parasitic bipolar current gain as well. Stella *et al.* noted that in this structure, due to the p−/p++ reflecting barrier and collection by the active guard ring, the characteristic decay length $L_\mathrm{d}$ can be defined as [57]

$$L_\mathrm{d} = \frac{1}{\sqrt{\left(\frac{1}{L_\mathrm{n}}\right)2 + \left(\frac{\pi}{2T_\mathrm{p}}\right)2}},$$

where $L_\mathrm{n}$ is the diffusion length and $T_\mathrm{p}$ is the epitaxial thickness between the n-well guard ring and the p−/p++ interface.

Another approach is the active compensation through inversion of the signal to nullify its effect. A circuit solution to address both noise and CMOS latchup is the usage of 'active guard rings'. Winkler and Herzl introduced this concept to eliminate noise in mixed signal chips [45]. An active guard ring circuit is a circuit that senses an undershoot phenomenon in the chip substrate. The concept of the 'active guard ring' is to sense the negative change in the substrate potential; this negative signal is fed as an input into an inverting amplifier. The output of the inverting amplifier is then electrically connected to the substrate. In this fashion, the negative polarity undershoot voltage is inverted and amplified, whose signal is then reinjected into the same electrical region to nullify the undershoot voltage transient state [45, 58, 59].

Figure 7.31 shows the p+ noise suppression element electrically connected to the input of a differential amplifier. Two resistors are formed across the differential amplifier between the input and output of an amplifier, where the resistors use the silicon substrate itself. In addition, an active noise suppression circuit is also added with an input connected to p-region and an output connected to substrate contact. In the differential amplifier, one input is electrically connected to the p+ noise suppression element and the second input is at electrically connected to ground [58, 59].

Figure 7.32 shows an example of a capacitance-coupled implementation with ESD protection [58,59]. In this case, the differential input is coupled to the p+ guard ring through a capacitor. An *RC* network is formed on the differential input node. In addition, the differential amplifier output is electrically connected to the resistor feedback network through a capacitor on the output node, forming an *RC* network on the output node.

In the presence of multiple 'victim' circuits, there is a plurality of potential circuits that may be influenced by a single injection source. In the case of multiple collection regions, the parasitic bipolar npn transistor formed between the injector source and the victim circuit is a function of the victim circuit 'collector' area and relative distance to the injection source (e.g. effective base width). Each circuit will

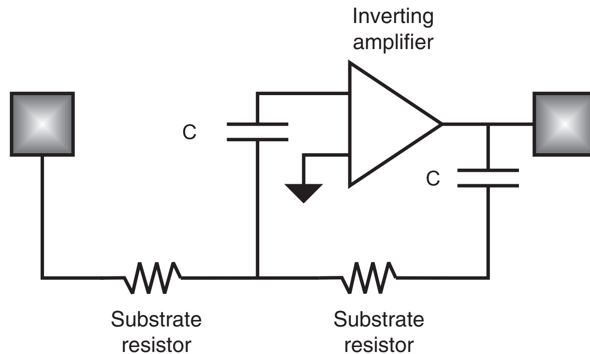



**Figure 7.31** Active guard ring noise and latchup suppression network.

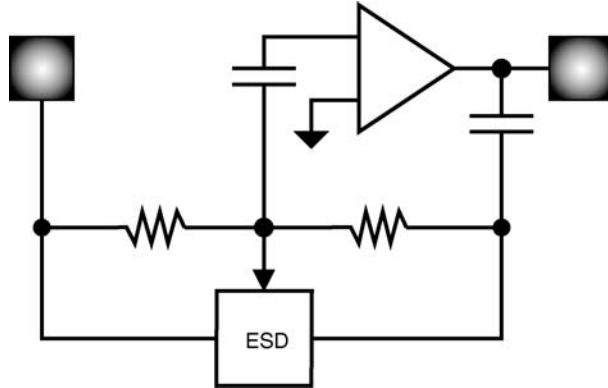



**Figure 7.32** Active guard ring noise and latchup suppression circuit with ESD network.

see a different level of current injection; hence, the level of active compensation will be different for each circuit. Compensation-based active guard ring structures can increase in sophistication with additional analog function circuits to provide the circuit compensation techniques [54]. Additionally, a key case of interest is that of large injection structures near large collecting structures. In this case, when the space between the injector and the collector is small, the distribution of the currents over the areas must address the spatial distances between points in the injector and the collector.

In the compensating active guard ring concept, the compensation current can be evaluated with the evaluation of segments of the 'collecting structures'. Compensating active guard rings are of interest for the case of large injecting sources and large collecting regions where the spacing between the two regions is small. There are many examples of these situations. A first example is the case of Huh *et al.*, as discussed in this chapter, where the injecting source was large I/O MOSFET OCD and the collecting structure was the n-well of decoupling capacitor network. In smart power, the injection from a DMOS injection transistor to the second adjacent collecting structure at close proximity can be estimated as a one-dimensional transport that has an exponential decay characteristic spatially. Horn expressed the injection in the form as follows [54]:

$$I_s(x) = \iint_{A_C} k_1 \exp\{-k_2' x\},$$

where $I_S$ is the current collected at the 'collector', $A_C$ is the collector area, and $k_1$ and $k_2$ are coefficients. The first coefficient $k_1$ is proportional to the injection current and diffusion properties, and the second coefficient $k_2$ is proportional to the inverse of the diffusion length in the $x$ direction [54].

This development was generalized in two dimensions, within the plane of the wafer. In the $x$ and $y$ dimensions, the equation is defined as

$$I_S(x, y) = \iint_{A_C} k_1 \exp\{-k_2 x - k_3 y\} \mathrm{d}x\mathrm{d}y,$$

$$I_S(x, y) = \iint_{A_C} k_1 \mathrm{e}^{-(k_2 x - k_3 y)} \mathrm{d}x\mathrm{d}y.$$

Given the collection region is a rectangle in the $x$ and $y$-dimensions, the derivation can be expressed as the integration over the $x$ and the $y$ directions as follows:

$$I_S = \int_{y_1}^{y_2}\int_{x_1}^{x_2} k_1 e^{-(k_2 x + k_3 y)} dx dy.$$

Placing the derivation in another form, there exists $x = x_n$, and $y = y_n$, where

$$\int_{y1}^{y2}\int_{x1}^{x2} f(x,y) dx dy = \frac{\int_{x1}^{x2} f(x,y_n) dx \int_{y1}^{y2} f(x_n,y) dy}{f(x_n,y_n)}.$$

Horn defined the following terms [54], where the first term is the current along the dimension $y = y_2$, by taking the integral of the collection over the $x$ dimension,

$$I_A = \int_{x1}^{x2} f(x,y_2) dx$$

and the second current is the integral along $x = x_2$, integrating over the $y$ dimension,

$$I_B = \int_{y1}^{y2} f(x2,y) dy,$$

and the current at the point $x = x_2$ and $y = y_2$ is

$$I_C = f(x_2,y_2).$$

In this form, the minority current collected can be expressed as

$$I_S = \int_{y_1}^{y_2}\int_{x_1}^{x_2} k_1 e^{-(k_2 x + k_3 y)} dx dy = \frac{I_A I_B}{I_C}.$$

Hence, an active compensation current can be evaluated from the terms $I_A$, $I_B$ and $I_C$. As shown by Horn [54], this method can be modified by using a strip in the $x$ direction from $x_3$ to $x_4$ (e.g. $x_3 < x < x_4$) and a second strip in the $y$ direction, $y_3$ to $y_4$ (e.g. $y_3 < y < y_4$). It can be show that

$$\int_{y_1}^{y_2}\int_{x_1}^{x_2} f(x,y) dx dy = \frac{\left\{\int_{y_3}^{y_4} dy \int_{x_1}^{x_2} (x,y_n) dx\right\}\left\{\int_{x_3}^{x_4} dx \int_{y_1}^{y_2} f(x_n,y) dy\right\}}{\int_{y_3}^{y_4}\int_{x_3}^{x_4} f(x,y) dx dy},$$

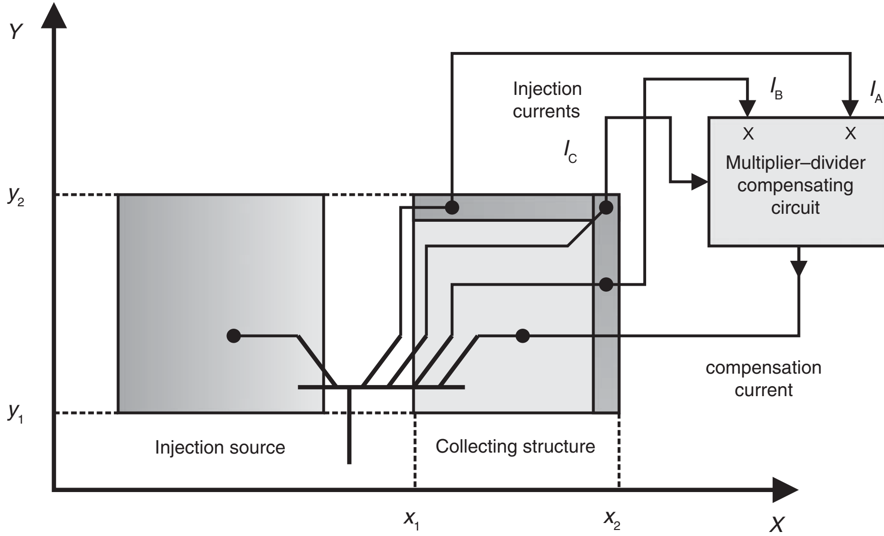



**Figure 7.33** Compensation-based active guard ring with analog multiplier–divider network.

where we redefine the three currents for the compensation current for the active guard ring according to the following relationships:

$$I_A = \int_{y3}^{y4}\int_{x1}^{x2} f(x, y)\mathrm{d}x\mathrm{d}y,$$

$$I_B = \int_{y1}^{y2}\int_{x3}^{x4} f(x, y)\mathrm{d}x\mathrm{d}y$$

and

$$I_C = \int_{y3}^{y4}\int_{x3}^{x4} f(x, y)\mathrm{d}x\mathrm{d}y,$$

where

$$\int_{y1}^{y2}\int_{x1}^{x2} f(x, y)\mathrm{d}x\mathrm{d}y = \frac{I_A I_B}{I_C}.$$

Figure 7.33 shows an example of a compensation-based active guard ring concept that introduces a multiplier–divider network. Each victim circuit is compensated based on its injection current. In the figure, the three currents $I_A$, $I_B$ and $I_C$ are shown.

A latchup circuit design practice is as follows:

- Passive guard rings are used to provide a means of collecting minority carrier current to the power supply rails.

- Passive guard rings can be used to lower the parasitic bipolar transistor current gain formed in the parasitic transistors between an injection source and a collection region.
- Active guard rings provide a means of collecting minority carriers and not injecting the current into the power supply rail, and reestablish the substrate potential.
- Active guard rings provide a means of introducing lateral electric fields to minimize transport between an injection source and a collecting region.
- Compensating active guard rings can process the injection signal and provide a means to neutralize the injection to reestablish the substrate potential.

## 7.9 TRIPLE-WELL NOISE AND LATCHUP SUPPRESSION STRUCTURES

Other methods to avoid noise and latchup interaction in circuits exist in triple-well technology and BiCMOS technology. Physical structures such as triple-well implants and high-dose buried layers can improve the latchup immunity of digital devices while providing isolation structures that provide noise isolation for both the digital and analog devices. HDBL or subcollector implants can be used to isolate physical chip sectors. These can be integrated with arrays of deep trench structures; an array of deep isolation trenches provides increased isolation between devices where needed.

## 7.10 SYSTEM-LEVEL ISSUES

Latchup can occur within systems leading to system-level failure. Latchup issues can occur in laptops, servers, personal computers, cell phones, handheld electronics to power systems. System-level latchup problems can occur as a result of the following interactions:

- improper power sequencing of system-level connections;
- system-level noise;
- transient systems;
- cable discharge events;
- board-to-chip capacitive coupling leading to overshoot and undershoot events on the semiconductor chip;
- semiconductor chips with low latchup immunity;
- improper assembly leading to reverse pin voltages (e.g. reverse polarity on electrical pin);
- improper installation of battery (e.g. reverse polarity of battery);
- poor grounding of ground supply;
- inductive load dumps;
- board-level inductance mismatch in a multiple-component system;
- bond-wire (wire-bond) mismatch in a multiple-component parallel system;

- bond-wire (wire-bond) bond breakage;
- nonuniform thermal shunting of power.

Solutions for system-level CMOS latchup can include the following:

- board-level diode anti-overshoot clamps;
- board-level diode anti-undershoot clamps;
- board-level active clamp networks;
- board-level spark gaps;
- voltage suppression cable elements (e.g. cables whose resistance increase at high current);
- cabling with discharge connectors (e.g. BNC connectors with shorting means);
- system-level 'touch pad' to discharge cables and personnel;
- Board-level transient voltage suppression devices (e.g. polymer voltage suppression shunt elements);
- sockets that prevent battery reversal;
- sockets that prevent negative polarity reversal;
- latchup robust semiconductor products;
- latchup-proof switches (e.g. SOI technology) [61];
- fault-protected switches [61];
- channel protector clamps [61].

## 7.11 SUMMARY AND CLOSING COMMENTS

This chapter focused on CMOS latchup and circuits, from circuit problems to circuit solutions. Today, there are hundreds of patent inventions in the area of circuits for latchup. In this chapter, the objective was to provide a small sample of the circuit art that exists today.

In the next chapter, latchup design rule and design integration are addressed. Chapter 8 will focus on latchup design rule checking (DRC) methodologies, graph model theory methods, parasitic extraction as well as verification CAD tools on a local and on a global perspective.

## PROBLEMS

1. In an integrated circuit, subfunctions necessary to construct a microprocessor design include I/O books, on-chip cache SRAM memory, core logic, PLL and DLL circuits. In addition, decoupling capacitors are used in between chip sectors. List the possible layout combinations that can occur.
2. In an integrated circuit, subfunctions necessary to construct a microprocessor design include I/O books, on-chip cache SRAM memory, core logic, PLL and DLL circuits. In addition, decoupling capacitors are used in between chip sectors. List the possible power rail to power rail combinations that can occur, which may lead to a CMOS latchup concern.

3. In a mixed signal chip, there are digital, analog and RF chip sectors. In the analog chip sector, the circuits may contain bipolar, CMOS or BiCMOS circuits. In the RF chip sector, npn bipolar transistors exist. What are the possible interactions between the chip sectors? What are the possible pnpn structures that can be formed?
4. In 'complementary bipolar', there are both pnp transistors and npn transistors in a common process. What are possible parasitics when the pnp and npn transistors interact that can lead to latchup? Given complementary BiCMOS, what are the possible parasitics formed between the pnp, the npn and the CMOS logic circuitry?
5. In a triple-well technology, what are the potential chip function interactions? Is triple well better or worse than dual-well CMOS for subfunction to subfunction latchup interactions?
6. Given a chip where there is a large digital chip sector, a smaller analog sector and very small RF sector, assume each is on a separate power supply. Assume the digital and analog sectors have the same $V_{DD}$ voltage levels but different power rails, and assume that the RF sector is at a higher $V_{CC}$ voltage. Assume the capacitance of each chip sector is proportional to its chip area. How would you sequence the semiconductor chip? Develop a circuit using circuit comparators and switches for the three power rail systems. Provide a solution extending the patent of Lin *et al*. [22].
7. Using a p-channel MOSFET well bias network, show the states of the input relative to the $V_{DD}$ power supply. Using a p-channel MOSFET well bias network, show each state when used between two power supply rails.
8. An insulated gate bipolar transistor (IGBT) device is constructed on a p+ substrate. The structure consists of an enclosed n+ MOSFET drain, within a MOSFET polysilicon gate structure. The n+ doped MOSFET drain is in a p-body region. The p-body region is contained within an n-region serving as the MOSFET source. The MOSFET gate structure overlaps the p-body and n-region providing a MOSFET lateral channel within the p-body. The n-source region is on a p+ substrate region. A vertical bipolar pnp is formed between the p+ source and the p-body region, and the n-region serves as the vertical bipolar pnp base region. The n-channel MOSFET serves as a switch and base current drive to the vertical pnp structure. Assume a load in series with the vertical pnp transistor. Draw the vertical cross section of the device and the circuit schematic. Show the existence of the vertical parasitic npn transistor and the parasitic pnpn device.
9. Assume $N$ parallel IGBT devices are electrically connected in parallel through wire bonds. Show how a wire-bond breakage can lead to nonuniform current distribution within a system. Assume $N$ IGBT devices in parallel, electrically connected to an inductor. Assume that the inductance of each inductor is not identical. Show how this leads to nonuniform current distribution and latchup *(Hint:* $DV = (L_i - L_j)\mathrm{d}I/\mathrm{d}t$).

## REFERENCES

1. C.T. Hsu, I.C. Lin, J.C. Tseng, M.D. Ker, Y.L. Chen, F.Y. Tsai, S.H. Yu, F.H. Chen and P.A. Chen. A pin latchup failure and the latch-up trigger current induced npn snapback effect in a high-voltage IC product. *Proceedings of the Taiwan Electrostatic Discharge Conference (T-ESDC)*, 2006. p. 53–57.
2. J. Salcedo-Suner, R. Cline, C. Duvvury, A. Cadena-Hernandez, L. Ting and J. Schichl. A new I/O signal latch-up phenomenon in voltage tolerant ESD protection circuits. *Proceedings of the International Reliability Physics Symposium (IRPS)*, 2003. p. 85–91.
3. Y. Huh, K. Min, P. Bendix, V. Axelrad, R. Narayan, J.W. Chen, L.D. Johnson and S. Voldman. Chip level layout and bias considerations for preventing neighboring I/O cell interaction-induced latchup and inter-power supply

latchup in advanced CMOS technologies. *Proceedings of the Electrical Overstress/Electrostatic Discharge (EOS/ESD) Symposium*, 2005. p. 100–107.

4. Y. Huh, K. Min, P. Bendix, V. Axelrad, R. Narayan, J.W. Chen, L.D. Johnson and Steven H. Voldman. Inter-circuit and inter-power supply CMOS latchup. *Proceedings of the Taiwan Electrostatic Discharge Conference (T-ESDC)*, 2005. p. 32–36.
5. A. Weger, S. Voldman, F. Stellari, P. Song, P. Sanda and M. McManus. A transmission line pulse (TLP) picosecond imaging circuit analysis (PICA) methodology for evaluation of ESD and latchup. *Proceedings of the International Reliability Physics Symposium (IRPS)*, 2003. p. 99–104.
6. C.J. Brennan, K. Chatty, J. Sloan, P. Dunn, M. Muhammad and R. Gauthier. Design automation to suppress cable discharge events (CDE) induced latchup in 90 nm ASICs. *Proceedings of the Electrical Overstress/Electrostatic Discharge (EOS/ESD) Symposium*, October 2005. p. 126–130.
7. S. Voldman. *ESD: Physics and Devices.* Chichester, UK: John Wiley & Sons, Ltd 2004.
8. S. Voldman. *ESD: Circuits and Devices.* Chichester, UK: John Wiley & Sons, Ltd 2005.
9. S. Voldman. *ESD: RF Technology and Circuits.* Chichester, UK: John Wiley & Sons, Ltd 2006.
10. S. Voldman. BiCMOS ESD circuit with sub-collector/trench-isolated body MOSFET for mixed signal analog/digital RF applications. US Patent No. 6,455,902 (September 24, 2002).
11. J. Pequignot, J.H. Sloan, D.W. Stout and S. Voldman. Electrostatic discharge protection networks for triple well semiconductor devices. US Patent No. 6,891,207, (May 10, 2005).
12. J. Lundberg. Low voltage CMOS output buffer. US Patent No. 4,963,766 (October 16, 1990).
13. R.D. Adams, R.C. Flaker, K.S. Gray and H.L. Kalter. CMOS off-chip driver circuits. US Patent No. 4,782,250 (November 1, 1988).
14. R.D. Adams, R.C. Flaker, K.S. Gray and H.L. Kalter. An 11ns 8K $\times$ 18 CMOS static RAM. *Proceedings of the International Solid State Circuits Conference (ISSCC)*, 1988. p. 242–243.
15. J.S. Austin, R.A. Piro and D.W. Stout. CMOS off chip driver circuit. US Patent No. 5,151,619 (September 29, 1992).
16. J. Hoffman, D. Jallice, Y. Puri and R. Richards. CMOS off chip driver for fault tolerant cold sparing. U.S. Patent No. 5,117,129 (May 26, 1992).
17. D.W. Dobberpuhl. Floating-well CMOS output driver. US Patent No. 5,160,855 (November 3, 1992).
18. S. Voldman. ESD protection in a mixed voltage interface and multi-rail disconnected power grid environment in 0.5- and 0.25-μm channel length CMOS technologies. *Proceedings of the Electrical Overstress/Electrostatic Discharge (EOS/ESD) Symposium*, 1994. p. 125–134.
19. S. Voldman. Electrostatic discharge protection circuits for mixed voltage interface and multi-rail disconnected power grid applications. US Patent No. 5,945,713 (August 1, 1999).
20. S. Voldman, Power sequence-independent electrostatic discharge protection circuits. US Patent No. 5,610,791 (March 11, 1997).
21. S. Voldman. A method, apparatus, and circuit for latchup suppression in a gate-array ASIC environment. US Patent No. 7,102,867 (September 5, 2006).
22. S.T. Lin, T.L. Yu and Y.C. Peng. Latch-up protection circuit for integrated circuits biased with multiple power supplies and its method. US Patent No. 6,473,282 (October 29, 2002).
23. L. DeClue and H. Muller. Active termination network for clamping a line signal. US Patent No. 3,937,988 (February 10, 1976).
24. E. Davidson and R. Lane. Low power transmission line terminator. US Patent No. 4,015,147 (March 29, 1977).
25. G. Slaughter. Non-reflecting transmission line termination. US Patent No. 4,943,739 (July 24, 1990).
26. J. Kosson. Voltage clamping with high current capability. US Patent No. 4,958,093 (September 18, 1990).
27. C. Petersen. High speed anti-undershoot and anti-overshoot circuit. US Patent No. 5,103,118 (April 7, 1992).
28. A. Furman. Zero power transmission line terminator. US Patent No. 5,227,677 (July 13, 1993).
29. E. Honningford. CMOS input voltage clamp. US Patent No. 5,528,190 (June 18, 1996).
30. S. Voldman and D. Hui. Switchable active clamp network. US Patent No. 6,075,399 (June 13, 2000).
31. S. Voldman. Silicon-on-insulator body- and dual gate-coupled diode for electrostatic discharge (ESD) applications. US Patent No. 6,034,397 (March 7, 2000).
32. B.W. Mashak, R.R. Williams, S. Voldman and D. Hui. Active clamp network for multiple voltages. US Patent No. 6,229,372 (May 8, 2001).

33. T. Li, C.H. Tsai, E. Rosenbaum and S.M. Kang. Substrate modeling and lumped substrate resistance extraction for CMOS ESD/Latchup circuit simulation. *Proceedings of the 36th ACM/IEEE Conference on Design Automation*, June 1999. p. 549–554.
34. B. Owens, B. Birrer, P. Adluri, S. Shreeve and R. Arunachalam. Strategies for simulation, measurement and suppression of digital noise in mixed-signal circuits. *Proceedings of the IEEE 2003 Custom Integrated Circuits Conference (CICC)*, September 2003. p. 361–364.
35. A. Koukab, K. Banerjee and M. Declercq. Analysis and optimization of substrate noise coupling in single-chip RF transceiver design. *Proceedings of the 2002 IEEE/ACM International Conference on Computer-Aided Design*, November 2002. p. 309–316.
36. H.H.Y. Chan and Z. Zilac. A practical substrate modeling algorithm with active guardband macro-model for mixed-signal substrate coupling verification. *Proceedings of the Eighth IEEE International Conference on Electronics, Circuits, and Systems*, September 2001. p. 1455–1460.
37. M. Nagata, J. Nagai and T. Morie. Quantitative characterization of the substrate noise for physical design guides in digital circuits. *Proceedings of the IEEE Custom Integrated Circuits Conference (CICC)*, May 2000. p. 95–98.
38. T. Li and S.M. Kang, Layout extraction and verification methodology for CMOS I/O circuits, *Proceeding of the Design Automation Conference (DAC)*, June 1998. p. 291–296.
39. B. Murari. Power integrated circuits: problems, tradeoffs, and solutions. *IEEE Journal on Solid State Circuits*, **13**(3), 1978, 307–319.
40. T.P. Chow, D. Patttanayak, B.J. Baliga, M.S. Adler, M.S. Hennessey and W.A. Logan. Interaction between monolithic junction-isolated lateral insulated-gate bipolar transistors. *IEEE Transactions on Electron Devices*, **38**(2), 1991, 310–315.
41. M. Bafluer, J. Buxo, M.P. Vidal, P. Givelin, V. Macary and G. Sarrabayrouse. Application of a floating well concept to a latchup-free low-cost smart power high-side switch technology. *IEEE Transactions on Electron Devices*, **40**(7), 1993, 1340–1342.
42. R. Peppiette. A new protection technique for ground recirculation parasitics in monolithic power IC's. *Sanken Technical Report*, **26**(1), 1994, 91–97.
43. M. Bafluer, M.P. Vidal, J. Buxo, P. Givelin, V. Macary and G. Sarrabayrouse. Cost-effective smart power CMOS/DMOS technology: design methodology for latchup immunity. *Analog Integrated Circuits and Signal Processing*, **8**(3), 1995, 219–231.
44. W.W.T. Chan, J.K.O. Sin and S.S. Wong. A novel crosstalk isolation structure for bulk CMOS power IC's. *IEEE Transactions on Electron Devices*, **45**(7), 1998, 1580–1586.
45. W. Winkler and F. Herzl, Active substrate noise suppression in mixed-signal circuits using on-chip driven guard rings. *Proceedings of the IEEE 2000 Custom Integrated Circuits Conference (CICC)*, May 2000. p. 356–360.
46. O. Gonnard and G. Charitat. Substrate current protection in smart power IC's. *Proceedings of the International Symposium on Power Semiconductor Devices (ISPSD)*, 2000. p. 169–172.
47. R. Zhu, V. Parthasarathy, V. Khemka and A. Bose. Implementation of high-side, high-voltage RESURF LDMOS in a sub-half micron smart power technology. *Proceedings of the International Symposium on Power Semiconductor Devices (ISPSD)*, 2001. p. 403–406.
48. O. Gonnard, G. Charitat, P. Lance, M. Susquet, M. Bafluer and J.P. Laine. Multi-ring active analogic protection (MAAP) for minority carrier injection suppression in smart power IC's. *Proceedings of the International Symposium on Power Semiconductor Devices (ISPSD)*, 2001. p. 351–354.
49. M. Schenkel, P. Pfaffli, W. Wilkening, D. Aemmer and W. Fichtner. Transient minority carrier collection from substrate in smart power design. *Proceedings of the European Solid State Device Research Conference (ESSDERC)*, 2001. p. 411–414.
50. V. Parthasarathy, R. Zhu, V. Khemka, T. Roggenbauer, A. Bose, P. Hui, P. Rodriguez, J. Nivison, D. Collins, Z. Wu, I. Puchades and M. Butner. A 0.25-μm CMOS based 70V smart power technology with deep trench for high-voltage isolation. *International Electron Device Meeting (IEDM) Technical Digest*, 2002. p. 459–462.
51. V. Parthasarathy, V. Khemka, R. Zhu, I. Puchades, T. Roggenbauer, M. Butner, P. Hui, P. Rodriquez and A. Bose. A multi-trench analog+logic protection (M-TRAP) for substrate crosstalk prevention in a 0.25-μm smart power platform with 100V high-side capability. *Proceedings of the International Symposium on Power Semiconductor Devices (ISPSD)*, 2004. p. 427–430.

52. J.P. Laine, O. Gonnard, G. Charitat, L. Bertolini and A. Peyre-Lavigne. Active pull-down protection for full substrate current isolation in smart power IC's. *Proceedings of the International Symposium on Power Semiconductor Devices (ISPSD)*, 2003. p. 273–276.
53. V. Khemka, V. Parthasarathy, R. Zhu, A. Bose and T. Roggenbauer. Trade-off between high-side capability and substrate minority carrier injection in deep sub-micron smart power technologies. *Proceedings of the International Symposium on Power Semiconductor Devices (ISPSD)*, 2003. p. 241–244.
54. W. Horn. *On the reverse-current problem in integrated smart power circuits*. PhD thesis, Technical University of Graz, Austria, April 2003.
55. S. Voldman, R.A. Johnson, L.D. Lanzerotti and S.A. St Onge. Deep trench-buried layer array and integrated device structures for noise isolation and latch up immunity. US Patent No. 6,600,199 (July 19, 2003).
56. S. Gupta, S.L. Kosier and J.C. Beckman. Guard ring structure for reducing crosstalk and lath-up in integrated circuits. US Patent No. 6,747,294 (June 8, 2004).
57. R. Stella, S. Favilla and G. Croce. Novel achievements in the understanding and suppression of parasitic minority carrier currents in p-epitaxial/p++ substrate smart power technologies. *Proceedings of the International Symposium on Power Semiconductor Devices (ISPSD)*, 2004. p. 423–426.
58. R. Singh and S. Voldman. Method and apparatus for providing ESD protection and/or noise suppression in an integrated circuit. US Patent No. 6,826,025 (November 30, 2004).
59. R. Singh and S. Voldman. Method and apparatus for providing noise suppression in an integrated circuit. US Patent No. 7,020,857 (March 28, 2006).
60. R. Zhu, V. Khemka, A. Bose and T. Roggenbauer. Substrate majority carrier induced LDMOS failure and its prevention in advanced smart power technologies. *Proceedings of the International Reliability Physics Symposium (IRPS)*, 2006. p. 356–359.
61. C. Redmond. Winning the battle against latchup in CMOS analog switches. *Analog Devices: Analogue Dialogue* **35**(5), 2001.

# 8 Latchup Computer Aided Design (CAD) Methods

## 8.1 LATCHUP CAD RULES

Since the discovery of latchup, with the introduction of CMOS as a commercial mainstream technology, it has been a latchup discipline to attempt to develop a set of design rules that can prevent latchup in a technology for any product built within that given technology [1]. In the early days of development of CMOS technology, there were no latchup CAD methods [2–4]. As technology evolved and semiconductor design tools became more advanced, the development of a checking and verification tool that can prevent both ESD and latchup in all products became the objective.

Today, with the migration of mixed voltage interface chips, mixed signal (MS) CMOS, system-on-chip (SOC) and network-on-chip (NOC) the applications are rapidly evolving that require more latchup rules and verification methods. SOC applications include high-voltage CMOS (HVCMOS), smart power chip sectors adjacent to analog function and CMOS digital logic; these chip functions can include digital-to-analog converters (DAC), analog-to-digital converters (ADC), band gap reference circuits, oscillators, voltage regulators, interface networks, charge pumps, digital logic, digital filters and power management supervisors. With the wide variety of applications within a given chip, new latchup rules are needed.

Additionally, with technology scaling, the differential voltage between the power, analog and digital domains voltage increase. The reason for the larger voltage gap (the differential voltage between any two power domains) is that HV applications do not scale, whereas low-voltage CMOS levels continue to scale. Additionally, the scaling of the size of the physical elements in the power domains, analog and digital CMOS circuits do not follow the same scaling parameters. For example, digital CMOS scales as MOSFET constant electric field scaling theory, with a continued scaling of MOSFET channel length and width. Smart power electronics scale the length of the LDMOS device for reduced on-resistance, $R_{ON}$, but do not scale in width. Hence, for external latchup issues, although the LDMOS power devices are 'injector sources' and the digital CMOS are the 'victim' circuits, the injector-to-victim area ratio is increasing (simplistically, the injectors are getting larger as the victim circuits are getting more sensitive). Hence, external latchup issues will increase.

*Latchup* Steven H. Voldman
© 2007 John Wiley & Sons, Ltd

This is a continuous challenge for CMOS latchup ground rule development. As the number of power domains and circuit types increases, the latchup rules must keep pace with the application space and scaling to avoid latchup.

## 8.1.1 Fundamental Latchup Design Rules

In defining a semiconductor technology, latchup design rules are used to prevent product failure. Design rules come in different forms of guidelines, recommended rules and required rules [2–4]. Design rule checking (DRC) and verification can consist of both physical spaces and electrical characteristics. DRC CAD methods typically evaluate only geometric spacings. There also exist checking methods that evaluate both the electrical conditions and the geometric spaces. For latchup, there are fundamental rules established in most semiconductor corporations. These latchup rules are universal and associated with the fundamental equations of latchup physics.

## 8.1.2 Local Latchup Rules

In latchup, the physical dimensions associated with the parasitic pnpn network are checked and verified. Fundamental design rule checks include the following (Figure 8.1).

*Minimum p+ to n-well space:* The physical space between the p+ diffusion and the n-well edge is set to some minimum value based on the desired p+/n+ minimum rule.

*Minimum n+ to n-well space:* The physical space between the n+ diffusion and the n-well edge is set to some minimum value based on the desired p+/n+ minimum rule.

*Maximum n-well resistance requirement:* The maximum n-well resistance is established based on the maximum allowed well shunt resistance. This is typically represented as a physical distance between the n-well contact and the p-channel MOSFET or any p-doped element in an n-well region.

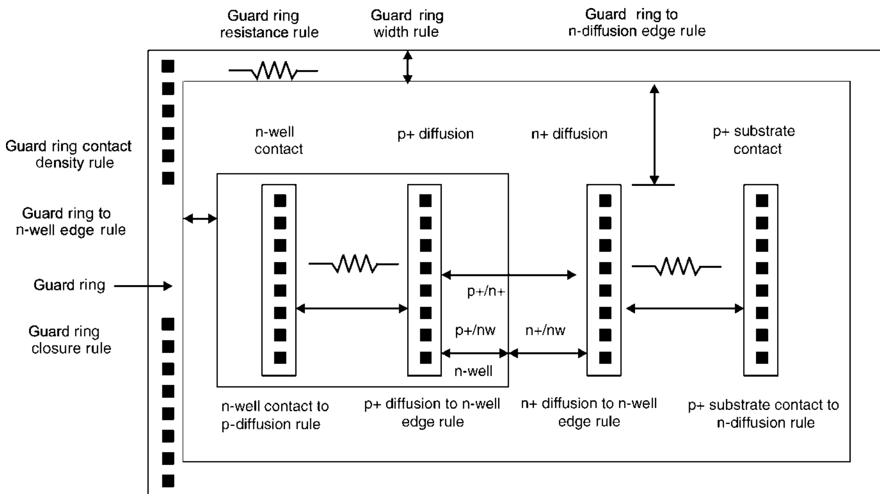



**Figure 8.1** Example of CMOS latchup ground rules.

*Maximum p− substrate resistance requirement:* The maximum substrate resistance is established based on the maximum allowed substrate shunt resistance. This is typically represented as a physical distance between the p-well contact and the n-channel MOSFET or any n-doped element in a p-well region.

*Guard ring type rule:* Design rules require guard rings for all elements electrically connected to an external node. The type of guard rings is a function of whether an element is p-doped or n-doped and the technology requirements.

*Minimum guard ring space rule:* Typically, the guard rings are spaced relative to the physical diffusion to allow electrical biasing without interaction. In addition, the spacing is optimized as to not be to close to elements to establish interaction, but at the same time at a distance too far where they do not collect minority carrier injection.

*Minimum guard ring width rule:* Guard ring width influences the guard ring efficiency of a guard ring structure. Hence, many technology guidelines will define the guard ring width or minimum width.

*Maximum guard ring resistance rule:* Guard ring design is either defined or a maximum guard ring resistance rule is established.

*Butted contact rules:* In many technologies, butted contacts are desired to minimize the resistance between a contact and the device, recommending that butted contacts should be utilized to minimize latchup concerns.

## 8.1.3 Voltage Condition Rules

Voltage guidelines are given in design manuals and cookbooks to provide circuit designers suggestions of situations to avoid electrical overstress. Other guidelines are methods to minimize the risk of CMOS latchup.

*Overshoot voltage criteria:* An overshoot requirement suggests to circuit designers to minimize the level of allowed forward bias (e.g. less than $V_{BE} = 0.7\,V$).

*Undershoot voltage criteria:* An undershoot requirement suggests to circuit designers to minimize the level of allowed forward bias (e.g. less than $V_{BE} = -0.7\,V$).

*Holding voltage criteria:* A holding voltage requirement is the holding voltage, $V_H$, value relative to the power supply voltage, $V_{DD}$.

$$V_H = \chi V_{DD}.$$

Given a technology is set so that the lowest $V_H$ is larger than $V_{DD}$, then the holding voltage can not be maintained by the power supply voltage source.

*Power sequence requirements:* A power sequence requirement provides an order of how the ground, the power rail and the signal pins are established in the 'power-up', or 'power-down' mode of operation to avoid latchup.

## 8.1.4 Off-Chip Driver Rules

In some design environments, there are specific rules for off-chip driver (OCD) circuits and bi-directional circuits (e.g. transmitter and receiver). Typical OCD CMOS latchup rules are associated with primarily guard ring rules specific to the OCD network.

*Continuous p+ guard ring around n-channel MOSFET OCD pull-down transistors:* Placement of a complete guard ring around all the sides of the MOSFET n-channel pull-down is desired for low contact resistance.

*Separate n-well for p-channel MOSFET OCD pull-up transistors:* Placement of p-channel MOSFETs in its own separate well is desirable to avoid interaction with other elements or circuits.

## 8.1.5 Gate Array Design Rules

In gate array design environments, unique design situations occur in the process of addressing the gate array design methodology that would not normally occur in other design methodologies. Two unique conditions can occur in the gate array environment. The unique states of these elements are as follows:

- unused OCD MOSFET finger electrical state;
- unused ballast resistor element electrical state.

In a 'gate array' design, a peripheral I/O design book may have one 'frame' size for a set of different OCD drive strengths and input impedances. With a given x and y dimension, a peripheral circuit has a single guard ring frame around the I/O book of a fixed dimension and fixed 'pitch'. The design methodology may have a 'customer' option of different drive strength for a given single OCD design book. The strength is achieved by changing the number of connected MOSFET gate fingers. With the electrical 'grounding' of the unused MOSFET segment, this introduces a grounded 'npn' element in the I/O cell. As a result, special rules may need to be developed to avoid latchup within the I/O OCD network for the unused segment of the design.

In a second gate array methodology, the circuit designer desires a given 'input impedance'. This input impedance is achieved by a set of ballast resistors. In this methodology, only one set of ballast resistors is used and all other sets are grounded. The grounding of the unused ballast resistor elements also leads to local parasitic npn elements within the gate array design book and may require special latchup rules.

## 8.1.6 Mixed Voltage Applications and Special Circuits Rules

In semiconductor chip applications, nonnative voltage power rails are introduced into applications, leading to unique latchup design rules. In addition, special circuits are introduced into a chip synthesis that can lead to latchup rules.

*Mixed voltage CMOS latchup rules:* In a mixed voltage semiconductor chip, with the presence of two different $V_{DD}$ power supply voltages, it is a common latchup design practice to do the following:

- increase space of p-well relative to higher voltage n-well;
- increase n-well to n-well spacing where at least one of the n-wells is at a higher bias voltage;
- increase p+/n+ space;
- wider guard ring structures;
- additional number of guard rings;
- special guard rings for the higher voltage power rails.

*Well bias generators and bootstrap circuits:* Well bias generators are on-chip circuits that generate a well bias voltage or voltage state. These well bias generator circuits provide well voltage states that

exceed the native power supply voltage of the technology or chip, and hence can initiate latchup. A typical latchup design guideline is that these circuits must satisfy all the requirements for latchup that are required on external I/O networks, as well as similar rules developed for nonnative voltage applications.

Bootstrap circuits are another type of circuit that generates a higher voltage state from a lower voltage state using capacitive coupling techniques. Circuits that bootstrap intentionally or unintentionally must allow for latchup margin by increasing the $V_H$ of the parasitic networks (e.g. pnpn).

*Voltage regulators:* Voltage regulators are on-chip circuit that provides a lower voltage or a quiescent voltage condition. Special guard ring are required or p+/n+ space rules are needed to avoid latchup.

*Substrate bias generators:* Substrate bias generators are on-chip circuits that generate a substrate voltage below the $V_{SS}$ ground potential. These networks are also referred to as ‘charge pumps’. Charge pump circuits can initiate latchup. A typical latchup design guideline is that these circuits must satisfy all the requirements for latchup that are needed on external I/O networks.

*NVRAM programming pin rules:* In a nonvolatile read-only memory (NVRAM), the programming power pin is at a significantly higher voltage than other circuitry in a semiconductor chip. For example, a 12 V program pin is introduced into a NVRAM product with a 5 V periphery power rail and a 3.3 V core circuitry. As a result of the programming pin having a voltage approximately 3 $\times$ the native power supply voltage, it is vulnerable to latchup events. To address the NVRAM program pin issue, the following latchup rules are required that are unique for this pin:

- additional guard ring structures;
- increased p+/n+ spacings.

*HVCMOS Latchup rules:* A semiconductor chip can consist of a HVCMOS and a low-voltage CMOS domain. The HVCMOS domain can be between 20 and 120 V range, whereas the low-voltage CMOS is in 1–5 V range. Additional rules are required between the power domains. In the HVCMOS section, there may be single or multiple power domains at different $V_{DD}$ voltages. Hence, the following additional rules are required:

- domain-to-domain guard ring rules;
- passive and active guard ring requirements;
- HVCMOS to HVCMOS device-to-device rules (e.g. 40–120 V HVCMOS rules);
- HVCMOS to LVCMOS rules device-to-device rules (e.g. 5–120 V rules);
- power-up and power-down sequencing rules.

### 8.1.7 Global Chip Level Rules

Global chip latchup concepts were focused on establishing a good ground plane and a good substrate contact, as well as separation of peripheral domains from internal core domains. These latchup rules include the following.

*Substrate ring:* Substrate ring is required around the complete semiconductor chip on the outer perimeter and should be electrically connected to ground.

*Peripheral I/O to core circuitry guard ring:* Guard ring is placed between the peripheral I/O and the core circuitry. Typically, the guard ring is an n-well guard ring structure to isolate I/O injection from core circuits.

*Peripheral I/O to memory core guard ring:* Guard ring is placed between the peripheral I/O and memory arrays.

*High-voltage power chip sector to low-voltage chip sector:* Passive and active guard rings can be placed between the high-voltage power sector and the low-voltage chip sector. The high-voltage chip section can be a HVCMOS and a low-voltage CMOS domain. The HVCMOS chip sector can be in 20–120 V range, whereas the low-voltage CMOS is in 1–5 V range.

*High-voltage power chip sector to medium high voltage chip sector:* Passive and active guard rings can be placed between the high-voltage power sector and the lower voltage chip sector (e.g. medium high voltage) within the high-voltage chip sector. The high-voltage chip section can be a first 120 V HVCMOS and a second voltage level at 40 V HVCMOS.

### 8.1.8 External Latchup Design Rules

External latchup rules can be established when the location of the external source can be identified. In the case where the design checking system can locate and define an injection source, a design rule can be established based on the relative distance and the injection source magnitude. The injection source can be initiated by a system level event, ionized particle or an on-chip circuit. The injection source can be local or global. In CMOS semiconductor chips, the injection source can include the following:

- I/O circuit;
- ESD circuit;
- n-wells connected to $V_{DD}$;
- triple-well regions connected to $V_{DD}$.

In smart power applications, the injection conditions from voltage transients can include the following:

- inductive load dump;
- reverse battery conditions;
- shorts to battery or ground;
- Negative pin operation.

The injection sources can be 120 V HVCMOS power devices (e.g. DeMOS, LDMOS and IGBT devices), 40 V HVCMOS, 25 V HVCMOS and low-voltage CMOS sectors. Hence, from these applications, injection conditions are established for evaluation of the impact on adjacent circuitry. There is a physical relationship between the relative distance between the injection source and victim circuit and the current required to initiate latchup. In Chapter 3, it was shown that the external injection source can influence the generalized tetrode condition for latchup. Hence, the following rules can be established (Figure 8.2).

*Well and substrate contact spacing versus the injector-to-circuit distance:* In case a sensitive circuit is close to an injection source, the spacing of the well and the contacts can be adjusted to avoid

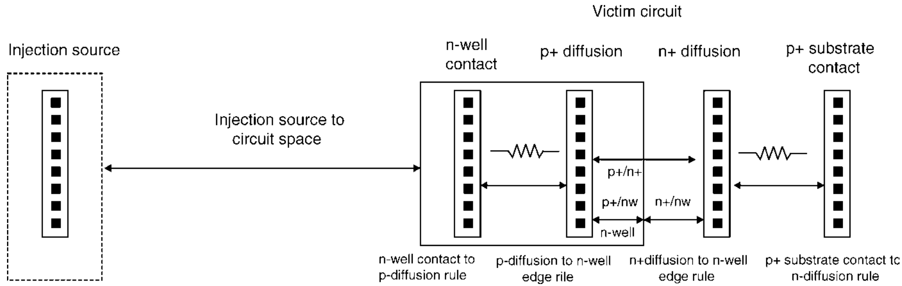



**Figure 8.2** Example of external latchup ground rules.

satisfying of the differential latchup criterion in the presence of an external source, as discussed in Chapter 3. As a result, the well and substrate contact spacing can be established based on distance from the external injection source.

## 8.2 DESIGN RULE CHECKING

DRC to verify the CMOS latchup rules is important to find the parasitic devices that can introduce latchup. In the following sections, the directions involve both internal and external latchup phenomena from simple circuits to full-chip environments.

### 8.2.1 Identification of Guard Rings

Historically, one of the key problems in latchup was the inability to identify the presence of guard rings in many semiconductor CAD methodologies. In some design methods, guard rings were physical shapes that were not identifiable or distinguishable from other physical design shapes; as a result, it was difficult to provide design checking and verification [5–7].

A CMOS latchup DRC methodology for guard ring identification is as follows:

- ‘virtual’ design level for guard ring identification [6];
- ‘built-in’ guard rings within a released set of library elements and macros [5];
- independent ‘parameterized cell’ (Pcell) guard ring [5,7].

## 8.3 COMPUTER-AIDED DESIGN EXTRACTION METHODOLOGIES – SEARCHING FOR THE pnpn

In the latchup design practice, a CAD methodology must identify the existence of parasitic transistors and determine which parasitic npn and pnp transistors as well as pnpn are important. As a result, considerable focus has been applied in the area of extraction and identification of parasitic devices [5–24]. Li’s thesis ‘Design automation for reliable CMOS chip I/O circuits’ [19] had considerable influence on some of the directions of future extraction tools, such as the work of Zhan *et al*. [20,23],

Venugopal *et al*. [21], Ramaswamy *et al*. [22], Habitz *et al*. [24] and others. In the next sections, the focus will shift toward extraction and verification methods instead of design rule checking.

Many CAD concepts have been applied to latchup to improve the ability to find parasitic devices. Model graphs (MG) have been introduced as a means to find the parasitic pnpn devices. Figure 8.3

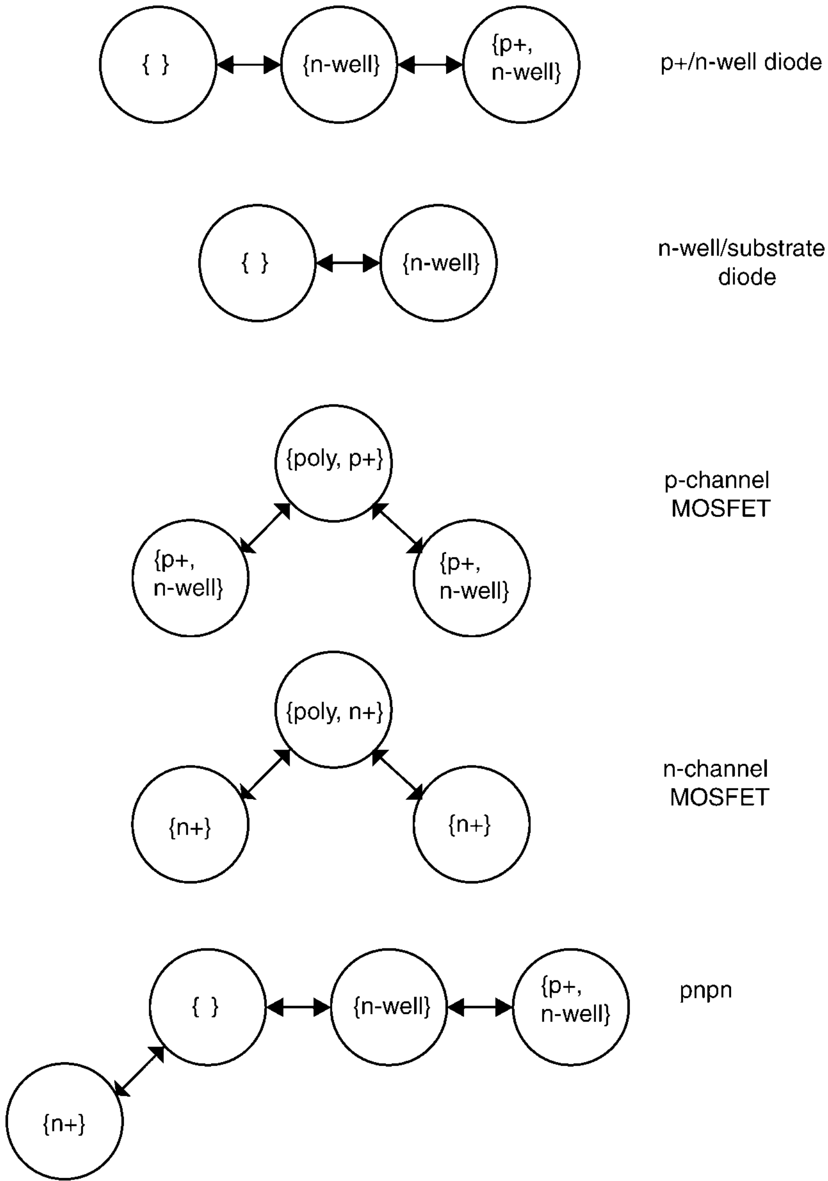



**Figure 8.3** Example of model graph.

shows a MG representation of typical CMOS elements. Once an MG representation is determined from standard devices, parasitic device MG can be developed for searching for the pnpn elements.

## 8.3.1 Extraction of the Parasitic pnpn Using Model Graphs (Li)

In a method by Li, a number of CAD methods have value for CMOS latchup (Figure 8.4). A layout extraction program extraction tool called iLEX (Illinois Layout Extraction) was developed with the following features [8,9,17,19]:

- *Device extraction* [8,9,15,19]: A method to extract actual devices.
- *Stress annotation* [19]: A method to determine the electrical potentials and classify the level of important stress conditions.
- *Bipolar junction transistor extraction* [9,19]: Extraction of parasitic bipolar transistors and means to determine the 'critical' transistors.

From the design layout graphics (e.g. GDSII, CIF and Cadence™), a circuit schematic was extracted. In the process, a 'stress annotation' was defined. The stress annotation was a method to

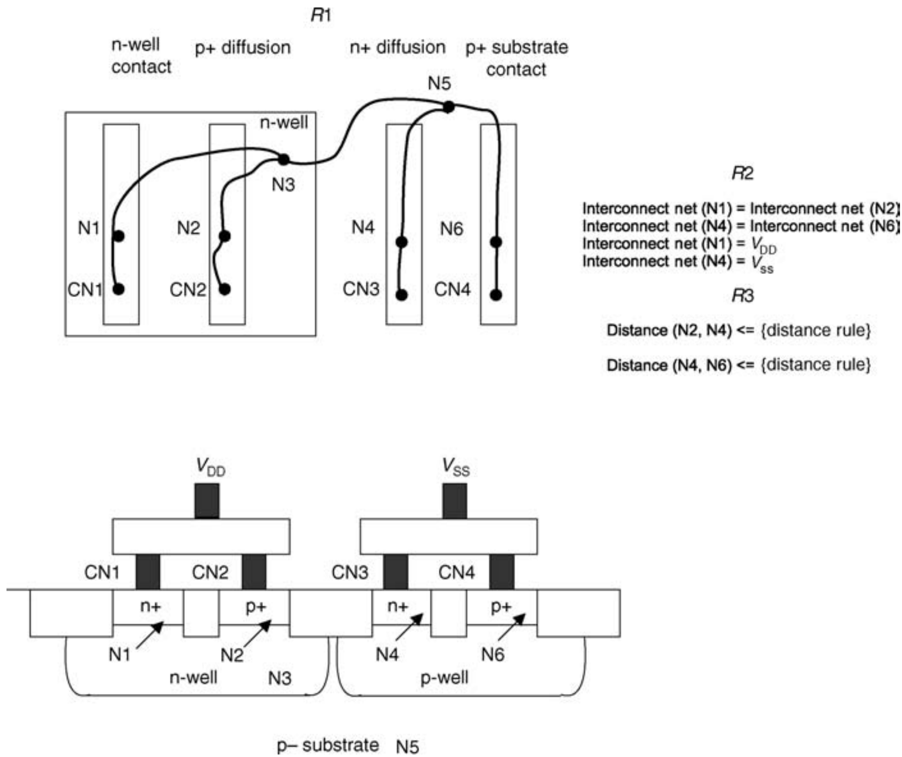



**Figure 8.4** Example of device graph.

determine the voltage states on the electrical connections to distinguish which electrical states are prone to forward bias parasitic transistors. In this method, the bias conditions were established. As a next step, the bipolar junction transistor parasitics were extracted. In the extraction of the bipolar transistors, 'reduction rules' were utilized to simplify and distinguish which bipolar transistor elements are most critical for latchup analysis (and ESD analysis).

*Device graph*: From the hierarchical structure of the graphics data, Li developed a generic device extraction methodology applying graph theory. A MG, or 'device graph' (DG), was formed to describe individual devices where the basic structure of the DG was defined as [19]

$$\mathrm{DG} = (N, R, s).$$

In this structure, $N$ is the set of nets in the device, $R$ is the set of relationships between the nets and $s$ is the seed net of the graph. Li noted that there are three types of relationships: $R_1$ associated with adjacency, $R_2$ associated with electrical connectivity and $R_3$ associated with geometrical positions. In this structure, any device of interest in CMOS can be defined, from a CMOS transistor to a pnpn. For completeness, it was found that the simple case of adjacency was not adequate for the purpose of latchup or ESD. In addition, this work extended the work of Burke on bipolar transistor representations [13].

This device graph DG $(N, R, s)$ was utilized in the extraction process as follows:

- identify the seed net by its type;
- construct the device graph DG from the adjacency relations, $R_1$;
- provide a check for electrical connectivity, $R_2$;
- provide a check for geometrical positions, $R_3$.

*Error device graphs*: Using the device graphs, the DRC procedure is efficient in the identification of 'error device graphs'. An example of an 'error device graph' can be a pnpn structure within the physical layout [19]. Hence, a design system can classify the DG from the error device graph to search for critical latchup concerns and sort them from the observable device graphs of the physical design.

*Stress annotation and stress strength*: Stress annotation is used in the extraction process as a means to classify the electrical state. Li proposed utilization of the static analysis technique for ESD and latchup analysis [19]. In this process, the circuit schematic is first extracted. The stress annotation is then propagated using the a search algorithm for the path of the electrical stress. Many search algorithms exist for evaluation of the optimum path between a first point and a second point. In the methodology of Li, the breadth first search (BFS) methodology was utilized to address the shortest path [19]. Alternative rapacious algorithms can be utilized, such as Dijkstra's algorithm and Prim's algorithm. A path is defined for all 'forward bias' devices in the electrical path but terminates at any reverse bias state. All elements in that path are part of designated and a 'stress strength' (SS) is defined. In this process, the stress strength is reduced for all resistive elements. For example, a stress strength of $\mathrm{SS} = 10$ is defined, but is reduced to a value of $\mathrm{SS} = 9$ through a resistor [19].

*Bipolar transistor extraction and reduction rules*: Li pointed out that in the extraction process of parasitic transistors in a physical design, there are a large number of parasitic transistors. In a multifinger MOSFET, each independent finger of the MOSFET can appear as a multifinger emitter or a multifinger collector. Given that there are $p$ emitters and $q$ collectors, there is a potential of $pq$ independent bipolar transistors that can be extracted. As a result of this complexity, there must be a simplification methodology to reduce the problem to a smaller set.

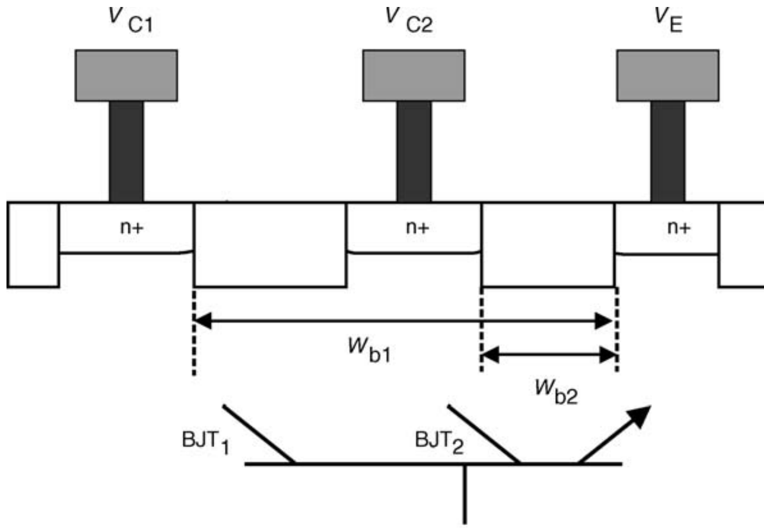



**Figure 8.5** Example of shared emitter rule.

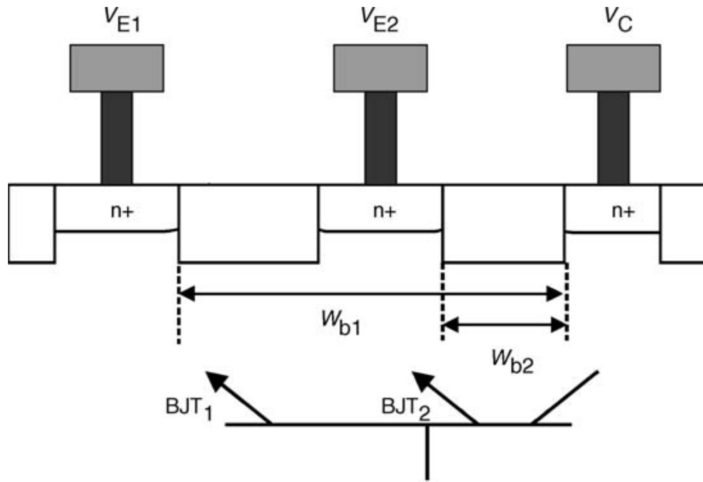
Li highlighted three practical reduction rule cases to simplify the extraction process. In these rules, they evaluate (1) the electrical state and (2) the bipolar current gain.

Figure 8.5 is a representation of the first reduction rule. The first reduction rule is the case of the 'shared emitter rule'. In a shared emitter rule, each independent collector has an independent voltage state and an independent bipolar current gain. The bipolar current gain can be obtained by determining the collector area and the geometric spacing relative to the emitter of interest. In this reduction process, the collectors that are farthest away and have the lowest voltages are removed [19].

Figure 8.6 is a representation of the second reduction rule, which is the case of the 'shared collector rule'. In a shared collector rule, each independent emitter has an independent voltage state and an independent bipolar current gain. The bipolar current gain again can be obtained by determining the emitter area and the geometric spacing relative to the collector of interest. In this reduction process, the emitters that are farthest away and have the lowest voltages are removed [19].

The third rule is a 'minimum bipolar current gain rule'. Given a bipolar current gain was less than a given value, the bipolar parasitic transistor is not evaluated. For example, given a bipolar current gain is less than unity, it can be removed [19].

It is clear from this framework and methodology that additional rules can be defined for pnpn elements that are the product of the bipolar current gains (or the sum of the collector-to-emitter transport factor, $\alpha$).




Reduction rules

IF $V_{E1} = V_{E2}$ AND $\beta_1 < \beta_2$ remove BJT 1

IF $V_{E1} > V_{E2}$ AND $\beta_1 = \beta_2$ remove BJT 1

IF $V_{E1} > V_{E2}$ AND $\beta_1 < \beta_2$ remove BJT 1

**Figure 8.6** Example of shared collector rule.

Hence, a CMOS CAD methodology is as follows:

- utilization of MG, or DG, for defining devices;
- utilization of error device graphs for finding parasitic devices;
- stress annotation as a method to quantify the important electrical states;
- search algorithm (e.g. BFS) methodology as a means to evaluate the propagation and extent of the current path for that given voltage state;
- utilization of reduction rules as a means of sorting out the important parasitic bipolar transistors for CMOS latchup.

## 8.3.2 CAD Extraction Methodology (Zhan–Feng–Wu–Chen–Guan–Wang Method)

In CAD extraction methodologies, one of the difficulties is to provide a method that is technology independent [20,23]. In today's semiconductor foundry environment, it is important to have the method that is technology and fabricator independent. Hence, a method that allows migration from technology to technology or fabricator to fabricator will provide the highest guarantee of success of preventing latchup. Zhan *et al.* focused on the utilization of MG theory to provide a

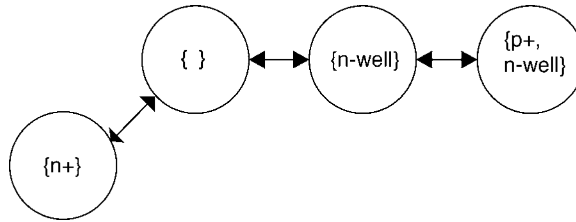



**Figure 8.7** Model graph representation of semiconductor pnpn element.

technology-independent CAD tool for ESD protection device extraction, called 'ESD extractor' [20]. In the process of extracting ESD parasitics, the MG theory methodology is able to extract parasitic pnpn elements. As in the methodology of Li, MG representations of the parasitics are found and identified, maintaining both the spatial and electrical states of the structure. Zhan *et al*. [20] showed that a pnpn element can be represented as a model graph as shown in Figure 8.7. In this representation, the pnpn regions are represented as circles with connectivity between the circle elements. From a latchup CAD practice, the usage of MG representation allows for identification and extraction of parasitic pnpn elements, which is technology independent and is also not customized to a given design method or environment.

### 8.3.3 CAD Extraction Methodology – Extraction of the Parasitic pnpn (Habitz–Galland–Washburn Method)

In the latchup design practice, a CAD methodology must identify the existence of parasitic pnpn in a semiconductor chip. In an advanced CMOS design, the density of parasitic pnpn are significant in number. A desire has been to develop a CAD tool that can identify the circuits in which latchup is likely to occur and quantify it efficiently, which is not limited by either analysis time, or computer processing time and data constraints.

In a method by Habitz *et al.* [24], one of the goals is to develop a CAD methodology that allows for identifying where latchup will occur and a method and tool apparatus that would both identify and quickly provide circuit analysis on that localized region to determine whether CMOS latchup is a potential problem in that chip section.

The CMOS latchup CAD methodology is as follows [24]:

- locating potential structures that meet predefined criteria for latchup (e.g. that may be a location for a pnpn parasitic);
- modifying the 'structure' to represent a pnpn device recognized by a circuit tool;
- construct a circuit model of the pnpn based on the geometrical factors and voltage conditions of the modified structure that influence CMOS latchup (e.g. base widths, shunt resistances, emitter areas and forward active conditions);
- perform latchup circuit analysis on the pnpn-modified structure circuit netlist for both dc and transient latchups;
- compare the results to a latchup criterion or semiconductor chip application specification;
- determine whether that specific modified structure will lead to CMOS latchup.

In the first step, the circuit designer or latchup engineer chooses an area of the semiconductor chip where latchup analysis is of interest. This is done to minimize the semiconductor software processing time and to allow the analysis to focus on a circuit, subfunction or chip sector. In the second step, the parasitic pnpn element identification is critical. In the Habitz–Washburn–Galland method, the key factors of a pnpn structure identification are as follows [24]:

- a p- and n-type diffusions are parallel;
- an n-diffusion edge parallel to an n-well edge where the n-diffusion is outside the n-well;
- a parallel p-diffusion edge parallel to an n-well edge where the p-diffusion is inside the n-well.

From this parasitic structure, a new physical model is formed using its geometrical factors. A trapezoidal region is cut from the original design that includes the parasitic pnp and the npn as well as the shunt resistors. In this method, the n-well shunt resistance is defined by identification of the n-well contact and the p-well shunt resistance is defined by identification of the p-well.

A parasitic pnpn transistor circuit is defined where a model is defined for the pnp transistor and the npn transistor on the basis of the geometrical factors extracted from the initial design. From the extracted trapezoidal regions, the physical pnpn model was constructed for analysis of CMOS latchup. Habitz *et al*. assumed the current flowing through the circuit is a function of the following [24]:

- bias condition across the pnp bipolar transistor junction;
- bias condition across the npn bipolar transistor junctions;
- bias that can be established at well regions under the junction area of the diffusions (e.g. addressing the shunt IR drops);
- n-diffusion width and p-diffusion width wherein the model used the smaller of the two dimensions of the parallel edges in the trapezoidal region;
- pnp base width;
- npn base width.

Having identified the four electrical nodes of the circuit, the model for the parasitic pnpn circuit can be used to evaluate the latchup sensitivity of the parasitic pnpn under both the dc and transient conditions in a circuit simulator. An extracted netlist of the networks is used to begin the circuit analysis.

## 8.4 CAD EXTRACTION METHODS – SEARCHING FOR THE GUARD RINGS

Extraction, verification and evaluation of guard rings has been a fundamental problem in semiconductor development, and even today is a fundamental reason latchup occurrences. Guard ring evaluation, identification and integration is critical. In the following section, new methods are discussed for finding the guard ring structures.

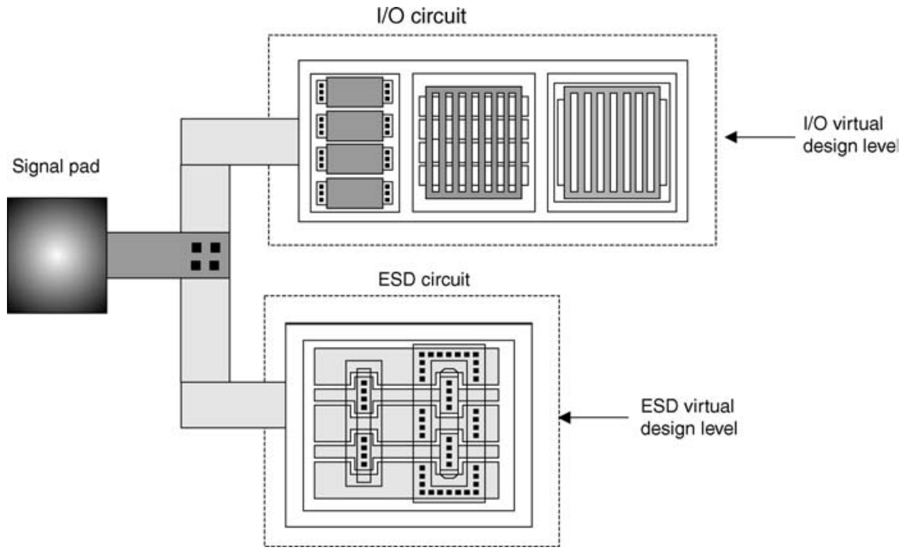



**Figure 8.8** Example of virtual design levels for ESD and I/O networks.

## 8.4.1 Virtual Design Level Methodology (Voldman–Sullivan–Nickel–Bass)

The concept to use 'virtual design levels' for guard rings was first addressed in the construction of customized I/O and ESD-related design rule checking methods but is suitable for latchup evaluation (Figure 8.8) [6]. In the CAD checking and verification methodology, virtual design levels were established for the I/O and ESD checking, referred to as 'dummy design layers'; these dummy levels were placed on the ESD devices and I/O networks [6]. In this process, the virtual design level was placed on the guard ring structure. This CAD methodology allowed for ESD-specific design rule checking by data compression of all data outside the ESD and I/O regions. Secondly, it allowed for verification of a one-to-one correspondence as a reminder to circuit designers that the 'dummy design layers' are placed on the guard rings. In this method, the full-chip data to the areas of interest in the I/O region are significantly compressed; this provided for easy implementation and time efficiency.

## 8.4.2 Built-in Guard Rings

A second methodology to ensure that the guard rings are contained in the designs is to release customized elements (all designs are released from a set of predesigned elements with built-in guard rings). This can be achieved on the following levels of design:

- guard rings within primitive semiconductor devices in a released library;
- guard rings within released circuit books;
- guard rings within the design methodology itself through graphical unit interfaces (GUI).

The problem with this methodology is that not all physical elements require guard rings. Secondly, the use of the guard ring is a function of the placement within the chip; the choice of the guard ring structure may not be suitable for both internal and external devices. Hence, the integration of a parameterized cell (Pcell) with a defined and fixed guard ring may not be advantageous.

Guard rings can be integrated into larger circuits or design books. For example, in an ASIC methodology, guard rings can be inherently integrated into the released peripheral I/O books. In this fashion, the peripheral book must be qualified in satisfying the latchup specification.

### 8.4.3 Guard Ring Parameterized Cells (Pcell) (Voldman-Perez)

In a third method, the guard ring itself can be a parameterized cell (Pcell) that is identifiable by the design environment [5,7]. With the establishment of a hierarchical Cadence™ based ESD design methodology, the opportunity to integrate guard rings into the design methodology provided built-in compliance, checking and verification. As a design methodology, Voldman, Strang and Jordan constructed a graphical layout, schematic and symbolic cell view representations of ESD networks that were hierarchical, which allowed both variable design size and circuit topology in a Cadence™ based system [25–28]. The elements are constructed of primitive order '1' $O[1]$ devices that were standard kit library items as well as ESD-optimized elements.

Guard ring Pcells can be constructed and compiled into the network to provide a higher order circuit, where the guard ring is detectable by the design environment in the layout, schematic and symbol 'cellviews'. In this fashion, the checking of the guard ring is evaluated by the identification of the guard ring Pcell in the net listing. To address design integration of guard rings in a Cadence™ based Pcell system, Perez and Voldman developed an independent guard ring parameterized cell [7]. The guard ring Pcell consists of a plurality of guard rings that can be integrated with the primitive $O[1]$ device elements or the higher order $O[n]$ hierarchical parameterized cells [5,7]. The guard ring Pcells are design such that the guard ring structures can be turned 'off' using switches in the GUI. In the guard ring Pcell, a large combination of rings can be switched 'on' or 'off', allowing significant design flexibility and cosynthesis of RF design and latchup optimization. An independent guard ring Pcell was defined that contains a plurality of consecutive ring structures types and number; this allows for the following:

- independent designing of a guard ring structure;
- choosing guard ring on the basis of guard ring efficiency requirements;
- a growable guard ring that expands on the basis of the identification of the element type;
- a growable guard ring that expands on the basis of the circuit or function block design input parameters.

This concept also allows for the following:

- designing of a guard ring structure with generated virtual design levels (e.g. guard ring virtual level, I/O virtual level and ESD virtual level);
- graphical and schematic representation view of a guard ring structure;
- designing of a guard ring structure symbology and symbol;
- designing of a guard ring structure symbology and symbol hierarchy that integrates the existing circuit design symbol view with the guard ring symbol view;

- checking of a guard ring structure by Pcell identification;
- verification of a guard ring structure by Pcell identification.

### 8.4.4 CAD Automation of Guard Ring Resistance

In the integration of guard rings into a semiconductor chip design, CAD methods are being used for latchup. Guard rings are placed around injection sources that can trigger latchup. Injection sources can be n+ diffusions, n-type resistors and n-wells (e.g. in a p-type substrate wafers). One of the primary design issues with guard rings is the effectiveness to collect the minority carrier injection. The guard ring effectiveness is dependent on the guard ring type, the guard ring depth, guard ring physical width and the relative distance from the injection source. Guard ring resistance is also a key criteria. The reasons this is a growing issue are the CMOS and BiCMOS dimensional scalings and an increase in the peripheral I/O density (e.g. I/O book width scaling). With the technology dimensional scaling, the physical dimensions of the p+, n+ and n-well have been reduced. With the scaling of the minimum n-well widths, the width of the n-well guard ring has been scaled. As a result, the resistance along the length has increased. The I/O circuit density increases with the increase in circuit density. ASIC environments have focused on reducing the width of the I/O peripheral book to allow more I/O circuits on the periphery of a semiconductor chip. In this process, the peripheral I/O length has been increased to compensate for the reduction in the peripheral I/O book width.

With the placement of an n-type guard ring in the p-type substrate, a metallurgical junction is formed that can collect the minority carrier electrons injected into the substrate. As an example, an n-type guard ring is biased to the power supply voltage, collecting the injection current.

At low injection currents, the electrons are collected by the reverse biased metallurgical junction formed between the substrate and the n-type guard ring. But, at very high injection currents, the series resistance between the power supply voltage and the guard ring is a key latchup design factor.

The injection source serves as an 'emitter', the substrate as a 'base' region and the guard ring as a 'collector'. When the emitter–base junction is forward active, the electrons are injected into the substrate region. When the collector is biased positive at the power supply voltage, the collector-to-emitter voltage is positive. In this state, the forward active parasitic transistor is formed between the injection source and the guard ring. When the resistance of the guard ring increases, a voltage drop occurs in the guard ring. The voltage drop is equal to the product of the guard ring resistance and the injection current

$$\Delta V = I_{\text{inj}} R_{\text{GR}},$$

where $I_{\text{inj}}$ is the injection current and $R_{\text{GR}}$ is the guard ring resistance between the point of injection and the power supply voltage. At the location of the injection, the voltage at the guard ring is equal to

$$V_{\text{GR}} = V_{\text{DD}} - \Delta V_{\text{GR}} = V_{\text{DD}} - I_{\text{inj}} R_{\text{GR}}.$$

As the voltage drop increases due to the injection current, the guard ring voltage at the point of injection will decrease. When the effectiveness of the guard ring to collect the current is minimized as a result of debiasing, the minority carrier electron current will flow to alternative structures (e.g. outside the guard ring). The following design parameters influence the resistance:

- guard ring sheet resistance (e.g. n-well sheet resistance or plurality of implants in the guard ring);
- guard ring width;
- guard ring contact density;

- guard ring contact resistance;
- guard ring silicide resistance;
- metal bus resistance;
- distance between the injection location and the power supply voltage source.

Historically, guard ring resistance was not a critical issue due to the technology, the ground rule dimensions and the I/O density. From 1980s to 2000, the guard rings used were typically n-well regions. During this time, the ground rules for both diffused or retrograde wells prevented narrow width n-well regions. As a result, the ground rules prevented scaling of the guard ring widths below some minimum dimension (e.g. typically wells could not be scaled below 3–7 μm). With the utilization of the n-diffusion and silicides (e.g. titanium silicide and cobalt silicide) and large contact dimensions, the resistance was very low. Additionally, due to wide 'wiring tracks' and peripheral I/O design, the power bus width was greater (e.g. 10–30 μm). In addition, the I/O density was low.

In this millennium, the vertical semiconductor process profile was scaled, leading to higher well sheet resistance. In addition, vertical scaling allowed for a decreased minimum well width requirement allowing a narrower guard ring structure in I/O design. In each technology generation, the number of I/O increase led to high aspect ratio I/O books that are long and narrow. In this case, the guard ring width is reduced, as well as length between injection sources and the power supply voltage is increased. In addition, with the metal scaling, the wire widths are reduced to allow a higher density of wire tracks. With all the scaling issues for both the semiconductor process and the semiconductor chip layout design, the resistance issue is more critical.

A CMOS latchup CAD system can be developed that addresses the guard ring resistance [29]. The CMOS latchup CAD evaluation must address a maximum resistance requirement for the guard ring resistance (Figures 8.9 and 8.10). The guard ring resistance can be evaluated as follows:

- identify injection source;
- identify the location near the guard ring structure;

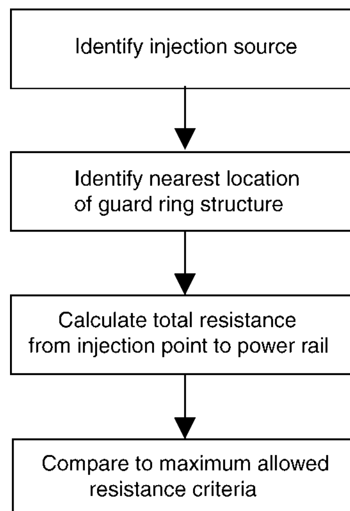



**Figure 8.9** Example of guard ring resistance evaluation methodology flow.

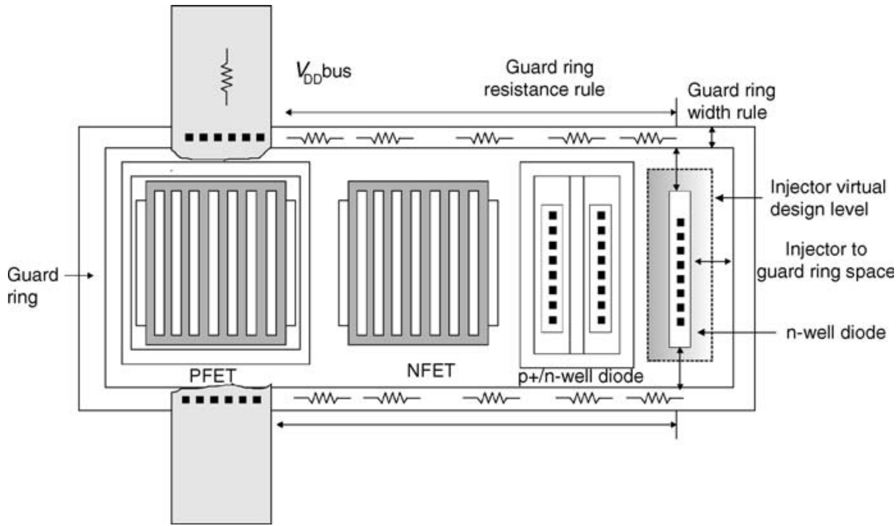



**Figure 8.10** Example of guard ring resistance evaluation.

- calculate the total resistance to the power supply $V_{DD}$;
- evaluate the maximum resistance allowed for the guard ring for the given conditions.

A key latchup design practice is as follows:

- Establish a guard ring structure under the condition of a maximum resistance criteria.
- Establish resistance criteria associated with resistance calculation from the location of injection sources and its intersection at the guard ring to the power supply voltage location that 'sinks' the latchup current. The effective resistance of all structures are used in the resistance calculation (e.g. sheet resistance, silicides, contact and metal bussing).

## 8.4.5 CAD Automation of Guard Ring Modification Methodology (Ker–Jiang–Peng–Shieh)

CAD methods to integrate guard ring structures (and contact density) can also be implemented in an integrated semiconductor chip design as a postprocessing methodology. Given a semiconductor foundry environment, the peripheral 'I/O books' can be predefined; this may include the OCD, the receiver, the ESD circuit and the guard ring structures. A given predefined I/O book may be mapped into different chip architectures and different technologies. Latchup sensitivity is influenced by the semiconductor technology and the integration into the power bus architecture; each design methodology may have different integration placement and practices. As a result, the placement of the guard rings relative to the power bus and the adequacy of the guard rings to minimize latchup requirements may need to be modified as a postprocessing of the peripheral I/O design or upon full-chip integration.

Ker *et al.* developed a CAD methodology that addresses this capability [30]. A design solution is a procedure that can be established to allow for the placement of additional guard rings within a given design. To satisfy the requirements of the latchup guard rings of a given latchup foundry, additional guard rings can be added to the preexisting design. To address integration with the power rails, additional shapes can be added under the power rails to electrically connect to the power rails (e.g. $V_{DD}$ and/or $V_{SS}$).

In this latchup design practice, the design method allows for the autogeneration of diffusions shapes serving as a guard ring under the power bus for $V_{DD}$ and $V_{SS}$. In this method, the designer has the ability to create n+ or p+ arrays under the power rail. For the case of a $V_{DD}$ power bus, n+ diffusion shapes, electrical contacts and vias are generated to serve as n+ guard ring attached. For the case of a $V_{SS}$ power bus, p+ shapes, contacts and vias are generated under the $V_{SS}$ power bus. To integrate the overlay of the guard ring, the shapes and the power bus, design layers are used to coordinate the integration. The procedure to auto-generate guard ring is as follows [30]:

- load the program;
- select to run $V_{DD}$ or $V_{SS}$ guard ring;
- select the guard ring type;
- input spacings and define variables;
- run simulation.

## 8.5 LATCHUP EXTRACTION METHODS AND TOOLS

Extraction methods and tools are important for semiconductor corporations today. Each tool has its advantages and disadvantages. In the following sections, different tools utilized for today's CMOS latchup design discipline will be highlighted.

### 8.5.1 Latchup Extraction Tool (Ramaswamy–Sinha–Kadamati–Gharpurey Method)

In some CAD methodologies, semiconductor extraction for ESD and latchup are integrated into a common methodology. The methodology of Ramaswamy *et al.* [22] combines the design environment to extract parasitic devices and its corresponding semiconductor process variables, analyzes the physical structures, stores the information and provides an output listing of the sensitive elements.

By taking advantage of the design environment, it is possible to provide good parasitic models from the technology rules file (e.g. also known as 'tech files') that contain the doping profiles and physical models. These can be used to extract netlists for analysis of the parasitic devices for either ESD or latchup issues. Given that these elements are extracted and identified, it is possible to store the key devices of interest with their physical circuit models for analysis in circuit simulation.

It was noted by Ramaswamy *et al*. [22] that there is a need for a coherent, low-cost CAD system and method of simulating ESD and latchup in the design stage, especially an extraction of parasitic devices and their location and correction. It was noted that the extraction should be simple, flexible and suitable for different technologies and processes.

In the Ramaswamy *et al*. [22] methodology, the design tool contains the following:

- input data generator (e.g. circuit layout, technology rules and process information);
- device extractor (e.g. for netlist generation, physical location and shunt resistances);
- verification database (e.g. store netlists of parasitics);
- translator (e.g. netlist generators, physical location and shunt resistances);
- output generator for identifying and displaying latchup sensitive circuits.

Similar to the work of Li [19], this design methodology utilized three extractors:

- a netlist extractor;
- a substrate resistance network extractor;
- a parasitic device extractor.

In this case, the netlist extractor took advantage of the technology information from the technology files, actual devices and the GDSII data. The substrate resistance network extractor also utilized the technology data as well as a distributed resistance representation. And as in other methods, the parasitic pnp, npn and pnpn devices are identified and extracted.

In this tool, the advantage of the design environment is utilized for the extraction of physical parameters and identification of parasitic transistors (npn and pnp) and parasitic pnpn elements. By utilizing the design information, the parasitic pnp and npn models can be defined on the basis of technology files and physical analytical models. The physical models must be able to extract all of its input information from the technological data and the geometrical data in the design layout. By remaining within the design environment, the extraction and identification process remains within the technology information which generates no additional work for the model development. In addition, given this is established, the tool is technology generation independent and can be utilized for all technology generations. In this tool, the integration with the existing design environment reduces additional work and provides a simple approach that will help provide latchup verification.

### 8.5.2 CAD Verification – Design Error Detector Methodology (Venugopal–Sinha–Ramaswamy–Duvvury–Prasad–Raghu–Kadamati)

As part of the latchup CAD discipline, design error detector CAD software tools have been added to assist the user in possible CMOS latchup concerns. Venugopal *et al.* constructed a design error detector CAD tool to help assist the circuit designer in analyzing and identifying latchup failures [21]. To the existing tool environment, a postprocessor can be added to further evaluate the sensitivity to CMOS latchup.

In this CAD tool, the features added to the existing system included the following [21]:

- element model;
- safe operation file;
- list of stress simulations;
- critical stress values.

The tool must first identify the parasitics and devices of concern. In this detector tool, the 'observed sensitivities' of the design elements to CMOS latchup are analyzed. These design elements have critical stress conditions applied to the circuit to judge whether these critical stress conditions can in fact induce latchup failure. Given that these elements 'fail' the critical stress conditions, the element failures are recorded. The detector tool then identifies the element and the physical locations of concern for latchup.

The preprocessor comprises the following 'information generators' [21]:

- information generator from the layout data;
- information generator operable to generate data related to a model of elements including parasitic latchup-sensitive elements;
- information generator operable to generate data related to safe circuit operating conditions relative associated with latchp stability;
- information generator operable to generate data related to customized stress simulation conditions relative to latchup applications;
- a translator operable to express the results.

In this methodology, the postprocessor CAD tool 'flags' circuit elements and devices that fail to withstand the quantitative latchup events. These elements are then placed in a user friendly report for interpretation for circuit designers.

## 8.5.3 CAD Verification Methodology (Kimura–Tsujikawa)

Latchup verification methods can provide many different methods of verifying the design immunity to CMOS latchup. Since latchup is a function of many layout design variables, as well as voltage and current conditions, different verification methods can be developed to check and double check a specific design for immunity of latchup. As an example of different factors to evaluate, Kimura and Tsujikawa established 25 different 'checks' to evaluate for latchup [31,32]. The Kimura–Tsujikawa verification cases for latchup are as follows [31,32]:

- independent separate assessment of the well region, the transistor region and the substrate contact;
- sequential assessment of the well region, the transistor region and the substrate contact;
- any assessment of the well region, the transistor region and the substrate contact (e.g. a structural feature);
- use condition of the transistor regions;
- positional relationship of the transistor, well and substrate contact;
- distance between the well region and the transistor region;
- distance between the well region and the substrate contact region;
- dimension of the transistor;
- dimension of gate length of the transistor;

- dimension of gate width of the transistor;
- current capability of the transistor regions;
- substrate contact as via hole contact region considerations;
- combination of a transistor structural feature and the electrical states of each region of the transistor;
- horizontal versus vertical orientation of substrate contacts;
- transistor contained or not contained within the contacted region.

In the Kimura–Tusjikawa verification methodology, different combinations and permutations of these conditions were obtained, generating at least 25 cases of interest for latchup design verification. With the combination and permutation of layout dimensions, spatial positioning, orientation, current carrying capability and electrical states, as well as the fundamental dimensions taken independently or sequentially, many different processes of verification are possible in a design flow.

One of the problems with many of the latchup CAD methods is that the design rule check conditions are all independent and not ‘coupled’. In this methodology, by addressing a sequence of successive checks and independent checks, the rules provide both independent geometric checks as well as combined geometric checks. From the theoretical work, it is clear that many of the latchup key metrics are coupled in the differential general tetrode relationship; hence, a system that couples the relationships of the geometrical shapes (e.g. p+ to n-well, n-well to n+ diffusion, well resistance, substrate contact resistance, silicide, orientation, emitter area, collector area, etc.) will provide a better system of verification against latchup failure. By sequentially executing many of the key metrics, a good method for verification is established.

### 8.5.4 CAD Design Methodology – Modification of Well and Substrate Contact Placement Based on Injection Source

Minority carrier injection into a semiconductor chip substrate can lead to latchup. As discussed in Chapter 3, the differential tetrode relationship can be shown to be a function of an external injection source. The latchup sensitivity of a circuit is a function of the p+/n+ spacing, the n-well contact to p-channel MOSFET space and the p+ substrate contact to n-channel MOSFET space. In addition, in the presence of an external injection source, as the injection source is closer to the circuit of interest, the injection current level increases; as a result, the relative distance between the circuit and the injection source is a design issue.

A latchup CAD methodology can address this issue in an automated fashion by modifying the contact spacings relative to the injection source [33]. A latchup analysis and parameter modification system can be developed that analyzes a circuit design for latchup sensitivity and allows for modifications of the circuit design to avoid latchup of the circuit.

Latchup of a circuit can be analyzed as follows:

- identifying an injector source;
- identifying a circuit;
- identify the relative distance between the injector source and the circuit;
- extract the n-well contact to p-channel MOSFET space;

- extract the p-well contact to n-channel MOSFET space;
- extract the p+/n+ space;
- evaluate the latchup sensitivity of the circuit;
- compare to a known latchup margin;
- modify the design parameters to provide adequate margin to CMOS latchup failure;
- re-evaluate the latchup sensitivity and margins.

In this methodology, injection sources can be identified by their relative interconnects to external pads or defined by virtual design levels [33]. The CAD tool can be utilized to identify the sources or define the sources using GUI. The 'user' can define or modify the sources to evaluate the injection source conditions, irrespective of whether it is a diode, a well or an ionizing radiation source (e.g. alpha particle, heavy ion).

In a CAD design methodology, the physical parameters can be adjusted in a number of ways. The parameters can be modified by improving on the reduction in the well and substrate shunt resistances (e.g. through process levels or layout) or the guard rings (through physical size or structure). For example, the parameters can be modified by adding subcollector implants, DT and TI structures to the perimeter of a structure. The modifications can be adding latchup guard rings (e.g. p+, n+, n-well, DT or TI) around the complete circuit. For example, in a Cadence™ design system, modifications can be initiated wherein the parameter is inherent in a primitive Pcell or a hierarchical Pcell. Additionally, the features can be implemented by forming a new higher order hierarchical Pcell from the existing hierarchical Pcell by autogenerating and compiling a new Pcell with the physical parameter of interest.

In today's design environment, these features can be utilized for modifying the existing designs of an injection source. In the Cadence™ design environments with parameterized cells formed in the design environment, all of these concepts are suitable and practical for design optimization (Figure 8.11).

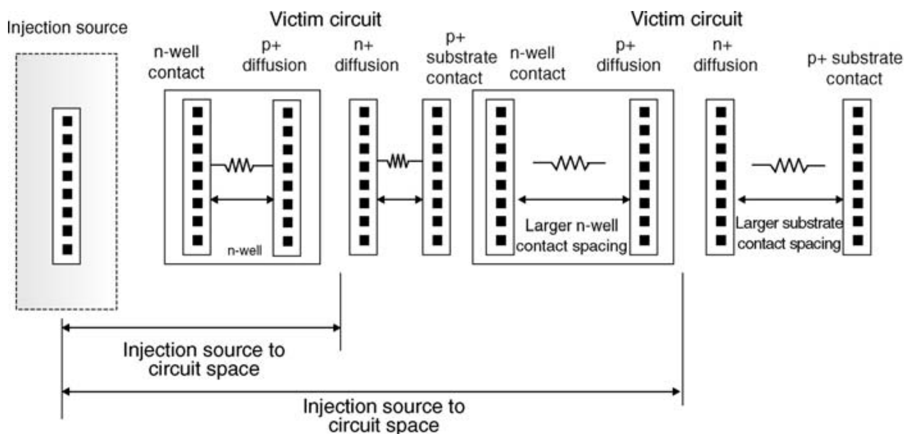



**Figure 8.11** Example of well and contact modification versus injection source distance.

### 8.5.5 Method of Well and Contact Spacing from External Source (Chatty-Brennan)

Latchup CAD methods to address external injection phenomenon can be addressed using experimental data and table-based evaluations as proposed by Chatty *et al*. [34,35]. In systematic design environments, it is also possible to simplify the CAD process in a given axis (e.g. $x$-axis or $y$-axis) to address the contact spacings in a given circuit.

In this methodology, the n-well spacings are adjusted in a quantized incremental fashion, so as to be consistent with the pitch of a design methodology, such as an ASIC environment. In a semiconductor chip with a dense circuit density, the adjustments to the well and substrate contacts must be compatible with the design environments to avoid impact to the wiring channels. As a result, the physical contact spacings may be 'quantized' in incremental steps to allow parameter adjustments. In a CMOS environment, the adjustments are limited to n-well contact space p-well contact space, and contact density. In the method of Chatty *et al*. [34,35], the following CAD method is proposed:

- identify one injection source;
- identify circuits sensitive to latchup;
- determine the number of contacts;
- determine number of contacts needed to suppress latchup;
- determine the varying distance of the contacts relative to the injection source.

In this method, a simple stepwise contact spacing can be defined in a design environment that is Euclidean in nature. This method is effective in that it can provide adjustments relative to a given source. The piecewise steps in the contact density provides advantages in that lookup tables and quantized steps can be utilized for easier design integration.

### 8.5.6 Method of Well and Contact Spacing for Array I/O

External CMOS latchup can occur when I/O networks are placed in an 'array' instead of the periphery of a semiconductor chip. External circuits are placed in the interior of a semiconductor chip in high pin count chips for the purpose of improved wire-ability into a semiconductor chip. In this process, I/O networks are placed within the regions of interior logic circuits. CMOS logic circuits interior to a semiconductor chip typically do not have guard ring structures to prevent latchup. Hence, the placement of I/O networks within the interior of a semiconductor chip can lead to interbook latchup failure at the boundaries or to external latchup failure due to external injection sources from the I/O circuit. One of these sources is the cable discharge event (CDE). Brennan *et al*. inserted well and substrate contacts with varying periodicities around susceptible cells according to the rules derived from an analytical latchup model [34–36].

A simplistic model was used to define the contact periodicity. The model assumed that latchup occurred when the injection current exceeded the sum of two currents. The first current is the current needed to forward bias the pnp transistor based on the Kirchoff current loop in the circuit,

$$I_{\mathrm{R,NW}} = \frac{(V_{\mathrm{be}})_{\mathrm{pnp}}}{R_{\mathrm{NW}}}.$$

The second current is the amount of base current in the pnp to bias the npn transistor. From the relationship of collector current,

$$I_{R,SX} = \frac{(V_{be})_{npn}}{R_{SX}} = \beta_{pnp}(I_B)_{pnp}.$$

Solving for the base current,

$$(I_B)_{pnp} = \frac{(V_{be})_{npn}}{\beta_{pnp} R_{SX}}.$$

Hence, the nodal analysis can then be evaluated at one of the nodes in the pnpn current feedback loop; the model then indicates latchup occurs for

$$I_{inj} \geq I_{R,NW} + (I_B)_{pnp}.$$

In this set of equations, the adjustment of the well and substrate contacts alters the values of $R_{NW}$ and $R_{SX}$. To address the spatial dependence of the injection current, a 'point source' hemispherical current distribution of injection is assumed. The current $I_{inj}$ collected by the n-well of the latchup structure at a distance $r$ from the injection point is modeled, assuming a steady-state injection current,

$$I_{inj} = k\left\{\frac{I}{2\pi r^2}\right\}\frac{WL}{2},$$

where $I$ is the current discharged into a signal pad, $W$ and $L$ are the dimensions of the collecting n-well between well contacts and $k$ is a fitting parameter. From this analysis, the maximum spacing between the well contacts as a function of the distance from the injection source was obtained. For a latchup, CAD methodology is as follows:

- identification of I/O cell for latchup evaluation;
- marking of the I/O cell using a virtual design shape.

A virtual design level can be used to identify the I/O cell itself, as was discussed in the Voldman–Sullivan–Nickel–Bass method for I/O identification [6]. In the analysis, a design system can utilize virtual design shapes and a specified region around each I/O cell for the evaluation of latchup.

Within the I/O cell, a second marker shape is needed to define the injector, using an 'injector shape'. Using an algorithm, the relative distance of a given circuit to the injector can be identified. In that process, the n-well contact 'cells' can be moved in order to lower the latchup sensitivity of the circuit. This can be modified until the latchup criteria are satisfied.

The problem with this methodology is that it is not always possible to modify the circuitry in a full-chip design. For many subfunctions, such as DRAM cells, SRAM cells and other design features, it is not advisable to modify the periodicity. Hence, this methodology is suitable for a specific environment where the periodicity of the contacts can be modified relative to an injection source. A second alternative is the placement of other types of circuits that are latchup insensitive; this will be discussed in a later section.

### 8.5.7 Transmission Probability Methodology (Voldman-Watson)

One of the difficulties in the analysis of external latchup is that a number of methodologies do not address the physical shapes and structures that are encountered in the minority carrier transport across a full-chip cross section. In a real semiconductor chip environment, it consists of metallurgical junctions, recombination surfaces, barriers and structures near the surface of the wafer. These modify the transport of the carriers and reduce the net transport across a semiconductor region.

One methodology that can be utilized is the Monte Carlo methods for minority carrier transport in a periodic arrays [37–42]. In this methodology, an injecting source establishes an 'initial condition' of the spatial location of the source. Carriers are then allowed to random-walk through a periodic three-dimensional structure. At each increment, the carrier could recombine, diffuse or get collected at a metallurgical junction. This Monte Carlo transport methodology can be applied to the external latchup problem for ionizing radiation or diffusion sources (e.g. diodes and n-wells). The analysis of the transport can be evaluated through Monte Carlo simulation to provide the actual transport probabilities and guard ring efficiency.

Another proposed approach is the conversion of the physical regions into stratified media of boundaries and regions [43]. This method is also known as the 'transfer matrix' methodology. Each physical region can be quantified as two surfaces and a region within the two surfaces. This can be converted to a 'transfer matrix'. To evaluate the transport from a number of regions, the transmission factor is obtained by the matrix multiplication of the different regions. These transfer matrices can also be determined by physical models, analytically or experimentally. In a complex set of shapes, this can be evaluated over a number of physical shapes. In the case of an array, using matrix algebra, simple calculations can predict transport through a complex array of many repeated structures.

In this fashion, a CAD method for external latchup injection sources can be developed, allowing evaluation of minority carrier transport through complex semiconductor designs. The CAD method evaluates electron transmission, reflection and absorption at geometric shapes that represent structures on the surface or near the surface of a semiconductor. A latchup CAD method for evaluating minority carrier transmission in a semiconductor chip design comprises the following steps [43]:

- Form a semiconductor chip computer design of geometric mask shapes.
- Define a first point of interest in the semiconductor chip design.
- Define a second point of interest in the semiconductor chip design.
- Define an arc between the first point and the second point in the said semiconductor chip design.
- Define a perpendicular interface (e.g. normal plane) at the intersection of the arc and the shape. Only define interfaces at boundaries in the silicon chip (e.g. ignore all metal shapes and only address isolation and silicon interface shapes).
- Define a region between each two surfaces (e.g. between the normal planes) along the arc path wherein the region is the volume and the boundaries of the region.
- Between each region, identify the transport properties within the region between the two interfaces.
- Define the absorption, reflection and transmission matrices for the domain.
- Calculate the transport between the two points of the arc as the product of the transmission matrices.
- Evaluate the total minority carrier transmission from said first point to said second point across the domains.

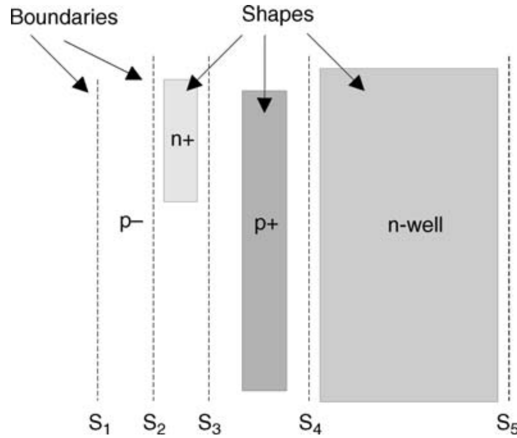



**Figure 8.12** Semiconductor chip design segmented into domains (e.g. regions and surfaces) for evaluation of transport properties through the regions.

In this process, the quantification of the different domains can be pre-evaluated using analytical tools, basic physical analysis, Monte Carlo analysis or experimentally. Any physical domain in a semiconductor chip can be defined as a set of simple domains:

- isolation region (shallow trench, deep trench or trench isolation regions);
- substrate region;
- well region;
- diffusions.

In these regions, diffusion transport and recombination characteristics (e.g. bulk or surface) can be defined. In a CAD tool, a substrate with a known doping polarity (e.g. p− substrate) is selected, and geometric shapes (or 'shapes' as used herein) may then be used to define or represent the components of a semiconductor device that are associated with a regions in the substrate. In the present context, 'regions' are taken to mean a collection of shapes located in an area of the semiconductor chip for which it is desired to determine the probability of transmission of electrons across the region or area. By defining the physical space into a set of domains as described herein, the characteristics of reflection, transmission and absorption operate as if the physical space (i.e. the path or trajectory) traversed by the electrons is a 'stratified media' of $N$ domains (Figure 8.12). From a matrix perspective, a forward transmission matrix allows the calculation of the transport from any point $A$ to point $B$ along the trajectory. An individual probability that an electron will traverse a given domain can be calculated for each individual domain along the path, and the total transmission probability (probability that the electron will travel from $A$ to $B$) is the product of the individual probabilities across each domain. The quantification of the probability of crossing an individual domain is dependent on either experimental measurements using test structures or the analytical knowledge of recombination and diffusion lengths. The transmission matrix for each domain can be defined as

$$T = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix}.$$

Notice that the transport characteristics of these domains can be defined and evaluated prior to the evaluation of the semiconductor chip analysis, where each domain has a stored evaluation based on the type of domain region and the technology information. In addition, in this methodology, the transmission characteristics of the transport can be preprocessed as (1) circuit macros, (2) parameterized cells, (3) array unit cells, (4) Cadence$^{TM}$ macros and (5) subfunctions or circuit blocks. In a Cadence$^{TM}$ environment, with parameterized cells, the transport matrix can be embedded into the parameterized cell (Pcell). The advantage of this method is the calculations that are carried out by the practice of the present invention can be prestored and contained in the Pcell and hierarchical Pcell prior to the chip design process. The Pcell transmission matrix may be coded into the design based on the Pcell design parameters and stored in the Pcell for each library element. In this case, the method would first evaluate the Pcell electron transmission process. In this process, the evaluation of the transmission factors can be achieved analytically, experimentally or theoretically. It can be contained in the library element deliverables prior to the calculation of the CMOS latchup evaluation process and inherently in the physical design allowing for cosynthesis of with CMOS latchup design the floor planning and chip design practice.

Hence, this CMOS latchup design methodology allows for an accurate quantification of the transport factors in a semiconductor chip to evaluate external CMOS latchup phenomenon (e.g. for any source), wherein all the physical shapes of the design are present.

## 8.5.8 Global Placement Methodology (Voldman)

With high-level integration, latchup continues to be a problem in MS, ASICs, SOC and NOC applications. Functional subfunction blocks that are being integrated into a semiconductor chip include memory, analog, digital and RF circuitries. In addition, in microprocessor design or ASIC implementations, subfunctions exist such as peripheral and array I/O networks, ESD elements, power books, breaker cells, voltage islands and core functions (e.g. clocks, PLL and DLL).

Few of the primary reasons why CMOS latchup is a concern today are as follows:

- latchup design is independently evaluated within a circuit, design subfunction or core prior to integration as a whole;
- latchup electrical testing and characterization are independently evaluated within a circuit, subfunction or core prior to final test on a full-chip level;
- latchup ground rules have historically been developed without considering the integration of different subfunctions and cores;
- latchup ground rules are focused on occurrence of internal and external latchup between the functional blocks that is observable only after full-chip design integration.

An example of this is the design and qualification of I/O designs. For example, in qualification of peripheral or array I/O books, the design and testing of the I/O ‘books’ are completed independent of integration into a semiconductor chip. A test site or a test vehicle may be established where all the I/O designs are tested for latchup prior to integration with a product or full chip as independent blocks. Cores, such as phase-locked loops (PLLs), are designed and tested for latchup independent of the final integration with other functional circuit blocks.

In chip integration, the floor planning of the semiconductor chip is typically completed based on performance, timing, wiring density and power distribution, or logical subfunction isolation. Latchup requirements are not integrated in the semiconductor chip optimization.

One solution to address interbook latchup concerns, due to internal or external latchup phenomenon, is the evaluation of the global placement of the chip subfunctions [44]. Internal latchup can occur on the chip boundaries between circuits of a first function and a second function that form a parasitic pnpn. External latchup extends beyond the physical boundaries and is dependent on the relative physical distances of the 'victim' pnpn and the injection source. An external latchup problem is a function of the relative placement of the injection sources from the vulnerable circuits.

A proposed CAD methodology is as follows [44]:

- a CAD chip subfunction design layout is defined;
- parasitic pnp, npn and pnpn structures are identified in the chip subfunction;
- identify pnp, npn and pnpn element density of at least one functional circuit block and element 'attributes' of elements associated with the at least one functional circuit block;
- an element density function parameterized from the element attributes is formed for pnp, npn and pnpn elements, where the element attributes contain parameters of interest to latchup (e.g. base widths, emitter widths and shunt resistances);
- the placement of the one functional circuit block is modified relative to other functional circuit blocks based on the element density function to substantially eliminate latchup between the functional blocks.

In many of the prior methods, the focus was on finding one pnp, or npn or pnpn to focus on a given device in a given region. For example, in the Habitz–Galland–Washburn method, the user searched for a given pnpn, extracted it, and analyzed the circuit response after the chip floor plan was completed [24]. In the method of Li, the 'strength' of the parasitic bipolars was quantified as a means of 'sorting' which pnpn elements will be the focus of the analysis and then integrate the pnpn into the circuit analysis [19]. In these cases, the analysis focuses on 'worst-case' pnpn, has significant focus on a given region or domain and does not address external latchup phenomenon.

In this method, for global analysis of chip design, this method provides a more 'full-chip analysis' methodology that can be processed prior to the assembly of the chip subfunction and then addressed with global integration.

In the step to identify element, the attributes could include identifying at least one of the elements sensitive to temperature, power, overshoot, undershoot, external sources, injection mechanisms, element pair density (e.g. pnp and npn), pnpn regenerative feedback strength, latchup resiliency, orientation and form factor in at least one functional circuit block and spatial density.

In the placement of the functional block relative to injection sources, the methodology can comprise steps of determining at least one of the following [44]:

- orientation and placement of an injecting source;
- distance between an ESD element and adjacent region of high pnpn density;
- form factors of the at least one functional circuit block;
- guard ring utilization between adjacent functional circuit blocks;
- npn density–number of npn per unit area;
- npn parameter strength (current gain);

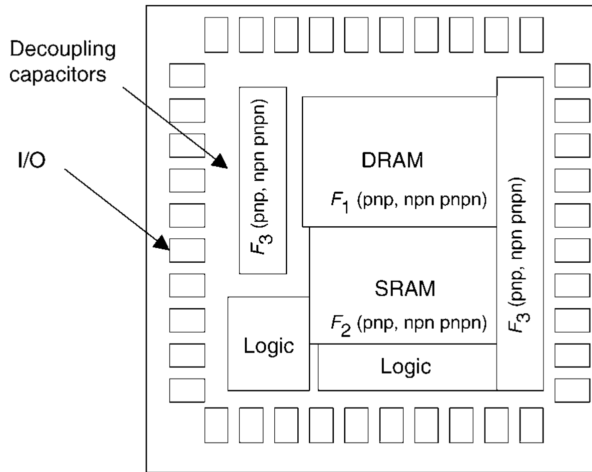



**Figure 8.13** Floor planning optimization based on pnp, npn and pnpn density functions.

- pnp density–number of pnp per unit area;
- pnp parameter strength (pnp current gain);
- parameterization of the npn density as a function of strength parameters statistics;
- parameterization of the pnp density as a function of the strength parameters statistics;
- pnpn density;
- pnpn parameter strength.

In the evaluation process of the strength parameters, this can be achieved by the geometrical variables from semiconductor design layout and the physical and electrical parameters from the technology data file. Analytical models can be used to determine the likelihood of latchup from injection sources.

In the floor planning of the semiconductor chips, chip subfunctions that have a low pnpn density can be placed near injection sources (Figure 8.13). Circuit subfunctions with high pnp density can be spaced from subfunctions that have high npn density. In this global chip design methodology, it is clear that the global floor plan of the semiconductor chip can be cosynthesized with the time analysis, wiring and performance objective in an efficient manner on a full-chip level.

## 8.5.9 Method of Latchup, Placement and Routing Cosynthesis (Cohn-Basaran)

With the introduction of mixed signal methodologies, which contain digital and analog functions, the issue of CMOS latchup was historically focused on the needs of the digital CMOS sector of the design. Analog cell placement and routing techniques and concepts to address these kind of issues have been developed in the Ph.D. thesis of Cohn (e.g. routing tool KOAN) [45,46]. With the growth of

analog–digital chips and SOC integration, there is a need for global methodologies that address the placement, routing and latchup [47,48]. Extending the Cohn methodology, Basaran *et al.* discuss the concept of constraints on the routing and placement algorithm that address latchup [47,48]. It is noted that many methodologies first place well and substrate contacts to define a 'zone of protection'. A 'halo' with a given radius from each well and substrate contact is formed and drawn based on the design rules of distance from each well and substrate contact. If any region of an active device is not within the 'halo' and the zone of protection, it is regarded as a latchup ground rule fail. Hence, a 'halo model' provides a region to establish the zone of protection of the devices within the physical region local to the contacts. Any device or part of a device falling outside the 'zone of protection' initiates a design rule failure. In many design methods, a postprocessing design tool provides additional contacts as a means to address the design failure after the design has been optimized. As noted by Basaran, in the presence of very large individual devices (e.g. analog elements or LDMOS power devices) or dense cells containing many devices, a postprocessing solution or any other *ad hoc* placement is problematic. The proposed method to address this issue is a three-step process [47,48]:

- device generation;
- device placement;
- well generation.

In this methodology, each device is a set of devices with alternate layouts available. In the second step, device placement is performed by a Gelatt–Jepsen annealing algorithm in which the elements are placed arbitrarily [47,48]. As a third step, the well is generated under the devices where these can be shrunk or expanded on the basis of the device placement.

In this methodology, the device generators created the actual geometry of the well and contact spacings. With this decision, the well and substrate contacts also can be a suite of dimensional layout (e.g. variants) with alternative layout shapes; it is as if the well and contacts are a 'device' in its generation step (e.g. as stated in Step 1). To reduce the complexity of the interaction, each device generated has a correspondence to the well and contacts suite, and as a second constraint, the contacts are 'run' around a 'bounding box' of the device generated. Hence, the contact placement is being treated almost as a 'device' itself, but has a one-to-one correspondence and geometric constraints to reduce the complexity. As a simplification of this method, a group of devices can be associated with the 'contact' under a similar constraint to allow sharing of well and substrate contact structures. In the method of Basaran *et al.*, the methodology allows for the optimization based on a subset of devices to reduce 'over-design'. Using the tools 'relocating' and 'reshaping' functions, the full global design can be optimized to include the needs of latchup contact placement optimization within the placement and routing algorithm. The cost functional algorithm is as follows [47,48]:

$$\begin{aligned} Cost = {} & W_L \; Wire\ Length + W_A \; Cell\ Area + W_M \; Mergeable\ Area\ Used \\ & + W_O \; Illegal\ Device\ Overlap + W_L \; Latchup\ Unprotected\ Devices. \end{aligned}$$

In this algorithm, the first three terms are minimized and the last two latchup-related terms are constraints that are driven to zero during the optimization process. The two latchup terms are associated with areas outside the 'zone of protection'. In this fashion, the latchup contact placement, routing and cell optimization can all be cosynthesized. This avoids the postprocessing methodology used in many foundries today. In this fashion, local and global optimizations can be addressed with cosynthesis of the latchup needs and product of area and performance optimization.

## 8.6 LATCHUP CAD SIMULATION

Latchup CAD methods also include both semiconductor device level and circuit simulations. Latchup CAD simulation can be completed at a semiconductor device level or circuit simulation. Latchup simulation of both device and circuit is a natural part of the CMOS latchup design discipline.

### 8.6.1 Latchup CAD Semiconductor Device Simulation

In the latchup design discipline, it is common to evaluate the geometric ground rules (e.g. p+/n+ spacing) and the semiconductor process variables using semiconductor device simulation. In latchup modeling, the electrical potential, current density and current contours can be evaluated to observe the spatial and temporal results during latchup events. As an example, Figure 8.14 shows the potential distribution of a four-stripe pnpn structure used to evaluate latchup. As a result, semiconductor device simulation can provide assistance to the following areas:

- semiconductor process definition;
- ground rule spacings definition;
- current and voltage responses;
- failure analysis;

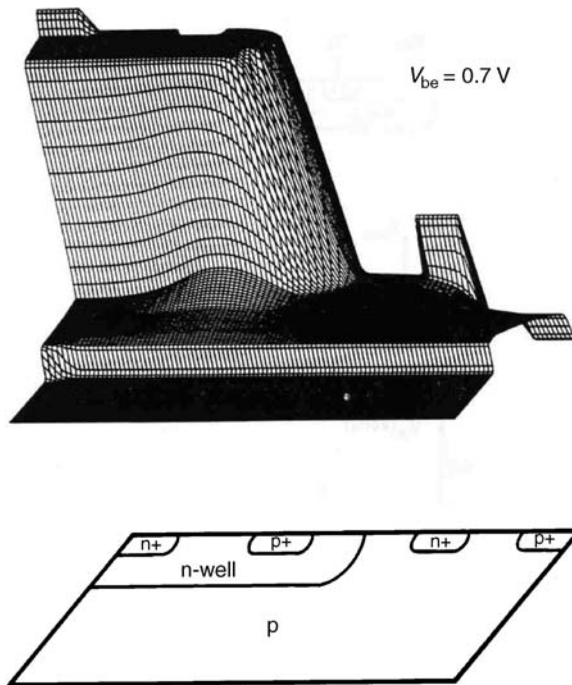



**Figure 8.14** Semiconductor device potential distribution of a four-stripe pnpn structure.

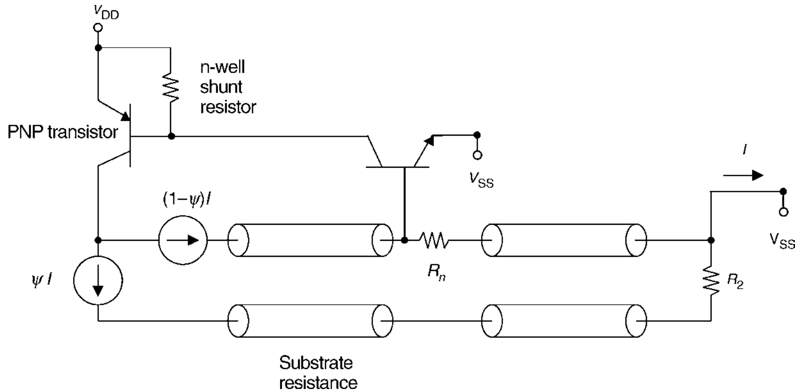



**Figure 8.15** Transmission line model representation of pnpn structure.

- root cause analysis;
- characterization-to-model calibration.

### 8.6.2 Latchup CAD Circuit Models

In the latchup design discipline, circuit simulation can also be used to evaluate the response of the circuit. Examples have been shown in the text to explain latchup events. The sophistication of the latchup models can be simple or complex. Figure 8.15 is a transmission line model of the pnpn structure where the substrate is represented as a transmission line. In this model, the model includes both a frontside and a backside substrate contact. Hence, one direction is to use circuit simulation tools to understand the transmission and reflection characteristics of a parasitic device.

In CAD circuit analysis, the structures used vary from simple four-stripe pnpn structures to full-circuit cross sections. In other directions, additional methodologies exist that focus on introduction of the parasitic devices into the simulation of the circuits themselves as proposed by Li [19]. In the Li method, the parasitic devices are sorted, extracted and then included into the active circuit itself. As a first step, the simulation tool extracts the pnpn structure or parasitic bipolar, distinguishes which ones are important and then includes the parasitic device into the circuit simulation of the active devices.

## 8.7 SUMMARY AND CLOSING COMMENTS

Chapter 8 addresses latchup design checking and verification methodologies; the methodologies address finding the parasitic pnpn elements, determining the presence of guard ring structures and verifying the substrate and well taps (e.g. contacts) to sophisticated graph model theory verification techniques. Simulation is used from semiconductor device simulators, circuit models of the pnpn, circuit simulation with attached parasitics to mixed-mode simulation of system/chip/circuit and a semiconductor device. Latchup can be integrated into today's design methods for chip optimization, wherein sophisticated methods using cost functionals, rapacious algorithms and cosynthesis can be instituted. In the last 20 years, the methods of chip designs have evolved, and as a result, the methods used today for latchup have also increased in technical sophistication.

## PROBLEMS

1. Given a latchup DRC methodology, what is the minimum number of rules needed to provide latchup sensitivity? What are the rules?
2. Given a CMOS inverter circuit, show the critical dimensions that are to be checked to latchup analysis.
3. Given an I/O cell, containing an n-channel pull-down MOSFET, n-doped ballast resistors, a p-channel pull-up, and a dual-diode ESD network (containing a p+/n-well diode and a n-well/p− substrate diode), show the guard ring configuration in the design to prevent latchup events.
4. In the Li methodology, strength factors were assigned to the parasitic pnpn structures by identifying the base widths of the lateral parasitic pnp and lateral npn transistors in a design layout. Show an inverter layout, identify the strongest pnpn regions, define the key parameters for the parasitic devices and add the parasitic pnpn circuit schematic to the CMOS inverter schematic.
5. Relate the strength factors criteria and the latchup criteria without inclusion of the well and substrate shunt resistors. Relate the strength factor criteria with the inclusion of the well and substrate extracted resistor elements.
6. Using the method of Zhan and Wang, show the MG for a parasitic pnpn element from a CMOS inverter layout.
7. Show the extracted MG of the parasitic pnpn elements from a dual-diode ESD network that contains a p+ diffusion in an n-well, an n-well diode adjacent to the first diode element and an n-well guard ring around both elements. Show all the parasitic diodes and transistors formed between the diodes, tubs and guard rings.
8. Given that all physical shapes can be defined as a region within two boundaries, each physical region in a design can be represented as a stratified medium. Using this representation, a transmission, absorption and reflection characteristic can be defined for all regions. Show a representation of a region in this formulation. From this transmission factor representation, what is the advantage of this representation?
9. Given a transmission factor can be defined for a repeating unit cell in a silicon chip, the transmission of a carrier across a periodic unit cell can be represented as a matrix product. Show how this can be utilized for prediction of minority carrier propagation in a semiconductor chip.
10. Given a methodology of chip design that addresses the density of pnpn parasitic elements in a given core or chip subsector, how would you floor plan the semiconductor chip assuming the chip contains I/O banks, DRAM and SRAM memory, silicon fill shapes, back-end only fill shapes, n-type decoupling capacitors and CMOS sea of gates.

## REFERENCES


1. R. Troutman. *CMOS Latchup in Semiconductor Technology: The Problem and the Cure*. New York: Kluwer Publications; 1985.
2. S. Voldman. *Latchup Design Rules Section*. CMOS IV Technology Manual. Burlington, Vermont: IBM Corporation; 1984.

3. S. Voldman. *Latchup Design Rules Section*. CMOS V Technology Manual. Burlington, Vermont, IBM Corporation; 1988.
4. A. Kapoor, D. Suh and P. Bendix. *ESD/latchup design rules*. Technical Report, LSI Logic; 1996.
5. S. Voldman, C.N. Perez and A. Watson. Guard rings: structures, design methodology, integration, experimental results and analysis for RF CMOS and RF mixed signal BiCMOS silicon germanium technology. *Journal of Electrostatics*, **64**, 2006, 730–743.
6. R.S. Bass Jr., D.J. Nickel, D.C. Sullivan and S.H. Voldman. Method of automated ESD protection level verification. U.S. Patent No. 6,086,627, July 11, 2000.
7. C.N. Perez and S. Voldman. Method of forming guard ring parameterized cell structure in a hierarchical parameterized cell design, checking and verification system. U.S. Patent Application No. 20040268284, December 30, 2004.
8. T. Li and S.M. Kang. Layout extraction and verification methodology for CMOS I/O circuits. *Proceedings of the IEEE/ACM Design Automation Conference* (*DAC*), 1998. p. 291–296.
9. T. Li, Y. Huh and S.M. Kang. Automated extraction of parasitic BJTs for CMOS I/O circuits under ESD stress. *Proceedings of the IEEE International Integrated Reliability Workshop* (*IRW*), 1997. p. 103–109.
10. U. Lather. An *O(NlogN)* algorithm for boolean mask operations. *Proceedings of the IEEE/ACM Design Automation Conference* (*DAC*), 1981. p. 555–562.
11. A. Gupta. ACE: a circuit extractor. *Proceedings of the ACM/IEEE Design Automation Conference* (*DAC*), 1983. p. 721–725.
12. K. Chang, S. Nahar and C. Lo. Time-efficient VLSI artwork algorithms in GOALIE2. *IEEE Transaction on Computer Aided Design*, **8**(6), 1989, 640–647.
13. E. Barke. A layout verification system for analog bipolar integrated circuits. *Proceedings of the ACM/IEEE Design Automation Conference* (*DAC*), 1983. p. 353–359.
14. T. Li, C.H. Tsai, E. Rosenbaum and S.M. Kang. Substrate resistance modeling and circuit level simulation of parasitic device coupling effects for CMOS I/O circuits under ESD stress. *Proceedings of the Electrical Overstress/Electrostatic Discharge* (*EOS/ESD*) *Symposium*, 1998. p. 281–289.
15. T. Li, C.H. Tsai, Y. Huh, E. Rosenbaum and S.M. Kang. A new algorithm for circuit level electro-thermal simulation under EOS/ESD stress. *Proceedings of the IEEE International Reliability Workshop* (*IRW*), 1997. p. 130–131.
16. T. Li, C.H. Tsai, E. Rosenbaum and S.M. Kang. Substrate modeling and lumped substrate resistance extraction for latchup/ESD circuit simulations. *Proceedings of the IEEE/ACM Design Automation Conference* (*DAC*), 1999. p. 549–554.
17. T. Li and S.M. Kang. Layout extraction and verification methodology for CMOS I/O circuits. *Proceedings of the IEEE/ACM Design Automation Conference* (*DAC*), 1998. p. 291–296.
18. T. Li, S. Ramaswamy, E. Rosenbaum and S.M. Kang. Simulation and optimization of deep submicron output protection device. *Proceedings of the Custom Integrated Circuits Conference* (*CICC*), 1997. p. 159–162.
19. T. Li. *Design automation for reliable CMOS chip I/O circuits.* Ph.D. thesis, University of Illinois Urbana-Champaign, UILU-ENG-98–2219; August 1998.
20. R.Y. Zhan, H.G. Feng, Q. Wu, G. Chen, X.K. Guan and A.Z. Wang. A technology-independent CAD tool for ESD protection device extraction: ESD extractor. *Proceedings of the International Conference on Computer-Aided Design*, 2002. p. 510–513.
21. P. Venugopal, S. Sinha, S. Ramaswamy, C. Duvvury, G.C. Prasad, C.S. Raghu and G. Kadamati. Integrated circuit design error detector for electrostatic discharge and *latch-up* applications. U.S. Patent No. 6,493,850, December 10, 2002.
22. S. Ramaswamy, S. Sinha, G. Kadamati and R. Gharpurey. Semiconductor device extractor for electrostatic discharge and latch-up applications. U.S. Patent No. 6,553,542, April 22, 2003.
23. R.Y. Zhan. *ESDCat: a new CAD software package for full-chip ESD protection circuit verification*, October 2005.
24. M.S. Galland, P.A. Habitz and S.E. Washburn. Method and apparatus for detecting devices that can latchup. U.S. Patent No. 6,848,089, January 25, 2005.
25. S. Voldman, S. Strang and D. Jordan. A design system for auto-generation of ESD Circuits. *Proceedings of the International Cadence Users Group* (*ICUG*), September 2002.
26. S. Voldman, S. Strang and D. Jordan. An automated electrostatic discharge computer-aided design (CAD) system with the incorporation of hierarchical parameterized cells in BiCMOS analog and RF technology for mixed signal

applications. *Proceedings of the Electrical Overstress/Electrostatic Discharge (EOS/ESD) Symposium*, October 2002. p. 296–305.

27. S. Voldman. Automated hierarchical parameterized ESD network design and checking system. U.S. Patent No. 6,704,179, March 9, 2004.
28. D.S. Collins, D.L. Jordan, S.E. Strang and S. Voldman. ESD design, verification, and checking system and method of use. U.S. Patent No. 7,134,099, November 7, 2006.
29. D. Collins, D. Hershberger and S. Voldman. US Patent Application; 2006.
30. M.D. Ker, H.C. Jiang, J.J. Peng and T.L. Shieh. Automatic methodology for placing the guard rings into a chip layout to prevent latchup in CMOS IC's. *Proceedings of the 8th IEEE International Conference on Electronics, Circuits and Systems* (*ICECS*), 2001. p. 113–116.
31. S. Kimura and H. Tsujikawa. Latch-up verifying method and latch-up verifying apparatus capable of varying over-sized region. U.S. Patent No. 6,490,709, December 3, 2002.
32. S. Kimura and H. Tsujikawa. Latch-up verifying method and latch-up verifying apparatus capable of varying over-sized region. U.S. Patent No. 6,718,528, April 6, 2004.
33. S. Voldman. Latch-up analysis and parameter modification. U.S. Patent No. 6,996,786, February 7, 2006.
34. K. Chatty, P. Cottrell, M. Muhammad, F. Stellari, A. Weger, P. Song and M. McManus. Model-based guidelines for cable discharge events (CDE) induced latchup in CMOS ICs. *Proceedings of the International Reliability Physics Symposium* (*IRPS*), 2004. p. 130–134.
35. K. Chatty, P. Cottrell and M. Muhammad. Method and structure to suppress external latchup. Patent Application No. 20050085028, April 21, 2005.
36. C.J. Brennan, K. Chatty, J. Sloan, P. Dunn, M. Muhammad, *et al*. Design automation to suppress cable discharge events (CDE) induced latchup in 90 nm ASICs. *Proceedings of the Electrical Overstress/Electrostatic Discharge (EOS/ESD) Symposium*, October 2005. p. 126–130.
37. G. Sai-Halasz, M.R. Wordeman and R.H. Dennard. Alpha-particle-induced soft error rate in VLSI circuits. *IEEE Transactions on Electron Devices*, **29**, 1982, 725–731.
38. G. Sai-Halasz and D.D. Tang. Soft errors rates in static bipolar RAMs. *International Electron Device Meeting (IEDM) Technical Digest*; 1983. p. 344–347.
39. S. Voldman and L. Patrick. Alpha particle induced single event upsets in bipolar static emitter coupled logic (ECL) cells. *Proceedings of the IEEE Nuclear and Space Radiation Conference*, Boulder Colorado; 1983.
40. S. Voldman and L. Patrick. Alpha particle induced single event upsets in bipolar static emitter coupled logic (ECL) cells. Special Edition on the 1984 Annual Conference on Nuclear and Space Radiation Effects, *IEEE Transactions on Nuclear Science*; 1984. p. 1196–1200.
41. S. Voldman, L. Patrick and D. Wong. Alpha particle effects on bipolar emitter-coupled logic static arrays. *Proceedings of the International Solid State Circuit Conference* (*ISSCC*), Vol. XXVII; 1985. p. 262–263.
42. S. Voldman, P. Corson, L. Patrick, K. Nguyen, L. Gilbert, R. Goodwin, T. Maffit and S. Murphy. CMOS SRAM alpha particle modeling and experimental results. *International Electron Device Meeting (IEDM) Technical Digest*; 1987. p. 518–521.
43. A.E. Watson and S. Voldman. Method of quantification of transmission probability for minority carrier collection in a semiconductor chip. U.S. Patent No. 7,200,825, April 3, 2007.
44. S. Voldman. Methodology for placement based on circuit function and latchup sensitivity. U.S. Patent No. 7,089,520, August 8, 2006.
45. J.M. Cohn. *Automatic device placement for analog cells in KOAN*. Ph.D. Thesis, Department of Electrical Engineering and Computer Science, Carnegie Mellon University, February 1992.
46. J.M. Cohn, D.J. Garrod, R.A. Rutenbar and L.R. Carley. KOAN/ANAGRAM II: new tools for device-level analog placement and routing. *IEEE Journal of Solid State Circuits*, **26**(3), 1991.
47. B. Basaran. *Latchup-aware placement and parasitic-bounded routing of custom analog cells.* M.S. Thesis, Department of Electrical Engineering and Computer Science, Carnegie Mellon University, May 7, 1993.
48. B. Basaran, R. Rutenbar and L. Carley. Latchup-aware placement and parasitic-bounded routing of custom analog cells. *Proceedings of the 1993 IEEE/ACM International Conference on Computer Aided Design* (*ICCAD*), 1993. p. 415–421.

# Index


Active clamp networks
bipolar active clamps, 357, 387, 388
CMOS active clamps and latchup, 388, 390
ESD-active clamp integration, 13, 27
latchup of ESD power clamps, 357, 364, 365, 366, 377
latchup sensitivity of ESD power clamps, 15, 69, 71, 192, 257, 420, 429, 430, 432
Active guard rings
compensating active guard rings, 397, 400
ESD, 192
inverting signal guard rings, 13, 396
latchup active guard rings, 393
lateral electric field assist, 91, 96
noise reduction network, 20
Alpha space
generalized alpha space latchup criteria, 22, 432, 441
Avalanche
avalanche breakdown, 14, 81, 82, 281, 282, 321
avalanche multiplication, 62, 82, 321

Ballasting
resistor ballasting, 247
Barak
single event latchup (SEL), 36
Beta space
beta space latchup criteria, 138, 142
BILLI (buried implanted layer for lateral isolation) structure
heavily doped buried layers, 289
Bipolar transistors
avalanche multiplication, 62, 82
current gain-second breakdown metric, 60
Gummel plot, 64
Johnson limit condition, 121
Kirk effect, 312
lateral transistor models, 85
triple well transistor models, 106–108
vertical transistor models, 106, 108, 110
Borland, J.
BILLI (buried implanted layer for lateral isolation) structure, 289–290
Boselli, G.
scaling, 24
substrate resistance, 302, 306
Buried grid (BGR) structure, 74, 257

Cable discharge event (CDE)
cable discharge event induced latchup, 34
event, 34, 35
Charged cable model, 34, 35, 217
Charged device model (CDM), 25
Cross section
latchup cross section, 44

Deep trench
bipolar transistors, 55, 60, 81




© 2007 John Wiley & Sons, Ltd

Deep trench (*Continued*)
ESD structures, 217
floating polysilicon-filled trench, 346
guard ring structures, 190, 232, 340, 422, 425
latchup, 206
polysilicon filled deep trench structures, 344
Design methodology
external latchup design methods, 435
internal latchup design methods, 211
substrate contact, 431
well contact, 195
well-substrate resistance space design plot, 138
Design rule checking (DRC)
design rules, 131, 193, 199, 408, 409, 438
guard rings, 413
placement of ESD networks, 421
virtual dummy ESD design levels, 421
virtual dummy latchup design levels, 421
Design systems
cellview, 422
checking systems, 412
guard ring parameterized cells, 422
hierarchical parameterized cells ESD design, 422
parameterized cells, 422
placement of ESD networks, 364
placement of ESD power clamps, 364, 365
symbol, 422
verification systems, 21
virtual dummy ESD design levels, 421, 422
virtual dummy latchup design levels, 421
Voldman–Strang–Jordan methodology, 422
Differential generalized tetrode relationship
external injection source, 201, 412
Diodes
diode equation, 58, 64
self-heating, 59
series resistance, 58, 59, 258, 360, 370
Domino effect
latchup criteria with external source, 173
latchup propagation, 9
Dual well CMOS, 25, 106, 107, 192, 257, 269, 278
Dutton, R.
Dutton–Whittier model, 103
metallurgical junction model, 98
Duvvury, C.
latchup design verification, 427–428

Electrical instability, 1, 3, 125
Electrostatic discharge (ESD)
failure mechanisms, 25–28
Electrostatic discharge (ESD) circuit induced latchup
cable discharge event, 216, 217
Electrostatic discharge (ESD) power clamps
field oxide device (FOD), 26, 362
n-channel RC triggered clamps, 258, 361, 385
p-channel RC triggered clamps, 27, 357, 385
Estreich, D.B.
field assisted lateral bipolar model, 91
lateral bipolar model, 91
sensitivity factors, 131
External latchup
alpha particle source, 173
cable discharge event (CDE), 431
computer aided design methods, 412–413
cosmic ray source, 39
injection source to victim distance, 202
injection source, 168, 202
latchup Domino effect, 9, 177
relative orientation, 202

Feedback
avalanche multiplication, 62, 82, 321
regenerative feedback, 125–131
Floor planning
core, 240
global placement, 435
guard ring placements, 189, 425
guard rings, 189, 425
peripheral I/O, 14, 366
well and substrate contact placement, 429

Ground rules
external latchup rules, 412
fundamental latchup design rules, 408
voltage condition rules, 409
Guard rings
design rule checking (DRC), 413
ESD and guard ring integration, 391
guard ring efficiency, 106, 231–232, 234
latchup, 236–237
n-well, 190, 259, 360, 368, 370
structures, 189–191, 196, 232, 263, 392, 420
Gummel plots, 64

Hargrove, M.
latchup simulation, 8
transmission line model, 8
Heavily doped buried layers (HDBL)
HDBL and substrate resistance, 184
HDBL bipolar transistor performance impacts, 351
HDBL MOSFET ESD impact, 184
HDBL recombination time, 292–293
High voltage CMOS (HVCMOS)
guard rings, 393
LDMOS, 275, 329, 393
Huh, Y.
inter-supply latchup 365, 366
Human body model (HBM), 33, 34

Instability
electrical instability, 1, 3, 125
regenerative feedback, 125–131
Internal latchup
n-well, 190
p-well, 190
substrate, 111
Isolation structures
deep trench (DT), 108, 329
dual depth shallow trench isolation, 395
LOCOS, 17, 85, 195, 274–275
shallow trench isolation (STI), 257, 275–277
trench isolation (TI), 9, 108, 202, 296, 317–328

Johnson limit
voltage relationship, 121
Jordan, D.
Cadence™ ESD design methodology, 422

Ker, M.-D.
RC-triggered PFET MOSFET power clamp latchup, 27, 358

Latchup
buried grid, 257
deep trench, 208, 209
heavily doped buried layer, 9, 184, 259
shallow trench isolation (STI), 17, 36, 196, 257, 275, 371
sub-collectors, 329
substrate doping scaling, 258, 287, 310
trench isolation, 17, 184, 196, 202, 257
LDMOS
active guard ring concepts, 23, 191, 329
inductive load issue, 28, 191, 203, 329
reverse current problem, 23
smart power, 10, 23, 28, 192, 203, 275, 329–330, 393, 407
Leakage mechanisms, 266
Lindmayer
lateral bipolar model, 89, 91, 353
Lindmayer–Schneider model, 86, 87
Linear energy transfer (LET)
linear energy transfer threshold, 23, 43
LOCOS isolation
LOCOS-defined ESD structure, 362
$n^+$/substrate diode, 247, 258, 357
$n^+$ to n-well lateral bipolar, 197, 290, 338
n-well-to-n-well lateral bipolar, 100
n-well-to-substrate diode, 247, 357
$p^+$/n-well diode, 26, 340, 358
thick oxide MOSFET, 362
Logic disturb
partial latchup, 131, 137

Machine model (MM), 34
Merged triple well
bipolar current gain, 109, 111, 113, 115, 117, 119
lateral bipolar transistor, 108
vertical bipolar transistor, 346
Models
cable discharge event (CDE), 431
charged device model, 25
human body model (HBM), 33
machine model (MM), 34
transmission line pulse (TLP) model, 24, 221
Morris, W.
buried grid (BGR), 74, 257
heavily doped buried layers (HDBL), 23, 24
MOSFET
avalanche breakdown and snapback, 81, 82
dielectric breakdown, 275
gate-induced drain leakage (GIDL), 267
MOSFET instability, 1

Noise
latchup and noise, 28
n-well design
diffused well, 261, 262
high energy well implant, 348, 400
retrograde well, 266

Passive guard rings
integration of active and passive guard rings, 391, 395
n-well rings, 392
$p^+$ substrate rings, 394
Power clamps
bipolar ESD power clamps, 11, 13, 357, 400
CMOS power clamps, 8, 270
diode string ESD power clamp, 363, 365
latchup events in power clamps, 357, 364
latchup prevention using power clamps, 375
Power-sequencing
ESD networks, 13
I/O, 358
mixed voltage interface, 11, 13, 357, 400
multiple power domains, 411
multiple power supplies, 30
p-well
connecting implant, 115, 270
dual well CMOS, 17, 269

Recombination and generation mechanisms
Auger recombination, 75, 184, 314
gold recombination centers, 78
lifetimes, 20
Shockley–Hall–Read (SHR) recombination, 73, 121
Retrograde wells
n-well, 261, 263, 267, 425
p-well, 22
retrograde well substrate modulation, 22, 23, 267
sheet resistance, 266
Rubin, L.
buried grid (BGR), 74, 258, 295
heavily doped buried layers (HDBL), 9, 24, 184
latchup simulation, 8, 273
Rung, R.D
deep trench CMOS, 207

Safe area
generalized alpha space safe area relationship, 154, 175
Salicide
cobalt salicide, 424
emitter resistance, 277
latchup stability criteria, 277
titanium salicide, 17, 424
Salicide block mask
bipolar, 277
diodes, 56
MOSFETs, 277, 278
resistors 384, 396
Scaling
deep trench scaling, 328
n-well scaling, 269
$p^+/n^+$ space scaling, 273
p-well scaling, 269
shallow trench isolation, 275
substrate doping concentration scaling, 258
trench isolation, 317
Sensitivity parameters, 132
Sequence dependent ESD networks, 362
Sequence independent ESD networks
floating-well ESD network, 374
pin-power sequencing, 29
power supply sequencing, 381
voltage islands, 435
Shallow trench isolation, 275
Shallow trench isolation ESD devices
$n^+$ to substrate diode 359
n-well to n-well lateral bipolar, 195, 290, 370
n-well-to-substrate diode, 357
$p^+$/n-well diode, 340, 359, 360, 370
Silicide
abrupt junctions, 79
Silicon controlled rectifiers
generalized tetrode relationship, 340, 342
holding current relationship, 137
regenerative feedback analysis, 125
Single event latchup (SEL)
alpha particle induced latchup, 178
cosmic ray induced latchup, 39, 42
heavy ion induced latchup, 40, 42
linear energy transfer (LET) threshold, 23
linear energy transfer (LET), 182
Smart power
active guard rings, 91, 191, 391, 393
compensating active guard rings, 397
external latchup, 168
guard ring, 189, 232, 234
high voltage CMOS, 275, 407
LDMOS, 392
passive guard rings, 392

Snapback
  bipolar, 259
  MOSFET, 363, 371
Stability
  electrical stability, 132
  generalized differential tetrode relationship, 154, 160
  thermal stability, 132
Standards
  ESD association transient latchup upset (TLU) standard, 24
  JEDEC latchup standard, 216, 217, 320
Stellari, F.
  latchup characterization, 240
  ESD circuit induced latchup, 240
Strang, S.E.
  Cadence-based ESD design methodology, 422
Sub-collectors
  heavily doped sub-collectors 312, 340
  lightly doped sub-collectors 323
Substrates
  epitaxial wafers, 259
  heavily doped substrates, 115, 180, 257, 258
  lightly doped substrates, 115, 180
  retrograde modulation effect, 267
System level
  battery reversal, 6
  inductive mismatch, 369
  reverse polarity issue 369
System level model
  cable discharge event (CDE), 431

Test
  external latchup characterization, 203
  internal latchup characterization, 16, 192, 205, 209, 213
  wafer level latchup characterization, 217
Test techniques
  chip level, 411
  product level, 214, 216
  system level, 368, 400
Transient latchup
  testing, 213
  theory, 160
Transient latchup upset (TLU), 223, 226
Transmission line pulse (TLP) testing, 222
Triple well
  buried layer, 9, 108–110
  vertical parasitic NPN, 121, 402
Troutman, R.R.
  epitaxial wafer model, 112
  generalized differential tetrode model, 142
  guard ring efficiency model, 231, 233
  SAFE model, 154
  transfer resistance method, 117
  transient latchup, 160
  transmission line model, 112, 114

Verification systems
  Habitz–Galland–Washburn, 419
  Ker–Jiang–Peng–Shieh, 425
  Kimura–Tsujikawa, 428
  Li, 415
  Ramaswamy–Sinha–Kadamati– Gharpurey method, 426
  Venugopal–Sinha–Ramaswamy–Duvvury–Prasad– Raghu–Kadamati, 427
  Voldman–Sullivan–Nickel–Bass, 421
  Zhan–Feng–Wu–Chen–Guan–Wang, 418
Voldman, S.
  active guard ring networks, 352
  biased deep trench (DT) structure, 9, 328
  Cadence-based methodology, 422
  computer aided design (CAD) latchup method, 407
  deep trench (DT), 328
  deep trench (DT)–log (substrate resistance) plots, 334
  deep trench (DT) guard rings, 329
  domino effect theory, 9, 173, 177
  dual depth shallow trench isolation (STI), 12, 275, 295
  edge implants, 295
  epitaxial grown sub-collector, 312
  ESD circuit induced latchup, 247
  global placement latchup CAD methodology, 436
  guard ring resistance CAD methodology, 423
  guard rings, 189–191, 233, 413, 422
  high resistivity substrate, 278
  implanted sub-collector, 24, 296, 323, 324, 326
  LOCOS isolation, 274–275
  merged triple well, 24, 108, 205, 279, 282, 288, 348
  n-well substrate modulation, 22, 23, 267

Voldman, S. (*Continued*)
  parameterized cell (Pcell) guard rings, 422
  p-well and scaling, 269, 270, 273
  retrograde wells, 263
  shallow trench isolation (STI), 275
  sub-collectors 312, 323, 340
  substrate resistance scaling, 111, 257, 306
  substrate resistance, 301
  TLP pico-second current analysis (PICA) tool, 24, 243
  trench isolation (TI), 317
  triple well, 277

Wang, A.Z.
  checking systems, 413
  extraction methods, 420, 426
  verification systems, 427, 428
Watson, A.
  deep trench (DT), 328
  deep trench (DT) guard rings, 329
  guard ring efficiency, 231
Weger, A.
  ESD circuit induced latchup, 247
  TLP-PICA tool, 246
Wells
  diffused wells, 261
  retrograde wells, 261, 267, 275, 424
  scaling, 19, 269, 273
  triple well and isolated MOSFETs, 277, 278, 279
  vertically modulated well, 23
  well ballast resistors, 192, 241, 410

Zappe, H.
  transient latchup, 160, 161, 164, 223, 225, 226
Zhan, R.Y.
  checking system, 412
  extraction, 412, 415, 418, 419, 420, 426
  model graph (MG), 10
  verification system, 10, 427, 428